J. CETACEAN RES. MANAGE. 18: 33–80, 2018

Guidelines for genetic data analysis

Robin S. Waples[1], A. Rus Hoelzel[2], Oscar Gaggiotti[3], Ralph Tiedemann[4], Per J. Palsbøll[5], Frank Cipriano[6], Jennifer Jackson[7], John W. Bickham[8] and Aimée R. Lang[9]

Contact e-mail: robin.waples@noaa.gov

ABSTRACT

The IWC Scientific Committee recently adopted guidelines for quality control of DNA data. Once data have been collected, the next step is to analyse the data and make inferences that are useful for addressing practical problems in conservation and management of cetaceans. This is a complex exercise, as numerous analyses are possible and users have a wide range of choices of software programs for implementing the analyses. This paper reviews the underlying issues, illustrates application of different types of genetic data analysis to two complex management problems (involving common minke whales and humpback whales), and concludes with a number of recommendations for best practices in the analysis of population genetic data. An extensive Appendix provides a detailed review and critique of most types of analyses that are used with population genetic data for cetaceans.

KEYWORDS: ABUNDANCE ESTIMATE; BREEDING GROUNDS; CONSERVATION; DNA FINGERPRINTING; FEEDING GROUNDS; GENETICS; HUMPBACK WHALE; MIGRATION; MINKE WHALE; REPRODUCTION; TAXONOMY

INTRODUCTION

Recently, guidelines were adopted for quality control of DNA data intended for use within the International Whaling Commission (IWC, 2009; 2015a). Once the data have been collected, the next step is to analyse the data and make inferences that are useful for addressing practical problems in the management of cetaceans. This is a complex exercise for two major reasons: (1) many methods can be used to analyse genetic data, and an equally wide range of computer software is available to conduct data analyses; and (2) a key objective is to inform those involved in cetacean management who do not have a background in population genetics. For these reasons, it has been suggested that it would be useful to have a document that provides guidelines for the analysis of population genetic data for use in a management context. Although it is not possible (nor is it desirable) to prescribe specific procedures for all analyses of population genetic data, it can help to provide general guidelines for some of the more common types of analyses conducted in a management context. The latter is the objective of this paper. Emphasis is on a general discussion of issues involved in genetic data analysis rather than detailed comments about specific computer software, but some popular programs will be discussed to emphasise particular points. Given the many analytical methods (and software packages) available, to focus on those most relevant to a particular study, the discussion has been organised around some common management problems one might try to address with genetic data. These problems are identified below. It is assumed that before the analyses considered here begin, the DNA data quality-control guidelines have been consulted and followed to the extent possible, and that any substantial deviations have been documented and explained.

As discussed in detail later, genetic information can provide insights relevant to many types of problems associated with conservation and management of living natural resources. Among other applications, genetic data can be used to:

(1) identify and delimit biological species, subspecies and populations;

(2) provide or improve estimates of census population size (N) and effective population size (N_e);

(3) help track contemporary movements of individuals, as well as estimate long-term levels of connectivity among populations;

(4) quantify genetic diversity within populations and provide insights into past bottlenecks and population expansions;

(5) help resolve mixtures of individuals originating from different breeding populations; and

(6) track products through the marketplace.

However, the most widespread practical application, particularly in the IWC context, is for the study of stock structure (genetic differentiation among populations). Before discussing details of particular genetic analyses, some of the key issues involved with assessing stock structure are

[1] *NOAA Fisheries, Northwest Fisheries Science Center, Seattle WA 98112, USA.*
[2] *Department of Biosciences, Durham University, Durham DH1 3LE, UK.*
[3] *Scottish Oceans Institute, East Sands, University of St. Andrews, KY16 8LB, UK.*
[4] *Institute of Biochemistry and Biology, University of Potsdam, D-14476 Potsdam, Germany.*
[5] *Groningen Institute for Evolutionary Life Sciences, University of Groningen, The Netherlands.*
[6] *California Academy of Sciences, San Francisco CA 94118, USA.*
[7] *British Antarctic Survey, Cambridge CB3 0ET, UK.*
[8] *Texas A&M University, College Station, TX 77843.*
[9] *NOAA Fisheries, Southwest Fisheries Science Center, La Jolla CA 92037, USA.*

summarised and best practice for assessing issues that may be complex or challenging are highlighted.

Two major issues related to stock structure/population differentiation

Two major issues that arise in applying genetic data to problems in stock structure are: (1) identification of threshold levels of population differentiation that require separate-stock management to achieve stated objectives; and (2) using genetic and other data to determine whether the system under consideration is above or below this threshold.

Identifying threshold levels of population differentiation

Population differentiation occurs along a continuum (Fig. 1). At one extreme (completely random mating), every individual has an equal probability of mating with any other individual. This situation is referred to as panmixia; although this is an idealised scenario not known to exist in any natural population, panmixia is typically adopted as the null hypothesis against which to compare alternative hypotheses that involve various degrees of departure from random mating. The other extreme of the population-differentiation continuum is characterised by complete isolation among locally panmictic groups of individuals. In nature, populations in general are neither completely panmictic nor completely isolated; instead, they typically are characterised by intermediate levels of differentiation and linked by restricted but non-zero levels of migration. In addition, the degree of connectivity among populations often changes over time.

Several types of data, including genetic information, can help to determine where a particular species falls on the population-differentiation continuum depicted in Fig. 1. In some cetacean species, the level of genetic differentiation among geographic areas within major ocean basins is relatively low (toward the panmictic end of the continuum), whereas evidence for higher population differentiation is often found among populations from different oceans (e.g. sperm whales; see Alexander *et al.,* 2016). In other cases, unexpectedly high levels of population differentiation are found within and among oceans (e.g. bottlenose dolphins, Natoli *et al.,* 2005).

Common questions that arise in conservation and management of living natural resources include the following: Is the differentiation among groups of individuals strong enough that they should be considered separate populations or stocks? Is any statistically significant (e.g. $P < 0.05$) departure from panmixia sufficient to warrant recognition as separate stocks? If not, how strong must the differentiation be? Unfortunately, there are no generally applicable answers to these questions, since the relevant degree of population differentiation depends on the conservation/management objectives, the risks associated with adopting different management strategies, and society's tolerance of the resulting consequences. Some general management objectives/considerations include:

(1) 'Management units' or 'populations' or 'stocks' must be considered separately because of a legal mandate. In the US, federal laws that include this type of mandate include the Endangered Species Act, the Marine Mammal Protection Act, and the Magnuson-Stevens Fishery Conservation and Management Act (see Waples *et al*. 2008 for discussion). Canada has similar provisions in its federal Species at Risk Act (2002), as do the Biodiversity Law of Costa Rica (1992), Australia's Endangered Species Protection Act (2002), and South Africa's National Environmental Management Biodiversity Act (2004).

(2) Sustainable harvest for management stocks should be maximised while preventing/minimising impacts on stocks that cannot withstand harvest, because:

   (a) locally depleted stocks might take a long time to rebuild, and/or

   (b) local extirpation might represent an irreversible loss of biodiversity.

Although such general considerations are useful for providing context, they are qualitative rather than

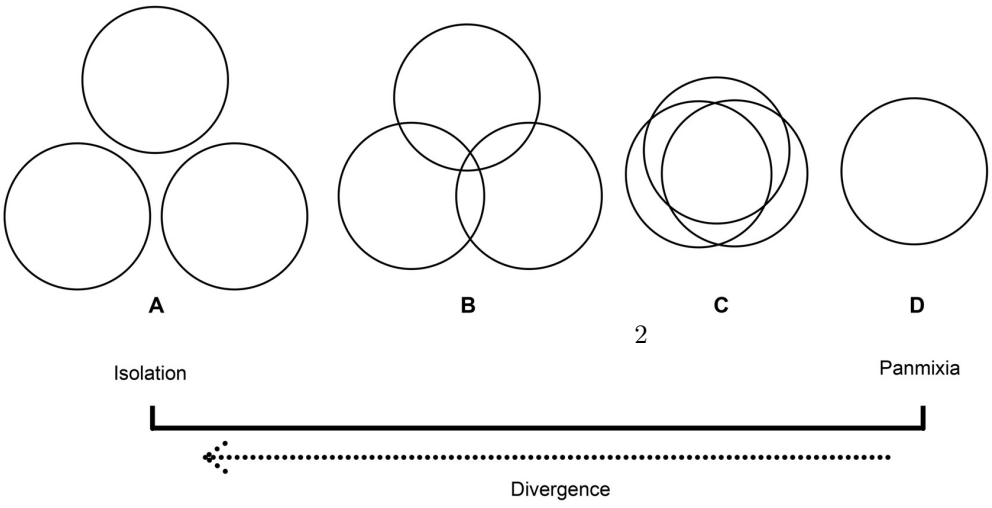

Fig. 1. The continuum of population differentiation. Each circle represents a group of individuals that might or might not be a separate population or stock. Four generic scenarios, with varying degrees of connectivity (geographical overlap and/or migration), are identified along the continuum: (A) Complete independence. (B) Modest connectivity. (C) Substantial connectivity. (D) Panmixia (circles are completely congruent). Reproduced from Waples and Gaggiotti (2006).

quantitative and by themselves will not produce repeatable outcomes. That is, you could not provide guidance this general to independent groups of scientists and expect them to produce similar results, even if they were analysing the same data. To make these objectives operational, it is necessary to identify thresholds on the continuum of population differentiation that do and do not make a difference with respect to the stated management objectives.

Most management problems related to stock structure involve population dynamics, and most stock-assessment models assume a single, closed population. If multiple stocks exist and they interact demographically to an appreciable extent, conclusions based on single-stock models can be misleading. Therefore, many of the stock-structure problems boil down to determining whether demographic linkages among populations are strong enough that any deviations from panmixia can be safely ignored. This has fostered interest in the concept of 'demographic independence,' which (under one common definition) occurs when immigration rates are low enough that population dynamic processes are determined primarily by local birth and death rates. According to this concept, units that are demographically independent require separate management because they strongly violate the standard assumption of a single panmictic unit.

Considering the general importance of this issue, it is surprising how little effort has been invested into defining the transition between populations that are demographically independent and those that are demographically linked. One study, conducted over two decades ago (Hastings, 1993), used a simple two-population model and found that with exchange rates greater than about 10% per generation, the two populations were demographically coupled (had correlated population trajectories), but at lower exchange rates the population trajectories were independent. McElhany *et al.* (2000) used this result to help define demographically-independent populations of Pacific salmon. Because they wanted to conduct separate viability analyses for each individual population, McElhany *et al.* (2000) refined the concept of demographic independence to mean that demographic linkages are small enough that they do not appreciably affect extinction risk over a 100-year time frame. However, demographic independence, like genetic differentiation, occurs along a continuum, and other concepts of demographic independence might lead to different quantitative criteria (see discussion in Waples *et al.,* 2008).

The IWC Scientific Committee has spent many years investigating this issue. In the IWC, the most common form of the stock-structure problem involves a situation where a species occurs in two or more geographic areas (areas A and B), but harvest occurs primarily or entirely in just one of the areas (area A; see Fig. 2). This corresponds to Stock-structure Archetype I as defined by TOSSM (Testing of Spatial Structure Methods; Martien *et al*., 2009; Lang and Martien, 2012). Assume for the moment that single-stock models show that, given assumptions about overall abundance and intrinsic growth rate, a harvest rate of *X*% per year is sustainable. The biological consequences of managing for this target harvest rate depend on what the true stock structure is. No problems are expected if the two areas actually are part of a single population or stock and harvest rate does not exceed *X*. If the stocks are partially isolated,

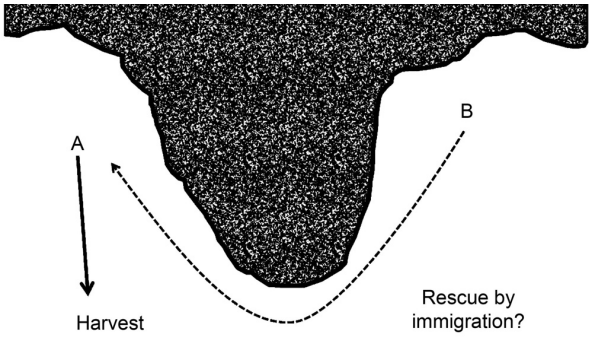

Fig. 2. Schematic representation of the generic harvest + stock structure problem. A species occurs in two space/time strata (A and B), but harvest is only taken in stratum A. If harvest rate is set based on total abundance of A+B, whether the harvest is sustainable will depend on whether immigration rates from B to A are high enough to rescue A from local depletion.

however, with regular, net immigration from area B into A at rate π per year, local harvest will be sustainable only if π is large enough to regularly offset removal of individuals from area A and hence demographically rescue that population. This would represent an example of source-sink population dynamics.

What is the tipping point – the migration rate below which demographic rescue is unlikely and management as separate stocks become necessary? For this simple two-stock problem, if we assume constant and symmetrical migration and equal population sizes, an approximate result is that the localised harvest becomes unsustainable when the net annual immigration rate drops below the annual harvest rate, less the local population's natural growth rate (see a worked example in fig. 1 of IWC, 2010). However, natural systems are seldom this simple, and the threshold can vary substantially depending on a variety of factors (e.g. relative sizes and productivities of the two stocks; variability in migration rates). Thus, although the condition where $\pi < X$ is a general warning sign that separate-stock management might be necessary, the threshold that applies to any given situation has to be evaluated on a case-by-case basis.

Largely for this reason, the IWC has resisted developing a single, quantitative threshold for differentiation that indicates separate stocks. This makes it difficult to apply the 'policy first, then science' approach advocated by Taylor and Dizon (1999), in which managers first identify a point on the stock-differentiation continuum that requires separate-stock management, and then scientists use genetic and other data to estimate how strong the stock differences are compared to this reference point. Instead, the IWC typically uses the following procedure:

(1) a range of 'plausible' stock-structure hypotheses (which specify the number of stocks and their distribution in space and time) is developed;

(2) these hypotheses are adjusted to ensure they are compatible with available information about abundance and stock composition in space and time, and hypotheses that are incompatible with the empirical data are rejected as implausible;

(3) projections are undertaken for various combinations of assumptions regarding the harvest regime, productivity, and exchange rates among populations; and

(4) results identify management regimes that perform adequately with respect to conservation goals.

Under this paradigm, there is no intrinsic interest in stock-structure *per se*; it only becomes important when failure to account for stock structure would lead to failure to achieve management objectives. In this guidelines document, we have made efforts to provide information relevant to this procedure. However, this might be considered a rather esoteric approach to the problem of stock structure, and even within the IWC there is considerable interest in a broader approach to stock definition that recognises its value in conservation and management. Therefore, this paper also considers the broader aspects of conservation/management problems associated with differentiation among populations. For example, an array of populations with diverse ecological and life history characteristics helps promote stability, resilience, and long-term sustainability, much as a diverse portfolio of investments reduces variability due to boom-or-bust cycles and produces more predictable returns; this has been termed the 'portfolio effect' (Schindler *et al.,* 2010).

Estimating rates of migration/levels of connectivity

Another major stock-structure problem that arises is how to convert estimates of levels of genetic differentiation into estimates of demographic connectivity, which (as discussed above) is generally what is needed by managers. This is challenging for several reasons. First, standard measures of genetic differentiation are most easily interpreted in terms of the product of migration rate per generation (*m)* and effective population size; the units of this combined term (mN_e) are the effective number of migrants per *generatio*n. In contrast, population dynamic processes depend primarily on the fraction of the population that is migrants each *year* (π). It is therefore often necessary to convert genetic estimates of the number of migrants per generation into estimates of the fraction of the population that migrates each year. This can in principle be done provided one can estimate N_e, N, and generation length, but this adds additional uncertainty to the final estimate.

A second major challenge is that the equilibrium relationship between genetic differentiation and migration is inverse and non-linear (Fig. 3). Even modest but constant rates of migration (a few individuals per generation) are sufficient to constrain levels of genetic divergence to low levels. Taking Hastings' (1993) result that the transition from demographic independence to demographic linkage occurs at approximately $m = 0.1$, and assuming that the population is moderately large (N_e ~1000), results in a tipping point at around $mN_e = 100$ migrants per generation. As seen in Fig. 3, values of F_{ST} that produce estimates that are above and below this value are all very small, which means that small errors in estimating levels of genetic differentiation can have a large effect on estimated levels of demographic connectivity. As a consequence, the ability of genetic data to inform the most common stock-structure problem for management is asymmetrical. If differentiation is strong (high F_{ST}), it generally will be easy to show that demographic linkages are small enough that separate stock management is required (left part of Fig. 3). However, if differentiation is low, it is very challenging, using genetic data alone, to determine whether migration is sufficient to allow demographic rescue of a population subject to local depletion. Unless a great deal of data is available, it will be difficult to demonstrate convincingly that migration is low enough that populations are demographically independent and the rescue effect is unlikely to occur, just as it will be difficult to demonstrate that migration is high enough to allow demographic rescue.

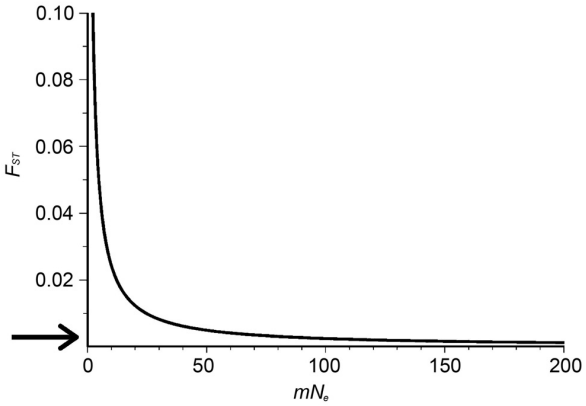

Fig. 3. Theoretical, inverse relationship between number of migrants per generation (mN_e) and F_{ST} (a measure of genetic differentiation), based on a common population genetics model. The 'tipping point' for levels of migration that do and do not make a difference for management often falls in the range indicated by the arrow, where small errors in correctly estimating F_{ST} have a large effect on the estimated level migration. For discussion of this issue, see Waples (1998), Palsbøll *et al.* (2007), and Waples *et al.* (2008).

Finally, the above discussions implicitly assume that migration rate is roughly constant for long enough for the system to reach a balance between gene flow (which promotes genetic similarity) and genetic drift (which increases genetic divergence). However, any given level of genetic differentiation (and any point on the continuum shown in Fig. 1) could also be produced by an isolation scenario that involves no current migration at all. Under an isolation model, F_{ST} is roughly proportional to the ratio of time since isolation (t, in generations) and N_e. This means that a small value of F_{ST}, which could be interpreted to imply high levels of equilibrium migration, could also be consistent with complete isolation for a few generations (if N_e is low) or perhaps many generations (if N_e is large).

Given these substantial challenges, perhaps the best approach for providing useful information related to stock structure to managers is to combine genetic and other data to characterise the full range of dispersal rates and isolation/migration models that are consistent with the empirical data. Some examples of how to do this can be found in Taylor and Martien (2004), Lang and Martien (2012), and van der Zee and Punt (2014). Armed with this information, managers can decide whether local depletions are likely and take appropriate actions.

To provide a sounder basis for making management decisions, we recommend the following be considered a high-priority research topic:

(1) review current and past IWC applications to identify 'tipping points' = levels of migration that do and do not make a difference for conservation outcomes under the

IWC - recent work on the maximum sustainable yield rate (MSYR) might provide some insights in this regard (IWC, 2014; 2015b);

(2) use simulations to determine under what circumstances it is likely to be feasible for genetic methods to distinguish between levels of migration that are above and below the tipping points (van der Zee and Punt, 2014 and van der Zee, 2014 are examples of these types of analyses); and

(3) once enough applications have been evaluated, it should be possible to determine whether these tipping points fall within a narrow range of connectivity/migration rates or are specific to each application – if the former, then it might be possible to empirically identify generic thresholds that are of practical relevance to IWC management.

Next we provide an outline of the in-depth material, followed by a discussion of some difficult IWC management problems for which genetic data can provide useful insights.

OUTLINE OF IN-DEPTH MATERIAL

Here we provide an outline of the material that is covered in more depth in Appendix 2. Appendix 1 provides a glossary.

(1) Species identification/delimitation

Issues related to alpha taxonomy come up consistently, especially regarding the boundary between populations and species of small cetaceans (so we expect some overlap with Section (4)). Because a standardised approach for DNA-based species identification of cetaceans already exists (Reeves *et al.,* 2004; see Baker *et al.,* 2003 and Ross *et al.,* 2003 for details about methodologies), this document will focus on analyses of intraspecific genetic diversity. *DNA Surveillance* and the comprehensive reference database, *Witness for the Whales*[10] does not delimit species but rather identifies specimens based on a reference database. Ross and Murugan (2006) presented results of a comparison of cetacean DNA sequences in *Witness for the Whales* and *GeneBank.*

(2) Analysis of diversity within populations

(a) Measures of genetic diversity, including rarefaction (controlling for sample size in estimating allelic richness).

(b) Information derived from tests of Hardy-Weinberg equilibrium (HWE).

(c) Information derived from tests of linkage disequilibrium (LD).

(3) Estimating population size and historical demography

(a) Census size, N
  (i) Genetic capture-mark-recapture of individuals
  (ii) Indirect capture-mark-recapture of individuals through the genetic identification of close relatives
  (iii) Identifying recent population bottlenecks

(b) Effective population size, N_e
  (i) Historical N_e
  (ii) Contemporary N_e

(c) The N_e/N ratio

(4) Analysis of diversity among populations (aka stock structure)

This is probably the most common type of management problem that utilises genetic data.

(a) Testing for heterogeneity
  (i) Putative populations defined *a priori*
  (ii) No *a priori* basis (or an uncertain basis) for grouping individuals into putative populations.
  In this case, the analyses are conducted on individuals rather than groups of individuals.
    • Standard clustering methods
    • Clustering based on ordination
    • Landscape genetics (units = individuals)
    • Analysis of close kin

(b) Describing population structure
  (i) Estimating degree of divergence
    • F_{ST} and related measures
    • Isolation by distance/landscape genetics (units = samples)

(c) Estimating migration
  (i) Methods that assume migration-drift equilibrium
  (ii) Isolation with migration models
  (iii) Methods that estimate contemporary migration
    • Assignment methods
    • Close-kin analyses

(d) Mixture analysis

(5) Generic/cross-cutting issues

Some issues will apply to many of the above analyses. Examples include:
(a) Choice of markers
(b) Ascertainment bias
(c) Multiple testing
(d) Mutation rates
(e) Sampling and experimental design
(f) Different approaches to statistical inference
(g) Monte Carlo issues
(h) Integrating genetic and non-genetic data
(i) Possible influence of selection
(j) Interpreting negative results

MANAGEMENT NIGHTMARES

This section highlights two examples that illustrate the complexity of stock-structure issues for cetaceans. For each example, the underlying biology and the key management questions are summarised, and then the major issues related to genetic data analysis are discussed, drawing on material discussed in the specific sections referenced in the Appendix.

North Pacific common minke whales

Background and management context

For a good general summary of this issue, see IWC (2013). The breeding areas for common minke whales in the western North Pacific have not been identified but are presumed to

[10] *http://www.cebl.auckland.ac.nz:9000/.*

5

occur to the south of Japan. Animals are observed and sometimes taken as they migrate along both coasts of Japan, as well as in oceanic waters farther to the east. The Okhotsk Sea is a primary feeding ground, but no common minke whales are harvested in this area. Convincing evidence (including differences in morphological and genetic traits and conception date) exists for the occurrence of at least two stocks, which have been termed O (more oceanic) and J (more coastal); however, rigorous characterisation of stocks is difficult because of the lack of samples from breeding areas. Existing population-genetic analyses are anchored by samples from areas thought to contain mostly pure J individuals or mostly pure O individuals. Some analyses also detected additional heterogeneity within either the J-like animals, the O-like animals, or both. One interpretation of these data is that this heterogeneity simply represents different mixture fractions of the same O and J stocks (Pastene *et al.,* 2012); in another view, this heterogeneity indicates the presence of 2–3 additional stocks, in addition to O and J (Wade and Baker, 2012).

Application of genetic data analysis guidelines

Several sections in the Appendix provide detailed treatment of issues that arise for North Pacific common minke whales. Examples include:

(1) The total number of individuals sampled is large (~2,500), which increases statistical power to detect heterogeneity that might not be biologically meaningful with respect to stock structure (Section (5)(j));

(2) Conversely, the lack of samples from breeding grounds raises questions about applicability of many standard methods (e.g. as described in Section (4)(a)(i)) that require *a priori* grouping of individuals into putative populations or stocks. Without those *a priori* groupings, it is necessary to use methods such as cluster analysis (Section (4)(a)(ii)) that generally require fairly strong genetic differences to produce robust results. This in turn tends to produce a substantial degree of uncertainty regarding underlying stock structure. This lack of power can in theory be overcome to some degree at least by substantially increasing the number of genetic markers (e.g. through development of single-nucleotide polymorphisms = SNPs), but by far the most robust and reliable method would be to obtain samples from individuals on or near their breeding grounds;

(3) These datasets and analyses provided ample opportunities to illustrate both the strengths and limitations of clustering methods like the program STRUCTURE (Section (4)(a)(ii)), which has been widely used both within and outside the IWC arena. They also provided an opportunity to illustrate the potential usefulness of a method based on principal components analysis (PCA; Section (4)(a)(ii));

(4) Some issues arose with respect to standardisation of analytical procedures from two different laboratories (Japan and Korea); the DNA Data Quality Guidelines document (IWC, 2009) discusses this issue;

(5) Most samples are believed to contain a mix of J-stock and O-stock individuals (and perhaps other stocks as well), which raises several issues regarding how to interpret results of tests of HWE and LD (Sections (2)(b) and (2)(c));

(6) Because most samples are believed to include mixtures, a variety of strategies were used to try to 'cleanse' the samples by removing individuals thought to belong to a particular stock. A number of caveats are associated with this type of approach;

(7) The samples were taken in a seasonal migratory corridor over a period of several decades, which also represents several generations for western North Pacific common minke whales. Therefore, it is important to consider the effects of temporal variation in genetic characteristics over time; and

(8) Several other types of data (e.g. conception date, morphology, identifying marks, etc.) are potentially relevant for stock structure of this species, so it is important to consider how best to integrate genetic and non-genetic data (Section (5)(h)).

Southern Hemisphere humpback whales

Background and management context

Southern Hemisphere humpbacks breed and calve at low latitudes during winter and feed in the Southern Ocean (SO) in summer (except for one small feeding ground in the Magellan Straits; Fig. 4). Southern Hemisphere humpback whale population assessments have been carried out with a focus on breeding ground abundance and the degree of recovery from past commercial whaling operations. Breeding grounds of Southern Ocean humpback whales are somewhat less genetically distinct than those in the Northern Hemisphere. Whaling data, discovery marks, and satellite telemetry data link Southern Hemisphere breeding regions to feeding grounds at a roughly similar longitude; however, feeding ground ranges are driven by both static (topographic) and dynamic (winter ice retreat rate) features of the Southern Ocean (Nicol *et al*., 2008), so the feeding ground destination of different SO humpback whale breeding stocks can vary greatly across SO longitudes between years and areas.

The central challenges for conducting Southern Hemisphere humpback assessments are:

(1) Choosing the appropriate spatial scale for assessment (e.g. stock or sub-stock). High population diversity and maternal fidelity to breeding grounds and probably also feeding grounds drives strong genetic sub-structure on both breeding grounds (e.g. Olavarria *et al*., 2007; Rosenbaum *et al*., 2009) and feeding grounds (Baker *et al.,* 2013). Genetic methods for measuring population composition are therefore central to stock delineation;

(2) To reach agreement on measures of breeding ground genetic differentiation that are most useful for management, given that all breeding grounds are still recovering from 20th century whaling and hence have not had time to achieve genetic/demographic equilibrium. Genetic methods for measuring population differentiation, in combination with population composition approaches (and possibly also multi-strata mark-recapture models to measure interchange), can be

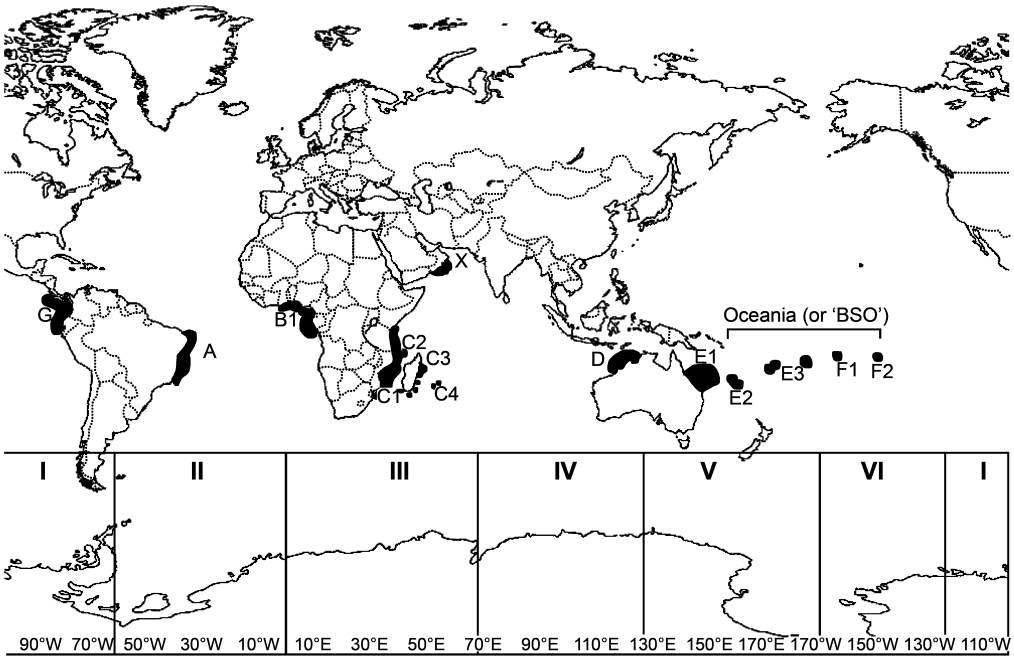

Fig. 4. Distribution of seasonal feeding areas off Antarctica (rectangles I–VI) and lower-latitude feeding areas (black shading) for Southern Hemisphere humpback whales. Letters A to G denote breeding/calving areas associated with genetically distinct breeding stocks (e.g. A, B) and sub-stocks (e.g. C2, C3).

used to determine the type of assessment model to apply (e.g. with and without interchange among stocks);

(3) Accounting for the temporal and spatial components of humpback use of breeding grounds in sample collection. Humpback whales are both temporally and spatially segregated by age and sex; there is evidence for age- and sex-stratified migratory routes (Chittleborough, 1965; Brown *et al.,* 1995), as well as different residence times on the breeding grounds for males and females. Careful survey design is therefore required to ensure that genetic samples collected from breeding grounds are fully representative of that ground; and

(4) To allocate whaling catches from high-latitude feeding grounds to their associated breeding sub-stocks. This requires quantifying the use of high-latitude feeding grounds by each breeding stock and is crucial for accurate assessment of population depletion and therefore recovery. The vast majority of whaling catches were taken on feeding grounds (Clapham and Baker, 2002), so population assessments can be very sensitive to how feeding ground catches are allocated to breeding grounds. Mixed stock analysis is therefore a useful tool for measuring feeding ground/breeding ground connectivity, though genetic samples currently available from high latitude feeding grounds are patchy in distribution and low in number at some longitudes (Pastene *et al.,* 2012). However, this approach also assumes that patterns of connectivity between feeding and breeding grounds have not substantially changed since the whaling period – an assumption that could be violated by shifts in oceanographic conditions or substantial range contractions following whaling.

Application of genetic data analysis guidelines

The challenges noted above raise a number of questions about stock structure of Southern Hemisphere humpback whales that are relevant to issues covered in the Appendix. For example:

(1) If one ignored geography and used only genetic and other relevant biological information, would the high-latitude catch allocation boundaries between breeding sub-stocks change? (see Section (4)(b)(i)).

(2) Do methods that assume long-term equilibrium conditions (see Section (4)(b)(2)) provide meaningful estimates of current levels of connectivity/migration?

(3) To what extent might differentiation measured from breeding ground samples be biased by limited genetic sampling (e.g. within one part of a large breeding ground, or one collection period within the season)? (see Section (5)(e) for experimental design considerations).

