Marine Mammal Science

MARINE MAMMAL SCIENCE, 33(SPECIAL ISSUE): 27–55 (June 2017)

DOI: 10.1111/mms.12409

Analytical approaches to subspecies delimitation with genetic data

This is the second of six papers forming a special issue of Marine Mammal Science (Vol. 33, Special Issue) on delimiting cetacean subspecies using primarily genetic data. An introduction to the special issue and brief summaries of all papers it contains is presented in Taylor et al. (2017a). Together, these papers lead to a proposed set of guidelines that identify informational needs and quantitative standards (Taylor et al. 2017b) intended to promote consistency, objectivity, and transparency in the classification of cetaceans. The guidelines are broadly applicable across data types. The quantitative standards are based on the marker currently available across a sufficiently broad number of cetacean taxa: mitochondrial DNA control region sequence data. They are intended as "living" standards that should be revised as new types of data (particularly nuclear data) become available.

Karen K. Martien,[1] Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A.; **Matthew S. Leslie,** Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A. and Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, Mailcode 0208, La Jolla, California 92037, U.S.A.; **Barbara L. Taylor, Phillip A. Morin, Frederick I. Archer, Brittany L. Hancock-Hanser,** Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A.; **Patricia E. Rosel,** Southeast Fisheries Science Center, National Marine Fisheries Service, NOAA, 646 Cajundome Boulevard, Lafayette, Lousiana 70506, U.S.A.; **Nicole L. Vollmer,** Southeast Fisheries Science Center, National Marine Fisheries Service, NOAA, 646 Cajundome Boulevard, Lafayette, Louisiana 70506, U.S.A. and National Systematics Laboratory, National Marine Fisheries Service, NOAA, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, U.S.A.; **Amélia Viricel,** Southeast Fisheries Science Center, National Marine Fisheries Service, NOAA, 646 Cajundome Boulevard, Lafayette, Louisiana 70506, U.S.A. and Littoral, Environnement et Sociétés, UMR 7266 CNRS – Université de La Rochelle, 2 rue Olympe de Gouges, 17000 La Rochelle, France; **Frank Cipriano,** Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132, U.S.A.

Abstract

The vast and remote distributions and large body size of most cetaceans make it difficult to obtain and maintain morphological collections adequate for advancing sound taxonomic arguments. Consequently, genetic data are playing an increasingly important role in cetacean species and subspecies delimitation. We review seven categories of analytical methods useful in delimiting subspecies based on genetic data.

[1]Corresponding author (e-mail: karen.martien@noaa.gov).

For each category, we summarize its utility in evaluating putative subspecies, the types of markers to which it can be applied, and potential challenges in interpreting results in a taxonomic context. We focus on the utility of each type of method to evaluate the critical features distinguishing subspecies from populations and species: the degree of diagnosability between putative taxa and the extent to which the putative taxa have diverged along separate evolutionary pathways. We conclude that diagnosability is best estimated with either assignment tests or multivariate methods, while evaluating the degree of divergence requires a synthesis of multiple lines of evidence derived from different analytical methods and different data types, including nongenetic data.

Key words: genetic differentiation, divergence, cetacean taxonomy, genetic analytical methods, subspecies definition.

Subspecies play an important role in cetacean conservation and evolutionary studies (see Taylor *et al*. 2017*b* for a detailed review). Although there has been considerable debate over the utility and validity of the subspecies concept over the last half decade (Patten and Unitt 2002, James 2010, Braby *et al.* 2012), the use of subspecies for understanding and conserving biodiversity has resulted in an increased focus on consistent and timely subspecies delimitation (Patten and Unitt 2002, Haig *et al.* 2006, Taylor *et al.* 2017*b*). Delimiting subspecies requires evaluating both the lower (population *vs*. subspecies) and upper (subspecies *vs*. species) boundaries of a subspecies concept. This in turn requires identifying those features that would lead to the conclusion that two putative taxa are sufficiently different to warrant subspecific status, but not so different that they should be elevated to full species status. We use the definitions of subspecies and species proposed by Taylor *et al*. (2017*b*): "a *species* is a separately evolving lineage composed of a population or collection of populations" and "a *subspecies* is a population, or collection of populations, that appears to be a separately evolving lineage with discontinuities resulting from geography, ecological specialization, or other forces that restrict gene flow to the point that the population or collection of populations is diagnosably distinct." Note that Taylor *et al*.'s species definition is de Queiroz's (2007) unified species concept modified slightly to reflect the fact that several cetacean species exist as single populations rather than metapopulations. The term "population" has many commonly used definitions spanning a wide range of levels of connectivity (Waples and Gaggiotti 2006). We use the term "population" broadly to refer to any Unit to Conserve below the subspecies level (Taylor *et al.* 2017*b*). These are the definitions we used to determine the features distinguishing these three levels.

Two types of metrics are needed to evaluate taxonomic cases based on these definitions: the degree of genetic differentiation (used for both boundaries) and the degree of diagnosability (used to distinguish subspecies from populations). Under our operational definition, subspecies can have some ongoing gene flow, so long as the rate of gene flow is low enough that diagnostic characters arise and are maintained, making it possible for unknown individuals to be assigned to the correct subspecies with high accuracy, though not necessarily 100% (Amadon 1949, Patten and Unitt 2002, Taylor *et al.* 2017*a*). However, they are not so diverged that they actually represent separate species. Though both taxonomic units are independently evolving, subspecies have not yet diverged to the point where their divergence is irreversible (Taylor *et al.* 2017*b*). If gene flow between them increased, they could still coalesce into a single taxonomic unit. Species, on the other hand, have diverged to the point where coalescence is no longer possible.

As noted by Taylor *et al*. (2017*b*), traditional morphological approaches to taxonomic analyses for cetaceans have proven difficult (*e.g.*, Milinkovitch *et al.* 2002, Reeves *et al.* 2004, Monaghan *et al.* 2009, Dupuis *et al.* 2012). Hence, herein, we review analytical approaches to delimiting cetacean subspecies using genetic data. Although genetic data can include data on any genetically heritable trait, we use the term to refer specifically to molecular data. We first review the features that delimit the lower and upper subspecies boundaries: diagnosability and the degree to which they are separately evolving entities. We then discuss seven categories of molecular genetic analytical approaches (Table 1) that can be used to evaluate these features. We discuss the genetic markers that work well with each approach, the advantages and disadvantages of each analytical approach, and give examples of how each approach has been applied in taxonomic studies. Given the rapid pace of development of molecular genetic analytical methods, we do not discuss individual algorithms in depth or even attempt to summarize all algorithms currently available in each category. Rather, we focus on identifying the types of analytical approaches that are best suited to subspecies delimitation using molecular data. Furthermore, we do not discuss the merits of different quantitative thresholds above which two taxa should be considered subspecies or species. Rosel *et al*. (2017*a*) apply many of the analytical methods we discuss to a large number of pairs of cetacean populations, subspecies, and species. Based on these comparisons, Taylor *et al*. (2017*a*) propose quantitative standards for delimiting cetacean subspecies. We refer readers to these papers.

We conclude with recommendations regarding the most promising analytical approaches to evaluate the upper and lower subspecies boundaries. To help the reader, we follow the logical progression of analysis for the case where the status of the taxonomic unit of interest is unknown: it could be a population, a subspecies, or a species. We assume that the researcher has already assessed the adequacy of their sample size and distribution (Zhang *et al.* 2010, Hale *et al.* 2012, Rosel *et al.* 2017*b*, Taylor *et al.* 2017*a*). Most subspecies definitions, including that of Taylor *et al*. (2017*b*), require a discontinuity between subspecies (Mallet 2004). We therefore also assume the researcher has done sufficient analyses to convince a reader that the units being compared are not from the ends of a cline (see Manel *et al.* 2003 for a review of appropriate analyses), as this would violate our subspecies concept.

Evaluating Diagnosability

The concept of diagnosability has long been central to morphological studies of subspecific variation (Amadon 1949, Mayr 1969). Archer *et al*. (2017*b*) provided a detailed review of the concept of diagnosability and its use in taxonomic studies. They defined diagnosability as "a measure of the ability to correctly determine the taxon of a specimen of unknown origin based on a set of distinguishing characteristics." Species are generally expected to be 100% diagnosable, meaning that individuals can be assigned to species without error based on the characters used for diagnosis. As noted above, on the other hand, diagnosability for subspecies is expected to be high but need not be 100%. The threshold value above which diagnosability is high enough to warrant subspecies delimitation is discussed in detail by Taylor *et al.* (2017*a*) and Archer *et al*. (2017*b*).

Although diagnosability has long proved useful as a criterion for delimiting subspecies, care must be taken in its application. As with most metrics, diagnosability can be either over- or underestimated as a result of inadequate or biased sampling.

Table 1. Summary of seven analytical approaches useful in genetic taxonomic studies. Methods are classified as to whether they can be used to estimate diagnosability, assess the evolutionary independence of lineages, generate taxonomic hypotheses and identify cryptic taxa, and test hypotheses generated from or corroborate patterns observed in other types of data (*e.g.*, morphology).

Analytical method	Compatible markers	Estimate diagnosability	Assess independence of lineages	Generate hypotheses	Corroborate/test hypotheses
Multivariate methods	Microsatellites, SNPs, AFLPs, sequence data (mtDNA or nuclear)	X			X
Assignment tests	Microsatellites, SNPs, AFLPs	X			X
Clustering methods	Microsatellites, SNPs, AFLPs			X	X
Divergence time estimates	Sequence data (mtDNA or nuclear)		X		X
Genetic distance estimates	Sequence data (percent fixed differences, percent divergence); Microsatellites, SNPs, AFLPs (private alleles, Euclidean distance)	X[a]	X		X
Tree-based methods	Sequence data (mtDNA or nuclear)		X	X	X
Dispersal rate Estimators	Microsatellites, SNPs, AFLPs				X

[a]Fixed differences and private alleles imply 100% diagnosability.