(4) Do current levels of differentiation between sub-stocks reflect long-term population structure, or have they been strongly affected by recent bottlenecks and fragmentation? (see Section (3)(a)(iii)).

SUMMARY AND CONCLUSIONS

The in-depth sections in the Appendix contain a great deal of detailed information that is, however, likely to be of interest primarily to specialists. Here we try to briefly capture some of the key conclusions and recommendations of practical relevance that emerge from joint consideration of the in-depth material and the more general conservation and management issues outlined in the Introduction.

Best practices

Although every empirical study of natural populations has its own idiosyncrasies, some core principles of genetic data should be followed in most or all studies. These include the following:

(1) Clarify goals and objectives and quantify expectations before the study begins, as that will help define the optimal experimental design;

(2) Follow appropriate data quality-control measures (see IWC, 2009) to determine that the data are reliable before conducting subsequent analyses and applying results to management questions;

(3) The first step in analysis of nuclear data should be to test for agreement with Hardy-Weinberg equilibrium (HWE). A table of F_{IS} values for each locus in each population is the most useful way to summarise this information. Significant departures in excess of those that can be attributed to chance should be examined for patterns related to specific loci or samples. Patterns of linkage disequilibrium should also be examined for evidence of linked pairs of markers;

(4) Other routine analyses include indices of genetic diversity, which can influence many downstream analyses;

(5) Statistical tests are a useful starting point, but by itself a significant *P* value provides no information other than that the null hypothesis can be rejected. Biological relevance of a statistical test depends on the magnitude of the effect size (e.g. genetic distance or migration rate);

(6) It is important to consider not only the point estimate of an effect size but the full range of values that are plausible given the empirical data;

(7) Absence of evidence is not evidence of absence. The strength of conclusions that can be drawn based on negative results is proportional to the rigorousness of a power analyses that determines how large an effect size could exist and still go undetected, given the amount of data collected;

(8) It is important to carefully examine implicit as well as explicit assumptions of the methods used and consider the consequences of (often inevitable) violations;

(9) Consider the distinction between scientific results, conclusions, and recommendations. Recommendations often involve consideration of normative factors such as societal values and risk tolerance and are most meaningful when expressed in the context of specific goals and objectives; and

(10) In the past, with at most a few dozen allozyme or microsatellite loci, it has been convenient to assume that all markers are independent (i.e. not genetically linked). However, as we move into the genomics era where tens of thousands of loci can be easily generated for non-model species, that assumption is no longer tenable, because the markers are necessarily situated on a small number of chromosomes. Effects of linkage will therefore be important to consider in genomics studies.

ACKNOWLEDGMENTS

This document was substantially improved by contributions from, comments by, and discussions with the following: Scott Baker, Mark Bravington, Doug Butterworth, Greg Donovan, Mike Double, Bill Perrin, Andre Punt, Howard Rosenbaum, and Hiroko Solvang.

REFERENCES

Alexander, A., Steel, D., Hoekzema, K., Mesnick, S.L., Engelhaupt, D., Kerr, I., Payne, R. and Baker, C.S. 2016. What influences worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? *Mol. Ecol.* 25:2,754–72.

Baker, C.S., Dalebout, M.L., Lavery, S. and Ross, H.A. 2003. DNA-surveillance: applied molecular taxonomy for species conservation and discovery. *Trends Ecol. Evol.* 18:271–72.

Baker, C.S., Steel, D., Calambokidis, J., Falcone, E.A., González-Peral, U., Barlow, J., Burdin, A.M., Clapham, P.J., Ford, J.K.B., Gabriele, C.M., González-Peral, U., Mattila, D., Rojas-Bracho, L., Straley, J.M., Taylor, B.L., Urbán-R., J., Wade, P., Weller, D., Witteveen, B.H. and M. Yamaguchi, 2013. Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. *Mar. Ecol. Prog. Ser.* 494:291–306.

Brown, M.R., Corkeron, P.J., Hale, P.T., Schultz, K.W. and Bryden, M.M.1995. Evidence for a sex segregated migration in the humpback whale (*Megaptera novaeangliae*). Proc. R. Soc. B 259:229–34.

Chittleborough, R.G. 1965. Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). *Aust. J. Mar. Fresh. Res.* 16:33–128. [DOI: 10.1071/MF9650033].

Clapham, P.J. and Baker, C.S. 2002. Modern whaling. pp.1,328–32 *In*: Perrin, W.F., Würsig, B., and Thewissen, J.G.M. (eds.) *Encyclopedia of Marine Mammals*. Academic Press, New York. i–xxxviii+1,414pp.

Hastings, A. 1993. Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. *Ecol.* 74:1,362–72.

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.)* 11:252–6.

International Whaling Commission. 2010. Report of the Scientific Committee. Annex I. Report of the Working Group on Stock Definition. *J. Cetacean Res. Manage. (Suppl. 2)* 11:252–7.

International Whaling Commission. 2013. Report of the Scientific Committee. Annex D1. Report of the Working Group on the *Implementation Review* for Western North Pacific Common Minke Whales. *J. Cetacean Res. Manage. (Suppl.)* 14:118–36.

International Whaling Commission. 2014. Report of the Fourth Intersessional Workshop on the Review of Maximum Sustainable Yield Rates (MSYR) in Baleen Whales, 26–28 March 2013, La Jolla, California, USA. *J. Cetacean Res. Manage.* (*Suppl.*) 15: 507–18.

International Whaling Commission. 2015a. Report of the Scientific Committee. Annex I. Report of the Working Group on Stock Definition. *J. Cetacean Res. Manage.* (*Suppl.)* 16:222–3.

International Whaling Commission. 2015b. Report of Sub-Committee on the Revised Management Procedure. *J. Cetacean Res. Manage*. (*Suppl*.) 15:87–111.

Lang, A.R. and Martien, K.K. 2012. Update on the use a simulation-based approach to evaluate plausible levels of recruitment into the Pacific Coast Feeding Group of gray whales. Paper SC/64/RMP4 presented to the IWC Scientific Committee, June 2012, Panama City, Panama (unpublished). 34pp. [Available from the Office of this Journal].

Martien, K.K., Gregovich, D., Bravington, M.V., Punt, A.E., Strand, A.E., Tallmon, D.A. and Taylor, B.L. 2009. TOSSM: an R package for assessing performance of genetic analytical methods in a management context. *Mol. Ecol. Res.* 9:1456–9.

McElhany, P., Ruckelshaus, M.H., Ford, M.J., Wainwright, T.C. and Bjorkstedt, E.P. 2000. Viable Salmonid Populations and the Recovery of Evolutionarily Significant Units. NOAA Technical Memorandum NMFS-NWFSC-42, US Department of Commerce. 157pp.

Natoli, A., Birkun, A., Aguilar, A., Lopez, A. and Hoelzel, A.R., 2005. Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*). *Proc. Roy. Soc. B* 272:1,217–26.

Nicol, S., Worby, A. and Leaper, R. 2008. Changes in the Antarctic sea ice ecosystem: potential effects on krill and baleen whales. *Mar. Fresh. Res.* 59:361–82.

Olavarria, C., Baker, C.S., Garrigue, C., Poole, M., Hauser, N., Caballero, S., Florez-Gonzalez, L., Brasseur, M., Bannister, J., Capella, J., Clapham, P., Dodemont, R., Donoghue, M., Jenner, C., Jenner, M.N., Moro, D., Oremus, M., Paton, D. and Russell, K. 2007. Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. *Mar. Ecol. Prog. Ser.* 330: 257–68.

Palsbøll, P.J., Berube, M. and Allendorf, F.W. 2007. Identification of management units using population genetic data. *Trends Ecol. Evol.* 22:11–16.

Pastene, L.A., Park, J.Y., An, Y.R. and Kanda, N. 2012. On the plausibility of stock structure hypotheses for western North Pacific common minke whales. Paper SC/64/NPM1 presented to the IWC Scientific Committee, June 2012, Panama City, Panama (unpublished). 9pp. [available from IWC Secretariat].

Reeves, R., Perrin, W., Taylor, B., Baker, S. and Mesnick, S. 2004 (eds.) Report of the workshop on shortcomings of cetacean taxonomy in relation to needs of conservation and management, April 30–May 2, 2004 La Jolla, California. NOAA Technical Memorandum NMFS-SWFSC-363, 94pp.

Rosenbaum, H.C., Pomilla, C., Mendez, M., Leslie, M.S., Best, P.B., Findlay, K.P., Minton, G., Ersts, P.J., Collins, T., Engel, M.H., Bonatto, S.L., Kotze, D.P.G.H., Meyer, M., Barendse, J., Thornton, M., Razafindrakoto, Y., Ngouessono, S., Vely, M. and Kiszka, J. 2009. Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. *PLoS ONE* 4(10): e7318. [doi: 10.1371/journal.pone.0007318].

Ross, H.A., Lento, G.M., Dalebout, M.L., Goode, M., Ewing, G., McLaren, P., Rodrigo, A.G., Lavery, S. and Baker, C.S. 2003. DNA surveillance: web-based molecular identification of whales, dolphins and porpoises. *J. Heredity* 94:111–114.

Ross, H.A. and Murugan, S. 2006. Using phylogenetic analyses and reference datasets to validate the species identities of cetacean sequences in GenBank. *Mol. Phylog. Evol.* 40:866–71.

Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A. and Webster, M.S. 2010. Population diversity and the portfolio effect in an exploited species. *Nature* 465:609–12.

Taylor, B.L. and Dizon, A.E. 1999. First policy then science: why a management unit based solely on genetic criteria cannot work. *Mol. Ecol.* 8 (Supplement S1): S11–S16.

Taylor, B.L. and Martien, K. 2004. Dispersal estimates for Western North Pacific minke whales. Report of the Second Workshop on North Pacific common minke whale Implementation Simulation Trials. SC/55/Rep3 Annex C. *J. Cetacean Res. Manage.* (*Suppl.*). 6:440–1.

Van der Zee, J.P. 2014. Assessing the performance of clustering methods under critical dispersal rates estimated from demographic simulations. Paper SC/A14/AWMP-RMP05 presented to the AWMP/RMP Joint Workshop on the stock structure of North Atlantic common minke whales, Greenland Representation, Copenhagen, 14–17 April 2014 (unpublished).

Van der Zee, J.P. and Punt, A.E. 2014. Use of RMP/AWMP-Lite to Evaluate Critical Dispersal Rates for the North Atlantic Minke Whales. Paper SC/A14/AWMP-RMP01 presented to the AWMP North Atlanntic minke whale workshop, Copenhagen, April 2014. (unpublished). 14pp. [Available from the Office of this Journal].

Wade, P.R. and Baker, C.S. 2012. High plausibility for Hypothesis III for stock structure of the western North Pacific minke. Paper SC/64/NPM11 presented to the IWC Scientific Committee, June 2012, Panama City, Panama (unpublished). 7pp. [Available from the Office of this Journal].

Waples, R.S. 1998. Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *J. Heredity* 89:438–50.

Waples, R.S. 2008. What is a population? pp.5–6 *In*: Eagle, T.C., Cadrin, S.X., Caldwell, M.E., Methot, R.D., and Nammack, M.F. (eds.) *Conservation Units of Managed Fish,Threatened or Endangered Species, and Marine Mammals.* Report of a workshop, February 14–16, 2006, Silver Spring, MD. U.S. Dept. of Commerce, NOAA Tech. Memo., NMFS-OPR-37, 100pp.

Waples, R.S. and Gaggiotti, O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* 15:1,419–39.

Waples, R.S., Punt, A.E. and Cope, J. 2008. Integrating genetic data into fisheries management: how can we do it better? *Fish. Fish.* 9:423–49.

Appendix 1

GLOSSARY OF TERMS

Words or phrases in italics have their own entries

adaptation: the process by which the frequency of *alleles* that enhance the survival and/or reproductive success (i.e. the *fitness*) of individuals in a given environment increases over time.

admixture: the result of interbreeding and *gene flow* between *genetically-differentiated populations*.

allele: one of two or more alternative forms of a gene or nucleotide sequence at a given *locus*.

allele frequency: the proportion of all *alleles* at a given *locus* that are of a specific type within the group being sampled.

allelic richness: a measure of the number of *alleles* per *locus* that uses *rarefaction* methods to minimise biases associated with unequal sample sizes.

allozyme: one of several variant forms of an enzyme coded by alternative *alleles* at a single genetic *locus*; variant forms typically differ in their charge and/or size and thus can be discriminated by gel electrophoresis.

ascertainment bias: a bias introduced by the use of an unrepresentative sample of individuals to identify *loci* for use in a *population* genetic study. For example, an ascertainment bias would be introduced if a *SNP* panel developed using samples from one *population* is then used to evaluate *genetic diversity* in another *population*.

assignment test: a statistical method using multi-*locus genotypes* to assign individuals to the *population* from which they most likely originated.

assortative mating: nonrandom mating, such that individuals prefer to mate with other individuals of a similar phenotype (**positive assortative mating**) or with other individuals that are of different phenotypes (**negative assortative mating**).

autosome: a chromosome that is not a sex chromosome.

balancing selection: a form of *natural selection* that acts to maintain *polymorphism* at a *locus* within a *population*.

bottleneck: a temporary reduction in *effective population size* that can result in a loss of *genetic diversity* due to the increased strength of *genetic drift*.

coalescence: the convergence of different *alleles* or lineages back through time to a common ancestral *allele*/lineage.

connectivity: the degree of exchange between two or more groups or *populations*. Connectivity can be *demographic*, in which case it relates to the degree to which *population* growth and vital rates are affected by *dispersal*; or genetic, in which case it refers to the exchange of genes (i.e. *gene flow*).

demographic: pertaining to processes that affect the size of a *population* (e.g. birth, death, *dispersal*).

diploid: having two sets of chromosomes. In sexually reproducing *populations*, one set is inherited from the mother and one from the father. At a given *diploid locus*, an individual can have two different *alleles* (*heterozygous*) or two identical *alleles* (*homozygous*). *Loci* with *autosomal* inheritance patterns are *diploid* (see *haploid*).

directional selection: a form of *natural selection* that consistently favors a particular phenotype or *allele*.

disruptive selection: a form of *natural selection* that selects for divergent values of a trait.

dispersal: movement of an individual away from its natal *population* and into another *population*. As used in this document, *dispersal* usually implies that the dispersing individual subsequently reproduces with members of the new *population*, resulting in *gene flow*; however, that is not always the case. In many species, but not cetaceans, passive *dispersal* of gametes or larvae is common.

effective population size (N_e): the size of an 'ideal' *population* that would experience the same rate of *genetic drift* as the *population* in question. In an ideal *population* (also called a *Wright-Fisher population*), generations are discrete, mating is random, and every individual has an equal probability of contributing genes to the next generation—in which case $N_e = N$. In most species, including cetaceans, N_e is typically smaller than the number of individuals in a *population* (see *population size*).

epistasis: the interaction between different genes, such that the expression of a given gene depends on the expression of one or more other genes.

fitness: a measure of the contribution of an individual, in terms of its *genotype* and/or *phenotype*, to the next generation's gene pool.

f**ixed:** in *population* genetics, this term is used to describe an *allele* that is found at a frequency of 100% within a *population*, such that no variation exists within the *population*.

***F_{IS}*:** a measure of whether the genotypic frequencies observed in a sample are compatible with those expected under *Hardy-Weinberg equilibrium*. Positive F_{IS} values indicate a deficiency of *heterozygotes* compared to *HWE*, while negative values indicate a deficiency of homozygotes compared to *HWE*.

***F_{ST}*:** a measure of the decrease in *heterozygosity*, relative to that expected under *random mating*, that occurs as a result of *population* structure. Low values of F_{ST} indicate that *allele frequencies* are similar among the groups being compared, while higher values indicate more *genetic differentiation* between groups.

gene flow: exchange of genes between *populations* or groups. *Gene flow* can result from an individual moving to a new *population*/group and successfully reproducing with members of that group, or through interbreeding between individuals of different *populations* or groups without any permanent movement of individuals (only gametes) between groups.

genetic distance: a measure of *genetic differentiation* between two groups.

genetic differentiation: the accumulation of genetic differences (*allele frequencies* or sequence *substitutions*) between groups. *Genetic differentiation* can occur due to limited *gene flow* as well as to *natural selection* on non-*neutral* genes in sympatric groups.

genetic divergence: the process of accumulating genetic changes (*mutations*) between two groups or lineages over time.

genetic diversity: genetic variation that occurs within individuals, within *populations*, and among *populations*.

genetic drift: random change in *allele frequencies* from one generation to the next. *Drift* is expected to have a greater effect as the *effective population size* of the *population* decreases.

genotype: the genetic makeup (*allelic* composition) of an individual, either of the entire genome or more commonly of a certain *locus* or set of *loci* (see *phenotype*).

haploid: having a single set of chromosomes, such that only a single copy of an *allele* or sequence exists at a given *locus*. In cetaceans, *mtDNA* is an example of a *haploid* marker, as it is inherited only from the mother. Sex-specific markers, such as Y-chromosome markers, also exhibit a *haploid* inheritance pattern (see *diploid*).

haplotype: the combination of *alleles* at *loci* that are found on a single chromosome or DNA molecule and thus tend to be inherited together. In cetaceans, *haplotype* typically refers to the *mitochondrial DNA* sequence held by an individual. Phased *nuclear alleles*, e.g. *SNP* variants physically located on the same chromosome, also constitute a *haplotype*.

Hardy-Weinberg equilibrium (HWE): an idealised state under which the genotypic frequencies in a *population* are simple products of *allele* frequencies. In theory, *HWE* is achieved in *randomly-mating populations* of infinite size that do not experience *migration*, *natural selection*, or *mutation*.

heterosis: a case when *hybrid* progeny have greater *fitness* than either of the parental organisms. Also referred to as *hybrid* vigor.

heterogeneity: the presence of multiple genetically or *demographically* distinct groups within a set of samples.

heterozygosity: a measure of the proportion of individuals in a group or *population* that carry two different *alleles* at a given *locus* (may also be averaged over *loci*). **Observed heterozygosity** (H_O) is the actual frequency of *heterozygote* individuals within a group, while **expected heterozygosity** (H_E, often referred to as *genetic diversity*) is the proportion of *heterozygotes* that would be expected in the group under *Hardy-Weinberg* expectations (e.g. *random mating*).

heterozygous: having two different *alleles* at a gene *locus* (e.g. Aa).

hitchhiking: a process that results in a change in the frequency of an *allele* at a selectively *neutral locus* due to physical *linkage* with an *allele* of another *locus* that is under *natural selection*.

homozygous: having two copies of the same *allele* at a gene *locus* (e.g. AA).

hybridisation: mating between individuals from two genetically distinct *populations* or species.

identity by descent: when two *alleles* are identical because they were both inherited from a common ancestor. In contrast, **identity by state** occurs when two *alleles* are identical for other reasons (e.g. convergent *mutation*).

inbreeding: mating between individuals that are more closely related than by chance alone. Inbreeding is expected to increase homozygosity because there is a greater probability that the *genotype* of an *inbred* individual will contain *alleles* that are *identical by descent* (inherited from a common ancestor).

inbreeding coefficient: the probability that two *alleles* found at a *locus* within an individual are *identical by descent* (i.e. were inherited from a common ancestor).

infinite sites mutation model (infinite alleles model): a model of evolution under which each *mutation* leads to a new *allele* in the *population,* and all *allele* types are equally different from each other (see *step-wise mutation model*).

introgression: incorporation of genes from one species or *population* into another through *hybridisation* and backcrossing.

intron: a region of non-coding sequence that is positioned between exons (coding regions) in a gene.

isolation by distance: a decrease in genetic similarity between pairs of individuals or *populations* as geographic distance increases.

landscape genetics: the study of the interaction between landscape or environmental features and *population* genetic parameters, such as *gene flow* and *genetic differentiation*.

lineage sorting: the process by which different genetic lineages within an ancestral taxon are lost by *genetic drift* or replaced by unique lineages evolving in different derived taxa. Incomplete lineage sorting is the persistence of an ancestral polymorphism through a speciation event, resulting in a shared lineage among different species.

linkage: a measure of the degree to which *alleles* of two *loci* do not assort independently. Two *loci* in close proximity on a chromosome have a higher probability of being inherited together than do two *loci* that are further apart and hence are said to be linked. Nonrandom associations of *alleles* at different *loci* can also occur by *natural selection, migration*, or *genetic drift* without physical *linkage.*

linkage equilibrium: the random association of *alleles* between *loci*. Also called gametic equilibrium.

linkage disequilibrium (LD): the nonrandom association of *alleles* between *loci,* often because the *loci* are located close together on the same fragment of DNA. Also known as gametic disequilibrium. Random LD also occurs in all *populations* due to *genetic drift*, with magnitude inversely proportional to *effective population size*.

locus (plural **loci**): a stretch of DNA at a particular place on a particular chromosome; often used to refer to a gene.

Mendelian inheritance: inheritance of traits or genes in accordance with the laws defined by Gregor Mendel, which include segregation of chromosomes, independent assortment and homologous exchange.

microsatellite: a genetic marker comprised of short DNA sequence units that are repeated multiple times (e.g. ATATATATAT). Although microsatellites can be found on sex chromosomes and in *mitochondrial DNA*, use of this term in cetacean *population* genetics typically refers to *loci* that are bi-parentally inherited and of *nuclear* origin. Microsatellite *alleles* are usually labeled according to the number of repeated units (and thus the size) contained in a given *allele*, as opposed to being directly sequenced.

migration: this term is commonly used in two different ways, to refer to: a) seasonal movements between two geographical areas that are related to the population's reproductive cycle, changes in their physical environment (e.g. ice formation), and/or prey availability; and b) movement of individuals between groups or *populations*, which might or might not result in successful reproduction and *gene flow.* Unless otherwise specified, as used in this document '*migration*' implies both movement between *populations* and *gene flow*.

mitochondrial DNA (mtDNA): a small, circular DNA molecule (in animals ~16–20 kbp long) found in the mitochondria (i.e. outside of the nucleus) of a cell. In cetaceans, mtDNA is inherited only from the mother and is thus an example of *haploid* inheritance.

mixture: a group of individuals that are aggregated within a common area but which are from more than one different breeding *population*. In cetaceans, a *mixture* of individuals often occurs when animals, such as humpback whales, from different breeding *populations* migrate to a common feeding area.

monomorphic: having only a single *allele.*

monophyletic: term used to describe a group of taxa or lineages that includes a common ancestor and all of its descendants (see *paraphyletic*).

mutation: a change to the genetic material of a cell. *Mutations* can include single nucleotide changes, deletions, and insertions, as well as duplications, losses, inversions, and translocations of segments of DNA sequence.

mutation-drift equilibrium: within a *population*, a state at which the loss in genetic variation due to *genetic drift* is balanced by an increase due to *mutation*.

natural selection: differential contribution of *genotypes* to the next generation due to differences in survival and/or reproduction.

nuclear DNA (nDNA): DNA found in a cell's nucleus. In cetaceans, *autosomal nuclear DNA* is bi-parentally inherited, such that an individual's *genotype* at a given *locus* contains one *allele* inherited from its mother and another *allele* inherited from its father. *Nuclear DNA* also includes DNA found on sex chromosomes.

nucleotide diversity: a measure of genetic variation calculated from DNA sequence data, which measures the average proportion of differences between all DNA sequences (i.e. the average difference between two randomly taken sequences) in a group.

neutral: not influenced by *natural selection.*

null allele: an *allele* that fails to amplify to detectable levels during PCR analysis.

numt: a segment of *mtDNA* sequence that has been transposed into the *nuclear* genome; numts are typically not expressed and hence constitute *pseudogenes*.

outbreeding: mating between unrelated individuals that is more frequent than would be expected on the basis of chance alone. *Outbreeding* can occur due to negative *assortative mating*.

overdominance: a condition in which *heterozygotes* have higher *fitness* than homozygotes.

panmixia: a state that occurs when all individuals within a *population* are mating randomly with each other. When used as an adjective, the term is **panmictic**.

paraphyletic: a term used to describe a group of taxa/organisms that share a common ancestor and some but not all of its descendants (see *monophyletic*).

phenotype: the observable or detectable characteristics of an individual, including physical and physiological traits, which are determined (to varying extents) by the individual's genotype and/or by environmental factors (see *genotype*).

phylogenetic: a term used to describe evolutionary relationships among taxa.

phylogeography: the study of how the genetic lineages of a taxon are distributed across the landscape, in order to better understand its evolutionary history (its origin and spread).

polymorphic: having more than one *allele* at a *locus*. This term is typically used to refer to a group/*population* rather than to an individual, which is considered to be *heterozygous* if more than one *allele* is present.

polyphyletic: a term used to describe a group of taxa or lineages that do not all share the same most recent common ancestor This scenario can occur if taxa or lineages are grouped together based on a shared trait that has evolved independently in multiple taxa in response to environmental or other *adaptation*.

population: a group of individuals that co-occur in space and time and freely interbreed. Terms that are often used synonymously with 'population' include 'subpopulation' and 'stock,' although the latter can also refer to units of management convenience that do not imply interbreeding

population size (*N*): the number of individuals in a *population,* often denoted as the census size (N_c). Commonly used to refer either to all individuals or only adults (see *effective population size*).

positive selection: *natural selection* for an *allele* that increases *fitness*.

probability of identity (*I*): the probability that two unrelated (randomly sampled) individuals would have an identical *genotype*. This probability can also be calculated assuming that full siblings are available to be sampled.

pseudo-overdominance: an increase in *fitness* of *heterozygotes* at a *neutral locus* due to *linkage disequilibrium* with another *locus* that is under *natural selection*.

pseudogene: a nonfunctional member of a gene family that has been derived from a functional gene.

random mating: mating in which the probability that any two individuals in a group will mate is the same for all possible pairs of individuals.

rarefaction: a method by which sample sizes are randomly reduced to the size of the smallest sample using simulations.

reciprocally monophyletic: a term used to describe two groups that are both *monophyletic* taxa or lineages and share a more recent common ancestor than either shares with any other taxa or lineage.

recombination: in sexually reproducing organisms, a process by which genetic material is exchanged between maternal and paternal chromosomes during meiosis, resulting in offspring that may have different combinations of genes from their parents. Genes that are located further apart on the same chromosome have a greater likelihood of undergoing recombination.

relatedness: a measure of the proportion of genes that are *identical by descent* (i.e. derived from a common ancestor) among two individuals.

single nucleotide polymorphism (SNP)**:** DNA sequence variation that occurs when a single nucleotide (A, T, C, or G) differs at a specific site among individuals or within an individual (for *diploid* markers).

step-wise mutation model: a model of *microsatellite* evolution which assumes that *mutations* result in the gain or loss of a single repeat unit, such that *alleles* that are more similar in size are more closely related (see *infinite alleles model*).

substitution: a *mutation* that results in the replacement, within a *population*, of one nucleotide with another nucleotide. In contrast to the spontaneous *mutation* rate, which reflects the probability of a *mutation* occurring between a parent and its offspring, the *substitution* rate only measures those *mutations* that accumulate over time in a surviving lineage.

underdominance: *natural selection* against *heterozygotes*.

Wahlund effect: a type of *heterozygote* deficiency that occurs when two or more *genetically differentiated* groups are sampled and included as part of the same strata in analyses.

Wright-Fisher model: an idealised, discrete-generation model of reproduction in which all individuals contribute equally to a very large gamete pool, which unites at random to form individuals in the next generation.

Appendix 2

(1) SPECIES IDENTIFICATION/DELIMITATION

There has been a long-standing debate about species concepts, some of which would lead to differing classifications with obvious complications for the objective of species delimitation (see de Queiroz 2007). The phylogenetic species concept (Eldredge and Cracraft, 1980) groups organisms according to a shared evolutionary history and has been promoted recently by the interest in DNA barcoding (see Monoghan *et al.,* 2009). Eldredge and Cracraft (1980) proposed a 'phylogenetic' species to be 'the smallest group of individuals that share a common ancestor and can be distinguished from other such groups'. Confidence in the assignment of a phylogenetic species can be provided through congruent reciprocal monophyly (where all specimens defined as that species fall into the same lineage) among phylogenies constructed from different genes (see Balloux, 2010). Exceptions to the strict monophyly rule can occur, for example when a population colonising a new geographic area evolves into a separate monophyletic species, rendering the ancestral range of the species paraphyletic (where some but not all members of a lineage share a common ancestor within that lineage). Other species concepts (of which there are many, see Mayden, 1997) are often more difficult to test (such as the 'biological species concept' which focuses on reproductive isolation; see Mayr 1942). Efforts to define a 'unified' species concept have included ideas associated with relative fitness (most consistent with the biological species concept; Hausdorf, 2011) and phylogenetic lineages (de Queiroz 2007). There has been a tendency for conservation biologists to favor the phylogenetic approach, as it typically identifies a greater number of species – Agapow *et al.* (2004) conducted a meta-analysis focused on this issue and discussed the implications for conservation management. Here, given the nature of the available data and the key objectives (effective conservation and management), we will focus on phylogenetic concepts of species delimitation.

Cetacean taxonomy is difficult, in part due to a recent radiation of diversity, especially among Odontocete taxa, but also due to the lack of good quality type specimens in many cases. A good example is that of the beaked whales, a group in which some named species were identified on the basis of just a few damaged bones, and later phylogenetic assessments based on both nuclear and mtDNA genes have led to some significant revisions (Dalebout *et al.,* 2004, 2008). There has also been incongruence between morphological taxonomic and molecular phylogenetic assessments of species delimitation for several groups of delphinid taxa, especially in the genera *Tursiops*, *Delphinus*, *Stenella,* and *Lagenorhynchus*. Both *Tursiops* and *Delphinus* have been subdivided into up to 20 named species at one point, then collapsed back down to one, only to be divided again (but this time to 2–4 species). It is clear that *Tursiops* in particular can be divided into multiple lineages, and several of these are now recognised as separate named species, with further resolution and clarification likely necessary (e.g. Wang *et al.,* 1999, Natoli *et al.,* 2004, Charlton-Robb *et al.* 2011; Moura *et al.* 2013a). There is a consistent pattern of differentiation among habitat or resource specialists, such that some nearshore and offshore populations of *Tursiops* (Hoelzel *et al.,* 1998a) and *Delphinus* spp. (Natoli *et al.,* 2006) form reciprocally monophyletic lineages (though not all), as do resource specialists in the genus *Orcinus* (see Hoelzel *et al.,* 2002, Morin *et al.,* 2010). However, in *Orcinus*, mtDNA phylogenies show reciprocal monophyly among matrifocal populations, while nuclear markers suggest ongoing male-mediated gene flow (Hoelzel *et al.,* 2007, Pilot *et al.,* 2010) and a genomic phylogeny based on 1.7MB nuclear DNA showed a different topology compared to the mtDNA tree (Moura *et al.*, 2015). This emphasises the importance of using multiple markers for phylogenetic studies at and below the species level. Various stochastic processes can lead to random differences among single gene trees, so several molecular markers are needed to characterise the genetic structure of species.

The analysis of phylogenetic lineages for the identification of robust monophyletic groups is useful from a conservation and management point of view regardless of the eventual taxonomic rank assigned, because it serves to identify cohesive and distinguishable units. It is expected that the conservation of these defined units of diversity will facilitate the potential for species to respond through natural selection to a changing environment, and this has become a core objective in conservation genetics. This approach is reflected in policy in the US, where 'distinct population segments' representing genetic populations that are discrete from other conspecific populations and represent a significant component of the overall diversity can be protected under the Endangered Species Act (ESA) (Waples, 2006). The killer whale (*Orcinus orca*) provides a case in point, where a regional population in decline (the 'southern resident' community in Puget Sound, WA) was protected under this provision of the ESA. In fact, in spite of some evidence for ongoing gene flow, killer whale populations show a level of differentiation that merits regional protection (including the recognition of sympatric populations of ecotypes, see Hoelzel *et al.,* 1998b), illustrating the importance of a mechanism to protect diversity prior to official recognition of named taxonomic units (e.g. species and subspecies).

Given a mechanism to protect diversity at the population level, a more important application of species delimitation studies is the identification of cryptic species. Specifically, this means the recognition of relatively deep and monophyletic lineages that have not been previously recognised or named. Recent cetacean examples have included populations that are morphologically similar to named species. For example, the specimens eventually named *Balaenoptera omurai* were collected in Indo-Pacific waters near the Solomon Islands in the 1970s. They were initially recognised as distinct from other species in this genus based on allozyme studies (Wada *et al.,* 1991) and classified as a small form of the Bryde's whale (*Balaenoptera brydei*). Later investigations more comprehensively assessed their morphology (Wada *et al.,* 2003) and molecular phylogeny (based on mitochondrial and

nuclear DNA sequence data; Sasaki *et al.* 2006) and named them *B. omurai*, proposing a further classification for another small form of the Bryde's whale as *B. edeni* (Wada *et al.,* 2003, Sasaki *et al.,* 2006). Based on a morphological study, Wada *et al.* (2003) described *B. omurai* as actually most similar to fin whales (*Balaenoptera physalus*), but smaller. Although the full classification of this group of whales remains uncertain and controversial, the lineage representing *B. omurai* is clearly divergent from other named species in the genus and warrants independent conservation status.

The analytical methodology for species delimitation is based on finding a robust representation of evolutionary histories, typically using a phylogenetic reconstruction approach. A phylogeny is essentially a hypothesis about the pattern of evolutionary histories among taxonomic units, and its accuracy is dependent on the quality and quantity of the data, together with an appropriate choice of evolutionary model and set of taxonomic units to compare. For a given set of taxonomic units there is an exponentially rising number of possible trees as the number of taxa increases. The objective is to find the best tree for a given method, and to then compare among methods and among markers (preferably including both morphological and multiple genetic markers). There are two primary approaches to phylogenetic tree construction, one based on similarities (e.g. DNA sequence similarity resulting in a 'phenetic' tree), and optimality methods, based on some criteria (such as finding the shortest tree using the 'maximum parsimony' method, or the most probable, 'maximum likelihood' tree). For a review of the various methodologies see Lemey *et al.* (2009). Recent multi-locus phylogenies of delphinid species have supported some traditional taxonomic classifications, while raising questions about others, even at the genus and family level (e.g. Harlin-Cognato *et al.,* 2006, Steeman *et al.,* 2009). These studies have made clear the necessity for further work to help establish appropriate units of conservation for whales and dolphins.

In the context of cetacean management issues, phylogenetic reconstruction methods help to establish boundaries among management units that merit conservation, and in the case of cryptic species, to recognise diversity that was previously unmanaged. Once taxonomic units are established, the same general methodology can be used to facilitate the identification of species from specimens that cannot easily be recognised by morphology (for example, processed materials in commercial trade). The process of identifying species using one or a few loci and phylogenetic methods has been termed 'DNA barcoding' and has found increasing applications in conservation biology (see Kress *et al*., 2015).

(2) ANALYSIS OF DIVERSITY WITHIN POPULATIONS

(a) Measures of genetic diversity

A common estimator of genetic variation within a population is average heterozygosity (H), which is the fraction of individuals that are heterozygotes (and have two different alleles at a gene locus). Individuals with two of the same alleles are homozygotes (frequency 1–H). Typically, H reflects an average across a number of gene loci. Just enumerating the heterozygote individuals yields the observed heterozygosity, H_o. One can also calculate the expected heterozygosity (H_e), which is the fraction of individuals expected to be heterozygous under the assumption of random mating, as

$$H_e = 1 - \Sigma\, p_i^2,$$

where p_i is the frequency of the i^{th} allele at a locus. Conceptually identical to expected heterozygosity is gene diversity (often used for mtDNA and then called haplotype diversity, h), which is the likelihood that the alleles of two individuals randomly taken from the population are identical. Both expected and observed heterozygosity and genotype diversities range from 0 (all genotypes homozygous and identical) to 1 (all genotypes heterozygous and different). Estimates of observed and expected heterozygosity based on finite samples of size n individuals are biased, and unbiased estimates can be obtained by multiplying the empirical values by the factor $n/(n–1)$ (Nei 1978).

At polymorphic microsatellite loci, many different alleles often occur, which produce high heterozygosities. The number of alleles N_A found in a sample is an additional measure of genetic diversity. N_A is not an unbiased measure, as it scales with sample size (up to a certain point, which can be estimated by rarefaction). To overcome this problem, allelic richness A_R is calculated through rarefaction: sample sizes are randomly reduced by Monte Carlo simulation to the size of the smallest sample (e.g. Thompson *et al.* 2016). A related measure is the effective number of alleles, $\tilde{n}$, calculated according to

$$\tilde{n} = 1/(1 - H_e)$$

Both A_R and $\tilde{n}$ provide estimators of allele diversity not biased by sample size (in the case of $\tilde{n}$ provided that H_e has been corrected for sample size; see above).

The measures above are primarily used to characterise levels of diversity in contemporary populations. Another measure of genetic diversity, θ (theta), is widely used to draw inferences about historical demographic processes, based on the theoretical expectation that, at mutation-drift equilibrium, theta is expected to be a simple function of the mutation rate (μ) and effective population size (N_e):

$$\theta = 4N_e\mu$$

See Section (3)(b)(i) for more information about this relationship. Usually neither N_e nor μ are known for natural populations. A variety of methods are used to estimate θ from genetic data. For practical reasons, we only cover a few common approaches here: For sequence data, the most common approach to estimate θ utilises the fact that (assuming an infinite site mutation model) $\theta = \pi$, where π is the nucleotide diversity. π is calculated according to

$$\pi = \Sigma\Sigma p_i p_j \pi_{ij}$$

where p_i and p_j are relative frequencies of the alleles i and j in the population and π_{ij} is the proportion of polymorphic nucleotides in pairwise comparison among i and j.

For microsatellites, a strictly stepwise mutation model is typically assumed, which implies that alleles with a similar number of tandem repeats are more likely to be identical by descent (IBD; Slatkin 1995). Under this mutation model, the expectation of θ is

$$E(\theta_v) = (2/(n-1))\sum_{i=1}^{n}(a_i - \bar{a})^2$$

where n is the number of investigated gene copies (for diploid loci, twice the sample size), a_i is the repeat number of the i^{th} allele, and $\bar{a}$ is the average repeat number. This is equivalent to

$$\theta = 2\sigma^2 ,$$

where σ^2 is the variance in repeat number among all sampled gene copies.

(b) Information derived from tests of Hardy-Weinberg equilibrium

An initial step in analysis of most population genetic datasets is to test for agreement between observed genotypic frequencies and those expected according to the Hardy-Weinberg principle. Often this is referred to as testing for Hardy-Weinberg Equilibrium (HWE; reviewed by Waples 2015), but HWE does not represent a true equilibrium; instead, it describes the expected relationship between allele frequencies and genotypic frequencies in a randomly-mating population that is not experiencing any evolutionary forces (drift, migration, selection, mutation). Under these conditions, the frequencies of genotypes are simple functions of allele frequencies. For example, if the population frequency of allele 'a' is p, under HWE the frequency of the 'aa' homogzygote genotype will be p^2. Statistical tests of HWE that account for effects of sampling error can help to identify samples or genetic markers for which a) one or more of the above assumptions are not met, or b) sampling or laboratory artifacts have affected the data.

Agreement with HWE can be measured by Wright's inbreeding coefficient, F_{IS}:

$$F_{IS} = (H_e - H_O)/H_e$$

Positive values of F_{IS} indicate deficiencies of heterozygote individuals compared to expected HWE proportions, and negative values indicate excesses. As discussed below and by Waples (2015), deviations from HWE can be caused by laboratory artifacts and/or biological processes that depart from HW assumptions. Here we assume that HW evaluations have been conducted as part of the DNA data quality control step. If a sample consistently deviates from HWE-expectations at multiple loci, this can be indicative of population-level processes such as migration or interbreeding (see Table A1).

Heterozygote deficiency (positive F_{IS}): A significant deficiency of heterozygous individuals can arise from true or apparent null alleles, which in turn can be related to sample quality or laboratory procedures (see Tiedemann *et al.,* 2012). These effects are generally locus-specific. A deficit could also occur through non-random sampling, if heterozygotes are less likely to be sampled because of some aspect of their phenotype. Some biological causes of heterozygote deficiency are:

- Undetected population structure. What is assumed to be a sample from a single population might include individuals from more than one gene pool. This could occur because of seasonal mixing (e.g. sampling different breeding stocks on a common feeding ground) or because of incorrect population definition (either spatially – the sample covers the range of more than one population, or temporally – lumping individuals from different seasons or years). This type of heterozygote deficiency is known as the Wahlund effect, and it should produce a positive correlation between F_{IS} and F_{ST} across all loci (Waples 2011, 2015; see also discussion of null alleles in De Meeûs, 2018 and Waples, 2018).
- Disruptive selection/local adaptation at or near the evaluated marker. Marker loci or loci closely linked to them might be affected by selection. While this might be of interest in itself in some contexts, it can compromise the validity of population genetic estimators that assume selective neutrality. However, even strong directional selection does not necessarily lead to departures from HWE (Lewontin and Cockerham 1959; Waples 2015). This effect, when it occurs, should be locus-specific.
- Inbreeding due to positive assortative mating. If individuals tend to mate with close relatives, or with individuals with similar phenotypes to theirs, the result can be a deficiency of heterozygotes. This should affect loci across the genome.

Heterozygote excess (negative F_{IS}): Data quality problems in the laboratory can produce heterozygote excesses (see Koehn, 1972 for an extreme example), as can sampling that favors collection of heterozygotes. Possible biological causes include the following:

- Recent population admixture. If two genetically dissimilar populations interbreed, the result will be an excess of heterozygotes in the offspring. This would affect many loci across the genome, but it would be transient, as it takes only a single generation of random mating to re-establish HWE.
- Selection favoring heterozygotes (balancing selection). An excess of heterozygous individuals can indicate a selective advantage of the heterozygous state. A classic example of heterozygote advantage involves the locus for sickle-cell anemia in humans, where the heterozygote form has higher resistance to malaria. Heterozygote excesses that have been observed at immune system loci could also be due to selection favoring heterozygotes (Hedrick 2002). In addition to these locus-specific effects, selection could also favor individuals that generally have higher heterozygosity across the genome, in which case heterozygote excesses might be found at many gene loci.
- Outbreeding. This occurs when mates are consistently less related than expected at random. Hedrick *et al.* (2016) described an example involving wolves from Yellowstone National Park (USA), where more matings than expected by chance occurred between individuals with different coat colors. Another possible cause of outbreeding is gender-specific dispersal, when one sex (in mammals often the female) disperses much less than the other. Outbreeding would affect loci across the genome.
- A small number of parents. This is the basis for the heterozygote excess method of estimating effective population size (Pudovkin *et al.,* 1996); when it occurs, it should affect loci across the genome.

Note that failure to reject HWE does not necessarily imply statistical support for HWE, as not getting a significant result can be caused by low statistical power. Fairbairn and Roff (1980) showed that the power to reject HWE generally is relatively low, even for large sample sizes. Although statistical tests of HWE are useful, more important than the resulting P value are the magnitudes of any departures and what they mean for (a) the biology of the species, and (b) any downstream genetic analyses that assume HWE. Both of these latter topics are in need of further research (discussed in Waples, 2015).

(c) Information derived from tests of linkage disequilibrium (LD)

Linkage disequilibrium is the non-random association of alleles at different gene loci. If two loci are physically 'linked' because they are close together on the same chromosome, they do not assort independently, which leads to LD. When only a handful of allozyme or microsatellite markers were available, it was convenient to ignore LD and assume the markers were unlinked. This was probably not unreasonable, given two facts: 1) the probability that any random pair of loci are on different chromosomes (and hence not physically linked) is (c–1)/c, where c is the number of chromosomes, and 2) even if two loci are on the same chromosome, they could be far enough apart that they are nearly independent. Currently, however, it is possible to generate many thousands of SNP loci for any species. Increasing the number of loci does not increase the proportion of pairs that are linked, but it does ensure that at least some pairs of loci will be physically linked (Waples *et al.,* 2016). Linkage can affect both bias and precision of population genetic analyses, and the consequences of this are still being evaluated; some analyses are much more sensitive to linkage than others. If detailed information is available about the arrangement of genes on the chromosomes, one locus of a linked pair could be removed so the analysis focused only on unlinked markers.

LD can also arise from a variety of processes even when markers are not physically linked (see Table A1), including:

- A small number of parents. This is the basis for the LD method of estimating effective population size (see Section (3)(b)(ii)); this type of LD should affect loci across the genome.
- A mixture of individuals from more than one population. Mixture LD is the two-locus analogue of heterozygote deficiency (Wahlund effect) caused by population mixture. This type of LD should produce a positive correlation between F_{IS} and F_{ST} across all loci.
- Admixture (interbreeding of individuals from different populations). Unlike HWE, which is restored with a single generation of random mating, LD at unlinked loci only decays by 50% each generation. The rate of recombination and decay of LD is much lower for physically linked markers. This means that genetic signatures of past interbreeding events can potentially be detected many generations later.
- Natural selection. Genes rarely work in isolation; their effects depend on the genetic background of the entire genome. Effects of one gene on another are called epistasis. Epistatic interactions of genes on different chromosomes can cause LD, but these associations are difficult to maintain because half of them decay by recombination every generation.
- Hitchhiking. Loci that are tightly linked to a new advantageous mutation can rapidly increase in frequency through a process called hitchhiking (Barton, 2000). Hitchhiking effects are important to consider in evaluating results of methods to detect natural selection.

Table A1

The most common causes of linkage disequilibrium (LD) and departures from Hardy-Weinberg proportions. Reproduced from Waples (2015).

Observation/possible cause	Locus[a] specific?	Sample specific?	Comments
Positive F_{IS}			
Positive assortative mating	Yes	Perhaps	Effect is expected only if phenotype is correlated with genotype
Self fertilisation	No	Yes	–
Wahlund effect	Yes[b]	Yes	[b] F_{IS} should be positively correlated with F_{ST}
True null alleles	Yes	No	–
Apparent null alleles	Yes	Perhaps[c]	[c]Could depend on sample quality
Non-random sampling	Perhaps	Perhaps	Expected if heterozygotes are less likely to be sampled
Underdominance	Yes	Perhaps	Selection against heterozygotes
Negative F_{IS}			
Negative assortative mating	Yes	Perhaps	Effect is expected only if phenotype is correlated with genotype
Non-random sampling	Perhaps	Perhaps	Expected if heterozygotes are more likely to be sampled
Overdominance	Perhaps[d]	Perhaps	Selection favors heterozygotes
Selection differs in M and F	Yes	Yes	Allele frequency differences between sexes cause het excess
Sex linkage	Yes	Perhaps[e]	Allele frequency differences between sexes cause het excess
Small N_e	No	Yes	–
LD; r^2 significantly > 0			
Small N_e	No	Yes	–
Wahlund effect	Yes[f]	Yes	[f] r^2 should be positively correlated with $F_{ST(1)}$*$F_{ST(2)}$
Epistasis	Yes	Perhaps	A wide range of patterns is possible
Hitchhiking	Yes	Perhaps	–

[a]Recognising that random variation will occur among loci, even if 'No' is indicated in this column. [b]See 'Comments' column. [c]See 'Comments' column. [d]If heterozygote advantage is due to general heterosis, locus-specific effects are not expected. [e]See Marshall *et al.* (2004) for an example of sample-specific departures from HWP due to sex linkage. [f]See 'Comments' column.

(3) ESTIMATING POPULATION SIZE

Estimating the number of whales found in a specific geographic area (whether these are termed 'stocks', 'populations', or 'sub-populations', etc.) is a critical component of IWC considerations, and such estimates can be obtained from sighting surveys as well as through use of genetic analyses. In addition to 'census' population size (typically denoted by N or N_c; i.e. the number of individuals), population and conservation genetic inferences are also concerned with the 'effective' population size (N_e).

(a) Census size, N

Census population size can be estimated in two ways using genetic data; by capture-mark-recapture methods and rarefication curves.

(i) Genetic capture-mark-recapture of individuals

If the genotype for a sufficient number of variable genetic markers is available, then each individual will end up with a unique genetic 'profile' (multi-locus genotype). Individual whales can then be uniquely identified from their genetic profiles, in which case the genetic data become equivalent to any other tagging data (e.g. external tags or photographic identification) and can be used to estimate abundance by standard capture-mark-recapture methods (Garrigue *et al.,* 2004; Palsbøll *et al.,* 1997; Seber, 1982).

Genetic capture-mark-recapture is subject to the same general caveats that apply to other tagging methods used for capture-mark-recapture estimation of abundance, including capture heterogeneity (different capture probabilities among individuals), underlying assumptions (e.g. open or closed population, mixing between sampling events etc.), and tag loss and mis-identification (Hammond *et al.,* 1990; Schwarz and Seber, 1999). For genetic capture-mark-recapture, the most fundamental criterion is to ensure that individuals are identified correctly, which implies that (i) each individual has different multi-locus genotypes and that (ii) different tissue samples from the same individual all have identical genetic profiles. The first issue is resolved by genotyping a sufficient number of loci, and the second is correlated with the genotyping error rate, which in turn is a function of DNA quality as well as the overall laboratory quality control. Measures to minimise errors and recommended ways to report error rates are detailed in the guidelines document for DNA quality control (Tiedeman *et al.,* 2012).

Assessing how many loci are required to ensure that every individual has a unique genetic profile is typically determined on the basis of the so-called 'probability of identity' (I, Paetkau *et al.,* 1995), which denotes the probability that two individuals have an identical genetic profile by chance. I is estimated from the observed allele frequency distribution. One potentially important aspect of I is the fact that closely related individuals have higher Is due to elevated levels of identity by descent in comparison to unrelated individuals. The issue of related individuals is typically circumvented by genotyping sufficiently many loci to ensure that, even if all individuals were related as full-siblings (which have the highest expected I), the expected number of *different* individuals with identical genetic profiles will be below an *a priori* threshold (Waits *et al.,* 2001). This approach is perhaps overly cautious as not all specimens are likely to be related; except in very small populations, the fraction of related individuals will be far less than that of unrelated individuals (Paetkau, 2004; Rew *et al.,* 2011). The estimation of I is based upon an assumption of random mating and the observed allele frequencies. Consequently, ignoring significant genetic sub-structuring (sub-populations, or social structure, such as maternally related pods) within samples might result in an under-estimate of the number of loci necessary to reliably identify individuals. No general assessment has been conducted to determine what levels of sub-structuring cause problems in this regard. If such issues are suspected, then it is advisable to undertake simulations to estimate the potential bias of I and to assess the effect upon the individual genetic identification and the final estimate of N.

It should be noted that simply genotyping more loci in each specimen to satisfy an I of high relatedness is not without problems either, because the error rate per individual is positively correlated with the number of loci used for the individual genetic profile. Accordingly, the optimal number of loci is a balance between the rigor of individual identification and the chance of mis-matching genotypes due to errors (Paetkau, 2004; Rew *et al.,* 2011). This optimal number will typically be a few more loci than the bare minimum suggested by the estimate of I (for unrelated individuals). In such cases, mis-matches due to low error rates are readily identifiable as nearly identical genetic profiles (i.e. two samples with identical genetic profiles at all but a single locus), which then can be retyped at the loci in question.

(ii) Indirect capture-mark-recapture of individuals through the genetic identification of close relatives

A previously captured individual genetic profile can be 're-captured' indirectly by the identification of parts of its genome in other individuals, *i.e.* the identification of close relatives to the captured individual. This can be done, for instance, using paternity analysis during which a previously sampled individual male genetic profile can be re-captured in offspring he has sired. Each of such paternities assigned to previously captured males constitutes a 'recapture event' of the male in question. The final data can be used to conduct a capture-mark-recapture estimate of census population size (Pearse *et al.,* 2001; Palsbøll, 1999; Nielsen *et al.*, 2001; Garrigue *et al.,* 2004). Such indirect recapturing of individual genetic profiles through the identification of their close relatives can, in principle, be extended to any degree of relatedness. However, in contrast to more traditional capture-mark-recapture assessments, relating the number of recaptures to census population size is less straightforward for this kind of indirect genetic recaptures. The re-capture probability of a captured and marked individual depends not only on population size but also on the demographic parameters such as survival, birth and dispersal rates as well as the reproductive variance in addition to the degrees of relatedness targeted in the estimation. (Palsbøll *et al.,* 2010; Fountain *et al.*, 2017; Bravington *et al.*, 2016a, b). Often tissue samples are collected over extended longer time periods (i.e. not during one sampling session) which adds further complexity to such an estimation.

There are a number of different statistical approaches to estimating the degree of relatedness among individuals (Blouin *et al.,* 1996). However, the LOD method appears to be the least biased (Milligan 2003). Typically, this kind of estimation requires more loci to be genotyped in each individual compared to the re-capture of individuals (Section (3)(a)(i) above), since identification of related individuals is subject to greater uncertainty. The number of loci required (*i.e.* the statistical power in the observed data to detect pairs of individuals of the targeted relatedness category) can be assessed *in silico* (Wang, 2006). As with the identification of individuals, estimation of relatedness among individuals is based on observed allele frequencies assuming Mendelian transmission of alleles and random mating. No general statistical framework has yet been developed for this approach, and hence this kind of estimation will typically require some level of simulation to calculate abundance and the associated statistical uncertainty (e.g. Palsbøll *et al.,* 2005).

Nielsen *et al.* (2001) employed paternity to estimate male abundance in North Atlantic humpback whales and a recent implementation of this approach in a large marine fish population (southern bluefin tuna) illustrates many of the issues involved (Bravington *et al.*, 2016a, b). An example using cetaceans is the study by Garrigue *et al.* (2004).

(iii) Identifying recent population bottlenecks

When a population decreases in abundance, so does the effective population size (N_e – see Section (3)(b)). Genetic diversity, in turn, decreases due to random genetic drift at a rate that is inversely proportional to the new, lower N_e. However, the loss of genetic diversity is not instantaneous. The time lag until the population attains mutation-drift equilibrium corresponding to the post-bottleneck N_e is a function of pre-bottleneck N_e, as well as the nature and degree of decrease. During this transitional phase, rare alleles, heterozygosity, and several other diversity indices (*e.g.* nucleotide diversity and the variance in number of repeats) decline at different rates, and these differences form the basis for a general class of tests of deviation from mutation-drift equilibrium (aka "bottleneck" tests). The general principle in these tests is to compare a ratio (or sum as in the case of Tajima's *D*, Tajima 1989) of diversity indices against the expectation under mutation-drift equilibrium, which serves as the null hypothesis.

The two most commonly-used methods to detect a recent bottleneck are; (*i*) a test for heterozygote 'excess' (Cornuet and Luikart, 1996), and (*ii*) the M-ratio test (Garza and Williamson, 2001). Both tests are based upon microsatellite genotypes. In the case of (*i*), coalescent-based simulations are used to generate the null distribution of heterozygosity, given the observed number of alleles under mutation-drift equilibrium. During a bottleneck, alleles are lost faster than heterozygosity declines. If the observed heterozygosity is higher than expected given the number of alleles, a bottleneck is inferred. The M-ratio test (*ii*) compares the size range of alleles to the number of alleles (the 'M-ratio'). During a bottleneck, the number of alleles decreases faster than the size range. Coalescent simulations are also employed to estimate the null distribution of the M-ratio (i.e. under equilibrium conditions). A lower than expected M-ratio is indicative of a bottleneck. The outcome of both approaches relies heavily on the underlying mutation model (essentially the fraction and mean length of mutations larger than a single repeat; Peery *et al.,* 2012; Williamson-Natesan, 2005). With both tests, an incorrect mutation model can generate results suggesting a bottleneck when none has occurred. In addition, the power of these tests varies substantially depending on the nature of the bottleneck (pre-bottleneck population size, and the nature and degree of the reduction), as well as the amount and kind of data; therefore, absence of a statistically significant result cannot be interpreted as indicating that no bottleneck has occurred. Since both tests are sensitive to the above issues, it is important to test the rigor of the observed outcome by simulations before drawing final conclusions (Peery *et al.,* 2013).

Another class of analytical methods aims at estimating the time and relative change in abundance over time. These approaches all employ coalescent theory, which implies that they are likely sensitive to the above issues regarding the underlying mutation model (mode as well as rate). BATWING (Wilson *et al.,* 2003) and MSVAR (Beaumont, 1999) allow the use of microsatellite genotypes, whereas other approaches are based solely on DNA sequence data (e.g. skyline plots, Drummond *et al.,* 2005; stairway plots, Liu and Fu, 2015).

The underlying population models in these methods are generally simple and might ignore (and thus violate) key aspects of the biology of real species. One example of a potentially important parameter is migration, which is ignored in estimation procedures that are based upon a single closed population model.

In the event that the desired population model and history are not available in any of the existing methods, coalescent-based simulation software, such as SIMCOAL, (Excoffier *et al.* 2000) can be utilised in an Approximate Bayesian Computation ('ABC', Beaumont *et al.,* 2002) framework to estimate the posterior probability distributions of user-defined demographic scenarios across pre-defined prior ranges of the relevant parameter values (Cornuet *et al.,* 2008). The goodness-of-fit of different demographic scenarios can also be evaluated using model selection criteria (Cornuet *et al.,* 2010). This framework might be used to estimate the mutation model as well, thereby potentially circumventing the above issues with the mutation model. In a recent study of Antarctic fur seals using microsatellite and mtDNA data, Bayesian skyline and ABC approaches strongly supported a recent, severe bottleneck due to sealing, while little support was found using empirical methods, or through analysis of mtDNA alone. The ABC approach was able to estimate the magnitude and time of the sealing bottleneck, giving ranges that were robust to expansion in the mutation rate prior and exclusion of mitochondrial DNA (Hoffman *et al.,* 2011). However, in practice ABC methods often use simple models to simulate the data, and hence they can be subject to the same types of biases from model-misspecification as other types of analyses (Peery *et al.,* 2012; 2013).

Constraints on bottleneck size are of interest for informing population trajectory-based assessments of recovery at the IWC (IWC, 2007). Lower bound constraints are currently imposed on single-stock assessments using the number of mtDNA control region haplotypes in the contemporary population as a proxy for the minimum number of surviving mothers. This is multiplied by three as a conservative *ad hoc* lower boundary, considering a number of factors (IWC, 2012). First, the contemporary sample usually provides only a subset of all haplotypes (maternal lineages) found in the population, and more are sometimes yielded at longer sequence lengths. Second, mtDNA lineages alone do not reflect the additional males and immature individuals present in the population at the time of the bottleneck. Third, there are likely to be multiple mothers with the same haplotype at the bottleneck (rather than one haplotype for each mother as assumed here) (Jackson *et al.,* 2011). The impact of these different correction factors on population assessment outcomes has been explored for southern right whales and Antarctic blue whales (Branch and Jackson 2008; Jackson *et al.,* 2008; Jackson *et al.*, 2009). This approach can provide an independent test of population dynamic model predictions using genetic data; that is, do population dynamic bottlenecks give abundances that are significantly smaller than those inferred from genetic data? However, this approach cannot yet adequately account for population sub-structuring and dispersal effects.

(b) Effective population size, N_e

The effective population size (N_e) is defined as the size of an 'ideal' population (a random mating, closed population unaffected by mutation and natural selection and with non-overlapping generations) that would show the same rate of loss of diversity over time as the observed population (Wright, 1938). The ecological interpretation of the effective population size is (roughly) the number of breeding individuals, but this does not capture variation among individuals in reproductive success. Effective population size determines the rate of loss of genetic variation from a population/species, which has implications for the short- and long-term persistence due to the adverse effects of breeding among genetically similar individuals (inbreeding depression) and loss of evolutionary potential. Small N_e also leads to the random loss of adaptive traits due to the dominance of random processes at small effective population sizes (where typical levels of selection are insufficient to overcome random effects).

(i) Historical N_e

Coalescent theory defines a relationship between the expected number of mutations (denoted θ) between any two random gene copies as the product of the per-generation mutation rate (μ) and the effective population size (N_e) in a single idealised population, at a single diploid Mendelian-inherited locus (Ewens, 1972; Hudson, 1990; Watterson, 1984):

$$\theta = 4N_e\mu$$

Consequently, if an estimate for μ is available, then it is possible to infer the long-term effective population size from an estimate of θ when the assumptions underlying the relationship between θ, N_e, and μ are upheld. The estimation is subject to a large number of assumptions and it is not always clear what time point the final estimate applies to (see Palsbøll *et al.* 2013 for a detailed review).

The value of θ can be estimated using various summary statistics such as the heterozygosity, nucleotide diversity, the number of segregating sites and the variance in repeat number at microsatellite loci, as well as more complex coalescent-based estimation methods. The various estimation methods can be divided into single or multi-population models, and these can be further sub-divided into approaches that do or do not assume equilibrium (with respect to mutation and genetic drift, and in some cases migration). The relevant timeframe is the time since the most recent common ancestor (TMRCA). The TMRCA is expected to be $4N_e$ generations for an autosomal diploid Mendelian-inherited locus, and $2N_ef$ for a mitochondrial locus, where N_ef denotes the N_e of females (Ewens, 1972; Hudson, 1990). The estimate for θ reflects the single most likely value for the entire genealogy (i.e. since TMRCA) and hence is not restricted to a specific time or even location, unless the population is closed to immigration or migration. Violation of the assumption of a closed population can lead to an overestimate the local N_e, such that the mean estimate of a single population could reflect instead the mean N_e for the global population (see Hudson, 1990). Using this general method, it is possible to estimate long-term N_e, assuming that the population has been isolated and has reached equilibrium between mutation and genetic drift (e.g. as implemented in the program migrate-n; see Beerli and Palczewski, 2010).

Some additional refinements to the above approach attempt to assess historical N_e during more defined periods of time, or even changes in N_e over time. The first of these approaches, 'isolation with migration' (see Hey and Nielsen, 2004 and Section (4)(b)(ii)2), provides estimates of N_e for populations following an estimated time of divergence, together with an estimate of the ancestral population size. Because this approach also attempts to infer the divergences among populations, the parameter estimates are likely improved when (as is probably the case for most studies) populations are not in migration-drift-mutation equilibrium. However, the final parameter estimates are, as in the above analyses, the most likely single estimates across the 'estimation period', which is before and after the populations diverged. An important assumption is that there are no un-sampled 'ghost' populations exchanging migrants with the study populations (or among ancestral populations). Such populations might even be extinct and hence their existence might be unknown or samples not available. Violation of this assumption can inflate the estimated ancestral N_e. Some recent modifications of the method (allowing for the inclusion of multiple populations) are intended to help address this problem (see Hey, 2010).

A coalescent analysis involves looking backwards in time through gene genealogies to where lineages come together ('coalesce'). There is, furthermore, a relationship between the time intervals between the coalescent events and the effective population size (Kingman, 1982). Historical dynamics can be assessed by taking advantage of this

relationship. The general approach is known as a 'skyline plot' (Pybus *et al.,* 2000). Generating a skyline plot is a two-step process – first a genealogy is estimated, then the population history is estimated from the genealogy (exploiting the information on coalescent intervals derived from the genealogy). Refinements to the method (reviewed in Ho and Shapiro, 2011) now permit a reconstruction of historical dynamics that can be quite accurate (based on tests against simulated datasets; Heled and Drummond, 2008), especially if chronological data are included (e.g. by incorporating sequences from ancient DNA, as implemented in the program BEAST; Drummond and Rambaut, 2007). The method assumes a closed population, however, and while empirical data suggest that it is fairly robust to violations of this assumption, this has not been fully tested. An alternative method for tracking historical trends is to compare modeled scenarios by generating simulations and assessing the relative support of different scenarios among those simulations through 'Approximate Bayesian Computational' – 'ABC' analysis (see Beaumont *et al.,* 2002). This method was recently used to identify a population bottleneck in the Greek population of common dolphins (*Delphinus delphis*) estimated to have been within the last 50 generations (Hoffman *et al.,* 2011, Fontaine *et al.,* 2012, Moura *et al.,* 2013b).