Underclassification errors (*i.e.*, failing to confer subspecific status when it is warranted) can occur when the genetic markers used in estimates of diagnosability are inadequate to resolve structure at the subspecies level. This can be the result of low diversity (*e.g.*, killer whale, *Orcinus orca* ecotypes; Hoelzel *et al.* 1998, LeDuc *et al.* 2008), recent and rapid divergence (*e.g.*, delphinids, Kingston and Rosel 2004; song sparrows, *Melospiza melodia*, Pruett and Winker 2010), or large effective population size *(e.g.*, spinner dolphins, *Stenella longirostris*; Galver 2002). Care must also be taken when interpreting estimates of diagnosability derived from a large number of characters. All of these issues are discussed at length by Archer *et al*. (2017*b*).

Mitochondrial DNA sequences have formed the basis of many phylogeographic and taxonomic studies. Appendix 1 discusses the many advantages of mtDNA over nuclear markers in phylogeographic studies, as well as its drawbacks. One objection to reliance on mtDNA that is particularly relevant to assessing diagnosability is that it will fail to reflect male-mediated gene flow due to its strict maternal inheritance, and could lead to overclassification errors. Such errors are less likely when gene flow occurs due to male migration (*i.e.*, organismal gene flow) because male immigrants to a population are liable to be sampled in their new population, reducing diagnosability at all genetic markers. However, in species such as humpback whales, *Megaptera novaeangliae* (Baker *et al.* 2013), where male-mediated gene flow occurs without permanent immigration of the males (*i.e.*, gametic gene flow; Larsen *et al.* 1996, Palsbøll *et al.* 1997, Peters *et al.* 2012), and in species with strong, matri-focal social structure such as killer whales (Barrett-Lennard 2000, Ford *et al.* 2011) and false killer whales, *Pseudorca crassidens* (Martien *et al.* 2014), mtDNA-based estimates of diagnosability between social groups and populations can be as high as those typically observed between subspecies or even species (Archer *et al.* 2017*b*). Therefore, if strong social structure is suspected or if nothing is known about the social structure, the pattern observed from mtDNA should be corroborated by an independent line of evidence, such as nuclear genetic markers, morphology, or distribution (*e.g.*, different ocean basins).

Evaluating Separate Evolution

Demonstrating that two putative taxa meet some threshold of diagnosability is only the first step in subspecies delimitation. To determine whether the taxa represent subspecies or full species, one must also evaluate the extent to which the taxa are separately evolving (Taylor *et al.* 2017*b*). Cetacean taxonomists face multiple challenges when attempting to delimit subspecies and species, including recent and rapid radiations (Kingston *et al.* 2009, Steeman *et al.* 2009, McGowen 2011, Amaral *et al.* 2012), variation in mutation rate across lineages, even for the same markers (Whitehead 1998, Dornburg *et al.* 2012), and difficulty of obtaining morphological and behavioral data. Many of these challenges are also common in other taxonomic groups (*e.g.*, Britten 1986, Smith and Donoghue 2008, Pruett and Winker 2010). Researchers typically employ a variety of proxies to infer evolutionary independence, most of which are based on estimating how different the putative taxa are for the characters under study (be they morphological, behavioral, or genetic) and comparing these divergence estimates to those typically seen between accepted species. Such a comparative approach to evaluating divergence is common in genetic taxonomic studies (*e.g.*, Caballero *et al.* 2007, Wolf *et al.* 2007). Comparisons must be undertaken with great care, however, because most measures of genetic divergence are influenced by the

effective population size of the groups being compared and the characteristics of the genetic markers used, such as mutation rate, mutation model, and whether the markers are from coding or noncoding portions of the genome (Appendix 2).

Attempts have been made to use data from a broad suite of sister taxa to identify cutoff values for various genetic estimates of divergence that could serve as criteria for evaluating species status. Hey and Pinho (2012) examined the distributions of two estimates of genetic divergence derived from the program IM (Wakeley and Hey 1998, Nielsen and Wakeley 2001) to look for a qualitative difference between species, subspecies, and populations. They failed to find any qualitative differences, finding instead that the distributions of divergence estimates were broadly overlapping for the different taxonomic levels. However, the lack of resolution of different taxonomic levels in Hey and Pinho's study may in part result from some of the confounding factors discussed in Appendix 2. Had they restricted their comparisons to only divergence estimates based on comparable loci and stratified them by effective population size, they may have found less overlap between taxonomic levels. Their results may also have been influenced by the broad range of taxa included and by possibly miscategorizing comparisons (*e.g.*, taxa currently considered subspecies that actually represent different species or *vice versa*) due to taxonomic errors or uncertainty in the literature they reviewed.

Rosel *et al*. (2017*a*) applied a similar comparative approach using a broader suite of genetic divergence estimates in an attempt to identify metrics that could be used to distinguish subspecies from species and populations. Unlike Hey and Pinho (2012), Rosel *et al*. (2017*a*) limited their study to a single taxonomic group (cetaceans), took greater care to ensure that their categorizations of units as populations, subspecies, or species were well-accepted, and calculated all metrics for a comparable locus. They also found broad overlap in estimates of most metrics between comparisons at different taxonomic levels. However, they identified two metrics—percent diagnosability from Random Forests (Archer *et al.* 2017*b*; see *Traditional Multivariate Methods* below) and d_A (Nei 1987, p. 276; see *Measures of Genetic Distance* below)—that were useful in delimiting subspecies. d_A performed well at distinguishing species from subspecies and populations, while percent diagnosability was useful for distinguishing populations from subspecies. Taylor *et al*. (2017*a*) use Rosel *et al*.'s results to develop quantitative standards for delimiting cetacean subspecies using mtDNA control region data, acknowledging that some comparisons will still fall in the "gray zone" of overlap between different taxonomic levels. Tobias *et al*. (2010) developed a point-based system to weight different lines of evidence (*e.g.*, morphology, acoustics, *etc*.) to evaluate the taxonomic status of avian taxa. Although currently available data are not sufficient to develop an equivalent approach for cetaceans (Taylor *et al.* 2017*a*), most cetacean studies aimed at describing new subspecies have, nonetheless, employed multiple lines of evidence in their taxonomic arguments (see review in Rosel *et al.* 2017*b*).

Analytical Methods for Evaluating Taxonomic Hypotheses

In the following sections, we review seven different approaches to evaluate the upper and lower subspecies boundaries using genetic data: traditional multivariate methods, assignment tests, clustering methods, divergence time estimation, measures of genetic distance, tree-based methods, and dispersal rate estimators. We highlight methods that are useful for directly addressing the critical factors that distinguish the

taxonomic levels. For a more general review of analytical approaches (both genetic and nongenetic) used in identifying and delimiting species see Sites and Marshall (2004).

Traditional Multivariate Methods

Multivariate methods combine data across multiple characters to classify samples into groups. They can be useful for generating and testing taxonomic hypotheses and, in particular, for estimating diagnosability. Numerous multivariate methods have been used to evaluate diagnosability from morphological data. The most commonly used are Principal Components Analysis (PCA, Gotelli and Ellison 2004), Discriminant Function Analysis (DFA, Fisher 1936, Klecka 1980), and Classification and Regression Trees (CART, Breiman *et al.* 1984). By taking advantage of the fact that PCA utilizes Euclidean distances between samples, which can be computed from genetic data, methods have now been developed that allow these traditional multivariate methods to be applied to a variety of genetic data types, including microsatellites, single nucleotide polymorphisms (SNPs), Amplified Fragment Length Polymorphism (AFLPs), and nuclear and mtDNA sequences.

Jombart *et al.* (2010) developed a method called Discriminant Analysis of Principal Components (DAPC), which combines DFA with PCA. DAPC can be used both to describe groups of genetically-related individuals *(e.g.*, Warmuth *et al.* 2013) and assign new individuals to those groups. The percentage of samples correctly assigned to the group from which they were sampled can serve as an estimate of diagnosability. Unknown samples can also be assigned to clusters by calculating their distance from the centroid of each cluster. DAPC can be applied to multiple data types, including microsatellites, SNPs and nuclear and mitochondrial DNA (mtDNA) sequences (by converting polymorphisms in the sequence data to SNPs; Jombart *et al.* 2010).

Similar to DAPC, Kingston and Rosel (2004) used an ordination technique closely related to PCA known as nonmetric multidimensional scaling (NMDS, Borg and Groenen 2005) to analyze AFLP data from two pairs of sister species from the family Delphinidae. They used the method to characterize genetic variation within and between the species. They also conducted species identification for six "unknown" samples (*i.e.*, samples for which the species identity was withheld from the researchers until after the analysis) and showed that all six were correctly identified. Although this demonstrates that NMDS can be used to classify samples and thereby estimate diagnosability, this method has the drawback that the samples to be classified must be included in the data set used to define the transformed variables used in the analysis. Thus, the characterization of genetic variability in the groups to which samples are to be assigned can vary depending on the properties of the unknown samples.

CART (Breiman *et al.* 1984) differs from PCA and DFA in that it attempts to form a classification algorithm based on the best set of predictor variables and rules that create the groups with the least amount of overlap. Archer *et al*. (2017*b*) developed a classification method for mtDNA data based on the Random Forests algorithm (Breiman 2001), which is in turn an extension of CART. Random Forests is an ensemble classification method in which multiple classification trees are developed based on subsets (a.k.a. "training sets") of the data. The trees collectively constitute a classification "forest" that can be used to categorize unknown samples. Random Forests shares the same goal as the DFA methods used in morphological studies, namely, to develop a classification model that maximizes the probability of correct classification.

Both Archer *et al*.'s implementation of Random Forests and Jombart *et al*.'s DAPC treat each site in a sequence alignment as an independent character and use all characters concurrently to classify samples. Although different sites in a neutral sequence locus are assumed to mutate independently, they are inherited as a single unit. This is different from the inheritance pattern of morphological characters, which may be correlated due to functional constraints. Multivariate methods such as PCA create sets of uncorrelated characters, thus appropriately defining functionally correlated characters. Random Forests, on the other hand, handles potential correlations through the stochastic selection of characters (*e.g.*, nucleotide sites) during tree building. Over enough trees in the forest, correlated characters will have equal contribution to classification power, and thus do not falsely enhance the diagnosability estimate from a Random Forests analysis.

Assignment Tests

Assignment methods are used to evaluate the group membership of samples based on allele frequencies. Like traditional multivariate methods, the primary utility of assignment tests in taxonomic studies is for estimating diagnosability, the critical parameter when evaluating the lower subspecies boundary. Unlike traditional multivariate methods, however, genetic assignment methods are based on models of genetic evolution and inheritance. Manel *et al*. (2005) review assignment methods and the types of questions they can address. They define assignment tests as methods that can identify the population of origin for unknown individuals. In a classic assignment test, a reference data set consisting of samples of known origin is used to characterize the allele frequencies of groups that are defined *a priori*. Samples of unknown or uncertain origin are then assigned to the groups. Assignment methods can also be used to perform "self-assignment" tests or "re-assignments" in which samples of known origin are excluded from the reference set and then assigned to groups. This approach is typically used to estimate the assignment error rate, identify first generation migrants, or assess the level of differentiation between two strata *(e.g.*, LeDuc *et al.* 2008). For subspecies delimitation, the percent correct self-assignment can be used as an estimate of diagnosability (*e.g.*, Matocq 2002, Froufe *et al.* 2003, de Oliveira *et al.* 2008). If all samples assign back to their group of origin, the groups are 100% diagnosable. If only 90% of the samples assign to their group of origin, the groups are 90% diagnosable. Diagnosability can be calculated independently for different groups. For example, when comparing two groups, A and B, one could end up with 90% of samples from A assigning to A, but only 60% of the samples from B assigning to B. Evaluating the results of an assignment test with respect to diagnosability can be complicated by the fact that most assignment methods assign samples probabilistically, leaving the researcher to decide whether a sample that has, for instance, an assignment score of 0.55 to its taxon of origin and 0.45 to a different taxon should be considered correctly assigned (for further discussion see Archer *et al.* 2017*b*).

Assignment tests rely on detecting differences in allele frequencies between strata. Consequently, they are most powerful when applied to highly polymorphic multilocus data sets. When statistical power is low due to the use of too few loci, diagnosability will be underestimated. This could lead to failure to recognize true subspecies as such. The high mutation rate and resultant large numbers of alleles at many microsatellite loci make them ideal for use in assignment tests. SNPs can also be used in an assignment-testing framework. However, the di-allelic nature of SNPs means

that large numbers of loci are often needed to achieve good discrimination between groups, and may also require larger sample sizes for known strata (Liu *et al.* 2005, Narum *et al.* 2008, Morin *et al.* 2012). For instance, Morin *et al*. (2012) found that nearly six times as many SNP loci than microsatellite loci were needed in order to accurately assign samples from two geographically disjunct populations of bowhead whales. Assignment tests can also be applied to AFLP data. Although we are not aware of any studies in which AFLP data and assignment tests have been used to assess diagnosability of subspecies, AFLP assignment tests have been used to examine rates and patterns of introgression and hybridization between species (Vallender *et al.* 2007, Kingston *et al.* 2009, McKinnon *et al.* 2010) and to assign unknown specimens to species (Picard *et al.* 2011).

The fact that the mitochondrial genome is typically inherited as a single, nonrecombining locus precludes the use of mtDNA data in standard assignment tests. When haplotype distributions are completely nonoverlapping, assignments can be made with certainty, but shared haplotypes lead to the trivial result that the probability of an individual coming from each potential source population is equal to the frequency of its haplotype in each of the sources. Problems arise in cases where an individual has a novel haplotype (*i.e.,* one not previously documented in any source), as these cannot be assigned. Unlike mtDNA data, nuclear sequence data can be generated for multiple loci, making it more suitable to assignment tests than mtDNA sequence data. However, the low mutation rate associated with most nuclear loci and the fact that alleles are often shared across species means that a large number of loci would likely be needed in order to reliably assign individuals to sources.

Clustering Methods

Clustering methods are similar to assignment tests in that samples are sorted into groups on the basis of multilocus genotypes. They can be applied to the same data types (microsatellites, SNPs, AFLPs) that are used in assignment tests. In fact, some programs that conduct assignment tests can also be used to perform clustering analyses (*e.g.*, STRUCTURE; Pritchard *et al.* 2000, Falush *et al.* 2003, Hubisz *et al.* 2009). Clustering analyses differ from assignment tests in that samples are not stratified *a priori* into groups. Rather, the groups are identified by the analysis based on the genetic data (see Manel *et al.* 2005 for further discussion of clustering analyses *vs.* assignment tests). The lack of *a priori* stratification means that clustering methods cannot be used to estimate diagnosability, and therefore cannot be used to evaluate the lower subspecies boundary directly. The concept of diagnosability in taxonomy is based on the need to determine unambiguously to which taxon a sample of unknown origin belongs, which requires an unambiguously determined training set of samples. If there are no predefined putative taxa, it is impossible to evaluate how often samples were clustered correctly. Additionally, clustering algorithms do not produce assignment algorithms independent of the data. If a clustering analysis is re-run when a new sample is added to a data set, it is possible that the clustering of some of the original samples could change. Nonetheless, a clustering analysis in which the resulting groups perfectly correlate with geography, morphology, or some other line of evidence indicates that the groups would be 100% diagnosable if stratified appropriately.

Clustering methods are often used in taxonomic studies as a means of generating taxonomic hypotheses and looking for concordance between genetic data and other lines of evidence. Because they do not rely on *a priori* stratification, clustering

methods can be helpful in detecting cryptic taxa (Natoli *et al.* 2006, LeDuc *et al.* 2008, Wang *et al.* 2008). Natoli *et al*. (2006) used a clustering analysis to reveal strong differentiation between two groups of long-beaked common dolphins (*Delphinus delphis bairdii* and *Delphinus capensis*) that were considered at the time to be the same species. LeDuc *et al*. (2008) used the same method to show that pygmy blue whales (*Balaenoptera musculus brevicauda*) from the Pacific and Indian Oceans are as different from each other as they are from true blue whales (*B. m. intermedia*). Methods that do not require *a priori* stratification also have been developed for the explicit purpose of identifying putative species. For instance, Population Aggregation Analysis and other DNA bar coding methods (Davis and Nixon 1992, DeSalle *et al.* 2005) attempt to group individuals into putative taxa based on mtDNA sequences. The multilocus Field For Recombination method (Doyle 1995) identifies groups that do not share any alleles across multiple nuclear loci.

When taxonomic hypotheses are based on nongenetic data (*e.g.*, morphology or distribution), clustering methods can be used to test the hypothesis of concordance between genetic and nongenetic data. For instance, de Olivera *et al*. (2008) used clustering analysis to divide South American fur seals (*Arctocephalus australis*) into two groups based on genetic data and found that they closely corresponded to groups defined based on morphology. Subsequent assignment tests showed the groups to be highly diagnosable, leading to subspecies designations. Möller *et al*. (2008) used clustering to confirm the differentiation of a suspected undescribed taxon of bottlenose dolphins (*Tursiops* sp.) in Australia. Although an assignment test was not done, the perfect correspondence between geography and the groups identified by the clustering analysis indicates that an assignment test in which taxa are defined based on geography would demonstrate 100% diagnosability. Such correspondence between multiple lines of evidence makes for a strong argument in support of the taxonomic hypothesis (Padial *et al.* 2010).

Divergence Time Estimation

Estimates of the time since two putative taxa diverged are sometimes used to assess the independence of lineages by determining whether they have been separated long enough to develop subspecies- or species-level differences. Estimates of divergence time can be made from either mtDNA or nuclear sequence data. Like most other estimates of divergence, divergence times are bound to overlap widely among taxonomic groups. The divergence times of sister species vary widely within and among large taxonomic groups (Weir and Schluter 2007), and recent reviews have shown broad overlap in the distributions of divergence time estimates between species, subspecies, and populations (Hey and Pinho 2012). There are many examples of recognized species with very recent divergence times (*e.g.*, North Pacific and North Atlantic right whales <1 million years ago; Steeman *et al.* 2009), and evolutionary mechanisms by which subspecies or species can arise rapidly (Via 2009). Thus it may not be possible to set divergence time thresholds on the population/subspecies or subspecies/species boundaries. Nonetheless, an estimated divergence time of many millions of years would constitute strong evidence that lineages are separately evolving.

Divergence time estimates have been based on the assumption of a consistent and linear accumulation of genetic changes (referred to as a "molecular clock") since DNA sequences were first employed in phylogenetics (*e.g.*, Brown *et al.* 1982). Early estimates of divergence time were based on simple models and often lacked confidence intervals. These methods applied a mutation rate to a measure of genetic distance,

which had been corrected to account for multiple mutations at a site and different rates for transitions and transversions, and assumed that the mutation rates were constant across categories for all vertebrates, or at least across all taxa in the analysis (Brown *et al.* 1982). As sequence data have become more available, it has become clear that mutation rates vary among loci (Hodgkinson and Eyre-Walker 2011) and that substitution rates can vary among lineages, resulting in an overall increase or decrease in rates along different branches of the evolutionary tree (Dornburg *et al.* 2012). To accommodate this kind of rate variation, divergence time estimation has evolved to include more complex models of substitution rate to estimate rates for different loci (Duchêne *et al.* 2011, 2013) and taxonomic groups (*e.g.*, Jackson *et al.* 2009, McGowen *et al.* 2009, Xiong *et al.* 2009, Ho and Lanfear 2010). The choice of fossil-based calibration points can also influence mutation rate estimates, and therefore inference of divergence dates, and need to be appropriate to the time scales of a given study (Ho *et al.* 2008, Ho and Lanfear 2010).

The most widely used and accepted methods for inferring divergence dates are Bayesian phylogenetic inference methods, implemented in the programs BEAST (v1.7, Drummond *et al.* 2012) and MrBayes (v3.2, Ronquist *et al.* 2012). Previous methods (maximum likelihood, parsimony, distance-based) used a specified model to generate a phylogenetic tree, and a specified substitution rate could be used to infer divergence times from the branch lengths. Bayesian methods make use of additional information to help calibrate the trees and incorporate uncertainty into the substitution model parameters to explore the parameter space for various mutation and phylogenetic models (*e.g.*, relaxed clock models). They can also make use of multiple types of data (*e.g.*, DNA, proteins, and morphology) in a single analysis. These methods are rapidly improving, providing expanded model and parameter choices and integrated ways to identify the best models for different data partitions or portions of the phylogenies.