Another study (Ho *et al.,* 2008) applied the Bayesian skyline plot method (Drummond *et al.,* 2005) to bowhead whales (*Balaena mysticetus*), building on data originally published by Rooney *et al.* (2001) and Borge *et al.* (2007). The bowhead has been heavily exploited in the past, especially between the 17th and early 20th centuries, and so tracking their historical population size to assess the extent of the impact and the nature of their recovery could provide important information in support of conservation efforts. An essential question is how the pattern of population dynamics relates to particular historical events; to make that possible, an appropriate estimate of the substitution rate is required (see Ho *et al.,* 2008 and Section (5)(d)). Incorporating data for 99 dated ancient DNA samples (from Borge *et al.,* 2007) Ho *et al.* (2008) calculated a substitution rate of 0.159 subs/site/Myr (95% HPD: 0.051–0.272 subs/site/Myr) for bowheads. Then, based on three separate analyses—one using only ancient DNA data, one using only modern data (from Rooney *et al.* 2001), and one incorporating both modern and ancient sequences – they found evidence for a population expansion in the late Pleistocene (after the last glacial maximum). All three analyses showed the same pattern (Ho *et al.,* 2008). A more recent pattern of contraction was not observed, and the confidence limits were large, but this is to be expected, given the genetic material available (a single mitochondrial DNA locus). Very recent events are unlikely to be revealed by skyline-plot analyses; however, inclusion of multiple loci using the same methods will increase the accuracy and the chance of revealing more recent trends (see Heled and Drummond, 2008, Hoffman *et al.,* 2011). This becomes increasingly practical with the development of next generation sequencing methods and DNA capture (see Horn, 2012). Genome sequencing can also provide sufficient data for much deeper N_e profiles (dating back up to a million years or so) using the pairwise sequentially Markovian coalescent (PSMC) model. An example is provided by a recent study that revealed a severe bottleneck at the time of the last glacial maximum in killer whales (*Orcinus orca*) based on whole killer whale genome sequences (Moura *et al.,* 2014a).

In general, the long-term estimates of N_e (or estimates of the average N_e since some point of divergence) have the potential to be useful in cetacean management applications provided that the temporal context is well understood. These estimates are averages with relevance over an extended period of time, not necessarily reflecting contemporary or even recent N_e (see next section). As modern methods increasingly facilitate generation of data from large numbers of loci, estimates of historical population trends (for example, using skyline-plot methods) become more accurate. It will likely become more practical to estimate these trends over time frames that allow harvesting or other anthropogenic impacts to be tracked; however, it is important to appreciate the limitations of these methods and the difficulty associated with tracking very recent trends using genetic data.

(ii) Contemporary N_e

The term 'contemporary N_e' generally refers to effective population size for a time period roughly encompassed by the samples being analysed. The key variables required to calculate contemporary N_e are N (census population size) and the mean and variance of number of offspring produced by each individual ($\bar{k}$ and V_k, respectively). A standard formula for effective size for species with discrete generations (Crow and Denniston 1988, Caballero 1994) is:

$$N_e = \frac{\bar{k}N - 2}{\bar{k} - 1 + V_k / \bar{k}}.$$

With modern molecular techniques, it is possible to genetically match offspring to their parents with high precision, which allows direct evaluation of reproductive success (reviewed by Jones *et al.* 2010) and calculation of effective size. However, N_e depends on the mean and variance of *lifetime* reproductive success, and such data are very challenging to collect for long-lived species like cetaceans. As a consequence, it has become common to use genetic methods to indirectly estimate contemporary N_e based on genetic indices that are influenced by random genetic drift.

Genetic methods for estimating contemporary N_e all depend on a signal of genetic drift that is a function of $1/N_e$, and they all depend on sampling a finite number of individuals (S), which introduces random sampling error of approximate magnitude $1/S$. Therefore, all else being equal, best results are obtained when relatively large samples are used to study relatively small populations, in which case the drift signal is relatively large compared to sampling error. Once N_e becomes relatively large (>500–1000 or so), the drift signal is so small that it is hard to distinguish from that produced by a very large or infinite population. Methods that estimate contemporary N_e assume that genetic drift is the primary evolutionary factor responsible for the genetic signal, which means that these methods typically assume the

genetic markers are selectively neutral, the target population is closed to immigration, and mutation is unimportant over the time frames considered.

Two main genetic approaches are used to estimate contemporary N_e: single-sample methods and two-sample (temporal) methods. The temporal method estimates changes in allele frequency (or occasionally heterozygosity) over time within a single population and hence requires at least two samples spaced in time. The premise for the temporal method is that a measure of the standardised variance of allele frequency over time due to genetic drift (F) is a simple function of N_e and elapsed time in generations (t): $E(F) \approx t/2N_e$. The smaller the N_e, the larger the change in allele frequencies over time. After accounting for effects of sampling error on estimates of F ($\hat{F}$), it is easy to rearrange the above equation to obtain an estimate of $\hat{N}_e$: $N_e \approx t/[2(\hat{F} - 1/S)]$). This so-called 'standard' temporal method was developed 40 years ago (Krimbas and Tsakas, 1972) and is still widely used, although likelihood based methods (e.g. Wang, 2001; Anderson, 2005) have been developed more recently. Because it is based on the variance in allele frequency over time, the temporal method provides an estimate of variance N_e. If N_e varies over time, then the temporal method estimates the harmonic mean N_e for the time period between temporal samples.

As their name implies, single-sample methods require only one sample and produce an estimate that applies to a single point in time (generally this is N_e in the parental generation, although in some cases it might be affected by N_e in other recent generations; see Waples, 2005). The most widely-used single-sample method is that based on linkage disequilibrium (LD; see Section (2)(c)). In the LD method, the inter-locus associations are estimated by the squared correlation coefficient r^2, and theory indicates that r^2 due to random genetic drift in a finite population is a simple function of $1/N_e$ and $1/S$. Several other single-sample methods have recently been developed. The software package OneSamp (Tallmon *et al.,* 2008) is an approximate Bayesian computation (ABC) method that simulates data using a variety of true N_e values and finds the value that produces the best match to the empirical data. OneSamp's summary-statistic approach uses a variety of genetic indices, but the most important is r^2. The heterozygote excess method (Pudovkin *et al.,* 1996) is based on the fact that a small number of parents cause small departures from Hardy-Weinberg expectations. The molecular coancestry method (Nomura, 2008) is based on the pattern of shared alleles among individuals, and the sibship reconstruction method of Wang (2009) estimates the value of N_e that is most consistent with the estimated pattern of relationships among the sampled offspring. A related method (parentage analysis without parents; Waples and Waples, 2011) uses a matrix of pairwise relatedness estimates to reconstruct the distribution of offspring number per parent, which allows calculation of N_e from a standard demographic formula. For this latter approach to be applicable, data (genetic or behavioural) need to be available to rigorously identify half- and full-sibling pairs. For a given inferred pedigree, the sibship and parentage-analysis-without-parents methods provide essentially the same estimate (Ackerman *et al.,* 2016).

Some evaluations of sensitivity to model assumptions have been conducted, but much work remains to be done. Wang and Whitlock (2003) developed a modified temporal method that can account for immigration and jointly estimates N_e and migration rate (m). However, it assumes a rather improbable migration model (an infinite source population with known and constant allele frequencies sends a fixed proportion of migrants to the focal population every generation), which limits its general usefulness. Waples and England (2011) showed that the LD method is relatively robust to migration that is constant enough to produce an equilibrium between migration and drift, but it can be strongly affected by non-equilibrium, pulse migration (or, of perhaps more relevance for cetaceans, inclusion of migratory individuals from another population in a local sample). Gilbert and Whitlock (2015) did extensive simulations of performance of both single-sample and temporal estimators under various types of migration. They concluded that the Wang/Whitlock temporal method and the LD method were most robust to departures from the assumption of a closed population. Wang (2016) conducted extensive simulations to evaluate single-sample methods and concluded that the sibship and LD methods performed much better than the molecular-coancestry or heterozygote-excess methods. Waples and Yokota (2007) evaluated performance of the temporal method when it is applied to species with overlapping generations; this can bias $\hat{N}_e$ either up or down, depending on the species' life history and the life stage(s) sampled. The modified temporal method of Jorde and Ryman (1995) can estimate N_e directly in iteroparous species, but it requires samples from consecutive cohorts and information about age-specific vital rates. A new version (Jorde, 2012) extends this model to allow samples from cohorts spaced any number of years apart. Waples *et al.* (2014) evaluated performance of the LD method for iteroparous species with overlapping generations, including simulations based on a life table for the bottlenose dolphin. This study (as well as Waples *et al.,* 2013) showed how the relationship between effective size per generation (N_e) and the effective number of breeders in one year (N_b) depends on the species' life history.

Markers that are physically linked (and hence have recombination probability <0.5) create more LD than unlinked markers, and this adds power (for estimating N_e) to the analysis if the recombination rate is known. In most non-model species this information is not available, however, in which case the default assumption of no linkage will lead to downward bias in $\hat{N}_e$ if LD due to linkage is misinterpreted as due to drift. This problem might not be serious for most current studies based on at most a few tens of markers, but will require careful consideration as large numbers (100s or 1000s) of SNP loci become more generally available. Lack of independence of markers should not bias estimates obtained with the temporal method, but it will lead to an overestimate of precision because the information content of the markers is not as large as assumed. Selection can take a wide variety of forms and it is possible to imagine all sorts of effects on $\hat{N}_e$, but effects on estimates of contemporary N_e have seldom been explored. One option researchers can use to address the effect of

selection is to estimate N_e using both the complete dataset and a reduced dataset that omits loci thought to be under selection.

Results of simulations with 'known' (specified in the model) N_e show that, although inclusion of numerous rare alleles can increase bias for some methods, low precision is more likely than bias to limit practical usefulness of methods that estimate contemporary N_e. Reduced precision for large N_e can be mitigated to some extent by using more genetic markers (in which case the average drift signal will be closer to the true $1/N_e$) and/or taking larger samples, which reduces the noise from random sampling error (Waples *et al.*, 2018). For more discussion, see Waples and Do (2010) and Waples (2016a).

Because the drift signal $1/N_e$ is nonlinear, the distribution of $\hat{N}_e$ is skewed, which has two important consequences. First, it means that confidence intervals for are $\hat{N}_e$ asymmetrical, being wider on the upper end than the lower end. Second, the skew means that the arithmetic average is not a good measure of central tendency. If one wants to average a number of estimates of N_e, the harmonic mean should be used; if the number of estimates is large enough, the median will generally be close to the harmonic mean.

Likelihood-based methods use more of the data than the moment-based temporal and LD methods and should be more precise, and this is generally reflected in tighter confidence intervals. However, because they are so computationally demanding, performance of these methods has not been as rigorously evaluated, especially for large populations. All of the likelihood and ABC methods require the user to specify a maximum N_e for the target population, and this can give the illusion of greater precision regarding the upper bound to N_e. However, it should be remembered that this upper limit is a user input and not an output of the model.

Because the various methods have different strengths and weaknesses, it can be useful to use more than one method with a particular dataset and compute a weighted harmonic mean across methods to arrive at an overall estimate, as described in Waples and Do (2010). In doing so, however, one should be careful about combining estimates that apply to different time periods, as discussed below (N_e/N ratios) and in Waples (2005). Good reviews of genetic methods for estimating contemporary N_e can be found in Wang (2005) and Luikart et al. (2010), with the latter including discussion of several of the newer single-sample methods developed after Wang's (2005) review was published.

Summary and relevance to cetaceans:
Current molecular methods cannot reliably distinguish between contemporary effective sizes that are relatively large ($N_e \sim 1000$) and very large ($N_e > 10000$, say), but they have much more power to distinguish between relatively small and relatively large populations (N_e – 100 vs $N_e = 1000$). As the N_e/N ratio can be considerably less than 1 (see next section), these methods could potentially provide useful information for a wide range of cetacean populations. The temporal method will be of limited practical use, unless 1) DNA from archived samples can be accessed (to allow a number of generations of genetic drift between samples) or 2) individuals can be aged accurately and age-specific vital rates can be estimated, in which case the Jorde and Ryman (1995) modification for age-structured populations can be used. Single-sample methods should be of more general value, although more work is needed to help interpret results for age-structured species like cetaceans. The LD method is the most widely-used single-sample estimator; the ABC method of Tallmon *et al.* (2008) and the sibship method of Wang (2009) have considerable potential but performance needs to be evaluated across a wider range of realistic scenarios.

(c) The N_e/N ratio

The N_e/N ratio has attracted a good deal of attention in the literature (e.g. Nunney, 1993; Frankham, 1995; Hedrick, 2005; Waples, 2016b). This ratio is of interest because N_e is difficult to measure or estimate, particularly for long-lived iteroparous species such as cetaceans, so if the ratio is known and N can be estimated, a reasonable estimate of N_e can also be obtained. Conversely, in some cases it is easier to estimate N_e from genetic methods than it is to obtain an unbiased estimate of total population size. In that situation, an estimate of abundance can be obtained from an estimate of N_e and knowledge of the ratio N_e/N.

Simple rearrangement of the standard demographic formula for N_e (see Section (3)(b)(ii)) leads to the following, which assumes that population size is stable (so $\bar{k} = 2$):

$$\frac{N_e}{N} = \frac{2}{1+V_k/2}.$$

This formula shows that the ratio N_e/N is inversely related to the variance in reproductive success among individuals (V_k). If $V_k > 2$, N_e/N is less than one. An uneven sex ratio is one factor that increases V_k and reduces N_e: since half the genes for the next generation have to come from males and half from females, members of the less numerous sex on average have higher reproductive success. However, the sex ratio has to be highly skewed to have a substantial effect on N_e.

Age structure is another factor that influences N_e. A modification to the discrete-generation formula that accounts for overlapping generations is (Hill, 1972):

$$N_e = \frac{4LN_1}{2+V_{k\bullet}}.$$

where L is generation length, N_1 is the number of offspring in each cohort, and $V_{k\bullet}$ is the lifetime variance in reproductive success among all individuals in a cohort. $V_{k\bullet}$ is difficult to measure in natural populations because it is necessary to integrate reproductive success across the entire lifetime of individuals. For species with constant survival and fecundity with age 20 (as might be approximately true for many cetaceans), adult lifespan depends on the adult mortality rate (d = annual probability of death), generation length is given by $L = 1/d + \alpha - 1$ (where α is age at maturity), adult census size is $N = N_1/d$, and lifetime variance in reproductive success is given by $V_{k\bullet} = 4(1-d) + 2\phi$ (Waples, 2016b). The parameter ϕ indicates the ratio of the variance to the mean reproductive success in one time period of individuals of the same age and sex. For example, if all 10-year old males act like a mini Wright-Fisher population with random variation

in reproductive success in that year, then $\phi = 1$. Putting this all together and substituting into the above equation yields this expression for the effective: census size ratio for iteroparous species with constant vital rates (Waples, 2016):

$$N_e / N \approx \frac{1+d(\alpha-1)}{(3+\phi)2-d}.$$

This equation shows that the N_e/N ratio depends on 3 key parameters: d, α, and ϕ. Approximations for the first two can be made based on cetacean biology. Annual adult mortality must be about 0.1 or lower in a long-lived species, and for many cetaceans, age at maturity will be about 5–10 years. If we take $\alpha = 10$ and $d = 0.1$ as examples, then the above equation simplifies to:

$$N_e / N \approx \frac{1.9}{(3+\phi)/2-0.1},$$

which can be approximated by $N_e / N \approx \frac{4}{3+\phi}$. Note that if $\phi = 1$, N_e is approximately the same as N.

Very little information is available about ϕ in natural populations of any species. Realistically, female cetaceans can only produce one or zero offspring per year. In this situation, it can be shown that female ϕ cannot exceed 0.5, so female N_e/N is probably about 1 or a little higher. This suggests that the overall N_e/N ratio in cetaceans depends heavily on the value for male ϕ, about which there is essentially no information. If reproductive success of males of the same age is random ($\phi = 1$), then male N_e/N is approximately 1. Male N_e/N drops to 0.5 when male $\phi = 5$ and to ~0.1 when male ϕ is as high as 35–40.

What scenarios produce $\phi = 5$ or 40? First, we need to identify the denominator of ϕ, which is the mean number of offspring produced per individual ($\bar{k}$). In a stable population with constant vital rates, the mean number of offspring produced each year by an individual is twice the annual mortality rate, d (Waples, 2016), so in our example $\bar{k} = 0.2$. Under these conditions, it can be shown that ϕ is approximately 5 when only one in every 25 males successfully produces offspring in a given year, and ϕ is approximately 40 when only one in every 200 males successfully reproduces. Paternity studies to date have generally not found a large variance in reproductive success among male cetaceans (Nielsen *et al.,* 2001; Cerchio *et al.,* 2005).

The Hill (1972) model used in the above analyses assumes that survival and reproduction are independent across years. That is not true in female cetaceans, who generally skip one or more years after giving birth. This 'skip breeding' scenario was evaluated by Waples and Antao (2014), who showed that, for a life history like that of the bottlenose dolphin, skip breeding by females could reduce N_b per year by up to 10–20% (because fewer females are available to breed each year). However, skip breeding increases N_e slightly in the bottlenose dolphin (by < 10%) because it tends to equalise lifetime reproductive success. Therefore, the above analyses of N_e/N are largely robust to skip breeding. Lee *et al.* (2011) evaluated a situation in which individuals were consistently either good or bad at producing offspring. They found that persistent differences in reproductive success reduced N_e/N; the maximum reduction they found was from about 0.5 to 0.1–0.2.

Collectively, these results suggest that, if the assumptions of the Hill model are met, rather extreme scenarios involving overdispersion of male reproductive success would be required to reduce cetacean N_e to as low as 50% of N. Accounting for skip breeding has a relatively small effect. However, if reproductive success is persistently centered on a relatively small number of individuals, the N_e/N ratio could potentially be reduced several fold.

A key issue in computing the ratio N_e/N is definition of N. N generally refers to the census population size, but which individuals are included in the census? Should it be all individuals alive at a given point in time, all individuals that have reached age at maturity, or only individuals that are still reproductively active (hence excluding senescent individuals)? Each of these definitions has been used in the literature, and each might be appropriate, depending on the species and the research question at hand. This should be kept in mind when comparing estimates of N_e/N for different populations or species. In all of the above analyses, N is taken to be the number of adults alive at a given point in time. If juveniles or subadults are counted as part of the census size, N_e/N ratios would be lower.

Another important consideration is to ensure that estimates of N_e and N apply to the same time period; if not, the resulting ratio can be misleading for populations in which N_e and or N vary over time (see Waples 2005 for details). For example, most single-sample methods estimate N_e in the generation prior to the generation that is sampled, in which case comparing $\hat{N}_e$ with current census size is not particularly meaningful.

Statistical analysis of ratio data is tricky. In general, the median value or geometric mean is a better measure of central tendency than the arithmetic mean. Ratios of N_e/N across multiple generations also require some consideration. Long-term N_e is the harmonic mean across the years under consideration. Some authors (e.g. Frankham, 1995) have computed long-term N_e/N as the harmonic mean N_e divided by the arithmetic mean N. However, mixing harmonic and arithmetic means creates a statistical artifact, such that any variation in N_e or N across generations reduces the long-term ratio N_e/N even if the ratio is the same every generation (see Kalinowski and Waples, 2002). Using the harmonic mean of both N and N_e to calculate the long-term ratio solves this problem.

Summary and relevance to cetaceans:

Knowledge of the N_e/N ratio can provide important insights into demographic and evolutionary processes in cetacean populations. Analyses conducted here show that variance in male reproductive success and whether some individuals consistently dominate reproduction are the key factors that determines this ratio. Assuming N represents the number of mature adults, N_e/N in long-lived cetaceans should be close to 1.0 unless (a) the variance in male reproductive success per year is greatly overdispersed, or (b) individual differences in reproductive success persist over time. Therefore, collecting more data on these parameters should be a high priority.

(4) ANALYSIS OF DIVERSITY AMONG POPULATIONS (STOCK STRUCTURE)

(a) Testing for heterogeneity

The most commonly applied approach to detect 'stock structure' is to assess whether there is some degree of genetic heterogeneity among the sampled specimens. Such an assessment can be conducted with or without an *a priori* partition of the specimens into putative stocks or populations. Partitioning of specimens is typically geographic or temporal (e.g. to pre-defined management units).

(i) Putative populations defined a priori

Detection of possible stock structure among *a priori* partitioned samples is typically conducted using a contingency table test of allele or genotype counts (Hudson *et al.,* 1992; Moritz, 1994) to assess the probability that different samples are drawn from the same underlying distribution (Sokal and Rohlf, 1995). If the probability of homogeneity in a single test is estimated at below a pre-defined α-value (typically 0.05), homogeneity is rejected and genetic structuring is typically inferred (Brosi and Biber, 2009; Morin *et al.,* 2009; Palsbøll *et al.,* 2007; Waples and Gaggiotti, 2006). This general approach has recently attracted criticism Waples, 1998; Waples and Gaggiotti, 2006; Fallon, 2007; Palsbøll *et al.*, 2007; Bernard *et al.,* 2009). The problem is that the ability to detect heterogeneity is a function not only of the effect size (i.e. the degree of genetic divergence among sample partitions), but also of the amount of data (the number of loci and specimens analysed) but only the effect size is of management relevance. Consequently, a homogeneity test might fail to detect heterogeneity (i.e. detect stocks) even when the genetic divergence among stocks is high (and hence migration rates low), stemming from low statistical power due to insufficient data. Accordingly, the failure to detect statistically significant levels of genetic heterogeneity does not necessarily imply the absence of heterogeneity (i.e. different stocks); it could simply be due to insufficient statistical power (Taylor and Gerrodette, 1993). Conversely, in cases where statistical power is high (large number of genetic markers analysed in many specimens), statistically significant levels of heterogeneity can be detected even when the effect size is very small and of little management relevance (i.e. at low genetic divergence, Waples, 1998; Palsbøll *et al.,* 2007; Waples and Gaggiotti, 2006). Therefore, when significant levels of heterogeneity have been detected, the biological relevance of the magnitude of genetic divergence should also be assessed. Another important factor to consider is whether the pattern of heterogeneity makes sense in terms of stock structure, or is better explained by other factors, such as age structure.

A number of statistics have been developed and implemented to conduct homogeneity tests using *a priori* partitioned genetic data. One of the most common assessments is a simple randomised χ^2 test, which, appears to have high statistical power (Hudson, 1992; Hudson *et al.*, 1992; Roff and Bentzen 1989). Often, the test statistic is one of the more common estimators of genetic divergence, such as Wright's F_{ST} (Wright, 1931) or related measures such as Φ_{ST} (Excoffier *et al.,* 1992) and K_{ST} (Hudson *et al.,* 1992). The latter two statistics were developed for use with DNA sequence data where the haplotype phase is known.

Advantages of homogeneity tests for stock structure:

- Well-established framework and conceptually simple;
- Implemented in many standard packages;
- Statistical power increases with more data without too much additional computational overhead.

Disadvantages include:

- No measure of relevant parameters (e.g. migration);
- The high power with rich data sets implies rejection of homogeneity at biologically insignificant levels;
- With other kinds of genetic structuring the analysis can produce rejection of homogeneity;
- Requires the user to define the strata to be tested a priori, so less useful for problems involving mixtures.

(ii) No a priori *basis (or uncertain basis) for grouping individuals into populations*

Standard clustering methods

Clustering programs attempt the very challenging task of decomposing a mixture of individuals into component groups or populations, all without using any *a priori* information about potential source populations (i.e. without any baseline information such as that used in genetic mixture analysis or assignment tests). The genetic clustering programs that are now in widespread use (such as STRUCTURE, Pritchard *et al.,* 2000, Hubisz *et al.,* 2009) are model-based in the sense that they assume random mating and linkage equilibrium within source populations (some versions allow explicit consideration of linkage). The programs then shuffle individuals among potential groups in an attempt to find the combination that minimises departures from Hardy-Weinburg equilibrium and linkage equilibrium. With anything but very small datasets, the number of potential solutions is so vast that MCMC (or other algorithms for sampling from probability distributions, see Section (5)(g)) are essential. In general, several replicates with different starting points and long 'burn-in' (initial steps that are discarded, typically in the tens of thousands) and very large numbers of steps should be run to check for 'convergence' (different replicates all arrive at a stationary distribution). For each presumed number of populations or gene pools ('*k*'), output of STRUCTURE and similar programs includes the estimated allele frequencies in each population, estimated F_{ST} values among populations, and the fraction of each individual's genes derived from each population. In addition, the relative likelihoods of the model 'fits' for different presumed values of *k* can be compared to determine the most likely number of 'populations'.

Application to a large number of real and simulated datasets has shown that STRUCTURE and similar programs can produce very impressive results, provided genetic differences among populations are at least moderately large. On the other hand, the ability to resolve mixtures breaks down as genetic differences among populations become small, generally at F_{ST} around 0.05 or less, although this threshold varies with sample size and number of markers.

In some cases, STRUCTURE can produce positively misleading results, where the program assigns a much higher likelihood to a scenario with the wrong *k*. A classic example of the latter is a STRUCTURE result that gives a high likelihood to $k = 2$ but concludes that every individual is a roughly 50:50 mixture of the two putative gene pools – a highly implausible result biologically that would only occur if two divergent populations hybridised and the entire mixture was composed of essentially all F_1 hybrids.

The authors of STRUCTURE admit that their approach to identifying the optimal value for *k* is *ad hoc* and suggest caution in its use – an admonition that is routinely ignored by users. Evanno *et al.* (2005) proposed an alternative algorithm, based on the second derivative of changes in the likelihood function with *k*. This method was designed to address the empirical observation that in many scenarios the likelihood creeps up indefinitely with increasing *k*, making it difficult to select a discrete optimal value and leading to overestimates of the number of populations. Waples and Gaggiotti (2006) examined both the standard and Evanno *et al.* methods for inferring k with a limited number of simulated datasets and did not find improved performance of the latter method, but the data were simulated with an island model while Evanno *et al.* suggest their method should be most useful for evaluating hierarchically-structured populations. It is important to note that because the Evanno method evaluates the rate of change in the log likelihood as *k* increases, it cannot be used to test the hypothesis that $k = 1$ (i.e. a single panmictic population).

Options for analysis include choosing allele frequencies that are correlated, or not and 'admixed', or not. In general, analysing with admixture and correlated allele frequencies both selected is appropriate if the putative populations are thought to be linked by gene flow, while no admixture and uncorrelated frequencies are appropriate for mixtures of pure individuals (as on a feeding grounds). However, the authors suggest trying both options as sometimes better resolution is achieved with the 'wrong' model.

HWLER (Pella and Masuda, 2006) is a recently developed clustering program that uses a different MCMC sampling method and can outperform STRUCTURE in at least some circumstances. However, it has not been widely used or extensively evaluated. GENELAND (Guillot *et al.*, 2005) and TESS (Chen *et al.,* 2007) are similar to STRUCTURE but can accept spatially explicit sampling information for each individual, which is then used to help guide formation of clusters. This option would not be likely to be useful, and might produce misleading results, for analysis of migratory individuals whose sampling locality is not indicative of its natal origin.

Summary and relevance to cetaceans:
Results discussed above indicate that although clustering programs can be very powerful, they often struggle to produce valid results under the low levels of population differentiation that characterise many cetaceans. More research is needed to better characterise the parameter space under which these programs can and cannot be expected to produce reliable results for cetaceans. Steps a researcher can take to maximise usefulness of these methods while minimising the likelihood of spurious results include the following:

- evaluate convergence by testing whether different starting points and run times produce comparable results
- check whether the inferred genetic makeup of individuals is consistent across runs
- consider whether the estimated pattern of genetic structure is biologically plausible

Clustering based on ordination
Over thirty years ago, Menozzi *et al.* (1978) introduced the use of Principal Component Analysis (PCA) to the study of genetic data. PCA is a general method for representing high-dimensional data, such as individual genotypes or population allele frequencies, in a smaller number of dimensions. In essence, PCA searches for orthogonal axes (uncorrelated components) along which projected objects (individuals or populations) show the highest variance and then returns the position of these objects along those axes or principal components (PCs). Typically, most of the variance is explained by the first few PCs so that they can be used to explore the structure of variation in the sample(s).

Cavalli-Sforza *et al.* (1994) used PCA to summarise allele-frequency data collected from worldwide populations of humans. The results were visualised using so-called PC maps depicting how the PC values of sampled population vary across geographic space (with each PC displayed on a separate map). The observed spatial patters were then explained in terms of population processes such as past population expansions and admixture events. Although this particular use of PCA is currently being questioned (Novembre and Stephens, 2008; Francois *et al.,* 2010), PCA became a standard statistical tool for the analysis of population genetics data. Initially PCA was used as a data-exploration tool but recent theoretical developments have extended its use to test for population structure and to cluster genetic samples (Patterson *et al.,* 2006; Jombart *et al.,* 2010). In this section we focus on its use to test for heterogeneity.

The simplest but least rigorous approach to investigate heterogeneity using PCA is through the spatial representation of PC scores on maps, as done in the seminal work of Cavalli-Sforza *et al.* (1994). Gaggiotti and Gascuel (2011) applied this approach to investigate the stock structure of North Pacific minke whales using individual genotype data. They used the *adegenet* package (Jombart, 2008) for the R software (R development Core Team, 2008) to carry out the PCA and then constructed maps of PCA scores using Kriging as implemented in the R package *fields*. Areas with similar PCA scores can be considered genetically homogenous and abrupt changes in PCA scores represent genetic discontinuities. This visual approach needs to be followed by a formal test. Gaggiotti and Gascuel (2011) use the Tracy-Widom test developed by Patterson *et al.* (2006). This test is based on rigorous statistical theory that also allows one to estimate how much data will be required to find population structure, given a level of genetic divergence. The test is applied to each eigenvalue and provides an approximate way of estimating the number of populations. In principle, given a large enough number of markers and assuming there are *K*

distinct populations represented in the sample, we expect to find $K-1$ eigenvalues and $K-1$ corresponding eigenvectors that are significant. Thus, we can reject panmixia if the first eigenvalue is significant, which would mean that our sample contains individuals from a least two populations. The test has been implemented in the program SMARTPCA and included in the software package EIGENSOFT 3.0[1].

A more recent method to estimate the number of genetic clusters combines Discriminant Analysis (DA) and PCA (Jombart *et al.*, 2010). PCA is not really appropriate for obtaining a clear picture of between-population variation because it aims to summarise the overall variability among individuals, which includes both within- and among-groups genetic variability. What is needed is a method that focuses on between-group variability. This is what DA does. More precisely, DA produces synthetic variables that maximise among-groups variation while minimising within-group variation. However, DA requires the number of alleles (i.e. the variables) to be less than the number of individuals (the sample size). Also, it requires alleles and loci to be uncorrelated. To resolve this problem, Jombart *et al.* (2010) proposed to transform the data using PCA and then use the PCA scores to carry out the DA. When there is no prior information about groups, it is possible to use *K*-means clustering of PCs to identify groups of individuals. In this case, it is then possible to use the Bayesian Information Criterion (BIC) to infer the number of genetic clusters.

Landscape genetics, units = individuals

Landscape genetics merges the fields of molecular population genetics, landscape ecology and spatial statistics to study how features of the landscape affect processes of microevolution such as gene flow, natural selection, local adaptation and genetic drift (Manel *et al.*, 2003; Storfer *et al.,* 2007; Holderegger and Wagner, 2008; Segelbacher *et al.,* 2010). Unlike most traditional approaches to the study of population subdivision, landscape genetics does not necessarily require the *a priori* assignment of individuals to populations, although that can be done as well. Often in landscape genetics the individual is the operational unit, and thus the experimental design can be free of assumptions (and their potential biases) about population structure. This also allows for studies on a much finer scale than is possible for analyses that require individuals to be grouped together into putative populations. Landscape genetics typically focuses on contemporary evolutionary processes; in this respect, it differs from phylogeography (another approach that links genetics to geography), which seeks to identify phylogenetic lineages and thus looks at the consequences of population structure over considerable periods of time.

The general aim of landscape genetics is to identify genetic discontinuities (i.e. population structure) that correlate with landscape and/or environmental features. From a conservation biology perspective, understanding the functional connectivity of populations across landscapes is an important goal (Van Dyck and Baguette, 2005). One primary application is to aid in the identification of management units, and the method can be particularly useful in cases where it is difficult to assign individuals to populations *a priori*. This is especially important for continuously distributed species, for which any *a priori* grouping into putative populations risks being arbitrary. Spatial patterns that potentially can be identified include clines, genetic discontinuities, metapopulations, isolation by distance, and random patterns. In the final analysis, validation of patterns of functional connectivity is ultimately based on gene flow estimates (Cushman *et al.*, 2009). Obviously, obtaining accurate locality data for each individual is a critical requirement of sampling. Genetic and statistical methods are used to identify spatial genetic patterns and their correlations with landscape and environmental features. Landscape genetic analyses can use any of the typical molecular markers, including mtDNA, microsatellites, SNPs, AFLPs, and allozymes. Neutral markers are useful for identifying some ecological and demographic processes, but targeted genes undergoing selection can potentially provide novel information about the genetic basis of adaptation, adaptive differentiation, and speciation. Sampling across ecological gradients can help to identify environmental factors that might drive selection.