The use of Bayesian methods requires the researcher to invest more effort in the selection of appropriate prior distributions for calibration points and models for analysis. Dornburg *et al*. (2012) pointed out that inclusion of different taxonomic groups with different substitution rates can bias the results, as will the number and distribution of calibration nodes in the tree. The clock-like behavior and relative mutation rates of loci can also affect divergence time estimates (Duchêne *et al.* 2011). The placement and timing of fossils within the phylogeny is often subject to debate, and constraints on the node composition of extant taxa used as priors in the tree can also significantly affect outcomes (Dornburg *et al.* 2012). Finally, the model for branch splitting (*e.g.*, Yule speciation model *vs*. coalescent) should be evaluated based on Bayes factors to determine which model best fits the data, especially below the species level (Rannala *et al.* 2012). All of these factors will contribute to variation in divergence time estimates. Bayesian methods require that both the data and assumptions about the uncertainty are specified, which makes those assumptions more transparent, but not necessarily more correct.

Measures of Genetic Distance

Numerous measures have been used to evaluate whether the genetic distance between groups is consistent with species or subspecies status. Their utility for assessing the independence of lineages makes these types of measures important for assessing both the upper and lower subspecies boundaries. Most genetic distance measures used in taxonomic studies are based on mtDNA or nuclear sequence data. One of the

most common measures of genetic distance used in recent cetacean taxonomic studies is the number of fixed differences between sequences (mtDNA or nuclear; Rosel *et al.* 2017*b*). A fixed nucleotide difference develops between two groups when a mutation from the ancestral population is lost from one group and becomes fixed in the other, or when a new mutation arises in one group after the groups have split and then becomes fixed in the population where it originated. The presence of fixed differences thus provides evidence that two groups represent independently evolving lineages with no ongoing gene flow. However, Rosel *et al*. (2017*a*) showed that the number of fixed differences is not a good predictor of the current taxonomic status of cetacean subspecies and species.

Another common genetic distance measure used in taxonomic studies is the percent nucleotide divergence between sequences. A simple estimate of percent nucleotide divergence (*i.e.*, the average number of nucleotides that differ between individuals from different groups divided by the sequence length) does not account for within group variability or the probability that multiple mutations have occurred at some nucleotide sites and thus is not appropriate for taxonomic inference. Instead, within group variability should be accounted for by calculating net nucleotide divergence, denoted d_A (Nei 1987, p. 276). Because mutation rates vary across the genome, d_A values are only comparable when using sequences that meet the assumption that for the portion compared mutation rates are comparable. For example, d_A is expected to be higher comparing the highly mutable control region than when comparing the full mitogenome, which has large, highly conserved regions under selection. Rosel *et al*. (2017*a*) found that for cetaceans, d_A calculated from control region sequence was very effective at distinguishing between populations, subspecies, and species, with species generally exhibiting values of d_A greater than 0.02 and populations exhibiting values less than 0.004. Taylor *et al*. (2017*a*) propose using these values of d_A as quantitative thresholds for delimiting subspecies of cetaceans using mtDNA control region sequence data.

To account for multiple substitutions and back-substitutions when calculating nucleotide divergence, the distance between a pair of sequences should be calculated using the mutation model that best fits the data. Programs such as jModelTest2 (Posada 2008) can be used to identify the best mutation model. Divergence values calculated under different substitution models are not necessarily equivalent (Fregin *et al.* 2012), calling into question the validity of comparing values across studies, even when each study used the best mutation model for their data set. Thus, care must be taken that comparisons between pairs of taxa are only made for comparable loci and comparable measures of divergence (Fregin *et al.* 2012). Furthermore, authors should clearly state the exact metric used to calculate divergence, including a citation to the original description of the metric, as the terms "sequence divergence," "nucleotide divergence," and "genetic distance" are not always applied consistently in the literature (Rosel *et al.* 2017*b*).

Estimates of genetic distance can also be made from nonsequence data (*e.g.*, microsatellites, SNPs, and AFLPs), though they are used far less frequently in taxonomic studies. These include Euclidean distance, Cavalli-Sforza chord distance (Cavalli-Sforza and Edwards 1967), Reynold's distance (Reynolds *et al.* 1983), Nei's distance (Nei 1978), and $(\delta\mu)^2$ (Goldstein *et al.* 1995). Most of these statistics were originally developed for allozyme data and therefore assume an infinite-alleles mutation model, rendering them inappropriate for microsatellite data. However, the distance measure $(\delta\mu)^2$ was developed specifically for microsatellite data. The percentage of "private alleles" at microsatellite or SNP markers (Robineau *et al.* 2007, Wolf

et al. 2007) or unique bands for AFLP markers *(e.g.*, Wang *et al.* 2003) can also be used as estimates of genetic distance. The remaining measures we discuss apply only to sequence data.

Tree-based Methods

Trees have long been a staple of phylogenetic studies. By examining how samples are grouped in clades on a tree, researchers can make inferences about the evolutionary relationships of the samples. Phylogenetic trees can be used to address questions at a broad range of scales, from intraspecific studies of gene flow and dispersal patterns *(e.g.*, Avise 1992) to taxonomic studies aimed at elucidating the relationships between species and genera *(e.g.*, McGowen *et al.* 2009). When delimiting subspecies, trees can be useful for assessing the independence of lineages, which is necessary when testing both the upper and lower subspecies boundaries. Trees are also valuable for corroborating existing taxonomic hypotheses and for generating new ones. Phylogenetic trees are generated from mtDNA or nuclear sequence data.

Numerous analytical approaches have been proposed for using gene trees to delimit species, many of which are reviewed by Sites and Marshall (2004). Most of these methods are aimed at identifying reciprocally monophyletic groups. However, reciprocal monophyly has been shown to be a poor proxy for species status. Funk and Omland (2003) conducted a literature search across a broad suite of taxonomic groups and concluded that 23% of species did not show a pattern of reciprocal monophyly. Of the many potential causes of nonmonphyly, those that are most likely to affect species delimitation in cetaceans are introgressive hybridization, incomplete lineage sorting at neutral markers due to large effective population size, recent radiation, and rapid divergence due to selection (Appendix 2). Many of these issues are exhibited by members of the subfamily Delphininae, in which both species delimitation and resolution of higher-level taxonomic relationships have been challenging (LeDuc *et al.* 1999, Kingston *et al.* 2009, McGowen *et al.* 2009, Andrews *et al.* 2013). In some cases (*e.g.*, recent radiation), these problems can be overcome with longer sequences, which can improve phylogenetic resolution and lead to reciprocal monophyly *(e.g.*, Morin *et al.* 2010). However, relying on reciprocal monophyly to evaluate either the lower or upper subspecies boundary can also lead to overclassification errors. This can occur when analyzing mtDNA in a taxon with substantial male-mediated gametic gene flow (see Appendix 1) and in small populations where lineage sorting may occur before adaptive divergence occurs.

Not all tree-based methods for assessing divergence rely on complete reciprocal monophyly. Cummings *et al*. (2008) developed the genealogical sorting index (*gsi*) for quantifying the degree of genealogical cohesion or "exclusivity" for groups on phylogenetic trees. Cummings *et al*. (2008) also described a permutation test that can be used to assess the statistical significance of *gsi* values. They showed that this permutation test was able to reject the null hypothesis of random ancestry soon (*i.e.*, fewer than N_e generations) after two groups diverged, whereas monophyly developed much more slowly. The *gsi* permutation test appears to be a powerful tool for evaluating genealogical cohesion within groups, and thus avoiding the problem of overclassification due to the high levels of diagnosability achievable with larger numbers of markers. However, simply rejecting the null hypothesis of no divergence would not be considered sufficient justification for defining new species. Thus, as with most proxies for inferring separate evolutionary pathways to evaluate the subspecies/species

boundary, the challenge with *gsi* is determining whether a particular value of the statistic equates to species-level differences.

In recent years, there has been considerable work aimed at bringing coalescent theory to bear on problems of species delimitation (Knowles and Carstens 2007, Liu 2008, Monaghan *et al.* 2009, Ence and Carstens 2010, Hausdorf and Hennig 2010, Yang and Rannala 2010). Several of these methods were recently used to investigate phylogenetic relationships within the notoriously difficult cetacean subfamily Delphininae (Amaral *et al.* 2012), although with only moderate success in resolving relationships (Andrews *et al.* 2013). Fujita *et al*. (2012) review coalescent-based species delimitation and several methods available for using this approach. All these methods use gene trees from multiple loci to investigate patterns of speciation. The coalescent process is used to model stochastic lineage sorting at different loci to resolve conflicts between branching patterns from different gene trees and estimate the point at which the branching pattern switches from one consistent with an intraspecific coalescent to one consistent with separately evolving species. Consequently, these methods can be used to identify species boundaries even in the presence of extensive incomplete lineage sorting. Several coalescent methods are specifically designed to evaluate hypotheses regarding the species status of specific lineages (*e.g.*, Ence and Carstens 2010, Yang and Rannala 2010) and may be particularly valuable in evaluating the subspecies/species boundary, though there is concern that relying solely on these rapidly-developing methods to delimit species may lead to taxonomic inflation by conferring species status when it is not warranted (see Bauer *et al.* 2011).

Dispersal Rate Estimators

Gene flow is the primary force preventing allopatric populations from diverging due to the effects of genetic drift and selection pressures imposed by different habitats. Therefore, it is natural to look to estimates of gene flow when attempting to determine whether two groups represent independently evolving lineages. The most popular and simplest estimates of gene flow are the *F*-statistics, originally developed by Wright (1931), which can be estimated from allelic and genotypic frequency distributions of any nuclear or mitochondrial marker. In a recent review, Rosel *et al*. (2017*b*) found that 28% of papers delimiting new cetacean species and 83% of papers delimiting new cetacean subspecies included estimates of F_{ST}.