Landscape genetic studies in which individuals can be *a priori* assigned to populations can employ standard population genetic methods such as *F*-statistics and assignment tests (see Section (4)(b)). However, in cases where the individual is the operational unit, application of Wright's Neighborhood concept can be useful, particularly for detecting patterns of isolation by distance. For continuously distributed species, Wright (1943) defined the genetic 'neighborhood' as the basic unit of population structure. In Wright's model, neighborhood size is $4\pi D\sigma^2$, where D is population density and σ^2 is mean squared parent–offspring distance along one axis. Statistical procedures for use in landscape genetics include Mantel's test for isolation by distance (Mantel 1967). Regression analysis of genetic distance with geographic distance allows the estimation of neighborhood size and thus dispersal distance. Spatial autocorrelation analysis allows one to identify spatial patterns such as clines and test whether distance is the main determiner of population structure. Bayesian clustering methods (see Section (4)(a)(ii)) can be used to assign individuals to populations by a method that minimises HW or linkage disequilibria among groups. Multivariate analyses like Principal Components Analysis (PCA) are used to summarise variation among many loci across an area. Interpolation of the major components into a synthesis map gives insight into spatial patterns and can allow identification of patterns such as clines. Two popular methods (Barbujani, 2000) to identify genetic boundaries or discontinuities are Monmonier's algorithm, which visualises data contained in a genetic distance matrix on a geographical map to identify boundaries, and wombling, which 'locates boundaries across a surface for an interpolated variable (i.e. allele frequency surface) by searching for regions in which the absolute value of the surface slope is large' (Manel *et al.*, 2003).

Studies of landscape genetics often apply methods of Bayesian clustering to more objectively group individuals into populations. The program STRUCTURE (Pritchard *et al.,* 2000) is the most widely applied Bayesian clustering approach. However, because of limitations to this program and to the use of isolation by distance models in general, Segelbacher *et al.* (2010) suggest that a new trajectory for

[1]*http://helix.nih.gov/Applications/eigensoft.html*.

investigations in landscape geneticists will be the study of population genetic structure and isolation by distance under computer-simulated dynamical models.

Statistical approaches also exist that allow for the correlation of genetic patterns with environmental variables. Mantel tests are used to correlate genetic distance to a wide variety of variables. Alternatively, canonical correspondence analysis (CCA) can be used to relate genetic diversity to environmental factors, at the same time testing for environmental factors that explain variation in genetic diversity. Geographical information systems (GIS) visualise spatial genetic patterns. Because GIS allows overlay of environmental or landscape variables onto genetics data, it facilitates development of hypotheses about the cause of spatial genetic patterns.

There are limitations and constraints to the use of landscape genetics, as with all analytical methods. Some analyses assume random mating, which might be violated by many cetacean populations. In addition, gametic disequilibrium and departures from HW proportions can result from processes besides population structure, including small populations, bottlenecks, inbreeding and admixture; hence, these factors can obscure patterns related to population structure. Schwartz and McKelvey (2009) showed that the program STRUCTURE identified varying numbers of clusters which were dependent on sampling scheme. STRUCTURE can thus provide misleading population assessments because genetic gradients or similar patterns likely are common in nature. A more general limitation is that isolation by distance mathematical models strongly depend on the assumption of migration-drift equilibrium (Rousset, 1997). Studies of great whales will often deal with non-equilibrium populations. Indeed, as a final caveat, the historic over-harvest of great whale populations and its resulting bottlenecks could represent a significant violation of the assumptions of certain analyses, particularly those that assume demographic and genetic equilibrium.

Analysis of close kin

An alternative to data analyses that employ allele frequencies is approaches utilising measures of relatedness among individuals. It follows intuitively that a random, finite sample of individuals from a small population will contain more pairs of closely-related individuals than a similar-sized sample from a larger population. The same is the case for dispersal; when the dispersal rate is high, the probability increases that pairs of closely-related individuals are sampled in different populations, whereas that will be rare when the dispersal rate is low. Thus, the fraction of all pairs of individuals that are closely related within and among populations is a function of the population sizes as well as the dispersal rates among populations.

Kinship-based approaches have two main advantages compared to most other methods of population genetic analysis. First, detection of closely related individuals does not require any degree of genetic differentiation among putative populations. Most methods for detecting population structure perform poorly when allele frequency differences among the gene pools involved are low ($F_{ST} \ll 0.05$), a situation that occurs quite commonly with cetaceans. This is particularly true for clustering and assignment method, such as STRUCTURE, DAPC, and BAYESASS. Second, whereas most traditional methods for estimating dispersal depend on equilibrium assumptions and therefore integrate information across evolutionary time scales, kinship-based approaches provide information about contemporary evolutionary processes. The latter is typically more useful for conservation and management.

The expected proportions of pairs of randomly-sampled individuals that fall in different relatedness categories, within and among populations, can be derived analytically under simplifying assumptions (no migration, random reproductive success, etc.). Several implementations have been published during the recent years that rely upon restrictive assumptions like these (Peery *et al.,* 2008; Wang, 2014). The effects of these simplifications on estimates of population size and/or migration rates have not yet been explored in detail. Two factors are key to successfully applying kinship-based approaches: (1) the proportion of the population sampled must be sufficiently large that pairs of closely-related individuals are among the sampled individuals; and (2) a sufficient number of independent diploid loci must be genotyped to estimate relatedness with sufficient statistical rigor to achieve the study objectives (Wang, 2006; Kopps *et al.,* 2015).

Developing an experimental design that is both powerful and feasible is a non-trivial endeavor, as it depends on the underlying vital rates (survival and reproduction) as well as dispersal rates and population sizes – which usually are unknown parameters a study is attempting to estimate. For a given level of precision, the required sampling proportion is inversely correlated with population size and positively correlated with dispersal rates, but it also depends on the statistical methods used to estimate abundance or detect population structure. In the latter case, two different approaches are commonly employed: homogeneity testing or estimating dispersal rates.

Homogeneity testing relies on rejecting a homogenous distribution of close relatives across space or time. Økland *et al.* (2010) proposed this approach for identification of management units. Their assessment evaluated whether the number of pairs of individuals inferred to be closely related was higher than expected by chance. In this respect it is similar to a standard population genetic test of panmixia proposed by Moritz (1994) for allele frequency data. Økland and colleagues defined 'closely-related' to be either first-order (parent-offspring or full siblings) or second-order (half siblings or grandparent-grandoffspring). Using simulations, they showed that power to reject homogeneity was high (>90%) unless only a few (~5) genetic markers were used or migration rates were high ($m = 5 \times 10^{-3}$/year and a generation time of 20 years). Interestingly, the power to reject homogeneity does not increase linearly with the sampling fraction: power using 300 individuals sampled from populations of size 7,500 was roughly the same as for 400 individuals sampled from populations of size 15,000.

Estimating dispersal rates from the spatio-temporal distribution of pairs of closely-related individuals is also possible but relies on estimates (or simplified assumptions) of key demographic parameters, such as birth and mortality rates, as well as reproductive variance. Peery *et al*. (2008) estimated immigration rates in a seabird (the marbled murrelet, *Brachyramphus marmoratus*) from the number of

parent-offspring pairs contained in a sample of 271 birds taken from a population estimated at 660 individuals. A total of 70 pairs of individuals were inferred to be related as parent and offspring which yielded an estimate of an annual immigration rate at 4–6%. Wang (2014) published a likelihood estimator of *m* based upon genetic parentage, which he showed to have higher precision than BAYESASS and to be unbiased unless the number of parent and offspring pairs detected was very low (<10). However, the approach by Wang was based on a simple population and migration model (e.g. no migration by mature individuals), which is unlikely to apply to most cetacean populations. Kinship-based estimations of abudance have also been undertaken (see Section (3)(a)).

A key limitation in the above approaches is the reliance on assigning each pair of individuals to a specific relatedness category – typically parent and offspring, full siblings (common in many species but less likely to be relevant for cetaceans), half siblings, or unrelated. While parent-offspring and full-sibling pairs generally can be identified with reasonably high statistical power, power to assign individuals to other relatedness categories is typically much lower (Wang 2006; Skaug *et al*., 2010; Kopps *et al.*, 2015). Unfortunately, in cetaceans as in other organisms, the relatedness categories that can be detected with the highest certainty (parent-offspring pairs and full siblings) usually compromise a much smaller fraction of a random population sample compared to more distantly related pairs of individuals. Accordingly, adding more distant relatedness categories is likely to increase the accuracy and precision of kinship-based estimates of abundance and dispersal rates substantially. However, these more distant relatedness categories remains elusive; even with substantial 'genomic' scale datasets, the part of the genome that is identical by descent decreases rapidly with the degree of relatedness and consequently so does the ability to assign pairs of individuals to a specific relatedness category (Kopps *et al.,* 2015). Furthermore, realistic evaluations of how much statistical power will increase if hundreds or thousands of SNP loci are used must explicitly account for the fact that those loci must be packaged into a small number of chromosomes, and this substantially limits the total amount of information that can be extracted from the data (Thompson, 2013).

(b) Describing population structure

(i) Estimating degree of divergence

F_{ST} and related measures

The so-called fixation index, F_{ST}, is one of the *F*-statistics originally suggested by Sewall Wright (see, *e.g.* Hartl and Clark 1989). F_{ST} is a measure of population divergence under a scenario where one investigates whether a population is genetically subdivided into partially or completely isolated subpopulations. These putative subpopulations have to be defined *a priori*. F_{ST} then can be used to evaluate whether gene frequencies are indicative of restricted gene flow across these subpopulations (rejection of panmixia). For conservation and management issues, F_{ST} is often used as a tool for analysis of molecular data to identify 'stocks,' 'management units,' or 'units to conserve.'

A common formulation for F_{ST} is:

$$F_{ST} = \frac{H_{eT} - H_{eS}}{H_{eT}},$$

where H_{eT} is the expected heterozygosity in an entire population and H_{eS} is the mean expected heterozygosity within the subpopulations. The expected heterozygosity is the probability that a randomly chosen individual is heterozygous at a given locus. It is calculated from allele frequencies and ranges from 0 (all individuals are expected to be homozygous, indicating that there is only a single fixed allele at an investigated locus) to 1 (all individuals are expected to be heterozygous, indicating that all alleles at an investigated locus are different). The underlying logic of F_{ST} is as follows: Under the null hypothesis of no divergence among subpopulations, H_{eS} measures for single subpopulations are unbiased estimators of the heterozygosity H_{eT} of the entire population. Hence, $H_{eT} - H_{eS} = 0$ and as a consequence F_{ST} also equals 0. Maximum divergence is reached if every subpopulation is fixed for a different allele. Then, $H_e = 0$ in every subpopulation, so $F_{ST} = (H_{eT} - 0)/H_{eT} = 1$.

If more than one locus is investigated, different ways of calculating multilocus F_{ST} (either averaging single-locus F_{ST} values, or averaging expected heterozygosities across loci and using those average values in Equation 1) can yield different values.

Fixation indices calculated from haplotype and/or gene diversity are sometimes identified with a different symbol (G_{ST}), but – as formula and concept are identical – most authors have adopted F_{ST} as the 'one-fits-all' name, regardless whether it is calculated on the basis of heterozygosity, haplotype diversity, or gene diversity. Because the theoretical distribution of F_{ST} remains unknown, significance of any deviation from the null hypothesis of panmixia is typically evaluated by permutation analysis. Rejection of the null hypothesis means that F_{ST} is significantly larger than 0, indicating that the subpopulations exchange fewer individuals than are expected under totally random dispersal (= panmixia).

The strength of F_{ST} as a measure of population divergence is threefold: (a) it is easily calculated and tested for significance for various types of molecular data; (b) the interpretation is straightforward, i.e. adoption ($F_{ST} = 0$) or rejection ($F_{ST} > 0$) of panmixia; and (c) it provides a measure of the fraction of total genetic variance that is apportioned among subpopulations. This has made the F_{ST} a very commonly used measure of genetic divergence.

There are, however, a number of known limitations to F_{ST}:

(1) It is based on diversity estimators (i.e. heterozygosity, haplotype diversity) that solely classify alleles/genotypes according to identity/non-identity, without any consideration of the level of difference among the alleles/genotypes;

(2) The theoretical maximum of $F_{ST} = 1$ can only be reached if each subpopulation is fixed for a single unique allele. If there is variability within any subpopulation, the maximum F_{ST} is $(1 - H_{eS})$. Unfortunately, this limit to the maximum F_{ST} is often overlooked. The maximum value

for F_{ST} is smaller the more variable a marker is, and the effect can be especially dramatic for microsatellites, which often exhibit high H_{eS} (over 0.9, in which case the maximum F_{ST} is only 0.1). In the extreme (yet possible) scenario of two subpopulations completely divergent (i.e. not sharing a single allele), but both with H_{eS} approaching 1 (i.e. all individuals are expected to be heterozygous because of high allelic diversity), F_{ST} becomes meaningless, as its theoretical maximum is then 0 (see Fig. 1 in Jost 2008 for a graphical representation);

(3) While the significance of F_{ST} is a good indicator of deviations from panmixia, the same information can be obtained more directly by testing equality of allele frequencies in different samples (Section (4)(a)(i)); and

(4) Translation of F_{ST} into dispersal estimates, although mathematically straightforward for populations assumed to be in equilibrium between genetic drift and gene flow, is very challenging for most real-world situations (see Section (4)(b)(ii)1).

Alternative measures of divergence have been suggested to overcome some of the limitations of F_{ST}:

(1) Some measures of population divergence incorporate the degree of divergence among genotypes. The F_{ST}-related measure Φ_{ST} explicitly accounts for mutational differences (Excoffier et al. 1992). For sequence data, divergence can also be expressed as nucleotide diversity (π), see Section (2)(a):

$$\pi = \sum_{ij} p_i p_j \pi_{ij}.$$

Analogously to H_{eT} and $\hat{H}_{eS}$, one can calculate π_T as nucleotide diversity for the entire population and mean π_S as the average nucleotide diversity within subpopulations. Subtracting mean π_S from π_T yields a measure of average % sequence difference due to divergence among subpopulations (Quinn and White, 1987). For microsatellites, one can reasonably assume a step-wise mutation model, in which case alleles of similar allele size are likely to be more closely related. This is incorporated in the divergence measure R_{ST} (Slatkin, 1995):

$$R_{ST} = \frac{\overline{S} - S_w}{\overline{S}},$$

where S_w is the sum over all loci of twice the weighted mean of the within-population variances in allele size within subpopulations and $\overline{S}$ is the sum over all loci of twice the variance in allele size in the entire population.

(2) From a mathematical point of view, the theoretical range of possible F_{ST} values can be easily re-adjusted to the interval [0,1] by a correction factor (Hedrick, 2005):

$$F_{ST\ adjusted} = \frac{H_{eT} - H_{eS}}{H_{eT}(1 - H_{eS})}.$$

This correction, however, has not been widely applied. In addition, there is so far neither a theoretical nor an empirical evaluation of the impact of such a correction on the distribution of F_{ST}. As a consequence, it remains unclear whether $F_{ST\,\text{adjusted}}$ also suffers from the limitations mentioned under (3) specifically, that identical values do not necessarily imply an equal level of divergence.

(3) It has been recently argued that a true measure of differentiation should decompose overall diversity into two parts: (a) the variance within the subpopulations and (b) the variance because of divergence among subpopulations (Jost, 2008). To yield an unbiased measure of the divergence component in heterozygosity, this author suggests a formula very similar to that above, i.e.

$$H_{ST} = \frac{H_{eT} - H_{eS}}{1 - H_{eS}},$$

where H_{ST} is the heterozygosity component caused by divergence among subpopulations. To yield the final estimator of differentiation (D) with values on the interval [0,1], H_{ST} is multiplied by (n/(n – 1)) where n is the number of subpopulations (Jost, 2008). As neither H_{ST} nor D are implemented in the most widely used data analysis software packages, these measures are rarely used. In addition, Jost's D has been found (a) to be (like F_{ST}) also affected by levels of heterozygosity, and (b) not to be conceptually related to basic population genetics quantities (effective population size, gene flow), rendering it potentially inferior to F_{ST} as a standard measure of population differentiation (Ryman and Leimar, 2009).

In summary, F_{ST} is an easily calculated measure of population differentiation and a good indicator of the presence/absence of panmixia. It has, however, limitations due to (a) neglect of the amount of genotype divergence and (b) dependence of the divergence estimator upon the within-subpopulation diversity. Alternatives to F_{ST}, which could at least partially overcome these limitations are available, but some are only rarely used and/or not generally accepted (in particular, H_{ST} and D). Given its widespread application and its sound theoretical foundation, F_{ST} can be considered as a valuable measure to infer population structure from genetic data, if *a priori* information about putative subpopulations is available. However, it is recommended to bear the limitations of F_{ST} in mind and to treat numerical comparisons among F_{ST} values with caution.

Isolation-by-distance/landscape genetics, units = samples

If populations are inferred to have genetically diverged, the stratifying factors remain to be identified. When inferred populations are geographically separated, a possible stratifying factor is simply the geographic distance separating populations – a pattern known as isolation-by-distance.[29] To test for such isolation-by-distance, the correlation of a metric of pairwise genetic divergence between populations (e.g. F_{ST}) to their respective geographic distance is estimated. The pairwise distance measures (both genetic and geographic) are arranged in distance matrices of equal organisation and rank, and from these matrices, the standard Pearson's Parametric Product-Moment-Correlation Coefficient (r) can be calculated. As any population is included repeatedly in the analysis (as part of multiple

pairwise comparisons), statistical significance of r is evaluated by a matrix correlation test (Mantel test) by permutation analysis, i.e. randomly permutating the order of elements in the matrices (i.e. exchanging rows and columns in tandem) and comparing the resulting correlation coefficients to r derived from the original matrices.

Any genetic divergence measure (e.g. F_{ST}, R_{ST}, π) can be employed for this use. The geographic distance is – depending on landscape features – not necessarily the shortest geographic distance between two populations, but the shortest distance along a possible migration route for the organism under study. For cetaceans, this is considered the shortest distance by water (see Wiemann *et al.,* 2010 for a cetacean example).

Populations can be stratified by environmental factors other than geographic distance. The general aim of landscape genetics is to identify genetic discontinuities (i.e. population structure) that correlate with landscape and/or environmental features. With *a priori* defined populations, any environmental factor for which pairwise quantitative measures can be calculated across populations can be used to produce a matrix, which subsequently can be compared to genetic distance with a Mantel test along the lines outlined above.

Indirect methods based on population genetics models are an alternative to the estimation of migration rates by mark-release-recapture methods. They typically require far less effort in the field, although a carefully planned sampling design aimed at collecting tissue samples for DNA extraction and analysis is needed for the subsequent inference of population size and migration rates.

Estimation of migration rates requires use of model-based methods, typically maximum likelihood or Bayesian approaches, that make various assumptions about the demographic history of populations. Broadly speaking, two main groups can be distinguished: (1) methods that provide estimates of migration rates averaged across long-time scales (often referred to as 'evolutionary time') and (2) methods that estimate recent migration rates. The former group of methods can be further subdivided into (*i*) methods that assume an equilibrium between migration and genetic drift, and (*ii*) non-equilibrium methods that allow for recent population divergence. The latter are described in a separate section below, so in what follows we only describe the former type of methods.

(c) Estimating migration

(i) Methods that assume migration-drift equilibrium

The first group of migration-estimation methods considers a scenario where extant populations have diverged from the ancestral population sufficiently long ago that we can assume that an equilibrium between migration and genetic drift has been achieved. The earliest method in this category is based on the well-known relationship between F_{ST} and the effective number of migrants, mN_e (the product of the population effective size, N_e, and the per-generation migration rate, m): $mN_e = 1/4\ (1/F_{ST} - 1)$. This method, first proposed by Sewall Wright (1931), has come under criticism due to the simplistic assumptions (constant and equal effective sizes and migration rates across populations) of the underlying genetic model (e.g. Waples, 1998; Whitlock and McCauley, 1999). Recent progress in population genetics theory and statistics has led to the development of methods that avoid these assumptions. Note, however, that they still make other simplifying assumptions (see below).

Several of the methods that fall in this category are based on coalescent theory (Kingman, 1982). In a nutshell, the so-called 'coalescent' describes the properties of samples of genes based on their genealogical and mutational history (Rosenberg and Nordborg, 2002). The probability that two genes in a sample 'coalesce' as we go backwards in time depends on population size and the co-occurrence of genes in the same deme. Therefore, migration between populations influences the coalescence process. Analytical models have been developed to describe the properties of gene geneologies in subdivided populations (Hudson, 1990; Wakeley, 2004) and have been used to implement several statistical methods to estimate mN_e. All of these methods carry out the joint estimation of the gene genealogy, mN_e and several other parameters and, therefore, require computationally intensive methods such as MCMC (Brooks, 1998).

It is important to keep in mind that the estimates of migration provided by coalescent approaches can represent long-term averages across the evolutionary history of the species. Thus, they are not well suited to management situations that require estimates of contemporary migration.

Two sister programs, MIGRATE (Beerli, 2009) and LAMARC (Kuhner, 2006), allow the estimation of migration rates using a wide range of molecular markers, including microsatellites and mtDNA. MIGRATE estimates effective population size and migration rates between n populations, allowing asymmetric migration and different effective subpopulation sizes.

LAMARC has many of the features of MIGRATE, but can also estimate additional population genetics parameters, such as population growth, recombination rate, and selection. Both programs allow users to choose between maximum likelihood and Bayesian inference approaches and provide a wealth of detailed results ranging from point estimates (mean, mode) and confidence or credible intervals to profile likelihoods and curve files. Thus, users should be willing to climb a steep learning curve to take full advantage of these sophisticated methods.

Both methods provide estimates of migration and effective population sizes that are scaled by the mutation rate. Thus, effective population size is estimated by $\theta = xN_e\mu$ (where x = 4 for diploids and 2 for haploid species, N_e is effective population size and μ = the mutation rate). Migration rate is estimated as m/μ, where m is the migration rate per generation.

30

Summary and relevance to cetaceans:

Methods that assume migration-drift equilibrium provide estimates of migration rates that represent long-term averages over the evolutionary history of the species. An additional difficulty is that the estimates obtained are scaled by the mutation rate and, therefore, migration rate estimates useful for management can only be obtained if mutation rate estimates are available (and reliable).

(ii) Isolation with migration models

Evolutionary models that form the basis for genetic comparisons of different populations have typically been based on either of two extreme scenarios. One model considers populations that have been exchanging migrants at a constant rate for an effectively infinite period of time (an 'equilibrium migration model'). The other considers populations that descended from some common ancestral population at some point in the past, and have since then evolved independently without gene flow (an 'isolation model'). Many of the population comparisons in the literature use measures based on Wright's inbreeding coefficients (especially F_{ST}; Wright, 1965 – see section 4(b)(i) above). The problem is that, by itself, a measure of genetic distance such as F_{ST} can't differentiate between the two scenarios described above. A low F_{ST} could mean either a relatively high rate of gene flow over time, or the recent cessation of gene flow altogether (because differentiation by genetic drift or selection will take some time to accumulate). Isolation with migration (IM) models allow both cases to be considered together and therefore (potentially) distinguished. This is important from a management perspective when trying to distinguish population segments for which there is some ongoing connectivity from populations that have become isolated (and are potentially incipient species). Boundaries to gene flow can be difficult to identify in the ocean, and several examples of cryptic cetacean species have been identified in recent years (e.g. Wada *et al.,* 2003). There are also a number of poorly resolved radiations of species that are polyphyletic (e.g. species in the subfamily Lissodelphininae), and species that have been variously recognised as single species or divided into multiple sub-species or species (e.g. *Delphinus delphis* and *Orcinus orca*; see below for an example). IM can help resolve these questions associated with alpha taxonomy, and thereby better identify management units. IM models can also help identify the relevant mechanisms and processes that regulate gene flow, and thereby improve the efficiency of management strategies.

An early example of an IM model was described by Wakeley (1996), based on the observation that the variance of pairwise nucleotide differences (a measure of the extent of diversity) is smaller under isolation than under equilibrium migration. However, any single statistic such as this will necessarily leave out much of the complexity likely to be reflected in real demographic histories. This fact inspired an approach that could simultaneously estimate multiple population size/migration parameters, initially described by Nielsen and Wakeley (2001). It is a two-population, one-ancestor model, and the parameters estimated are the effective population size of each of the three populations (N_1, N_2, & N_A), the migration rate from population 2 into population 1 (m_1), the migration rate from population 1 into population 2 (m_2), and the splitting time between the two extant populations (t). Under the assumption that $N_1 = N_2$ and $m_1 = m_2$, there are four total parameters to estimate. Thus, the IM model simultaneously estimates parameters related to both the equilibrium-migration and isolation models. The input is sequence data from a single locus, and the model assumes an infinite allele model of mutation (all new mutations are novel), selective neutrality, constant population size, no recombination, and a closed system (no other populations exchanging alleles with the focal populations). The model was tested in both likelihood and Bayesian frameworks. A later version allowed for the application of a finite site model of evolution (Palsboll *et al.,* 2004). This approach is especially appropriate for the analysis of mtDNA sequence data (no appreciable recombination), but a finite site mutation model (such as HKY; Hasegawa *et al.,* 1985) is better suited to mtDNA where mutation rates are comparatively high. Another limitation of the original model was addressed in a paper by Hey and Nielsen (2004). The Nielsen and Wakeley (2001) method (implemented in the program MDIV) was based on a single locus, but single gene trees can be strongly affected by stochastic processes (including lineage sorting, Neigel and Avise, 1986) or natural selection, and become unrepresentative of the true history.

Hey and Nielsen (2004) extended the IM method to include multiple loci (implemented in the program IM). The parameterisation is also different in this model, where population size, migration rate, and splitting time are all scaled by mutation rate. It is assumed that the loci included have been drawn from all loci at random (that is, that none are atypical with respect to the depth of the gene tree or degree of gene flow), and as before, that they have evolved neutrally. Additional parameters are required to accommodate the additional loci (scalars to account for differences in mutation rate and the mode of inheritance). As described in a later paper (Hey, 2005), the IM program can also model populations that are expanding or contracting (through the addition of the 's' parameter). Three mutation models are available in the IM program: the infinite sites model (often appropriate for nuclear genes that evolve relatively slowly), the HKY finite-site model (better suited for relatively fast evolving sequences, such as mtDNA), and the stepwise mutation model (appropriate for microsatellite DNA loci). Note that mutation rates are input as per locus per year.

The programs that test these models employ Markov chain Monte Carlo (MCMC) simulations, and a major challenge with this approach is to verify that the output corresponds to something like reality. It is, unfortunately, quite easy to generate data that looks plausible but is in fact quite wrong. For MCMC analyses a number of precautions are necessary, including the need to run the same simulations for a sufficient number of steps, and at least three times. The first run allows some assessment of the effectiveness of the chosen input parameters, while the last two full runs allow confirmation that independent runs (with different random number seeds) give equivalent results. While a detailed explanation is beyond the scope of this summary, the key objective is for the simulation to converge on the 'stationary distribution' (the distribution that you want to estimate with your sample of parameter values). Convergence is necessary if the results are to be credible, and therefore a variety of indicators need to be assessed during the progress of the run. A description of these indicators and a more detailed explanation of the problem can be found in the support documents for the program IM. A consequence of this is that

the simulations sometimes need to run for a very long time (often weeks or more, depending on the speed of the computer processors). One feature in the program IM that helps is the ability to run a number of chains in parallel (called 'Metropolis coupling'). A further advance was implemented in a second version of the program (IMa; Hey and Nielsen, 2007), whereby approximations of the values for some parameters allows the program to progress more quickly.

A few applications of the IM approach have been published for cetacean species. In one study, fin whales inside and outside the Mediterranean basin were compared, and a model of ongoing gene flow (at about 2 females per generation) was shown to be better supported that the alternative of recent isolation (Palsboll *et al.,* 2004). Pastene *et al.* (2007) used IM to explore radiation and speciation of the common minke whale. In another study, the timing of the founding of populations of killer whales currently using coastal habitat was found to post-date the last glacial maximum (after habitat was released from under ice), and ongoing gene flow was indicated between populations of different ecotypes in the eastern North Pacific (Hoelzel *et al.,* 2007). These same ecotypes have recently been proposed as different species (Morin *et al.,* 2010), though this was based only on mtDNA (mitogenome) sequences. Nuclear markers suggested ongoing gene flow both from IMa (Hoelzel *et al.,* 2007) and individual genotype and parentage analyses (Pilot *et al.,* 2010).

An important limitation of the model (as applied in the program IMa) remains. The two-population, one-ancestor model means that there should not be other populations more closely related to the sampled populations than they are to each other, and that no unsampled ('ghost') populations have exchanged genes with either the focal populations or the ancestral population. Violations of these assumptions can inflate the apparent size of the ancestor population and make estimates of gene flow between the focal populations (which might in fact be signals of gene flow through intermediaries) misleading. To help address this problem, a new version of IMa (IMa2) has now been released which allows inclusion of up to 10 populations (including ghost populations; Hey 2010). The difficulty with using a multi-population model, however, will be the introduction of new parameters to estimate, as well as more expansive requirements for sampling. This analysis would need to be supported by a large number of highly informative loci to produce reliable results; even then, obtaining reliable estimates for closely related populations is likely to be very challenging. As described in Section (3)(b)(i), DNA capture methods and next generation sequencing will facilitate the acquisition of sequence data from large number of loci and the application of the IMa2 method.

(iii) Methods that estimate contemporary migration

Assignment methods

Efforts to estimate levels of connectivity from genetic data have traditionally relied on equilibrium models that integrate information over evolutionary time periods (see Section (4)(b)(ii)1). The last decade has seen increasing interest in so-called 'assignment methods' that do not require equilibrium assumptions and instead can estimate contemporary patterns of migration over time frames encompassed by the samples. 'Assignment tests' (Paetkau *et al.,* 1995; Manel *et al.,* 2005) are a type of discriminant function analysis in which the discriminant functions are based on genetic traits that differ in frequency among potential source populations. Multilocus genotypes are used to 'assign' individuals to the most likely source population, guided by learning samples collected from potential sources. If an individual is assigned to a population other than the one it was sampled from, it can be inferred that the individual is a recent migrant (Waser and Strobeck 1998; Berry *et al.* 2004). The program *GeneClass* (Piry *et al.,* 2004) includes several different assignment test methods and offers the user various options for attempting to identify first-generation migrants. Other programs attempt to identify second-generation migrants (Wilson and Rannala, 2003) or estimate the fraction of genes in each individual that are derived from each population (Pritchard *et al.,* 2000).

Assignment methods have some advantages for estimating migration: they don't require one to assume migration-drift equilibrium, as do most standard models; they can potentially provide very detailed information about connectivity (both magnitude and direction); and they provide information about contemporary dispersal, which might be of interest for a variety of reasons. However, assignment methods also have some substantial limitations for studying dispersal. First, these methods provide information about movement of individuals but not reproductive success of the migrants; therefore, they do not provide any direct information about gene flow. Second, assignment methods provide information about dispersal only for the time frames encompassed by the sampling. Because dispersal is a stochastic process, samples taken from only one or a few years might not provide a representative picture of migration. This can be contrasted with equilibrium models, which can provide an estimate of long-term patterns of gene flow from samples taken at a single point in time. For any given application, these two factors might or might not represent serious limitations, depending on the nature and objectives of the research program.

A third factor – statistical power – is potentially a more general limitation on use of assignment methods to study contemporary dispersal. Power to detect migrants with genetic methods depends on two things: the amount of data one has (samples of individuals, gene loci, and alleles), and the magnitude of genetic differences among populations. The researcher has control over the former but not the latter, and therein lies a conundrum: power is highest when genetic differences among populations are large, but in that case migrants will be rare and difficult to detect without a very ambitious sampling program; conversely, if migration is high enough to provide reasonable prospects for finding migrants, the resulting levels of gene flow should erode most differences among populations, making it difficult to genetically distinguish migrants from residents.

Two examples illustrate the inherent difficulty related to power. Paetkau *et al.* (2004) used computer simulations to evaluate power to detect first generation migrants. They

found that even with fairly large amounts of data (50 individuals sampled per population; 20 microsatellite-like gene loci), power to detect true migrants was <50% when gene flow rates were high enough ($mN \geq 5$) to keep F_{ST} values below about 0.05. These conditions would apply to a substantial fraction of potential applications for cetaceans.