Although F_{ST} and its analogues are extremely valuable for addressing demographic questions at the population level, they are of limited utility in subspecies and species delimitation. For analyses where the putative taxa could be species, it is plausible that the dispersal rate is actually zero, which invalidates calculations of F_{ST} and related statistics. This is particularly true when estimates are based on microsatellite loci, whose high mutation rate and complicated mutation model violate the assumptions of F_{ST} calculations (Meirmans and Hedrick 2011). Although the effect of mutation rate is typically assumed to be trivial and ignored (Hedrick 2005), the high mutation rates associated with highly variable markers such as microsatellites will result in F_{ST} equilibrating at a value less than one even in the absence of gene flow. For instance, the values of F_{ST} and its various analogues calculated from microsatellite data between two recognized subspecies of blue whales are low and variable, ranging from 0.008 to 0.252, depending on which analogue is used and which samples are included in the comparison (Table 2; LeDuc *et al.* 2007). Even leatherback turtles (*Dermochelys coriacea*) and green turtles (*Chelonia mydas*), two species believed to have diverged over 100 million years ago (Duchêne *et al.* 2012), exhibit an F_{ST} value based

Table 2. Values of *F*-statistics between pairs of subspecies (first three examples) and species (last five examples) calculated from microsatellite data. Sample sizes are given in parentheses after each subspecies/species. Divergence metrics for each data set were generated using the strataG package in R (R Development Core Team 2011, Archer *et al.* 2017*a*). The divergence metrics were F_{ST} (Weir and Cockerham 1984), G_{ST} (Nei 1987), G'_{ST} (Hedrick 2005), G''_{ST} (Meirmans 2006) and D (Jost 2008). F_{ST} was converted into an estimate of N_em using Wright's (1931) formula $F_{ST} = 1/(4N_em + 1)$. For microsatellite data generated at the Southwest Fisheries Science Center (first two and last three examples), we genotyped 1–3 individuals of each species pair together for each microsatellite to assign allele sizes, then normalized all genotypes from each species pair using the program Allelogram (Morin *et al.* 2009).

Species pair	No. of loci	F_{ST}	G_{ST}	G'_{ST}	G''_{ST}	D	N_em
True blue whale (46) *vs.* Indian Ocean pygmy blue whale (36)[a]	7	0.082	0.037	0.224	0.252	0.200	2.799
True blue whale (46) *vs.* Pacific Ocean pygmy blue whale (28)[a]	7	0.030	0.008	0.055	0.063	0.061	8.083
Hector's dolphin (159) *vs.* Maui dolphin (58)[b]	9	0.178	0.103	0.369	0.428	0.123	1.154
Domestic sheep *vs.* Bighorn sheep[c]	8	0.237	0.134	0.790	0.815	0.633	0.805
California sea lion (16) *vs.* Galapagos sea lion (367)[d]	22	0.210	0.103	0.523	0.567	0.322	0.940
Green turtle (92) *vs.* Leatherback turtle (125)[e,f]	3	0.244	0.127	0.485	0.543	0.203	0.775
Beluga whale (299) *vs.* Harbor Porpoise (220)[g,h]	4	0.154	0.083	0.831	0.844	0.496	1.373
False killer whale (108) *vs.* common bottlenose dolphin (142)[i,j]	5	0.122	0.063	0.418	0.453	0.326	1.799

[a]LeDuc *et al.* (2007).
[b]Hamner *et al.* (2012).
[c]Forbes *et al.* (1995).
[d]Wolf *et al.* (2007).
[e]Dutton *et al.* (2013).
[f]Roden *et al.* (2013).
[g]O'Corry-Crowe *et al.* (2010).
[h]Chivers *et al.* (2002).
[i]Martien *et al.* (2014).
[j]Martien *et al.* (2012).

on microsatellite data that is consistent with gene flow at a rate of nearly one migrant per generation (Table 2).

A recent review found similarly poor performance of F_{ST} for discriminating between taxonomic levels (Hey and Pinho 2012). The weak correlation between F_{ST} and taxonomic level likely reflects the dependence of F_{ST} estimators on effective population size and mutation rate (Appendix 2). Because of their dependence on mutation rate, comparing F_{ST} values across taxa or loci with different levels of heterozygosity is not meaningful. The exception is the F_{ST} analogue Φ_{ST} (Excoffier *et al.* 1992), which can be calculated for sequence data. Because it explicitly incorporates the mutation process by taking into consideration the genetic distance between haplotypes, it is not similarly biased by increased mutation rate (Kronholm *et al.* 2010). However, Φ_{ST} does depend on having the genetic distances estimated by the appropriate mutation model (see discussion above). We refer readers to Meirmans and Hedrick (2011) for a thorough review of the limitations of F_{ST} analogues.

Although F_{ST} analogues are commonly used as measures of differentiation, they are rarely converted into actual estimates of dispersal rate in cetacean taxonomic papers (Rosel *et al.* 2017*b*). However, alternative methods of estimating dispersal rates, such as those used in the programs IM (Wakeley and Hey 1998, Nielsen and Wakeley 2001), Migrate (Beerli and Felsenstein 1999, 2001; Beerli and Palczewski 2010) and LAMARC (Kuhner 2006), are sometimes used *(e.g.*, Wang *et al.* 2008). The model-based approaches used in these programs allow for simultaneous inference of multiple demographic parameters, including N_e, m, and divergence time, thereby avoiding some of the problems associated with F_{ST} analogues. However, Hey and Pinho (2012) reviewed estimates of dispersal rate generated by the program IM for a wide variety of species, subspecies, and population pairs and found that the dispersal rate estimates showed even lower correlation with taxonomic status than estimates of F_{ST}.

Conclusions

The two key features that must be evaluated in a taxonomic argument for subspecies delimitation are the degree of diagnosability between putative taxa (Amadon 1949, Patten and Unitt 2002) and the extent to which they are separately evolving. Of the analytical approaches we reviewed, only two can be used to produce diagnosability estimates from genetic data (Table 1)—assignment tests and genetic implementations of traditional multivariate methods. Taylor *et al*. (2017*a*) include an estimate of diagnosability from a multivariate method (Random Forests; Archer *et al.* 2017*b*) in their quantitative guidelines for delineating subspecies based on mtDNA control region sequence. Taylor *et al*. give thorough consideration to the threshold above which diagnosability is considered high enough to warrant subspecies designation. They discuss the merits of both an empirically-based threshold (80%) and a higher threshold (95%) that is arbitrary, but consistent with the threshold used for normally distributed morphological characters. While acknowledging the merits of an empirically derived threshold, Taylor *et al*. ultimately recommend the continued use of the higher (95%) threshold for consistency with past studies.

Although they cannot be used to directly estimate diagnosability, the other methods we reviewed are, nonetheless, also potentially useful in genetic studies aimed at evaluating the population/subspecies boundary. For instance, clustering methods can be quite valuable in identifying cryptic taxa or testing hypotheses of divergence

generated *via* other means (*e.g.*, de Oliveira *et al.* 2008, Möller *et al.* 2008). Similarly, estimates of gene flow can be used to test for male-mediated gene flow in cases where subspecific status is indicated by mtDNA.

Evaluating the extent to which two groups represent separately evolving lineages is often more challenging than evaluating diagnosability. Though diagnosability can be directly estimated, assessing evolutionary independence requires the use of proxies. Because the expected level of evolutionary independence is lower for subspecies than for species, evaluating evolutionary independence at the subspecies/population boundary will generally be easier than at the subspecies/species boundary. For example, clustering of the diagnosable groups on a phylogenetic tree, distinct from groups of similar taxonomic rank, would indicate that they had been evolving relatively independently for long enough to allow the *in situ* generation of new mutations (*e.g.*, Kershaw *et al.* 2013). Estimates of divergence time or genetic distance (*e.g.*, d_A; Rosel *et al.* 2017*a*) could also be used to assess whether a group has been isolated long enough to warrant subspecies status.

When judging whether the degree of divergence between two putative taxa is consistent with subspecies or species designation, the strongest arguments will be those that use an integrative approach to combine the results from several different analytical methods and data types (DeSalle *et al.* 2005, Padial *et al.* 2010, Tobias *et al.* 2010). Such an integrative approach is consistent with the recommendations of a 2004 Taxonomy Workshop aimed at addressing the shortcomings in current cetacean taxonomy, which state that at least two independent lines of evidence (which could be two different types of genetic markers, such as mtDNA and nuclear sequence data) are necessary for delineating new species (Reeves *et al.* 2004).

Comparative studies have shown that taxonomic status is not strongly correlated with estimates of divergence time, *F*-statistics, or the number or proportion of fixed difference between groups (Hey and Pinho 2012, Rosel *et al.* 2017*a*). However, Rosel *et al*. did find that d_A performed well at distinguishing between populations, subspecies, and species for cetaceans, with subspecies predominantly falling into the range $0.004 < d_A < 0.02$. Although these values are specific to the marker type (mitochondrial control region sequence) and taxa Rosel *et al*. examined, the results provide important threshold values for the most commonly used type of genetic data in cetacean taxonomic studies and suggest that similar threshold values of d_A could be identified for other marker types and taxonomic groups. Taylor *et al*. (2017*a*) used Rosel *et al*.'s results to develop a set of qualitative and quantitative standards designed to assist researchers in evaluating multiple lines of evidence in cetacean taxonomic studies. Quantitative standards such as those proposed by Taylor *et al*. for control region sequence data should bring consistency and reproducibility to the construction of arguments for subspecies and species designations.

Increases in the quantity and quality of neutral markers and the addition of selected markers will bring new opportunities and challenges to the analysis and interpretation of genetic data for improving taxonomy. In this exciting period of rapid changes, it becomes all the more important to develop and evolve standards for interpreting genetic data in a taxonomic context to lend order and consistency to the discovery process. By focusing on the features that distinguish subspecies from the taxonomic units above (*i.e.*, species) and below (*i.e.*, populations) them and choosing methods that evaluate those critical features, researchers can make more convincing and consistent arguments for subspecies delimitation.

Acknowledgments

We would like to thank Bill Perrin, Annalisa Berta, and Sarah Mesnick for their thoughtful contributions to the ideas presented in this paper. We are also indebted to Aimee Lang, John Hyde, Bill Perrin, Michael Schwartz, and two anonymous reviewers for their comments on the manuscript.

Literature Cited

Amadon, D. 1949. The seventy-five per cent rule for subspecies. The Condor 51:250–258.

Amaral, A. R., J. A. Jackson, L. M. Möller, L. B. Beheregaray and M. M. Coelho. 2012. Species tree of a recent radiation: The subfamily Delphininae (Cetacea, Mammalia). Molecular Phylogenetics and Evolution 64:243–253.