Second, the power issue sets up an inherent tradeoff between Type I errors (incorrectly labeling a resident as a migrant) and Type II errors (failing to detect a true migrant), either of which can seriously bias estimates of migration. Consider this hypothetical example: a group of populations with N = 100 individuals each are connected by 1% migration per generation (m = 0.01). This leads to mN = 1 (a low level of gene flow) and relatively large genetic differences among populations. Optimistically, assume that these large differences lead to ~100% power to detect migrants using assignment methods (as found by Paetkau *et al.,* 2004 for data-rich scenarios). So, a large sample would on average contain 1% true and correctly-identified migrants. But if the standard tolerance for Type I error is used (α = 0.05), then 5% of the sample would also be incorrectly identified as migrants. In this case, even with perfect statistical power, the estimate of migration rate (0.01 + 0.05 = 0.06) would be six times the true level. The only solution to this problem is to adopt a very low α level, but doing so is likely to compromise power unless genetic differentiation is very strong.

The conundrum regarding power does not necessarily represent an insurmountable problem for using assignment methods to study contemporary dispersal – for example, Berry *et al.* (2004) reported reasonably good agreement between genetic and mark-recapture estimates of dispersal in a series of populations of the grand skink, *Oligosoma grande,* for which F_{ST} values ranged between 0.04 and 0.11. However, the issues discussed above do indicate that careful attention to experimental design is essential, as is a realistic assessment of prospects for producing useful information. Three general strategies can help improve performance. First, in theory at least, it is possible to achieve high power to identify migrants among populations with very low levels of genetic differentiation, provided that arbitrarily large numbers of loci and alleles can be scored. The ability to do this with non-model species is rapidly increasing. Second, adopting a very low tolerance for Type I errors (e.g. $\alpha \leq 0.01$) can help reduce some of the most serious sources of potential bias, but this will likely compromise power unless genetic differences are moderately large and/or very large amounts of data are available. Third, the major challenges to these methods arise from uncertainty in identifying individual migrants. Using an analogue to Genetic Stock Identification (resolution of mixed-stock fisheries using genetic data – Shaklee *et al.,* 1999; Manel *et al.,* 2005), if focus is shifted from identifying individual migrants to estimating an overall migration *rate*, then uncertainty about origins of individuals might not preclude precise and accurate estimates of migration. This would require developing, or at least refining, some new analytical techniques. One software program, *BayesAss* (Wilson and Rannala, 2003) does actually attempt to estimate migration rate, but its performance with weakly differentiated populations has not been encouraging (Faubet *et al.,* 2007). Another Bayesian program (BIMr; Faubet and Gaggiotti, 2008) estimates the fraction of immigrant genes in subdivided populations and finds environmental variables associated with patterns of migration.

Finally, the conundrum regarding the inverse relationship between the level of migration and genetic differentiation largely disappears if the system one is analysing involves populations that historically have been strongly isolated (and hence are well differentiated genetically) but which are currently exchanging sizeable numbers of migrants. This non-equilibrium situation cannot persist for long unless the migrants have little or no reproductive success, but in the interim could provide a large number of migrants to sample *and* high power to distinguish them from residents. This scenario, in fact, is one that the Wilson and Rannala (2003) program was designed to study. It is not clear how often this scenario might occur with cetaceans.

Summary and relevance to cetaceans:
Assignment methods have considerable potential to provide insights into contemporary movement of individuals without resorting to equilibrium assumptions, as many widely-used models do. They are most powerful for studying genetically divergent populations, in which case individual migrants can be identified with a high degree of certainty. Two general scenarios are conducive to this type of analysis: (1) divergent populations that have only recently come into genetic contact; (2) analysis of mixtures of migrant individuals that co-occur but do not interbreed (e.g. if sampled on feeding grounds). For populations that regularly exchange even a modest fraction of individuals that successfully interbreed, genetic differences will generally be small, and considerable care is needed to evaluate results in the context of the inherent tradeoff between Type I and Type II errors in identifying migrants.

Close-kin analyses
See Section (4)(a)(ii) for a discussion of use of close-kin analyses to estimate dispersal.

(d) Mixture analysis
This method, often called Genetic Mixture Analysis (GMA) or Mixed-stock Analysis (MSA) or Genetic Stock Identification (GSI), uses samples from potential source populations (the 'baseline') to resolve mixtures of individuals from different populations. In general this does not involve hybridisation or 'admixture' (interbreeding); rather, it is used to analyse mixtures like those found on feeding grounds of cetaceans or in oceanic catches of salmon. If the goal is to identify the source populations for specific individuals (e.g. for a forensic application), assignment methods (see Section (4)(b)(ii)3.A) would generally be the method of choice. However, assignment methods allocate entire individuals to the most likely source, and this ignores uncertainty associated with each assignment. Therefore, if the underlying question is, 'What fraction of the mixture comes from each source population', then a form of GMA/MSA/GSI is better. These mixture programs deal with uncertainty by fractionally

allocating each individual based on its relative likelihood of coming from different sources.

The original maximum likelihood GSI model developed almost 40 years ago (Grant *et al.,* 1980) to analyse mixed-stock catches of salmon was constrained by the assumptions that (1) baseline allele frequencies were known without error, and (2) all stocks that might contribute to the mixture are represented by samples in the baseline. Smouse *et al.* (1990) developed an unconstrained model that relaxed both assumptions, by allowing baseline allele frequencies to vary based on estimated composition of the catch and by allowing a test of the hypothesis that one or more unsampled stocks are present in the mixture. However, this model was not widely adopted because it was computationally very demanding for computers available at the time. All of the likelihood and Bayesian models currently in widespread use (e.g. Pella and Masuda, 2001; Kalinowski *et al.,* 2007) now treat baseline allele frequencies explicitly as samples, but most are still constrained by the second assumption. STRUCTURE, however, has completely relaxed the assumption about comprehensive baseline samples, as no baseline samples are required at all. However, power is considerably reduced as a result, and even if the mixture can be successfully resolved into component gene pools it is not clear how to use this information to guide allocation decisions.

Performance of several MSA programs has been extensively evaluated using simulated mixtures (e.g. Koljonen *et al.,* 2005). Here is a brief summary of results:

(1) The main factors that influence precision are the magnitude of genetic differences among source populations and the sizes of the baseline and mixture samples.

(2) Because stock composition estimates are constrained to the biologically-plausible domain [0–1], estimates for stocks that are rare or absent from mixtures tend to be biased upwards. Although this bias might be small in absolute terms (typically a few per cent or less), it can mean that the estimate from that stock is several times its true value. This bias is reduced if the low-contributing stocks are genetically very distinctive. However, even with perfect power to genetically resolve the mixture, strong inferences about contributions of small stocks is limited by the size of the sample from the mixture.

(3) Misallocations most frequently are allocated to other populations that are genetically similar. When the desired level of precision is not achievable at the individual population level, one commonly used option is to create larger "management" or "reporting" groups, whose allocation is the sum of the allocations to stocks within the group (Hess *et al.* 2011).

(4) Care is needed in using simulations to assess power to resolve mixtures involving specific natural populations that have been sampled to obtain baseline frequencies. In these situations, overly optimistic conclusions about power can be reached unless proper cross-validation is used (Anderson *et al.,* 2008).

Bootstrapping or other resampling methods can be used to produce confidence intervals around point estimates. In the case of Bayesian GSI methods (e.g. Pella and Masuda, 2001; Gaggiotti *et al.,* 2004) it is possible to obtain credible intervals (the equivalent of frequentist confidence intervals). Some of the methods can accommodate haploid markers like mtDNA and/or other characters potentially useful for stock identification, such as scale patterns, otolith microchemistry (Smith and Camapana, 2010), or incidence of parasites.

(5) GENERIC/CROSS-CUTTING ISSUES

(a) Choice of markers

The choice of molecular markers depends on the purpose of a study. Markers can be classified by:

- inheritance mode (autosomal vs. unisexual)
- degree of polymorphism
- reliability of genotyping
- likelihood to be affected by natural selection

With regard to the inheritance mode, autosomal markers (autosomal microsatellites, autosomal SNPs, other sequences on autosomes) are inherited through both sexes and can hence provide a comprehensive assessment of population structure. For sex-specific assessment (e.g. for the estimation of gender-specific dispersal), markers located on the mitochondrial DNA can be used for inference of female population structure and dispersal, as mitochondrial DNA is inherited maternally. If a study aims at a female-specific analysis, male mtDNA should be excluded, because in long-lived species like cetaceans, inclusion of males will potentially bias dispersal estimates (Tiedemann *et al.,* 2000). In mammals (like cetaceans), male-specific dispersal can be analysed by targeting the Y-chromosome, although polymorphic Y-chromosome markers are not yet available for all species of interest.

For population genetic analysis, the most informatively polymorphic mtDNA region in most mammals (including cetaceans) is the Control Region (= D-loop). Cytochrome *b* has also been frequently targeted. With the advent of cheap sequencing technologies, entire mitochondrial genomes are now being used for population genetic analysis, yielding the most reliable phylogenetic inference of mitochondrial lineages. Yet, because mtDNA is effectively inherited as a single locus, even full mitochondrial DNA analyses cannot overcome all the limitations of single locus studies, such as cross-species introgression and ancestral lineage sorting. The very same issues apply also for the Y-chromosome, but an additional complication there might be a lack of knowledge regarding polymorphisms in the species of interest.

The most commonly applied autosomal markers are microsatellites and single nucleotide polymorphisms (SNPs). A single microsatellite locus potentially provides more

information than a SNP, because a single microsatellite can exhibit many alleles and their size difference may contain additional divergence information (see above). Therefore, microsatellites are still useful, especially when budgets are limiting. However, Next-Generation-Sequencing (NGS) methods such as restriction associated DNA sequencing (RADseq; Baird *et al.*, 2008) are being increasingly used. They provide much higher resolution (thousands of loci), the potential to investigate loci under selection, and do not require a genome reference sequence for the study organism. Furthermore, reference genomes are quickly being produced for non-model organisms (including cetaceans), which facilitates the re-sequencing of whole genomes and the interpretation of RADseq data. SNP analyses can be upscaled and automated relatively easily, and the limited amount of information provided by a single nucleotide polymorphism is overcome by the high number of SNPs that can be analysed per specimen.

Unless a study is targeted on adaptive processes, the general assumption for molecular markers to be used for population genetic inference is selective neutrality, i.e. none of the genotypes/alleles is favoured by natural selection. There are known examples for occasional positive selection among mtDNA genotypes. One should remember this possibility and test for selection, if mtDNA analysis yields unexpected results inconsistent with the results of microsatellites and other nuclear markers.

Despite a few indications of selective disadvantages of particular microsatellite alleles, microsatellite variation can be generally considered selectively neutral. A microsatellite locus can, nonetheless, be situated in close proximity to a functional locus, such that selection at that locus will impact the variation at the linked microsatellite locus. If a microsatellite locus is suspected or inferred to be affected by selection, it can be excluded from the population-genetic analyses. A powerful approach is to look for selective sweeps affecting particular loci with a genome scan of the respective region. Significant deviation from HWE can also point to selection, although HWE deviation can occur for a variety of other reasons as well (see Section (2)(b)).

SNPs can be situated in any part of the genome. Hence, they can be positioned both in non-coding and coding regions. It is generally assumed that – because SNP analyses typically cover very many SNPs – that selection at single loci will not significantly impact population genetic inference. If there is, however, sufficient evidence that a single polymorphic site is under positive selection, this locus might be of interest on its own right (potential for local adaptation), but can be excluded from analyses that assume selective neutrality.

If SNP data sets comprise hundreds or thousands of loci, there is high likelihood for linkage disequilibrium among particular sets of SNP data. Researchers analysing SNP data in a population genetic context should be aware of whether their analytical method of choice assumes linkage equilibrium and how sensitive the method is to violations of that assumption. For such methods, linked (= 'phased') SNPs can be combined to haplotypes/alleles which are subjected to further analysis. Alternatively, only one SNP out of an identified linkage group can be retained for subsequent analysis.

(b) Ascertainment bias

Ascertainment bias means a bias introduced by the specimens from which molecular markers are derived, prior to their application in population genetic studies. The bias can be particularly strong when the investigator selects markers based on levels of genetic diversity. Different populations often have different levels of diversity at different markers, so the markers that are developed depend to some extent on which populations are included in the initial screening panel. The biases arise when the resulting suite of markers is treated as representative of the entire genome, or when they are used to describe patterns of diversity in populations not included in the original panel.

The mitochondrial DNA of animals is rather small and well described. Within cetaceans, protocols are available to readily analyse any part of the mitochondrial genome in any species, such that ascertainment bias is a non-issue. In microsatellite development, people generally target polymorphic loci. Here, ascertainment bias can occur, if specimens of only one population or one geographic region are used for marker development. Furthermore, if only loci which are polymorphic are retained in these specimens, there will be a bias towards overestimation of genetic diversity in this population. Heterologous microsatellite loci (= loci developed for another species) are likely to be more prone to this effect than species-specific ones. Comparisons of levels of diversity among populations or species are most meaningful when they are based on random sets of markers, including monomorphic ones.

The most serious issue is ascertainment bias during SNP development. As the sole criterion for a SNP is to be polymorphic (with a threshold frequency for the minor allele, often ≥ 5%; e.g. Hao *et al.,* 2008), the choice of specimens for SNP development constrains the SNPs found to be those polymorphic among these specimens. If the screened specimens are not representative of the entire geographic region to be targeted, there will be a strong bias towards overestimation of diversity in the populations for which the SNPs were established and underestimation in the others. In addition, many potentially informative SNPs will remain undetected, if the specimens used for development are not truly representative. Some simulation methods take into account the potential confounding effect of ascertainment bias (e.g. Guillot and Foll, 2009). Nonetheless, to minimise these effects, it is highly recommended to (1) make the best effort to develop markers from a panel of specimens representative for the planned study, and (2) to keep track of this panel and its eventual limitations, should the markers later be used in a different context than they were originally developed for. In some SNP methods (e.g. RAD sequencing), the initial data analysis outputs both monomorphic and polymorphic loci. If both data sets are retained for later analysis, initially monomorphic loci could be repeatedly checked for polymorphisms when additional specimens are added to the analysis.

(c) Multiple testing

Some analyses routinely involve multiple tests of the same hypothesis (e.g. tests of HWE and LD, or pairwise tests of heterogeneity between populations). 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. Two points should be kept in mind when using this type of correction for multiple tests.

- The Bonferroni correction is widely known to be conservative and hence will fail to detect some actual departures from the null hypothesis.
- If the correction is performed, then the expectation is (with probability 1–*P*) that the number of adjusted significant tests will be zero. Therefore, even a single adjusted significant test cannot easily be attributed to chance and requires an explanation.

If a multiple testing correction is to be performed, a better option might be the 'false discovery rate' (FDR: the fraction of tests in which the null hypothesis is falsely rejected; Benjamini and Hochberg, 1995), 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 (see Waples, 2015).

(d) Mutation rates

The parameter $\theta = 4N_e\mu$ plays a key role in both theoretical and applied population genetics. θ is a composite parameter, proportional to the product of effective population size (N_e) and mutation rate (μ). Although this fact adds complexity to some analyses, it can be used to advantage by a simple rearrangement of the above equation:

$$N_e = \theta/(4\mu).$$

This means that if θ can be estimated from population genetic data (as is routinely done with microsatellite profiles, mitochondrial and nuclear DNA sequence data or single nucleotide polymorphism (SNP) data), then insights into N_e can be obtained if one can also estimate mutation rate. The effective population size that is estimated in this way is a long-term, or 'historic', N_e that depends (among other things) on the assumption that measured levels of genetic diversity reflect an equilibrium between mutation and genetic drift (see Section (3)(b)(i)). This general approach has a variety of practical applications, such as estimating historical effective population size; estimating divergence times between populations or species; and estimating population demographic patterns over time (see Beaumont and Rannala, 2004, Nielsen and Beaumont, 2009). Several factors, however, contribute to uncertainty and limit the practical usefulness of these approaches.

First, only four kinds of DNA bases occur (termed A, T, C, G for short), and DNA sequences are typically compared by counting the fraction of sites at which they have different bases. Once a mutation has occurred at a particular site (e.g. from A to T), a subsequent mutation at that site will still result in only a single difference compared to the reference sequence (if the mutation is from T to G or C) or will negate the original change (if the mutation is a back mutation from T back to A). This 'saturation effect' is of particular relevance for estimates of mutation rates for mtDNA, which are typically obtained by the 'phylogenetic method' that involves comparing sequences from different species. In addition to making duplicate mutations more likely, this approach introduces potential sources of error in developing calibration points for divergence times – typically derived from the fossil record, which is relatively poor for cetaceans.

Second, mutation rates can vary considerably among species and among regions of the genome within species. For many years, a '2% rule' was used for mtDNA, based on calculations using the phylogenetic method suggesting that, for vertebrates, the average rate of base substitution was about 2% per million years (Wilson *et al.,* 1985). However, rates vary among regions of the mtDNA molecule, and for the mitochondrial control region, the estimates are considerably higher (12% to 38% per million years in humans; see review in Henn *et al.,* 2009). Furthermore, the mtDNA control region itself is heterogeneous for mutation rate, with the central, very conserved, region being flanked by two 'hypervariable' regions (HVR1 and HVR2).

Finally, recent estimates of mutation rate over shorter time frames for intraspecific comparisons often differ greatly from those based on the phylogenetic method. An extensive analysis (Howell *et al.,* 2003) provided an estimate for the human HVR1 of 95% per million years (0.95 changes/site/million years). Similar approaches applied to other species, including *C. elegans* (Denver *et al.,* 2000) and Adélie penguins, *Pygoscelis adeliae* (Millar *et al.,* 2008), have also produced estimates that are 1–2 orders of magnitude higher than suggested by the '2% rule'. Henn *et al.* (2009) suggested that for humans, the high mutation rates decay after about 15,000 years, but for penguins the elevated rate seemed to extend back further in time (Millar *et al.,* 2008). These elevated rates have been reported for a broad range of species, including cetaceans (see Ho *et al.,* 2007). This effect can be associated with the time associated with slightly deleterious mutations segregating prior to fixation or loss.

From the form of Equation 6, it is easy to see that errors in estimating the mutation rate directly translate into proportional errors of the same magnitude when estimating N_e. For example, if mutation rate is underestimated by a factor of two, N_e will be overestimated by the same amount. This fact, together with the wide range of published estimates of mutation rates, has helped spawn much of the controversy that has surrounded some attempts to estimate historical N_e based on existing levels of genetic diversity. For example, Roman and Palumbi (2003) calculated that there must have been many more whales in pre-whaling oceans than had previously been thought, based on an estimate of mutation rate in the cetacean mtDNA control region derived from the phylogenetic method (e.g. Hoelzel *et al.,* 1991) – about 2% per million years. However, the relevant timeframe suggests that the much higher rate estimates derived from intra-specific genealogies might be more appropriate (e.g. Henn *et al.,* 2009, Millar *et al.,* 2008). If those higher rates were instead applied, the cetacean population size estimates would fall in line with what had been previously interpreted from historical catch data. The problem was illustrated in a recent phylogeny based on whole mitochondrial DNA genomes for delphinid species, focusing on the genus *Tursiops* (Moura *et*

al., 2013a). Both bio-geographic and fossil calibrations were incorporated into the tree, and it was clear that node dates were biased upwards when only the fossil dates were used, and downwards when only the bio-geographic calibration was used.

In summary, current levels of genetic diversity and other patterns in the DNA of contemporary populations contain information about historic size and demographic processes. However, deciphering this information is tricky and depends heavily on obtaining a reliable estimate of mutation rate. It is not enough to have an estimate of mutation rate for the focal species; it is also important to have estimates for the regions of the genome that produced the genetic data being analysed, and to apply the correct mutation rate to the relevant time frame – higher rates for more recent events (Ho *et al.,* 2005).

(e) Sampling and experimental design

Experimental design involves shaping key features of an experiment to maximise the likelihood of distinguishing among competing hypotheses. A poor experimental design can result in failure to reject an incorrect hypothesis or rejection of the true hypothesis because of (i) insufficient amounts of data or (ii) incorrect assumptions in the analysis. Such pitfalls can be avoided by conducting a prospective assessment, typically utilising computer simulations to generate 'virtual' genetic data under each competing hypotheses. The *in silico* data generated in this manner are then subjected to the planned data analysis, thereby enabling an assessment of the accuracy and statistical power in the planned analyses. In other words: is the planned study designed in a manner that will permit a reasonably rigorous assessment of the competing hypotheses?

In the context of the IWC SC's work, genetic methods are primarily employed to define management units and to estimate connectivity among populations, as well as historical or present abundance. As this section is for illustrative purposes, we will use the detection of management units, or stocks, as an example, but the same general principles apply to other aspects. Conducting a prospective assessment has the additional advantage that it forces the investigator to formulate the hypotheses that are to be tested in an explicit and qualitative manner. Key parameters, such as migration rates, population sizes, and mutation rates, as well as temporal and spatial changes in these parameters, need be specified in order to generate the *in silico* genetic data. Such translation of management objectives into demographic rates, and further their effects upon the population genetic make-up, might not be straightforward and can involve substantial amounts of work.

One main obstacle in conducting a reasonably exhaustive prospective assessment is that the number of possible combinations of parameter values increases rapidly, and it quickly becomes infeasible to assess the entire range of parameter values. For this reason, it is advisable to limit the extent of a prospective assessment as much as possible. Estimates of the relevant parameters from available data can be used to select reasonable ranges of parameter values. In addition, there are also relatively hard limits upon the number of tissue samples that can be collected, as well as the number of loci which can be analysed, although this latter aspect has changed dramatically over the last few years.

As an illustration, for stock identification a prospective assessment could consist of estimating the statistical power to reject the hypothesis of a single stock across a range of different levels of genetic divergence, which in turn is a function of the population history and dispersal as well as mutation rates of the loci modeled. Alternatively, and perhaps more productively, would be to identify those dispersal rates for which management recommendations would change. Once such 'tipping points' have been identified, a prospective assessment can be used to guide how many tissue samples and how many genetic markers are necessary to determine whether the actual degree of genetic divergence is above or below the critical dispersal rates. A relatively simple example would be: how much data (individuals, loci, samples) would be required to reject homogeneity if the genetic divergence is at $F_{ST} = 0.02$?

A large body of published assessments have evaluated the statistical power (or precision) in more common population genetic inference methods, which might serve as a guide to a specific study (Bjorklund and Bergek, 2009; Cornuet *et al.,* 1999; Faubet *et al.,* 2007; Gaggiotti *et al.,* 1999; Larsson *et al.,* 2009; Morin *et al.,* 2009; Paetkau *et al.,* 2004; Ryman *et al.,* 2006; Waples and Gaggiotti, 2006). It might therefore be that the results of these more general assessments are sufficient to rule out some hypotheses without a need for computer simulations.

The simplest (and most common) approach used to generate *in silico* population genetic data are based upon coalescent theory (Kingman, 1982). The efficiency of coalescent-based methods lies in the simple, underlying (but standard) Wright-Fisher population genetic model (Wright, 1969) and the fact that only the coalescent of the sampled gene copies are modeled (as opposed to all past and present individuals as is the case in individual-based simulations). The individuals within each population are assumed to mate randomly with uniform reproductive success. Generations are discrete and non-overlapping. Population sizes can be modeled as constant or as changing, typically either instantaneous or exponential population changes. Migration (and population growth) rates are kept constant during each time phase in a simulation.

Although such 'ideal virtual' populations likely will differ substantially from the *in vivo* populations under study, the outcome of such simulations nevertheless provides valuable insights into the data requirements and experimental set-up necessary to discriminate among the targeted management hypotheses. Coalescent simulations are an efficient means by which to get a rough estimate of statistical power, which more extensive simulations under realistic population models might not greatly improve upon. Hence, coalescent-based simulations might constitute a reasonable starting point (Hudson, 2002; Laval and Excoffier, 2004), which, due to their efficacy, facilitates the opportunity to test a range of evolutionary models and parameters values within a realistic time frame.

If the assumptions underlying the coalescent are problematic, the population history is difficult to model, or the statistical assessment planned is impossible using

the coalescent, then individual-based simulations, as implemented in programs such as SPIP (Anderson and Dunham, 2005), RMetaSim (Strand, 2002) and TOSSM (Martien *et al.,* 2009) constitute a useful alternative. Individual-based approaches often permit greater flexibility in terms of demographic parameters (whereas the mutation models for the genetic markers often are simpler). The main drawback of individual-based simulations is that they are considerably slower compared to coalescent-based simulations and expectations might be unknown, making it difficult to assess whether the model behaves correctly.

In summary, coalescent simulation models are very efficient and well suited to approximate a wide range of population models (Hoban *et al.* 2012). If populations are too complicated to model in a coalescent framework, then individual-based models constitute a useful alternative. It can be useful to combine the approaches: use a coalescent-based simulation to 'seed' individual-based simulations with data at migration-drift-mutation equilibrium (Martien *et al.,* 2009; Williamson-Natesan, 2005; Peery *et al.,* 2012).

Only rarely are assumptions underlying the inference methods likely to match the reality, so a prospective power assessment might be required to assess the effect and magnitude of biologically realistic deviations from the assumptions made.

The central parameter in coalescent simulations is θ, which denotes the expected number of mutational events between two gene copies in a single constant-sized random mating population (Hudson, 1998; Kingman, 1982). The size of θ is a function of the generational mutation rate (μ) and the effective population size (N_e). In other words, as N_e increases so does the expected time to the most recent common ancestor of a random pair of two gene copies, and with that the number of mutational events separating the two gene copies. Of course, an increase in μ also elevates the average number of substitutions between the sampled gene copies. Accordingly, the expected level of genetic variation is a function of the product of the effective population size and generational mutation rate. Migration is scaled as the number of (effective) migrants per generation, or mN_e. Population size changes are typically modeled as instantaneous or at a fixed growth rate. These parameters (θ, mN_e and growth rate) can all be estimated from preliminary data. Difficulties arise when trying to separate m, N_e, or μ, which some simulation programs require. In such cases, one needs to ensure that the product (*i.e.* θ or mN_e) is similar to the observed estimates.

In contrast, individual-based simulation programs typically require specification of a large number of non-composite parameters, such as birth rates, survival rates, population sizes, reproductive variance, *etc.,* which allows for greater flexibility in fitting more complex demographic models, but also necessitates relatively precise estimates of several demographic parameters.

In both cases, it is advisable to pick some observed statistics (e.g. number of haplotypes or local genetic diversity) by which to check that the simulations behave similarly to the observed data, keeping in mind that such checks do not constitute a validation *per se* and that a number of different parameter value combinations can yield similar observed values (but which could lead to substantially different inferences).

(f) Different approaches to statistical inference

Statistical inference is aimed at drawing conclusions from datasets arising from systems subject to random variation. There are two main schools of thought in statistical inference, **frequentist** and **Bayesian**.

Frequentist inference is best suited to experimental settings where any given experiment can be considered as one of an infinite sequence of independent repetitions of the same experiment. Statistical significance testing is a typical example of this type of approach. Here one focuses on a test statistic and proposes a 'null hypothesis'. It is then possible to decide whether or not the null hypothesis (no difference between the samples compared) is supported by calculating the *p*-value, the probability of obtaining a test statistic as extreme as the one actually observed, due to random choice of which samples are analysed, even though there is no actual difference. We reject the null hypothesis if the *p*-value is less than the significance level α, typically set at 0.05 (a 5% chance of concluding there is a difference when there is none, or 0.01 (a 1% chance of making such a conclusion). The 'α level' (or False Positive Rate) is usually set in this range to balance it with the converse 'β level', the probability of concluding there is no difference when in fact there is (False Negative Rate).

Bayesian inference is best suited to situations in which experimental manipulations are not possible and we have to rely on observational data that have arisen from the natural dynamics of the system. Instead of being concerned with rejecting the null hypothesis, Bayesian statistics focuses on estimating the 'posterior probability' that the null hypothesis is true after the relevant evidence is taken into account. It is then possible to make an informed decision about the plausibility of the hypothesis. An important difference between frequentist and Baysian methods is that the former relies only on the evidence as a whole, while the latter combines new evidence (from the data) with prior beliefs through the application of Bayes' rule (a mathematical approach for deciding how one should change existing beliefs in the light of new evidence – see Gelman *et al.* 1995 for a full description). Below we describe two so-called 'likelihood' approaches that represent examples of the two schools of thought as applied to genetic analysis: maximum likelihood (ML) and Bayesian.

Likelihood-based methods proceed by assuming that the observed data arose from some probabilistic model with unknown parameters. Their objective is to use the data to estimate the parameters of the model, and to assess the degree of uncertainty associated with these estimates. The core of the method consists in the calculation of the probability $P(G|\Theta)$ of observing the genetic data G if the parameters of the model take the value $\Theta = \theta$. This probability is the so-called likelihood function, $L(\Theta|G)$, which by definition is a function of Θ. For example, in the context of a Genetic Stock Identification method, the data G are the individual genotypes observed in the sample from the genetic mixture and the allele frequency distributions observed in the source populations. The model parameters

that we want to estimate are proportions that each source population contributes to the genetic mixture, which can be denoted by the vector $\mathbf{x} = \{x_i\}$, where x_i is the contribution of source population i.

(i) Maximum Likelihood inference

Maximum likelihood (ML) inference consists of finding the value of Θ that maximises the likelihood function $L(\Theta|G)$. One problem with this approach is that the uncertainty associated with the estimate is expressed by a 95% confidence region that has a rather obscure interpretation. The precise interpretation is that the probability that the confidence region contains the true value of Θ is 0.95. Note that this is not equivalent to saying that the probability that Θ lies in the confidence region is 0.95. An important advantage of ML inference is that for large sample sizes, the maximum likelihood estimate, $\hat{\Theta}$, will have an approximately normal distribution centered on the true parameter value Θ. Thus, an approximate 95% confidence interval can be calculated as the range of $\hat{\Theta}$ values that are within two log-likelihood units of the maximum log-likelihood. Additionally, we can test whether the maximum likelihood estimate is significantly different from another fixed value, θ_0, using the likelihood ratio test. This test uses the fact that the log-likelihood ratio statistic,

$$A = -2\log\frac{L(\theta_0)}{L(\hat{\theta})},$$

has asymptotically a chi-squared distribution, if θ_0 is the 'true' value of θ. Λ can then be assessed for statistical significance using standard χ^2 significance levels.

ML estimates can be obtained analytically for simple models, but application of this method in population genetics typically leads to complex likelihood functions that need to be explored using computer approaches such as the Expectation Maximisation algorithm or MCMC.

(ii) Bayesian inference

To make probability statements about the parameter Θ given the data G, we must begin with a model providing a joint probability distribution for Θ and G. The joint probability mass, $P(\boldsymbol{\Theta},\mathrm{G})$, can be written as a product of two probability distributions, the *prior* distribution $P(\boldsymbol{\Theta})$, and the *sampling* distribution, given by the likelihood function $L(G|\boldsymbol{\Theta})$: $P(\boldsymbol{\Theta},\mathrm{G}) = P(\boldsymbol{\Theta})L(G|\boldsymbol{\Theta})$ (Gelman *et al.* 1995). Using Bayes' theorem, we obtain the post-data or *posterior* distribution,

$$P(\Theta \mid G) = \frac{L(\Theta \mid G)P(\Theta)}{P(G)}.$$

The posterior distribution represents our knowledge about the parameters, taking into account both our prior information (represented by the prior distribution) and the observed data. The primary task of any specific application is to develop the model $P(\Theta,\mathrm{G})$ and perform the necessary computations to summarise $P(\Theta|\mathrm{G})$ in appropriate ways.

Visual inspection of the posterior distribution provides information that is not available when using ML estimation. Additionally, this distribution can be described by point estimates such as the mode, median or mean. The uncertainty around the estimate is expressed by the 95% credible region for Θ. The intuitive interpretation of this region is that the probability that Θ lies in it is 0.95. Another advantage of Bayesian over ML estimation approaches is that the former does not rely on asymptotic arguments, and therefore is valid in situations where the standard likelihood theory fails.

Simple problems in estimation lead to closed form solutions for the posterior distribution, but typical applications in population genetics require the use of numerical integration methods such as MCMC (e.g. Brooks, 1998).

The effect of the prior distribution: In the context of population genetic analysis, the prior distribution can be interpreted as a population of possible parameter values, from which the parameter of interest has been drawn (Gelman *et al.,* 1995). Thus, it should include all plausible values of the parameter but it need not be centered around the true value (which in any case is unknown) because the information contained in the data will (it is hoped) far outweigh any reasonable prior probability specification. Frequentists view prior distributions as subjective judgements of opinion that cannot be rigorously justified (Williamson, 2010). However, objective prior distributions can be obtained from analyses of previously available data. It is also possible to obtain objective priors using independent data, as exemplified by methods that incorporate environmental, ecological, and demographic data to estimate population parameters from genetic data (e.g. Gaggiotti *et al.,* 2002; Foll and Gaggiotti, 2006). If no reasonable basis exists for assigning priors to parameters, non-informative priors (which assume a uniform distribution) can be used.

Summary and relevance to cetaceans:

A large number of Maximum Likelihood (ML) and Bayesian methods can help answer important questions concerning cetacean populations (estimation of migration rates, stock delimitation, estimation of effective population sizes, etc.). Both approaches lead to similar results if the amount of data is large. However, they could differ if this is not the case. In particular, the asymptotic theory underlying ML methods can become invalid if sample sizes are small. Both ML and Bayesian approaches use the same computational statistics methods to obtain the parameter estimates. These sophisticated and complicated methods require users to get acquainted with their underlying principles and for these reasons, ML and Baysian approaches should not be used as 'black boxes'.

Finally, it is important to understand that both are model-based methods that make several assumptions concerning the population biology of species. Violations to these assumptions can lead to biases that could invalidate the results. It is, therefore, important to carefully read the articles describing the methods to make sure that the species under study fits those assumptions.

(g) Monte Carlo issues

The Markov chain Monte Carlo (MCMC) method is a generic term that refers to several algorithms that allow one to sample from complex distributions (Gelman *et al.* 2003).

MCMC is commonly used in both Bayesian and frequentist settings. In the former case, the aim is to sample from the posterior distribution of parameters of a complex probabilistic model; in the latter case, MCMC is used to estimate the likelihood surface to obtain the maximum likelihood estimate of model parameters.

The idea of MCMC is very simple (see Brooks 1998). Suppose that we want to estimate the parameter of a model (e.g. migration rate) whose posterior distribution, $\pi(x)$, is known only up to some multiplicative constant. This is the so-called 'target' distribution. In order to obtain samples from such a distribution, we construct a Markov chain whose stationary distribution is $\pi(x)$. Then, if we run the chain for a sufficiently long period of time, simulated values from the chain can be treated as a dependent sample from the target distribution and used for summarising the main features of $\pi(x)$. Thus, the main task when implementing an MCMC approach is to construct a Markov chain that combines a number of different transition kernels describing the probability of moving between any two states of the chain. These kernels are usually referred to as MCMC updates or transitions. The most standard transition kernel is the Gibbs sampler (Gelman *et al.,* 1995), which can be considered as a special case of the more general Metropolis-Hastings transition.

A number of important implementation issues are associated with these techniques. As a consequence, Bayesian and MLE methods based on MCMC cannot be used as 'black boxes' and require a substantial period of tests before the final results can be generated. The first issue is that before reaching the steady state (i.e. the target distribution $\pi(x)$), the states reached by the Markov chain are strongly influenced by the starting values given to the chain. For this reason, samples generated during this 'burn-in' are discarded. The main problem is then to determine the length of the burn-in. Several diagnostic tests are available and many of them have been implemented in CODA[2]. The most sensitive ones are those that compare the output of two or more independent chains (sometimes called replicate MCMC runs or simply replicates).

Another important issue is the number of additional iterations that need to be run after the burn-in to obtain an independent and identically distributed (IID) sample from the target distribution. The problem here is that random samples from a Markov chain are correlated. Although the magnitude of this correlation depends on the problem at hand, it is always the case that it decreases as the number of MCMC iterations between sampled values increases. Thus, it is almost always necessary to thin the observations by saving only every k^{th} observation. The value of k is usually referred to as the 'thinning interval'. The number of additional iterations that a Markov chain needs to run after the burn-in is obtained by multiplying the desired size of the sample (e.g. 10,000) by the thinning interval. The total length of the MCMC run is obtained as burn-in + thinning interval * sample size.

Developers of statistical methods based on MCMC also need to consider several other important technical issues that arise during the development of the methods, but typical users do not need to be concerned about them. Nevertheless, one particular issue that users do need to be aware of refers to so-called 'mixing' problems. These problems plague methods that focus on models with a very large number of parameters, which is often the case in population genetics. Highly dimensional problems lead to joint posterior distributions that have multiple local modes. The Markov chain can become stuck in these local modes for many iterations, in which case it is impossible to obtain a representative (IID) sample from the target distribution unless the MCMC is run for an unrealistically large number of iterations. To help minimise mixing problems, developers implement simulated tempering or Metropolis-coupled updating schemes (see Brooks, 1998). These schemes use several transition kernels, say m, and then run m simultaneous chains, each one using one of the transition kernels. At each iteration, all chains are updated and then the state of one pair is swapped with a probability that ensures that the target distribution is preserved. Several statistical genetics methods implement these techniques (e.g. LAMARC, MIGRATE, IMa2), so users should acquire the basic knowledge that will allow them to take full advantage of these methods to overcome mixing problems when they arise. Mixing problems can be identified by looking at the acceptance rates (the proportion of proposed values that are accepted by the algorithm) included in the output of the computer programs implementing the statistical methods. As a rule of thumb, acceptance rates should be between 20% and 40%. Some programs calculate an effective sample size that can be used to establish if there are mixing problems (see below).

For all the above issues, users should carry out several test runs before the production runs. The objectives of this testing period are described below.

- Determine the length of the burn-in period. This is achieved by carrying out several independent runs and using CODA[3] to test for convergence. Another useful and complementary approach is to visually compare the posterior distributions from independent runs using TRACER[4].

- Determine the additional number of iterations needed after the burn-in period. This can be determined by looking at the effective sample size (ESS) obtained for the different parameters. Some programs (e.g. IMa, MIGRATE) output the ESS. When this is not the case, it is possible to use TRACER to calculate it. Runs for which ESS <200 should not be used to produce the final results of the analyses. To increase the ESS it is necessary to increase the sample size. It might also be necessary to increase the thinning interval to avoid producing unnecessarily large output files.

- Make use of the simulated 'tempering' or Metropolis-coupled MCMC techniques implemented in the computer programs to solve mixing problems. In the case of Metropolis-coupled MCMC, this requires

[2]*http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml.*

[3]*http://www.mrc-bsu.cam.ac.uk/bugs/documentation/coda03/cdaman03.html.*
[4]*http://tree.bio.ed.ac.uk/software/tracer/.*

choosing the number of searches to run and their temperatures (e.g. see LAMARC user manual[5]).

(h) Integrating genetic and non-genetic data

For the purpose of this section we define 'non-genetic' data as demographic (e.g. population census size, carrying capacity, growth rate, etc.), environmental (e.g. temperature, salinity, rainfall, strength of oceanic currents and their direction, etc.), or ecological (e.g. diet, prey distribution, predator distribution, etc.). There are three main reasons why one might want to combine genetic and non-genetic data: (i) to make inferences about demographic, evolutionary, and ecological processes; (ii) to help overcome the problems generated by the lack of sufficient information in the genetic data (e.g. weak genetic differentiation among populations); and (iii) to distinguish among possible factors that can be influencing the evolution of stock structure.

In all cases the incorporation of these data should decrease the variance of the estimates without biasing the results of the analysis. Bayesian methods provide the framework needed for achieving these goals. Below we briefly present three of these methods.

The first one, COLONISE (Gaggiotti *et al.,* 2004; Foll and Gaggiotti, 2005), implements a Genetic Stock Identification approach originally aimed at making inferences about the demographic and environmental factors that control migration processes (Gaggiotti *et al.,* 2002). This works by quantifying the contribution of the different source populations to a genetic mixture (mixed fishery or newly colonised population) and simultaneously explaining those contributions in terms of demographic, ecological, or environmental factors. The method uses a hierarchical Bayesian model to combine genetic data (incorporated in the likelihood function) with non-genetic data (incorporated in the priors used for the parameters). Genetic data comprise multilocus genotypes from the genetic mixture and allele counts from the source populations. Gaggiotti *et al.* (2004) evaluated the performance of the method under different conditions. They focused on the effect of model parameters such as the number of source populations and the degree of genetic differentiation among sources, and fixed the number of loci to 9 and the sample size per population to 150. Under these conditions, results are very accurate even when genetic differentiation is as low as F_{ST} = 0.01, as long as the number of source populations is seven or more. Foll and Gaggiotti (2005) further evaluated the effect of sample size and number of loci with only three source populations but strong genetic differentiation (F_{ST} = 0.25). The results are accurate even with 10 loci and small sample sizes (10 individuals per population), but the uncertainty increases as data quality deteriorates.

GESTE (Foll and Gaggiotti, 2006) is a hierarchical Bayesian method of wide applicability aimed at making inferences about the demographic, ecological, and environmental factors that influence the spatial structuring of genetic differentiation. This method quantifies the degree of genetic isolation of each local population as measured by population-specific F_{ST} values and evaluates the influence of non-genetic factors on them. As in COLONISE, genetic data (allele counts in each of the local populations) are incorporated in the likelihood function and non-genetic data are incorporated in the prior distributions. Foll and Gaggiotti (2006) provide a fairly extensive evaluation of the performance that includes scenarios that deviate from the population genetic model on which the method relies. The performance is very good even for the lowest sample sizes (10 individuals per population) and number of loci (10) and for local-F_{ST} values ranging between 0.029 and 0.18). This is true for scenarios including at least five populations. Overall, the results indicate that the method can accurately estimate the parameters and identify the relevant environmental/demographic factors if the samples are at least of average quality.

BIMr (Faubet and Gaggiotti, 2008) is a hierarchical Bayesian method that uses multilocus genotype data to estimate contemporary migration rates between local populations and infers the environmental, ecological, and demographic factors that influence them. It allows for asymmetric migration and deviations from Hardy-Weinberg equilibrium in the local populations. Faubet and Gaggiotti (2008) provide an evaluation of the method, mainly aimed at identifying the region of parameter space where the method is and is not able to provide accurate estimates. Overall, the results indicate that reliable results can be obtained when the global level of genetic differentiation (F_{ST}) is 1%, the number of loci is at least 10, and sample sizes are of the order of 50 individuals per population. However, convergence problems are observed when the number of local populations is above 15.

Balkenhol *et al.* (2009) investigated the suitability of GESTE, BIMr and several other statistical approaches used in landscape genetics. They evaluated their statistical power, type-1 error rates, and their overall ability to lead researchers to accurate conclusions about landscape-genetic relationships. Although both GESTE and BIMr exhibit above average performance, the latter performed better, being ranked at the top with two other non-parametric methods. These three best methods provide a good balance between type-1 error and power, leading to correct conclusions for a high percentage of analysed datasets. GESTE was one of only two methods that had a corrected type-1 error rate of zero across all scenarios.

(i) Possible influence of selection

The use of population genetic data in a management context generally relies on the assumption that the genetic markers used are 'neutral'; that is, they are not influenced by natural selection. Neutral loci across the genome are similarly affected by demography and the evolutionary history of populations (e.g. Luikart *et al.,* 2003). Thus, they contain signatures of past demographic events and can be used to estimate parameters such as effective population size, migration and growth rates, etc. However, when the chosen set of molecular markers includes selected loci, the estimates obtained can be highly biased (see Landry *et al.,* 2002) because natural selection can be strong in wild populations (Luikart *et al.,* 2003). Thus, when undertaking analyses that assume neutrality, it is important to verify that data sets do not contain so-called 'outlier loci', which are subject to potentially strong selection.

[5]*http://evolution.genetics.washington.edu/lamarc/documentation/index.htm.*

Outlier loci can have a wide range of aberrant behaviours, including exceptionally high or low F_{ST} between populations, an excess or deficit of low frequency alleles, or an excess or deficit of heterozygous genotypes. Several population genetics methods focus on one or other of these quantitative measures to identify outlier loci. For example, several genome scan methods focus on F_{ST} (e.g. Beaumont and Nichols, 1996; Beaumont and Balding, 2004; Foll and Gaggiotti, 2008), while others focus on heterozygosity (e.g. Schlötterer, 2002). To apply these methods to identify outlier loci it is necessary to genotype individuals at many loci (ideally 100 or more). The larger the number of loci used, the more accurately the neutral baseline distribution of F_{ST} or heterozygosity will be. This in turn will provide more power to detect outlier loci. Screening large numbers of loci has become relatively easy with the development of next generation sequencing methodologies, where for example restriction associated DNA (RAD) sequencing methods can generate thousands of loci for population-level analyses at a reasonable cost (see Peterson *et al.,* 2012 and citations therein).

It could be argued that studies based on a moderate number of markers are unlikely to include selected loci. However, it is in precisely this scenario (small to moderate number of loci) that the effect of an outlier locus will be strongest (see White *et al.,* 2010). Thus, it is always important to test for their presence.

So far, the focus has been on the negative effects generated by outlier loci. One problem for which the use of outlier loci can be of great help is the identification of distinct demographic units (Allendorf *et al.,* 2010; Lowe and Allendorf, 2010). Indeed, outlier loci can represent an informative subset of loci for population assignment of individuals to source populations. This is typically the case when the outlier is due to divergent selection, which increases genetic differentiation among local populations. For example, Moura *et al.* (2014b) using the RAD method show that the pattern of differentiation among killer whale (*Orcinus orca*) populations differs for neutral and outlier loci (putatively under positive selection). They further show that outlier loci and those with fixed differences among ecotypes (populations with distinct foraging behaviours) are associated with genes reflecting functional characteristics consistent with ecotype differences (e.g. associated with digestion).

Finally, the identification of loci under selection is important for prioritising populations for protection. Loci under selection are likely to be responsible for adaptation to local environmental conditions and can greatly contribute to the among-population component of genetic diversity. Thus, the preservation of populations with rare adaptive variants could be a priority (Bonin *et al.,* 2007). Current sequencing technologies allow the examination of thousands of genetic markers with relative ease and, therefore, provide a means to fully characterise both neutral and selected genetic diversity. This in turn will allow more effective ranking of the conservation status and priority of management units.

It is also possible to derive useful inference about local adaptation from more traditional genetic methods. One example involved evidence for both balancing (retention of diversity through selection for heterozygotes or by frequency dependence) and directional selection at an immune system gene in several odontocete cetacean species (Vassilakos *et al.,* 2009) based on conventional DNA sequencing. This could suggest adaptation to local pathogen environments. However, genomic methods will increasingly be applied to resolve questions associated with local adaptation (see reviews in Nielsen, 2005; Jensen *et al.,* 2016), and while genome sampling comparing phenotypes (e.g. Moura *et al.* 2014b) can identify some relevant loci, a more inclusive assessment will be based on screening whole genomes, which is now beginning to include population-level comparisons using draft genomes sequenced to sufficient depth to ensure genotype accuracy (see de Manuel *et al.,* 2016).

Summary and relevance to cetaceans:
There are three important applications in the context of cetacean management. First, detecting outlier loci helps ensure that analyses requiring neutral assumptions are accurate. Second, outlier loci can be useful as population markers. And finally, evidence for selection in regional populations will likely increase with the increased application of new technologies, and this factor should be included in the design of management strategies for cetacean populations, especially when the pattern of differentiation differs for neutral compared to selected loci.

(j) Interpreting negative results

Interpretation of negative results (failure to find a significant result or to reject a null hypothesis) for genetic data should be guided by the same basic principle that applies more broadly to other types of analysis: absence of proof is not proof of absence. That is, just because differences are not found does not prove that no differences exist. Nevertheless, negative results still can be informative about a plausible upper limit to the hypothesised differences, provided a power analysis is conducted. Such an analysis would address the question, 'How large could a difference be and still escape detection, given the amount of data available?' For example, suppose a researcher has two samples of 50 and 75 individuals that have been analysed for 15 microsatellite loci. A test of heterogeneity of allele frequencies (as described in Section (4)(a)(i)) produces a P value >0.05, thus failing to reject the hypothesis that both samples could have been taken from the same population. However, this same result might also occur if the samples come from two populations that are not very different genetically. A power analysis can predict the probability of obtaining a non-significant result for populations that actually differ by a specified amount (the effect size) given a certain level of sampling. One program that will conduct power analyses for tests of heterogeneity using genetic data is POWSIM (Ryman and Palm, 2006[6]). With 15 loci having 5 alleles each and samples of 50 and 75 individuals, a run of POWSIM indicates that close to 100% of combined (across all loci) tests of heterogeneity should be significant if true average F_{ST} is 0.01. In this case it can be concluded that, if the samples actually do

[6]*http://www.zoologi.su.se/~ryman.*

represent different populations, they must differ by less than F_{ST} = 0.01. Additional runs with smaller true F_{ST} values could be run to determine the largest F_{ST} that is consistent with the negative results. POWSIM only models allele frequency data, but an analogous procedure could be used to evaluate power associated with analyses based on DNA sequence data.

A related issue is whether differences that are statistically significant are biologically meaningful. This issue is relevant because it is becoming increasingly possible to amass large quantities of genetic data, and as a result it is possible to demonstrate that very small genetic differences are statistically significant. To illustrate the effect, the human sex ratio at birth is close to 1:1 but slightly skewed toward males. Because this topic has been studied extensively for a long time, enormous sample sizes are available and there is no question that the small difference is highly significant statistically. But is it biologically meaningful? Skewed sex ratio reduces N_e by the fraction (1–4*mf*), where *m* and *f* are the proportions of males and females. Using empirical data for humans that indicate $m = 0.515$ and $f = 0.485$ at birth (Grech *et al.* 2002), and assuming for the moment that these ratios remained the same until sexual maturity, the reduction in N_e from unequal sex ratio would be $1–4(0.515)(0.485) = <0.001 = <0.1\%$. This change to effective population size would be trivial for most applications. However, this small difference (and subsequent changes in the ratio with age) would be of great interest to actuaries for life insurance companies and perhaps marketers of gender-specific products, and it might be enough to affect key behaviours involving mate choice, reproduction, etc.

Summary and relevance to cetaceans:

The common theme of these two contrasting examples is that it is not sufficient merely to know the *P* value associated with a statistical test; it is also important to consider the underlying effect size and its biological relevance. The latter could vary depending on the question of interest.

APPENDIX REFERENCES

Ackerman, M.W., Hand, B.K., Waples, R.K., Luikart, G., Waples, R.S., Steele, C., Garner, B.A., McCane, J. and Campbell, M. 2016. Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. *Evolutionary Applications* (published online early 17 October 2016 doi:10.1111/eva.12433.)

Agapow P.M., Bininda-Edmonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C. and Purvis, A. 2004. The impact of species concepts on biodiversity studies. *Q. Rev. Biol.* 79:161–79.

Allendorf, F.W., Hohenlohe, P.A. and Luikart, G. 2010. Genomics and the future of conservation genetics. *Nature Rev. Genet.* 11:697–709.

Anderson, E.C. 2005. An efficient Monte Carlo method for estimating Ne from temporally spaced samples using a coalescent-based likelihood. *Genetics* 170:955–67.

Anderson, E.C. and Dunham, K.K. 2005. SPIP 1.0: a program for simulating pedigrees and genetic data in age-structured populations. *Mol. Ecol. Notes.* 5:459–61.

Anderson, E.C., Waples, R.S. and Kalinowski, S.T. 2008. An improved method for predicting the accuracy of genetic stock identification. *Can. J. Fish. Aquat. Sci.* 65:1,475–86.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3(10): e3376.

Balkenhol, N., Waits, L.P. and Dezzani, R.J. 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* 32:818–830.

Balloux, F. 2010. The worm in the fruit of the mitochondrial DNA tree. *Heredity* 104:419–20.

Barbujani, G. 2000. Geographic patterns: how to identify them and why. *Hum. Biol.* 72:133–53.

Barton, N.H. 2000. Genetic hitchhiking. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 355:1,553–62.

Beaumont, M.A. 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153:2,013–29.

Beaumont, M.A. and Balding, D.J. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mole. Ecol.* 13:969–80.

Beaumont, M.A. and Rannala, B. 2004. The Bayesian revolution in genetics. *Nature Rev. Genet.* 5:251–61.

Beaumont, M.A. and Nichols, R.A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B.* 263:1,619–26.

Beaumont, M.A., Zhang, W. and Balding, D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2,025–35.

Beerli, P. 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs diffcult to use? pp. 42–79. *In:* Bertorelle, G., Bruford, M.W., Hauffe, H.C., Rizzoli, A., and Vernesi, C. (eds.) *Population Genetics for Animal Conservation.* Cambridge University Press, Cambridge, United Kingdom.

Beerli, P. and Palczewski, M. 2010. Unified framework to dvaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185:313–326.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Series. B.* 57:289–300.

Bernard, A.M., Ferguson, M.M., Noakes, D.L.G., Morrison, B.J. and Wilson, C.C. 2009. How different is different? Defining management and conservation units for a problematic exploited species. *Can. J. Fish. Aquat. Sci.* 66:1,617–30.

Berry, O., Tocher, M.D. and Sarre, S.D. 2004. Can assignment tests measure dispersal? *Mol. Ecol.* 13:551–61.

Bjorklund, M. and Bergek, S. 2009. On the relationship between population differentiation and sampling effort: is more always better? *Oikos* 118:1,127–9.

Blouin, M.S., Parsons, M., Lacaille, V. and Lotz, S. 1996. Use of microsatellite loci to classify individuals by relatedness. *Mol. Ecol.* 5:393–401.

Bonin, A., Nicole, F., Pompanon, F., Miaud, C. and Taberlet, P. 2007. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conserv. Biol.* 21:697–708.

Borge, T., Bachmann, L., Bjornstad, G. and Wiig, Ø. 2007. Genetic variation in Holocene bowhead whales from Svalbard. *Mol. Ecol.* 16:2,223–35.

Branch, T. and Jackson, J.A. 2008. Minimum bottleneck abundance of Antarctic blue whales based on current mtDNA diversity. Paper SC/60/SH10 presented to the IWC Scientific Committee, May 2008 (unpublished). 7pp. [Available from the Office of this Journal]

Bravington, M.V., Grewe, P.M. and Davies, C.R. 2016a. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications* 7, 13162.

Bravington, M.V., Skaug, H.J. and Anderson, E.C. 2016b. Close-Kin Mark-Recapture. *Statistical Science* 31:259–274.

Brooks, S. P. 1998. Markov chain Monte Carlo method and its application. *The Statistician* 47:69–100.

Brosi, B.J. and Biber, E.G. 2009. Statistical inference, Type II error, and decision making under the US Endangered Species Act. *Front. Ecol. Environ.* 7:487–94.

Caballero, A. 1994. Developments in the prediction of effective population size. *Heredity* 73:657–79.

Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. 1994. *The History and Geography of Human Genes.* Princeton University Press, Princeton NJ.

Cerchio, S., Jacobsen, J.K., Cholewiak, D.M., Falcone, E.A. and Merriwether, D.A. 2005. Paternity in humpback whales, *Megaptera novaeangliae*: assessing polygyny and skew in male reproductive success. *Animal Behav.* 70:267–77.

Charlton-Robb, K., Gershwin, L.A., Thompson, R., Austin, J., Owen, K. and McKechnie, S. 2011. A new dolphin species, the burrunan dolphin *Tursiops australis* sp. nov., endemic to southern Australian coastal waters. *PLoS One* 6:e24047. [doi: 10.1371/journal.pone.0024047].

Chen, C., Durand, E., Forbes, F. and Francois, O. 2007. Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Mol. Ecol. Notes* 7:747–56.

Cornuet, J-M. and Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2,001–14.

Cornuet, J-M., Piry, S., Luikart, G., Estoup, A. and Solignac, M. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153:1,989–2,000.

Cornuet, J-M., Ravigné, V. and Estoup, A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* 11:401.

Cornuet, J-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J-M., Balding, D.J., Guillemaud, T. and Estoup, A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *BMC Bioinformatics* 24:2,713–19.

Crow, J. F. and Denniston, C. 1988. Inbreeding and variance effective population numbers. *Evolution* 42:482–95.

Cushman, S.A., McKelvey, K.S. and Schwartz, M.K. 2009. Use of empirically derived source-destination models to map regional conservation corridors. *Conserv. Biol.* 23:368–76.

Dalebout, M.L., Baker, C.S., Mead, J.G., Cockcroft, V.G. and Yamada, T.K. 2004. A comprehensive and validated molecular taxonomy of beaked whales, Family Ziphiidae. *J. Hered.* 95:459–73.

Dalebout, M.L., Steel, D. and Baker, C.S. 2008. Phylogeny of the beaked whale genus *Mesoplodon* revealed by nuclear introns: implications for the evolution of male tusks. *Syst. Biol.* 57:857–75.

de Manuel, M., Kuhlwilm, M., Fransdsen, P., Sousa, V.C., Desai, T., Prado-Martinez, J., Hernanez-Rodiguez, J., Dupanloup, I., Lao, O., Hallast, P., Schmidt, J.M., Heredia-Genestar, J.M., Benazzo, A., Barbujani, G., Peter, B.M., Kuderna, L.F.K., Casals, F., Angedakin, S., Arandjelovic, M., Boesch, C., Kühl, J., Vigilant, L., Langergraber, K., Novembre, J., Gut, M. Gut, I., Navarro, A., Carlsen, F., Andrés, A.M., Siegismund, H.R., Scally, A., Excoffier, L., Tyler-Smith, C., Castellano, S., Xue, Y., Hvilsom, C. and Marques-Bonet, T. 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* 354:477–81.

De Meeûs, T. 2018. Revisiting F_{IS}, F_{ST}, Wahlund effects, and null alleles. *J. Hered.* 109:446–456.

de Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–86.

Denver, D.R., Morris, K., Lynch, M., Vassilieva, L.L. and Thomas, K. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* 289:2,342–4.

Drummond, A.J. and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. [doi:10.1186/1471-2148-7-214].

Drummond, A.J., Rambaut, A., Shapiro, B. and Pybus, O.G. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1,185–92.

Eldredge, N. and Cracraft, J. 1980. *Phylogenetic Patterns and the Evolutionary Process: Method and Theory in Comparative Biology.* Columbia University Press, New York, N.Y.

Evanno, G., Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14:2,611–20.

Ewens, W.J. 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 21:87–112.

Excoffier, L., Novembre, J. and Schneider, S. 2000. SimCoal: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *J. Hered.* 91:506–09.

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–91.

Fairbairn, D.J. and Roff, D.A. 1980. Testing genetic models of isozyme variability without breeding data: Can we depend on the $\chi 2$? *Can. J. Fish, Aquat. Sci.* 37:1,149–59.

Fallon, S.M. 2007. Genetic data and the listing of species under the US endangered species act. *Conserv. Biol.* 21:1,186–95.

Faubet, P. and Gaggiotti, O.E. 2008. A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics* 178:1491–504.

Faubet, P., Waples, R.S. and Gaggiotti, O.E. 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Mol. Ecol.* 16:1,149–66.

Foll, M. and Gaggiotti, O.E. 2005. COLONISE: a computer program to study colonization processes in metapopulations. *Mol. Ecol. Notes* 5:705–07.

Foll, M. and Gaggiotti, O. 2006. Identifying the environmental factors that determine the genetic structure of Populations. *Genetics* 174:875–91.

Foll, M. and Gaggiotti, O.E. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–93.

Fontaine, M.C., Snirc, A., Frantzis, A., Koutrakis, E., Öztürk, B., Öztürk, A.A. and Austerlitz, F. 2012. History of expansion and anthropogenic collapse in a top marine predator of the Black Sea estimated from genetic data. Proceedings of the National Academy of Sciences [doi:10.1073/pnas.1201258].

Francois, O., Currat, M., Ray, N., Han, E., Excoffier, L. and Novembre, J. 2010. Principal component analysis under population genetic models of range expansion and admixture. *Mol. Biol. Evol.* 27:1,257–68.

Frankham, R. 1995. Conservation genetics. *Annu. Rev. Genet.* 29: 305–27.

Gaggiotti, O.E., Brooks, S.P., Amos, W. and Harwood, J. 2004. Combining demographic, environmental and genetic data to test hypotheses about colonization events in metapopulations. *Mol. Ecol.* 13:811–25.

Gaggiotti, O.E. and Gascuel, F. 2011. Stock structure of North Pacific minke whales as revealed by Principal Component Analyses of microsatellite data. Paper SC/63/RMP23 presented to the IWC Scientific Committee, June 2011 (unpublished). 7pp. [Available from the Office of this Journal].

Gaggiotti, O.E., Lange, O., Rassmann, K. and Gliddon, C. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Mol. Ecol.* 8:1,513–20.

Gaggiotti, O.E., Jones, F., Lee, W.M, Amos, W., Harwood, J. and Nichols, R.A. 2002. Patterns of colonization in a metapopulation of grey seals. *Nature* 416:424–7.

Garrigue, C., Dodemont, W., Steel, D. and Baker, C.S. 2004. Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive autonomy of humpback whales on the wintering grounds of New Caledonia. *Mar. Ecol. Prog. Ser.* 274:251–62.

Garza, J.C. and Williamson, E.G. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* 10:305–18.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. 1995. *Bayesian Data Analysis.* Chapman and Hall, London. 552pp.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. 2003. *Bayesian Data Analysis, Second Edition*. Chapman and Hall, London. 690 pp.

Gilbert, K.J. and Whitlock, M.C. 2015. Evaluating methods for estimating local effective population size with and without migration. *Evolution* 69:2,154–66.

Grant, W.D., Milner, G.B., Krasnowski, P. and Utter, F.M. 1980. Use of biochemical genetic variants for identification of sockeye salmon (*Oncorhynchus nerka*) in Cook Inlet, Alaska. *Can. J. Fish. Aquat. Sci.* 37:1,236–47.

Grech, V., Savona-Ventura, C. and Vassallo-Agius, P. 2002. Unexplained differences in sex ratios at birth in Europe and North America. *BMJ.* (BMJ, NCBI/National Institutes of Health) 324 (7344):1010. [doi: http://dx.doi.org/10.1136/bmj.324.7344.1010].

Guillot, G. and Foll, M. 2009. Correcting for ascertainment bias in the inference of population structure. *Bioinformatics* 25:552–4. [doi: 10.1093/bioinformatics/btn665].

Guillot, G., Mortier, F. and Estoup, A. 2005. Geneland: A computer package for landscape genetics. *Mol. Ecol. Notes* 5:708–11.

Hammond, P.S., Mizroch, S.A. and Donovan, G.P. (eds.) 1990. *Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters.* Reports of the International Whaling, Commission, Special Issue 12, Cambridge, UK. 440 pp.

Harlin-Cognato, A.D. and Honeycutt, R.L. 2006. Multi-locus phylogeny of dolphins in the subfamily Lissodelphininae: character synergy improves phylogenetic resolution. *BMC Evol. Biol.* 6:87 doi:10.1186/1471-2148-6-87.

Hartl, D.L. and Clark, A.G. 1989. *Principles of Population Genetics*. Second Edition. Sinauer, Sunderland. xiii+682 pp.

Hasegawa, M., Kishino, H. and Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.

Hao, K., Schadt, E.E. and Storey, J.D. 2008. Calibrating the Performance of SNP Arrays for Whole-Genome Association Studies. *PLoS Genetics* 4(6):e1000109. [doi:10.1371/journal.pgen.1000109].

Hausdorf, B. 2011. Progress toward a general species concept. *Evolution* 65:923–31.

Hedrick, P.W. 2002. Pathogen resistance and genetic variation at MHC loci. *Evolution* 56:1902–08.

Hedrick, P.W. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633–8.

Hedrick, P.W. 2005. *Genetics of Populations*. Third Edition. Jones and Bartlett, Sudbury, MA. 737 pp.

Hedrick, P.W., Smith, D.W. and Stahler, D.R. 2016. Negative-assortative mating for color in wolves. *Evolution* 70:757–66.

Heled, J. and Drummond, A.J. 2008. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* 8:289. [doi: 10.1186/1471-2148-8-289].

Henn, B.M., Gignoux, C.R., Feldman, M.W. and Mountain, J.L. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol. Biol. Evol.* 26:217–30.

Hess, J.E., Matala, A.P. and Narum, S.R. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Mol. Ecol. Res.*11(s1):137–49. [doi: 10.1111/j.1755-0998.2010.02958.x].