Andrews, K. R., W. F. Perrin, M. Oremus, L. Karczmarski, B. W. Bowen, J. B. Puritz and R. J. Toonen. 2013. The evolving male: Spinner dolphin (*Stenella longirostris)* ecotypes are divergent at Y chromosome but not mtDNA or autosomal markers. Molecular Ecology 22:2408–2433.

Archer, F. I., P. E. Adams and B. B. Schneiders. 2017*a*. STRATAG: An R package for manipulating, summarizing, and analyzing population genetic data. Molecular Ecology Resources 17:5–11.

Archer, F. I., K. K. Martien and B. L. Taylor. 2017*b*. Diagnosability of mtDNA with Random Forests: Using sequence data to delimit subspecies. Marine Mammal Science 33 (Special Issue):101–131.

Avise, J. C. 1992. Molecular population structure and the biogeographic history of regional fauna: A case history with lessons for conservation biology. Oikos 63:62–76.

Baird, N. A., P. D. Etter, T. S. Atwood, *et al*. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376.

Baker, C. S., D. Steel, J. Calambokidis, *et al*. 2013. Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. Marine Ecology Progress Series 494:291–306.

Ballard, J. W. O., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13:729–744.

Barrett-Lennard, L. G. 2000. Population structure and mating patterns of killer whales (*Orcinus orca*) as revealed by DNA analysis. Ph.D. thesis, University of British Columbia, Vancouver, Canada. 97 pp.

Bauer, A. M., J. F. Parham, R. M. Brown, *et al*. 2011. Availability of new Bayesian-delimited gecko names and the importance of character-based species descriptions. Proceedings of the Royal Society B: Biological Sciences 278:490–492.

Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763–773.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences of the United States of America 98:4563–4568.

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

Borg, I., and P. Groenen. 2005. Modern multidimensional scaling: Theory and applications. Springer-Verlag, New York, NY.

Braby, M. F., R. Eastwood and N. Murray. 2012. The subspecies concept in butterflies: Has its application in taxonomy and conservation biology outlived its usefulness? Biological Journal of the Linnean Society 106:699–716.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1984. Classification and regression trees. Wadsworth Inc, Monterey, CA.

Britten, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1391–1398.

Brown, W. M., E. M. Prager, A. Wang and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. Journal of Molecular Evolution 18:225–239.

Caballero, S., F. Trujillo, J. A. Vianna, *et al*. 2007. Taxonomic status of the genus Sotalia: Species level ranking for "tucuxi" (*Sotalia fluviatilis*) and "costero" (*Sotalia guianensis*) dolphins. Marine Mammal Science 23:358–386.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. American Journal of Human Genetics 19:233–257.

Chan, K. M., and S. A. Levin. 2005. Leaky prezygotic isolation and porous genomes: Rapid introgression of maternally inherited DNA. Evolution 59:720–729.

Chivers, S. J., A. E. Dizon, P. J. Gearin and K. M. Robertson. 2002. Small-scale population structure of eastern North Pacific harbor porpoises (*Phocoena phocoena*) indicated by molecular genetic analyses. Journal of Cetacean Research and Management 4:111–122.

Cummings, M. P., M. C. Neel and K. L. Shaw. 2008. A genealogical approach to quantifying lineage divergence. Evolution 62:2411–2422.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen and M. L. Baxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499–510.

Davis, J. I., and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Systematic Biology 41:421–435.

de Oliveira, L. R., J. I. Hoffman, E. Hingst-Zaher, P. Majluf, M. M. C. Muelbert, J. S. Morgante and W. Amos. 2008. Morphological and genetic evidence for two evolutionarily significant units (ESUs) in the South American fur seal, *Arctocephalus gazella* [*sic*]. Conservation Genetics 9:1451–1466.

de Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology 56:879–886.

Degnan, S. M. 1993. The perils of single gene trees—mitochondrial versus single-copy nuclear DNA variation in white-eyes (Aves: Zosteropidae). Molecular Ecology 2:219–225.

DeSalle, R., M. G. Egan and M. Siddal. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 360:1905–1916.

Dornburg, A., M. C. Brandley, M. R. McGowen and T. J. Near. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). Molecular Biology and Evolution 29:721–736.

Doyle, J. J. 1995. The irrelevance of allele tree topologies for species delimitation, and a non-topological alternative. Systematic Botanay 20:574–588.

Drummond, A. J., M. A. Suchard, D. Xie and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29:1969–1973.

Duchêne, S., F. I. Archer, J. Vilstrup, S. Caballero and P. A. Morin. 2011. Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS ONE 6:e27138.

Duchêne, S., A. Frey, A. Alfaro-Núñez, P. H. Dutton, M. T. Gilbert and P. A. Morin. 2012. Marine turtle mitogenome phylogenetics and evolution. Molecular Phylogenetics and Evolution 65:241–250.

Duchêne, S., M. Molak and S. Y. Ho. 2013. ClockstaR: Choosing the number of relaxed-clock models in molecular phylogenetic analysis. Bioinformatics 30:1017–1019.

Dupuis, J. R., A. D. Roe and F. A. H. Sperling. 2012. Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. Molecular Ecology 21:4422–4436.

Dutton, P. H., S. E. Roden, K. R. Stewart, *et al*. 2013. Population stock structure of leatherback turtles (*Dermocherlys coriacea*) in the Atlantic revealed using mtDNA and microsatellite markers. Conservation Genetics 14:625–636.

Edwards, S., and S. Bensch. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Molecular Ecology 18:2930–2933.

Edwards, S. V., S. B. Kingan, J. D. Calkins, C. N. Balakrishnan, W. B. Jennings, W. J. Swanson and M. D. Sorenson. 2005. Speciation in birds: Genes, geography and sexual selection. Proceedings of the National Academy of Sciences of the United States of America 102:6550–6557.

Emerson, K. J., C. R. Merz, J. M. Catchen, P. A. Hohenlohe, W. A. Cresko, W. E. Bradshaw and C. M. Holzapfel. 2012. Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences of the United States of America 107:16196–16200.

Ence, D. D., and B. C. Carstens. 2010. SpedeSTEM: A rapid and accurate method for species delimitation. Molecular Ecology Resources 11:473–480.

Engelhaupt, D., A. R. Hoelzel, C. Nicholson, *et al*. 2009. Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (*Physeter macrocephalus*). Molecular Ecology 18:4193–4205.

Escorza-Trevino, S., F. I. Archer, M. Rosales, A. M. Lang and A. E. Dizon. 2005. Genetic differentiation and intraspecific structure of Eastern Tropical Pacific spotted dolphins, *Stenella attenuata*, revealed by DNA analyses. Conservation Genetics 6:587–600.

Excoffier, L., P. E. Smouse and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Falush, D., M. Stephens and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7:179–188.

Forbes, S. J., J. T. Hogg, F. C. Buchanan, A. C. Crawford and F. W. Allendorf. 1995. Microsatellite evolution in congeneric mammals: Domestic and bighorn sheep. Molecular Biology and Evolution 12:1106–1113.

Ford, M. J., M. B. Hanson, J. A. Hempelmann, *et al*. 2011. Inferred paternity and male reproductive success in a killer whale (*Orcinus orca*) population. Journal of Heredity 102:537–553.

Fregin, S., M. Haase, U. Olsson and P. Alström. 2012. Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae. Molecular Phylogenetics and Evolution 62:319–328.

Froufe, E., I. Knizhin, M. T. Koskinen, C. R. Primmer and S. Weiss. 2003. Identification of reproductively isolated lineages of Amur grayling (*Thymallus grubii* Dybowski 1869): Concordance between phenotypic and genetic variation. Molecular Ecology 12:2345–2355.

Fujita, M. K., A. D. Leaché, F. T. Burbrink, J. A. McGuire and C. Moritz. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology & Evolution 27:480–488.

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution and Systematics 34:397–423.

Funk, W. C., J. K. McKay, P. A. Hohenlohe and F. W. Allendorf. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution 27:489–496.

Galver, L. M. 2002. The molecular ecology of spinner dolphins, *Stenella longirostris*: Genetic diversity and population structure. Ph.D. thesis, University of California, San Diego, CA. 192 pp.

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza and M. W. Feldman. 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471.

Gotelli, N. J., and A. M. Ellison. 2004. A primer of ecological statistics. Sinauer Associates, Sunderland, MA.

Haig, S. M., E. A. Beever, S. M. Chambers, *et al*. 2006. Taxonomic considerations in listing subspecies under the US Endangered Species Act. Conservation Biology 20:1584–1594.

Hale, M. L., T. M. Burg and T. E. Steeves. 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS ONE 7:e45170.

Hamner, R. M., F. B. Pichler, D. Heimeier, R. Constantine and C. S. Baker. 2012. Genetic differentiation and limited gene flow among fragmented populations of New Zealand endemic Hector's and Maui's dolphins. Conservation Genetics 13:987–11002.

Hancock-Hanser, B. L., A. Frey, M. S. Leslie, P. H. Dutton, E. I. Archer and P. A. Morin. 2013. Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear sequencing for population genomics. Molecular Ecology Resources 13:254–268.

Hausdorf, B., and C. Hennig. 2010. Species delimitation using dominant and codominant multilocus markers. Systematic Biology 59:491–503.

Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:1633–1638.

Hey, J., and C. Pinho. 2012. Population genetics and objectivity in species diagnosis. Evolution 66:1413–1429.

Ho, S. Y. W., and R. Lanfear. 2010. Improved characterization of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. Mitochondrial DNA 21:138–146.

Ho, S. Y., U. Saarma, R. Barnett, J. Haile and B. Shapiro. 2008. The effect of inappropriate calibration: Three case studies in molecular ecology. PLoS ONE 3(2):e1615.

Hodgkinson, A., and A. Eyre-Walker. 2011. Variation in the mutation rate across mammalian genomes. Nature Reviewes Genetics 12:756–766.

Hoelzel, A. R., M. Dahlheim and S. J. Stern. 1998. Low genetic variation among killer whales (*Orcinus orca*) in the eastern North Pacific and genetic differentiation between foraging specialists. Journal of Heredity 89:121–128.