Hey, J. 2005. On the number of new world founders: a population genetic portrait of the peopling of the Americas. *PLoS Biology* 3:e193. [doi: 10.1371/journal.pbio.0030].

Hey, J. 2010. Isolation with migration models for more than two populations. *Mol. Biol. Evol.* 27:905–20.
Hey, J. and Nielsen, R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–60.
Hey, J. and Nielsen, R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci.* 104:2,785–90.
Hill, W.G. 1972. Effective size of population with overlapping generations. *Theor. Popul. Biol.* 3:278–89.
Ho, S.Y.W., Kolokotronis, S-O. and Allaby, R.G. 2007. Elevated substitution rates estimated from ancient DNA sequences. *Biol. Lett.* 3:702–05.
Ho, S.Y.W., Phillips, M.J., Cooper, A. and Drummond, A.J. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22:1,561–8.
Ho, S.Y.W., Saarma, U., Barnett, R., Haile, J., and Shapiro, B. 2008. The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One* 3(2): e1615 [doi 10.1371/journal.pone.0001615].
Ho, S.Y.W. and Shapiro, B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Res.* 11:423–34.
Hoban, S., Bertorelle, G. and Gaggiotti, O.E. 2012. Computer simulations: tools for population and evolutionary genetics. *Nature. Rev. Genet.* 13:110–22.
Hoelzel, A.R., Dahlheim. M. and Stern, S.J. 1998b. Low genetic variation among killer whales (*Orcinus orca*) in the eastern North Pacific, and genetic differentiation between foraging specialists. *J. Hered.* 89:121–28.
Hoelzel, A.R., Hancock, J.M. and Dover, G.A. 1991. Evolution of the cetacean mitochondrial D-loop region. *Mol. Biol. Evol.* 8:475–93.
Hoelzel, A.R., Hey, J., Dahlheim, M.E., Nicholson, C., Burkanov, V., and Black, N. 2007. Evolution of population structure in a highly social top predator, the killer whale. *Mol. Biol. Evol.* 24:1,407–15.
Hoelzel, A.R., Natoli, A., Dahlheim, M., Olavarria, C., Baird, R.W. and Black, N. 2002. Low world-wide genetic diversity in the killer whale (*Orcinus orca*); Implications for demographic history. *Proc. R. Soc. Lond. B.* 269:1,467–75.
Hoelzel, A.R., Potter, C.W. and Best, P. 1998a. Genetic differentiation between parapatric 'nearshore' and 'offshore' populations of the bottlenose dolphin. *Proc. R. Soc. Lond. B.* 265:1–7.
Hoffman, J.I., Grant, S.M., Forcada, J. and Phillips, C.D. 2011. Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. *Mol. Ecol.* 20:3,989–4008.
Holderegger, R. and Wagner, H.H. 2008. Landscape genetics. *Bioscience* 58:199–207.
Horn, S. 2012. Target enrichment via DNA hybridization capture. *Methods Mol. Biol.* 840:177–88.
Howell, N., Smejkal, C.B., Mackey, D.A., Chinnery, P.F., Turnbull, D.M. and Herrnstadt, C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum. Genet.* 72:659–70.
Hubisz, M.J., Falush, D., Stephens, M. and Pritchard, J.K. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Res.* 9:1,322–32.
Hudson, R.R. 1990. Gene genealogies and the coalescent process. pp. 1–44. *In:* Futuyma, D.J. and Antonovics, J. (eds.) *Oxfords Surveys in Evolutionary Biology.* Oxford University Press, Oxford.
Hudson, R.R. 1992. Reply to Roff and Bentzen. *Mol. Biol. Evol.* 9:969.
Hudson, R.R. 1998. Island models and the coalescent process. *Mol. Ecol.* 7:413–18.
Hudson, R.R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–8.
Hudson, R.R., Boos, D.D. and Kaplan, N.L. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9:138–51.
International Whaling Commission. 2007. Annex H: Report of the sub-committee on the Other Southern Hemisphere Whale Stocks. *J. Cetacean Res. Manage.* (*Suppl.*) 9:188–209.
International Whaling Commission. 2012. Report of the Scientific Committee. Annex I. Report of the Working Group on Stock Definition. *J. Cetacean Res. Manage.* (*Suppl.*) 13:217–20.
International Whaling Commission. 2015. Annex H: Report of the sub-committee on the Other Southern Hemisphere Whale Stocks. *J. Cetacean Res. Manage.* (*Suppl.*) 16: 196–221.
Jackson, J.A., Baker, C.S. and Branch, T.A. 2011. The use of N_{min} in population assessments: current status and future directions. Paper SC/63/SD3 presented to the IWC Scientific Committee, May 2011 (unpublished). 7pp. [Available from the Office of this Journal].
Jackson, J.A., Patenaude, N.J., Carroll, E.L. and Baker, C.S. 2008. How many whales were there after whaling? Inference from contemporary mtDNA diversity. *Mol. Ecol.* 17:236–51.
Jackson, J.A., Pinsky, M.L. and Baker, C.S. 2009. Erratum [to Jackson *et al.*, 2008]. *Mol. Ecol.* 18(4):134–5.
Jensen J.D., Foll, M. and Bernatchez, L. 2016. The past, present and future of genomic scans for selection**.** *Mol. Ecol.* 25:1–4.
Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–05.
Jombart, T., Devillard, S. and Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94.
Jones, A.G., Small, C.M., Paczolt, K.A. and Ratterman, N.L. 2010. A practical guide to methods of parentage analysis. *Mol. Ecol. Res.* 10:6–30.
Jorde, P.E. 2012. Allele frequency covariance among cohorts and its use in estimating effective size of age-structured populations. *Mol. Ecol. Res.* [doi: 10.1111/j.1755-0998.2011.03111.x].
Jorde, P.E. and Ryman, N. 1995. Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics* 139:1,077–90.
Jost, L. 2008. GST and its relatives do not measure differentiation. *Mol. Ecol.* 17:4,015–26.
Kalinowski, S.T., Manlove, K.R. and Taper, M.L, 2007. *ONCOR: a computer program for genetic stock identification.* Department of Ecology, Montana State University, Bozeman MT, 59717.
Kalinowski, S.T. and Waples, R.S. 2002. Relationship of effective to census size in fluctuating populations. *Conserv. Biol.* 16:129–36.
Kingman, J.F.C. 1982. On the genealogy of large populations. *J. Appl. Prob.* 19:27–43.
Koehn, R.K. 1972. Genetic variation in the eels: a critique. *Mar. Biol.* 14:179–81.
Koljonen, M-L., Pella, J.J. and Masuda, M. 2005. Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo salar*) catches from DNA microsatellite data. *Can. J. Fish, Aquat. Sci.* 62: 2,143–58.
Kopps, A.M., Kang, J., Sherwin, W.B. and Palsbøll, P.J. 2015. How well do molecular and pedigree relatedness correspond, in populations with diverse mating systems, various types and quantities of molecular and demographic data? *G3: Genes, Genomes, Genetics* 5:1,815–26.
Kress, J.W., Garcia-Robledo, C., Uriarte, M. and Erickson, D.L. 2015. DNA barcodes for ecology, evolution and conservation. *Trends. Ecol. Evolut.* 30:25–35.
Krimbas, C.B. and Tsakas, S. 1971. The genetics of *Dacus oleae*. V. Changes of esterase polymorphism in a natural population following insecticide control – Selection or drift? *Evolution* 25:454–60.
Kuhner, M.K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–70.
Landry, P.A., Koskinen, M.T. and Primmer, C.R. 2002. Deriving evolutionary relationships among populations using microsatellites and (deltamu)(2): All loci are equal, but some are more equal than others. *Genetics* 161:1,339–47.
Larsson, L.C., Charlier, J., Laikre, L. and Ryman, N. 2009. Statistical power for detecting genetic divergence-organelle versus nuclear markers. *Conservation Genetics* 10:1,255–64.
Laval, G. and Excoffier, L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20:2,485–7.
Lee, A.M., Engen, S. and Sæther, B-E. 2011. The influence of persistent individual differences and age at maturity on effective population size. *Proc. R. Soc. Lond. B.* 278:3,303–12. [doi: 10.1098/rspb.2011.0283].
Lemey, P., Salemi, M. and Vandamme, A-M. (Eds.). 2009. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press. 750pp.
Lewontin, R.C. and Cockerham, C.C. 1959. The goodness-of-fit test for detecting natural selection in random mating populations. *Evolution* 13:561–4.
Liu, X. and Fu, Y.X. 2015. Exploring population size changes using SNP frequency spectra. *Nature Genetics*. 47(5):555–559.
Lowe, W.H. and Allendorf, F.W. 2010. What can genetics tell us about population connectivity? *Mol. Ecol.* 19:3,038–51.
Luikart, G., England, P.R., Tallmon, D., Jordan, S. and Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature. Rev. Genet.* 4:981–94.
Luikart, G., Ryman, N., Tallmon, D.A., Schwartz, M.K. and Allendorf, F.W. 2010. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv. Genet.* 11:355–73.
Manel, S., Gaggiotti, O., and Waples, R.S. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends. Ecol. Evolut.* 20:136–42.
Manel, S., Schwartz, M.K., Luikart, G. and Taberlet, P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends. Ecol. Evolut.* 18:189–97.
Mantel, M. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–20.

Marshall, T.C., Slate, J., Kruuk, L.E.B. and Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7:639–55.

Marshall, A.R., Knudsen, K.L. and Allendorf, F.W. 2004. Linkage disequilibrium between the pseudoautosomal PEPB-1 locus and the sex-determining region of chinook salmon. *Heredity* 93:85–97.

Martien, K.K., Gregovich, D., Bravington, M.V., Punt, A.E., Strand, A.E., Tallmon, D.A. and Taylor, B.L. 2009. TOSSM: an R package for assessing performance of genetic analytical methods in a management context. *Mol. Ecol. Res.* 9:1456–1459. [doi: 10.1111/j.1755-0998.2009.02686.x].

Mayden, R.L. 1997. A hierarchy of species concepts: The denouement in the saga of the species problem. Pp. 381–424. *In:* Claridge, M.F., Dawah, H.A. and Wilson, M.R. (eds.). *Species: The Units of Biodiversity*. Chapman and Hall, London.

Mayr, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York.

Menozzi, P., Piazza, A. and Cavalli-Sforza, L. 1978. Synthetic maps of human gene frequencies in Europeans. *Science* 201:786–92.

Millar, C.D., Dodd, A., Anderson, J., Gibb, G.C., Ritchie, P.A., Baroni, C., Woodhams, M.D., Hendy, M.D. and Lambert, D.M. 2008. Mutation and evolutionary rates in Adélie Penguins from the Antarctic. *PLoS Genetics* 4. [doi:10.1371/journal.pgen.1000209].

Milligan, B.G. 2003. Maximum-likelihood estimation of relatedness. *Genetics* 163:1,153–67.

Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J.G., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G. and Vogler, A.P. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst. Biol.* 58: 298–311.

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., Bouffard, P., Abel Nielsen, S.C., Rasmussen, M., Willerslev, E., Gilbert, M.T.P. and Harkins, T. 2010. Complete mitochondrial genome phylogenetic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* 20:908–916. [doi: 10.1101/gr.102954.109].

Morin, P.A., Martien, K.K. and Taylor, B.L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Mol. Ecol. Res.* 9:66–73.

Moritz, C. 1994. Defining evolutionarily-significant-units for conservation. *Trends in Ecology and Evolution* 9:373–5.

Moura, A.E., Janse van Rensburg, C., Pilot, M., Tehrani, A., Best, P.B., Thornton, M., Plon, S., de Bruyn, P.J.N., Worley, K.C., Gibbs, R.A., Dahlheim, M.E. and Hoelzel, A.R. 2014a. Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. *Mol. Biol. Evol.* 31:1,121–31.

Moura, A.E., Nielsen, S.C., Vilstrup, J.T., Moreno-Mayar, J.V., Gilbert, M.T., Gray, H.W., Natoli, A., Möller, L. and Hoelzel, A.R. 2013a. Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. *Syst. Biol.* 62:865–77.

Moura, A.E., Natoli, A., Rogan, E. and Hoelzel, A.R. 2013b. Atypical panmixia in a European dolphin species (*Delphinus delphis*); implications for the evolution of diversity across oceanic boundaries. *J. Evol. Biol.* 26:63–75.

Moura, A.E., Kenny, J.G., Chaudhuri, R., Hughes, M.A., Welch, A., Reisinger, R.R., de Bruyn, P.J.N., Dahlheim, M.E., Hall, N., and Hoelzel, A.R. 2014b. Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Mol. Ecol.* 23:5,179–92.

Moura, A.E., Kenny, J.G., Chaudhuri, R., Hughes, M.A., Reisinger, R.R., de Bruyn, P.J.N., Dahlheim, M.E., Hall, N. and Hoelzel, A.R. 2015. Phylogenomics of the killer whale indicates ecotype divergence in sympatry. *Heredity* 114:48–55

Natoli, A., Cañadas, A., Peddemors, V.M., Aguilar, A., Vaquero, C., Fernández-Piqueras, P. and Hoelzel, A.R. 2006. Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). *J. Evol. Biol.* 19:943–54.

Natoli, A., Peddemors, V. and Hoelzel, A.R. 2004. Population structure and speciation in the genus *Tursiops* based on microsatellite and mitochondrial DNA analyses. *J. Evol. Biol.* 17: 363–75.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–90.

Neigel, J.E. and Avise, J.C. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. pp. 515–534 *In*: E. Nevo and S. Karlin (eds.) *Evolutionary Processes and Theory*. Academic Press, New York.

Nielsen, R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218.

Nielsen, R. and Beaumont, M.A. 2009. Statistical inferences in phylogeography. *Mol. Ecol.* 18:1,034–47.

Nielsen, R., Mattila, D.K., Clapham, P.J. and Palsbøll, P.J. 2001. Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* 157:1,673–82.

Nielsen, R. and Wakeley, J. 2001. Distinguishing migration from isolation. A Markov chain Monte Carlo approach. *Genetics* 158:885–96.

Nomura, T. 2008. Estimation of effective number of breeders from molecular co-ancestry of single cohort sample. *Evol. Appl.* 1:462–74.

Novembre, J. and Stephens, M. 2008. Interpreting principal components analyses of spatial population genetic variation. *Nature Genet.* 40:646–9. [doi:10.1038/ng.139].

Nunney, L. 1993. The influence of variation in female fecundity on effective population size. *Biological Journal of the Linnean Society* 59: 411–25. [doi:10.1111/j.1095-8312.1996.tb01474.x].

Økland, J-M., Haaland, Ø.A. and Skaug, H.J. 2010. A method for defining management units based on genetically determined close relatives. *ICES J. Mar. Sci.* 67:551–8.

Paetkau, D. 2004. The optimal number of markers in genetic capture-mark-recapture studies. *J. Wildl. Manage.* 68:449–52.

Paetkau, D., Calvert, W., Stirling, I. and Strobeck, C. 1995. Microsatellite analysis of population structure in Canadian polar bears. *Mol. Ecol.* 4:347–54.

Paetkau, D., Slade, R., Burden, M. and Estoup, A. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol. Ecol.* 13:55–65.

Palsbøll, P.J. 1999. Genetic tagging: contemporary molecular ecology. *Biol. J. Linnean. Soc.* 68:3–22.

Palsbøll, P.J., Allen, J., Bérubé, M., Clapham, P.J., Feddersen, T.P., Hammond, P., Jorgensen, H., Katona, S., Larsen, A.H., Larsen, F., Lien, J., Mattila, D.K., Sigurjonsson, J., Sears, R., Smith, T., Sponer, R., Stevick, P. and Oien, N. 1997. Genetic tagging of humpback whales. *Nature* 388:676–9.

Palsbøll, P.J., Bérubé, M., Aguilar, A., Notarbartolo di Sciara, G., and Nielsen, R. 2004. Discerning between re-current gene flow and recent divergence under a finite-site mutation model applied to North Atlantic and Mediterranean Sea fin whale (*Balaenoptera physalus*) populations. *Evolution* 58:670–5.

Palsbøll, P.J., Bérubé, M. and Allendorf, F.W. 2007. Identification of management units using population genetic data. *Trends Ecol. Evolut.* 22:11–16.

Palsbøll, P.J., Bérubé, M., Anderson, E.C. and Dunham, K.K. 2005. High levels of statistical uncertainty in 'gametic' recapture estimates of male abundance in humpback whales. *Mar. Ecol. Prog. Ser.* 295:305–7.

Palsbøll, P.J., Peery, M.Z. and Bérubé, M. 2010. Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence. *Mol. Ecol. Res.*10:797–805.

Palsbøll, P.J., Peery, M.Z. Olsen, M.T., Beissinger, S.R. and Bérubé, M. 2013. Inferring recent historic abundance from current genetic diversity. *Mol. Ecol.* 22:22–40.

Pastene, L.A., Goto, M., Kanda, N., Zerbini, A.N., Karem, D., Watanabe, K., Bessho, Y., Hasegawa, M., Nielsen, R., Larsen, F. and Palsbøll, P.J. 2007. Radiation and speciation of pelagic organisms during periods of global warming: the case of the common mink whale, *Balaenoptera acutorostrata. Mol. Ecol.* 16: 1,481–1,495.

Patterson, N., Price, A.L. and Reich, D. 2006. Population structure and eigenanalysis. *PLoS Genetics* 2:2,074–93.

Pearse, D.E., Eckerman, C.M., Janzen, F.J. and Avise, J.C. 2001. A genetic analogue of 'mark-recapture' methods for estimating population size: an approach based on molecular parentage assessments. *Mol. Ecol.* 10:2,711–18.

Peery, M.Z., Beissinger, S.R., House, R.F., Bérubé, M., Hall, L.A., Sellas, A. and Palsbøll, P.J. 2008. Characterizing source-sink dynamics with parentage assignments. *Ecology* 89:2,746–59.

Peery, M.Z., Kirby, R., Reid, B.N., Stoelting, R., Doucet-Bëer, E., Robinson, S., Vásquez-Carrillo, C., Pauli, J.N. and Palsbøll, P.J. 2012. Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology* 21:3403–3418. [doi:10.1111/j.1365-294X.2012.05635.x].

Peery, M.Z., Kirby, R., Reid, B.N., Stoelting, R., Doucet-Beer, E., Robinson, S., Vásquez-Carrillo, C., Pauli, J.N. and Palsboll, P.J. 2012. Reliability of genetic bottleneck tests for detecting recent population declines. *Mol. Ecol.* 21:3,403–18.

Pella, J. and Masuda, M. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. *Fish. Bull.* 99:151–67.

Pella, J. and Masuda, M. 2006. The gibbs and split-merge sampler for population mixture analysis from genetic data with incomplete baselines. *Can. J. Fish, Aquat. Sci.* 63:576–96.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E. 2012. Double Digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* 7:e37135. [doi:10.1371/journal.pone.0037135].

Pilot, M., Dahlheim, M.E. and Hoelzel, A.R. 2010. Social cohesion among kin, gene flow without dispersal and the evolution of population genetic structure in the killer whale (*Orcinus orca*). *J. Evol. Biol.* 23:20–31.
Piry, S., Alapetite, A., Cornuet, J-M., Paetkau, D., Baudouin, L. and Estoup, A. 2004. GENECLASS2: a software for genetic assignment and first generation migrant detection. *J. Hered.*95:536–9.
Pritchard, J.K., Stephens, P. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59.
Pudovkin, A.I., Zaykin, D.V. and Hedgecock, D. 1996. On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* 144:383–7.
Pybus, O.G., Rambaut, A. and Harvey, P.H. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155:1,429–37.
Quinn, T.W. and White, B.N. 1987. Analysis of DNA sequence variation. pp. 163–198. *In:* Cooke, F. and Buckley, P.A. (eds). *Avian Genetics*. Academic Press, London.
R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [Available at: *http://www.R-project.org.ISBN3-900051-07-0*].
Rew, M.B., Robbins, J., Mattila, D., Palsbøll, P.J. and Bérubé, M. 2011. How many genetic markers to tag an individual? An empirical assessment of false matching rates among close relatives. *Ecol. Appl.* 21:877–87.
Roff, D.A. and Bentzen, P. 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Mol. Biol. Evol.* 6:539–45.
Roman, J. and Palumbi, S.R. 2003. Whales before whaling in the North Atlantic. *Science* 301:508–510.
Rooney, A.P., Honeycutt, R.L. and Derr, J.N. 2001. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. *Evolution* 55:1,678–85.
Rosenberg, N.A. and Nordborg, M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Rev. Genet.* 3:380–90.
Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1,219–28.
Ryman, N. and Leimar, O. 2009. GST is still a useful measure of genetic differentiation – a comment on Jost's D. *Mol. Ecol.* 18:2,084–7.
Ryman, N. and Palm, S. 2006. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Mol. Ecol. Notes* 6:600–02.
Ryman, N., Palm, S., André, C., Carvalho, G.R., Dahlgren, T.G., Jorde, P.E., Laikre, L., Larsson, L.C., Palmé, A. and Ruzzante, D.E. 2006. Power for detecting genetic divergence: differences between statistical methods and marker loci. *Mol. Ecol.* 15:2,031–45. [doi:10.1111/j.1365-294X.2006.02839.x].
Sasaki, T., Nikaido, M., Wada, S., Yamada, T.K., Cao, Y., Hasegawa, M. and Okada, N. 2006. *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage. *Mol. Phylogenet. Evol*. 41:40–52.
Schlotterer, C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* 160:753–63.
Schwartz, M.K. and McKelvey, K.S. 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conserv. Genet.* 10:441–52.
Schwarz, C.J. and Seber, G.A.F. 1999. Estimating animal abundance: review III. *J. Stat. Sci.* 14:427–56.
Seber, G.A.F. 1982. *The Estimation of Animal Abundance and Related Parameters, Second Edition.* Charles Griffin and Company Ltd., London.
Segelbacher, G., Cushman, S.A., Epperson, B.K., Fortin, M-J., Francois, O., Hardy, O.J., Holderegger, R., Taberlet, P., Waits, L.P. and Manel, S. 2010. Applications of landscape genetics in conservation biology: concepts and challenges. *Conserv. Genet.* 11: 375–85. [doi:10.1007/s10592-009-0044-5].
Shaklee, J.B., Beacham, T.D., Seeb, L. and White, B.A. 1999. Managing fisheries using genetic data: case studies from four species of Pacific Salmon. *Fish. Res.* 43:45–78.
Skaug, H.J., Bérubé, M. and Palsbøll, P.J. 2010. Detecting dyads of related individuals in large collections of DNA-profiles by controlling the false discovery rate. *Mol. Ecol. Res.* 10:693–700.
Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–62.
Smith, S.J. and Campana, S.E. 2010. Integrated stock mixture analysis for continous and categorical data, with application to genetic-otolith combinations. *Can. J. Fish, Aquat. Sci.* 67:1,533–48.
Smouse, P.E., Waples, R.S. and Tworek, J.A. 1990. A mixed fishery model for use with incomplete source population data. *Can. J. Fish, Aquat. Sci.* 47:620–34.
Sokal, R.R. and Rohlf, F.J. 1995. *Biometry: The Principles of Statistics in Biological Research,* Third Edition. Freeman and Company, New York.
Steeman, M.E., Hebsgaard, M.B., Fordyce, E., Ho, S.Y., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner H., Sørensen, M.V. and Willerslev, E. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. *Syst. Biol.* 58:573–85.
Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., Delmelle, E., Vierling, L. and Waits, L.P. 2007. Putting the 'landscape' in landscape genetics. *Heredity* 98:128–42.
Strand, A.E. 2002. METASIM 1.0: an individual-based environment for simulating population genetics of complex population dynamics. *Mol. Ecol. Notes* 2:373–6.
Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95.
Tallmon, D.A., Koyuk, A., Luikart, G. and Beaumont, M.A. 2008. ONeSAMP: a program to estimate effective population size using approximate Bayesian computation. *Mol. Ecol. Res.* 8:299–301.
Taylor, B.L. and Gerrodette, T. 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conserv. Biol.* 7:498–500.
Thompson, E.A. 2013. Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics* 194:301–26.
Thompson, K.F., Patel, S., Baker, C.S. and Millar, C.D. 2016. Bucking the trend: genetic analysis reveals high diversity, large population size and low differentiation in a deep ocean cetacean. *Heredity* 116:277–285. [doi:10/1038/hdy.2015.99]
Tiedemann, R., Cipriano, F., Morin, P.A., Hoelzel, A.R., Palsbøll, P., Waples, R., Natoli, A., Bachmann, L., Postma, L., Double, M., Pampoulie, C., Skaug, H., Baker, C.S. and Jackson, J. 2012. Updated guidelines for DNA data quality control and error rate estimation, for genetic studies relevant to IWC management advice. Paper SC/64/SD2 presented to the IWC Scientific Committee, June 2012 (unpublished). 12pp. [available from IWC Secretariat]
Tiedemann, R., Hardy, O., Vekemans, X. and Milinkovitch, M.C. 2000. Higher impact of female than male migration on population structure in large mammals. *Mol. Ecol.* 9:1,159–63.
Van Dyck, H. and Baguette, M. 2005. Dispersal behaviour in fragmented landscapes: routine or special movements? *Basic Appl. Ecol.* 6:535–45.
Vassilakos, D., Natoli, A., Dahlheim, M. and Hoelzel, A.R. 2009. Balancing and directional selection at exon-2 of the MHC DQB1 locus among populations of odontocete cetaceans. *Mol. Biol. Evol.* 26:681–9.
Wada, S. and Numachi, K. 1991. Allozyme analyses of genetic differentiation among the populations and species of the Balaenoptera. *Rep. Int. Whaling. Commn.* 13(special issue):125–54.
Wada, S., Oishi, M. and Yamada, T.K. 2003. A newly discovered species of living baleen whale. *Nature* 426:278–81.
Waits, L.P., Luikart, G. and Taberlet, P. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Mol. Ecol.*10:249–56.
Wakeley, J. 1996. Distinguishing migration from isolation using the variance of pairwise differences. *Theor. Popul. Biol.* 49:369–86.
Wakeley, J. 2004. Metapopulations and coalescent theory. pp. 175–198 *In:* Hanski, H. and Gaggiotti, O. (eds.) *Ecology, Genetics, and Evolution of Metapopulations.* Elsevier, Oxford.
Wang, J. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetic Research* 78:243–57.
Wang, J. 2005. Estimation of effective population sizes from data on genetic markers. *Philos. Trans. R. Soc. Lond. B. Biol. Sci. Sciences* 360:1,395–409.
Wang, J.L. 2006. Informativeness of genetic markers for pairwise relationship and relatedness inference. *Theor. Popul. Biol.* 70:300–21.
Wang, J. 2009.A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Mol. Ecol.*18:2,148–64. [doi:10.1111/j.1365-1294X.2009.04175.x].
Wang, J. 2014. Estimation of migration rates from marker_based parentage analysis. *Mol. Ecol.* 23:3,191–213.
Wang, J. 2016. A comparison of single-sample estimators of effective population sizes from genetic marker data. *Mol. Ecol.* 25:4,692–711.
Wang, J. and Whitlock, M.C. 2003. Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 163:429–46.47
Wang, J.Y., Chou, L.S. and White, B.N. 1999. Mitochondrial DNA analysis of sympatric morphotypes of bottlenose dolphins (genus: *Tursiops*) in Chinese waters. *Mol. Ecol.* 8:1,603–12.
Waples, R.S. 1998. Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *J. Hered.* 89:438–50.
Waples, R.S. 2005. Genetic estimates of contemporary effective population size: To what time periods do the estimates apply? *Mol. Ecol.* 14:3,335–52.
Waples, R.S. 2006. Distinct Population Segments. pp. 127–149 *In:* Scott, J.M., Goble, D.D. and Davis, F.W. (eds). *The Endangered Species Act at Thirty: Conserving Biodiversity in Human-Dominated Landscapes.* Island Press, Washington, D.C.

Waples, R.S. 2011. Can evidence for spatial and/or temporal genetic heterogeneity of North Pacific minke whales be explained by different mixture fractions of the same two core stocks, or is it necessary to postulate an additional stock(s)? Paper SC/63/RMP7 presented to the IWC Scientific Committee, June 2011 (unplublished). 11 pp. [available from IWC Secretariat].

Waples, R.S. 2015. Testing for Hardy-Weinberg proportions: Have we lost the plot? *J Hered.* 106:1–19 [doi: 10.1093/jhered/esu062].

Waples, R.S. 2016a. Tiny estimates of the N_e/N ratio in marine fishes: Are they real? *J. Fish Biology* 89:2479–2504.

Waples, R.S. 2016b. Life history traits and effective population size in species with overlapping generations revisited: the importance of adult mortality. *Heredity* published online 8 June 2016; [doi:10.1038/hdy.2016.29].

Waples, R.S. 2018. Null alleles and $F_{IS} \times F_{ST}$ correlations. *J. Hered.* 109:457–461.

Waples, R.S. and Antao, T. 2014. Intermittent breeding and constraints on litter size: consequences for effective population size per generation (Ne) and per reproductive cycle (Nb). *Evolution* 68:1,722–34 [doi:10.1111/evo.12384].

Waples, R.S., Antao, T. and Luikart, G. 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics* 197:769–80.

Waples, R.S. and Do, C. 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol. Appl.* 3:244–62.

Waples, R.S., and England, P.R. 2011. Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* 189:633–644. [doi:10.1534/genetics.111.132233].

Waples, R.S. and Gaggiotti, O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* 15:1,419–39.

Waples, R.K., Larson, W.A. and Waples, R.S. 2016. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity* 2016:1–8. [doi:10.1038/hdy.2016.60].

Waples, R.S., Luikart, G., Faulkner, J.R. and Tallmon, D.A. 2013. Simple life history traits explain key effective population size ratios across diverse taxa. *Procceedings of the Royal Society. B.* 280:20131339 [doi:10.1098/rspb.2013.1339].

Waples, R.S., Grewe, P.G., Bravington, M.V., Hillary, R. and Feutry, P. 2018. Robust estimates of a high N_e/N ratio in a top marine predator, southern bluefin tuna. *Science Advances* 4:eaar7759.

Waples, R.S. and Waples, R.K. 2011. Inbreeding effective population size and parentage analysis without parents. *Mol. Ecol. Res.* 11(Suppl. 1):162–71.

Waples, R.S. and Yokota, M. 2007. Temporal estimates of effective population size in species with overlapping generations. *Genetics* 175:219–33.

Waser, P.M. and Strobeck, C. 1998. Genetic signatures of interpopulation dispersal. *Trends. Ecol. Evolut.* 13:43–4.

Watterson, G.A. 1984. Lines of descent and the coalescent. *Theor. Popul. Biol.* 26:77–92.

White, T.A., Stamford, J. and Hoelzel, A.R. 2010. Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). *Mol. Ecol.* 19:216–26.

Whitlock, M.C. and McCauley, D.E. 1999. Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity* 82:117–25.

Wiemann, A., Andersen, L.W., Berggren, P., Siebert, U., Benke, H., Teilmann, J., Lockyer, C., Pawliczka, I., Skora, K., Roos, A., Lyrholm, T., Paulus, K.B., Ketmaier, V. and Tiedemann, R. 2010. Mitochondrial control region and microsatellite analyses on harbour porpoise (*Phocoena phocoena*) unravel population differentiation in the Baltic Sea and adjacent waters. *Conserv. Genet.* 11:195–211.

Williamson-Natesan, E.G. 2005. Comparison of methods for detecting bottlenecks from microsatellite loci. *Conserv. Genet.* 6:551–62.

Williamson, J. 2010. Review of Bruno di Finetti. 'Philosophical Lectures on Probability'. *Philosophia Mathematica* 18:130–35.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, E.M., Sage, R.D. and Stoneking, R.D. 1985. Mitochondrial-DNA and 2 perspectives on evolutionary genetics. *Biol. J. Linnean Soc.* 26:375–400. [doi:10.1111/j.1095-8312.1985.tb02048.x].

Wilson, G.A. and Rannala, B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1,177–91.

Wilson, I.J., Weale, M.E. and Balding, D.J. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. R. Stat. Soc. Series B* 166:155–88.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* 87:430–31.

Wright, S. 1943. Isolation by distance. *Genetics* 28:114–38.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395–420.

Wright, S. 1969. *The Theory of Gene Frequencies*. The University of Chicago Press, Chicago and London.