Hoelzel, A. R., J. M. Hancock and G. A. Dover. 1991. Evolution of the cetacean mitochondrial D-loop region. Molecular Biology and Evolution 8:475–493.

Hohenlohe, P. A., P. A. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics 6:e1000862.

Hubisz, M. J., D. Falush, M. Stephens and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322–1332.

Jackson, J., C. S. Baker, M. Vant, D. J. Steel, L. Medrano-González and S. R. Palumbi. 2009. Big and slow: Phylogenetic estimates of molecular evolution in baleen whales (suborder Mysticeti). Molecular Biology and Evolution 26:2427–2440.

James, F. C. 2010. Avian subspecies: Introduction. Ornithological Monographs 67:1–5.

Jombart, T., S. Devillard and F. Balloux. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics 11:94.

Jost, L. 2008. *GST* and its relatives do not measure differentiation. Molecular Ecology 17:4015–4026.

Karl, S. A., R. J. Toonen, W. S. Grant and B. W. Bowen. 2012. Common misconceptions in molecular ecology: Echoes of the modern synthesis. Molecular Ecology 21:4171–4189.

Kershaw, F., M. S. Leslie, T. Collins, *et al*. 2013. Population differentiation of 2 forms of Bryde's whales in the Indian and Pacific Oceans. Journal of Heredity 104:755–764.

Kingston, S. E., and P. E. Rosel. 2004. Genetic differentiation among recently diverged delphinid taxa determined using AFLP markers. Journal of Heredity 95:1–10.

Kingston, S. E., L. D. Adams and P. E. Rosel. 2009. Testing mitochondrial sequences and anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group: Molecular systematics of the Delphininae (Cetacea: Odontoceti: Delphinidae). BMC Evolutionary Biology 9:245–263.

Klecka, W. R. 1980. Discriminant analysis. Sage Publications, Thousand Oaks, CA.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. Systematic Biology 56:887–895.

Kronholm, I., O. Loudet and J. de Meaux. 2010. Influence of mutation rate on estimators of genetic differentiation – lessons from *Arabidopsis thaliana*. BMC Genetics 11:33.

Kuhner, M. K. 2006. LAMARC 2.0: Maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22:768–770.

Larsen, A. H., J. Sigurjonsson, N. I. Oien, G. Vikingsson and P. J. Palsbøll. 1996. Populations genetic analysis of nuclear and mitochondrial loci in skin biopsies collected from central and northeastern North Atlantic humpback whales (*Megaptera novaeangliae*): Population identity and migratory destinations. Proceedings of the Royal Society B: Biological Sciences 263:1611–1618.

LeDuc, R. G., W. F. Perrin and A. E. Dizon. 1999. Phylogenetic relationships among the delphinid cetaceans based on full cytochrome *b* sequences. Marine Mammal Science 15:619–648.

LeDuc, R. G., A. E. Dizon, M. Goto, *et al*. 2007. Patterns of genetic variation in Southern Hemisphere blue whales and the use of assignment test to detect mixing on the feeding grounds. Journal of Cetacean Research and Management 9:73–80.

LeDuc, R. G., K. M. Robertson and R. L. Pitman. 2008. Mitochondrial sequence divergence among Antarctic killer whale ecotypes is consistent with multiple species. Biology Letters 4:426–429.

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24:2542–2543.

Liu, N., L. Chen, S. Wang, C. Oh and H. Zhao. 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genetics 6:S26.

Mallet, J. 2004. Poulton, Wallace and Jordan: How discoveries in *Papilio* butterflies led to a new species concept 100 years ago. Systematics and Biodiversity 1:441–452.

Manel, S., M. K. Schwartz, G. Luikart and P. Taberlet. 2003. Landscape genetics: Combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189–197.

Manel, S., O. E. Gaggiotti and R. S. Waples. 2005. Assignment methods: Matching biological questions with appropriate techniques. Trends in Ecology & Evolution 20:136–142.

Martien, K. K., R. W. Baird, N. M. Hedrick, *et al*. 2012. Population structure of island-associated dolphins: Evidence from mitochondrial and microsatellite markers for common bottlenose dolphins (*Tursiops truncatus*) around the main Hawaiian Islands. Marine Mammal Science 28:E208–E232.

Martien, K. K., S. J. Chivers, R. W. Baird, *et al*. 2014. Nuclear and mitochondrial patterns of population structure in North Pacific false killer whales (*Pseudorca crassidens*). Journal of Heredity 105:611–626.

Matocq, M. D. 2002. Morphological and molecular analysis of a contact zone in the *Neotoma fuscipes* species complex. Journal of Mammalogy 83:866–883.

Mayr, E. 1969. Principles of systematic zoology. McGraw-Hill, New York, NY.

McGowen, M. R. 2011. Toward the resolution of an explosive radiation—A multilocus phylogeny of oceanic dolphins (Delphinidae). Molecular Phylogenetics and Evolution 60:345–357.

McGowen, M. R., M. Spaulding and J. Gatesy. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Molecular Phylogenetics and Evolution 53:891–906.

McKinnon, G. E., J. J. Smith and B. M. Potts. 2010. Recurrent nuclear DNA introgression accompanies chloroplast DNA exchange in two eucalypt species. Molecular Ecology 19:1367–1380.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60:2399–2402.

Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11:5–18.

Meyer, M., and M. Kircher. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010: pdb.prot5448.

Milinkovitch, M. C., R. G. LeDuc, R. Tiedeman and A. E. Dizon. 2002. Applications of molecular data in cetacean taxonomy and population genetics with special emphasis on defining species boundaries. Pages 325–359 *in* P. G. H. Evans and J. A. Raga, eds. Marine mammals: Biology and conservation. Kluwer Academic/Plenum Publishers, New York, NY.

Möller, L. M., K. Bilgmann, K. Charlton-Robb and L. B. Beheregaray. 2008. Multi-gene evidence for a new bottlenose dolphin species in southern Australia. Molecular Phylogenetics and Evolution 49:674–681.

Monaghan, M. T., R. Wild, M. Elliot, *et al*. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Systems Biology 58:298–311.

Morin, P. A., and M. McCarthy. 2007. Highly accurate SNP genotyping from historical and low-quality samples. Molecular Ecology Notes 7:937–946.

Morin, P. A., C. Manaster, S. L. Mesnick and R. Holland. 2009. Normalization and binning of historical and multi-source microsatellite data: Overcoming the problems of allele size shift with ALLELOGRAM. Molecular Ecology Resources 9:1451–1455.

Morin, P. A., F. I. Archer, A. D. Foote, *et al*. 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Research 20:908–916.

Morin, P. A., F. I. Archer, V. L. Pease, *et al*. 2012. Empirical comparison of single nucleotide polymorphisms and microsatellites for population and demographic analyses of bowhad whales. Endangered Species Research 19:129–147.

Narum, S. R., M. Banks, T. D. Beacham, *et al*. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Molecular Ecology 17:3464–3477.

Natoli, A., A. Canadas, V. M. Peddemors, A. Aguilar, C. Vaquero, P. Fernandez-Piqueras and A. R. Hoelzel. 2006. Phylogeography and alpha taxonomy of the common dolphin (*Delphinus* sp.). Journal of Evolutionary Biology 19:943–954.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 76:379–390.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation. A Markov change Monte Carlo approach. Genetics 158:885–896.

Nosil, P., and D. Schulter. 2011. The genes underlying the process of speciation. Trends in Ecology & Evolution 26:160–167.

O'Corry-Crowe, G., C. Lydersen, M. P. Heide-Jørgensen, L. Hansen, L. M. Mukhametov, O. Dove and K. M. Kovacs. 2010. Population genetic structure and evolutionary history of North Atlantic beluga whales (*Delphinapterus leucas*) from West Greenland, Svalbard and the White Sea. Polar Biology 33:1179–1194.

Oremus, M., M. M. Poole, D. Steel and C. S. Baker. 2007. Isolation and interchange among insular spinner dolphin communities in the South Pacific revealed by individual identification and genetic diversity. Marine Ecology Progress Series 336:275–289.

Padial, J. M., A. Miralles, I. De la Riva and M. Vences. 2010. The integrative future of taxonomy. Frontiers in Zoology 7:16.

Palsbøll, P. J., J. Allen, M. Bérubé, *et al*. 1997. Genetic tagging of humpback whales. Nature 388:767–769.

Patten, M. A., and P. Unitt. 2002. Diagnosability versus mean differences of sage sparrow subspecies. The Auk 119:26–35.

Perrin, W. F. 1990. Subspecies of *Stenella longirostris* (Mammalia: Cetacea: Delphinidae). Proceedings of the Biological Society of Washington 103:453–463.

Perrin, W. F., G. D. Schnell, D. J. Hough, J. W. Gilpatrick Jr. and J. V. Kashiwada. 1994. Reexamination of geographic variation in cranial morphology of the pantropical spotted dolphin, *Stenella attenuata*, in the eastern Pacific. Fisheries Bulletin 92:324–346.

Peters, J. L., K. A. Bolender and J. M. Pearce. 2012. Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: The role of male-biased dispersal in a Holarctic sea duck. Molecular Ecology 21:3562–3575.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

Picard, C. J., M. H. Villet and J. D. Wells. 2011. Amplified fragment length polymorphism confirms reciprocal monophyly in *Chrysomya putoria* and *Chrysomya chloropyga*: A correction of reported shared mtDNA haplotypes. Medical and Veterinary Entomology 26:116–119.

Posada, D. 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25:1253–1256.

Pritchard, J. K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pruett, C. L., and K. Winker. 2010. Alaska song sparrows (*Melospiza melodia*) demonstrate that genetic marker and method of analysis matter in subspecies assessments. Ornithological Monographs 67:162–171.

Puritz, J. B., J. A. Addison and R. J. Toonen. 2012. Next-generation phylogeography: A targeted approach for multilocus sequencing of non-model organisms. PLoS ONE 7: e34231.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rannala, B., T. Zhu and Z. Yang. 2012. Tail paradox, partial identifiability, and influential priors in Bayesian branch length inference. Molecular Biology and Evolution 29:325–335.

Reeves, R. R., W. F. Perrin, B. L. Taylor, C. S. Baker and S. L. Mesnick, eds. 2004. 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. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-363. 94 pp.

Reynolds, J., B. S. Weir and C. C. Cockerham. 1983. Estimation of the coancestry coefficient: Basis for short-term genetic distance. Genetics 105:767–779.

Robineau, D., R. N. P. Goodall, F. Pichler and C. S. Baker. 2007. Description of a new subspecies of Commerson's dolphin, *Cephalorhynchus commersonii* (Lacepede, 1804), inhabiting the coastal waters of the Kerguelen Islands. Mammalia 71:172–180.

Roden, S. E., P. A. Morin, A. Frey, G. H. Balazs, P. Zarate, I.-J. Cheng and P. H. Dutton. 2013. Green turtle population structure in the Pacific: New insights from SNPs and microsatellites. Endangered Species Research 20:227–234.

Rohland, N., and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research 22:939–946.

Ronquist, F., M. Teslenko, P. van der Mark, *et al*. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542.

Rosel, P. E., B. L. Hancock-Hanser, F. I. Archer, *et al*. 2017*a*. Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences. Marine Mammal Science 33 (Special Issue):76–100.

Rosel, P. E., B. L. Taylor, B. L. Hancock-Hanser, *et al*. 2017*b*. A review of molecular genetic markers and analytical approaches that have been used for delimiting marine mammal subspecies and species. Marine Mammal Science 33(Special Issue):56–75.

Seehausen, O., K. Egbert, M. V. Schneider, *et al*. 2002. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. Proceedings of the Royal Society of London Series B: Biological Sciences 270:129–137.

Shaw, K. L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences of the United States of America 99:16122–16127.

Sites, J. W., and J. C. Marshall. 2004. Operational criteria for delimiting species. Annual Review of Ecology, Evolution and Systematics 35:199–227.

Smith, S. A., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86.

Steeman, M. E., M. B. Hebsgaard, R. E. Fordyce, *et al*. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. Systems Biology 58:573–585.

Taylor, B. L., F. I. Archer, K. K. Martien, *et al*. 2017*a*. Guidelines and quantitative standards to improve consistency in cetacean subspecies and species delimitation relying on molecular genetic data. Marine Mammal Science 33(Special Issue):132–155.

Taylor, B. L., W. Perrin, R. R. Reeves, *et al*. 2017*b*. Why we should develop guidelines and quantitative standards for using molecular genetic data to delimit subspecies for data-poor organisms like cetaceans. Marine Mammal Science 33(Special Issue):12–26.

Tobias, J. A., N. Seddon, C. N. Spottiswoode, J. D. Pilgrim, L. D. C. Fishpool and N. J. Collar. 2010. Quantitative criteria for species delimitation. Ibis 152:724–746.

Vallender, R., R. J. Robertson, V. L. Friesen and I. J. Lovette. 2007. Complex hybridization dyamics between golden-winged and blue-winged warblers (*Vermivora chrysoptera* and *Vermivora pinnu*) revealed by AFLP, microsatellite, intron and mtDNA markers. Molecular Ecology 16:2017–2029.

Via, S. 2009. Natural selection in action during speciation. Proceedings of the National Academy of Sciences of the United States of America 106 Supplement 1:9939–9946.

Wakeley, J., and J. Hey. 1998. Testing speciation models with DNA sequence data. Pages 157–175 *in* R. DeSalle and B. Schierwater, eds. Molecular approaches to ecology and evolution. Birkhäuser Verlag, Basel, Switzerland.

Wang, Z., G. E. Hill, A. J. Baker and S. V. Edwards. 2003. Reconciling actual and inferred population histories in the house finch (*Carpodacus mexicanus*) by AFLP analysis. Evolution 57:2852–2864.

Wang, J. Y., T. R. Frasier, S. C. Yang and B. N. White. 2008. Detecting recent speciation events: The case of the finless porpoise (genus *Neophocaena*). Heredity 101:145–155.

Waples, R. S., and O. Gaggiotti. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15:1419–1439.

Warmuth, V., A. Manica, A. Eriksson, G. Barker and M. Bower. 2013. Autosomal genetic diversity in non-breed horses from eastern Eurasia provides insights into historical population movements. Animal Genetics 44:53–61.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Weir, J. T., and D. Schluter. 2007. The latitudial gradient in recent speciation and extinction rates of birds and mammals. Science 315:1574–1576.
Whitehead, H. 1998. Cultural selection and genetic diversity in matrilineal whales. Science 282:1708–1711.
Wolf, J., D. Tautz and F. Trillmich. 2007. Galápagos and Californian sea lions are separate species: Genetic analysis of the genus *Zalophus* and its implications for conservation management. Frontiers in Zoology 4:20.
Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.
Xiong, Y., M. C. Brandley, S. Xu, K. Zhou and G. Yang. 2009. Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. BMC Evolutionary Biology 9:20.
Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences of the United States of America 107:9264–9269.
Zhang, A. B., L. J. He, R. H. Crozier, C. Muster and C.-D. Zhu. 2010. Estimating sample sizes for DNA barcoding. Molecular Phylogenetics and Evolution 54:1035–1039.
Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17:2107–2121.

Appendix 1. The role of mtDNA sequence data in subspecies delimitation.

Mitochondrial DNA (mtDNA) sequence data have played a major role in taxonomic and phylogeographic studies in the past several decades (Zink and Barrowclough 2008). In recent years, there has been considerable backlash against the use of this marker for these types of studies (Ballard and Whitlock 2004, Edwards *et al.* 2005, Edwards and Bensch 2009, Dupuis *et al.* 2012). Much of this criticism stems from the use of mtDNA in studies aimed at elucidating the relationships among species. Reliance on mtDNA in such cases has proven problematic *(e.g.*, Duchêne *et al.* 2011). Here we focus on the utility of mtDNA in taxonomic studies at and below the species level—those aimed at delimiting subspecies and species rather than determining the relationships among species and higher-level taxa.

The prevalence of mtDNA in taxonomic studies results from the fact that it has many characteristics that make it particularly well suited to such studies, including those aimed at delimiting subspecies (Zink and Barrowclough 2008). The four-fold smaller effective population size of the mitochondrial genome compared to the nuclear genome allows lineage sorting to proceed more rapidly in the mitochondrial genome and reciprocal monophyly to be achieved sooner, on average. Consequently, mtDNA is often better suited to resolving recent divergences (but see Karl *et al.* 2012), a characteristic that is particularly important when attempting to delimit subspecies in cetacean taxa, many of which radiated rapidly and recently (Kingston *et al.* 2009, Steeman *et al.* 2009). Mutation rates within the mitochondrial genome are often well-tuned to the problem of phylogeographic inference at and below the species level *(e.g.*, Duchêne *et al.* 2011). Microsatellite loci typically mutate rapidly and therefore exhibit considerable homoplasy,[2] a problem that is exacerbated by a mutation model that results in high levels of back mutation, while nuclear introns and other noncoding and coding sequences typically mutate at much slower rates than

[2]Homoplasy is acquisition of the same trait in unrelated lineages. Individuals can have the same number of repeats (identical by state) but have acquired that state through independent mutational histories (not identical by descent).

mtDNA, resulting in insufficient variation at these loci to be able to resolve recent divergences.

The mitochondrial genome is inherited as a single, nonrecombining locus, making it possible to generate gene trees from which to make phylogenetic inferences. Mitochondrial gene trees have proven extremely powerful in phylogenetic and taxonomic studies. Although gene trees can also be generated for nuclear sequence data, doing so presents considerable technical challenges when dealing with the possibility of recombination and the diploid nature of the loci (but see Puritz *et al.* 2012). Finally, because most mammalian cells contain thousands of copies of the mitochondrial genome compared to only a single copy of the nuclear genome, when dealing with degraded, old, or poor quality samples, mtDNA is much easier to extract and amplify than is nuclear DNA. The ability to obtain data from less-than-ideal samples is critical in studies of cetaceans due to the difficulty in collecting samples from many species. It should be noted that nuclear SNP loci have been shown to work reliably with poor quality samples, including historical samples (Morin and McCarthy 2007), and the advent of Next Generation Sequencing (Meyer and Kircher 2010, Rohland and Reich 2012) may make it possible to obtain nuclear sequence data from poor quality samples as well.

Control region and cytochrome *b* sequence data are likely to continue to predominate in cetacean taxonomic studies because the technology to generate data from them is readily available to most researchers. Furthermore, there are many existing control region and cytochrome *b* data sets that can be added to as samples become available. Although these markers are likely to continue to provide valuable insights into taxonomic questions, they are vulnerable to errors in certain instances. Due to the stochastic nature of the evolutionary process, the gene tree reconstructed from mtDNA may not accurately reflect the species tree (Davis and Nixon 1992). Consequently, accidents of lineage sorting may result in the most recent common ancestor of two lineages substantially predating the actual species (or subspecies) divergence. Conversely, past introgressive hybridization can result in species appearing much more recently derived than they actually are (Seehausen *et al.* 2002, Shaw 2002, Chan and Levin 2005), or can cause general lack of concordance between the mtDNA gene tree and the actual species tree *(e.g.*, Degnan 1993). These limitations likely contribute the difficulty in inferring phylogenetic relationships and delimiting species and subspecies within the subfamily Delphininae (Kingston *et al.* 2009, Amaral *et al.* 2012).

The strict maternal inheritance of mtDNA, while advantageous with respect to effective population size and generating gene trees, means that mtDNA will fail to reflect male-mediated gene flow. Some cetaceans exhibit male-biased dispersal (*e.g.*, Escorza-Trevino *et al.* 2005, Oremus *et al.* 2007, Engelhaupt *et al.* 2009), while others exhibit strong matri-focal social structure *(e.g.*, Barrett-Lennard 2000). These scenarios can result in strong differentiation in mtDNA but sufficient gene flow in the nuclear genome to prevent divergence worthy of subspecific status. When gene flow occurs due to male-biased dispersal (*i.e.*, organismal gene flow), it is possible that gene flow could be detected in a mtDNA data set if migrant males are included in the sample set. However, if gene flow occurs only through mating between groups (*i.e.*, gametic gene flow), it will not leave any signal in the mtDNA gene pool. Thus, when using mtDNA to evaluate taxonomic status researchers must convince readers that their results will not be subject to unwarranted delimitation of subspecies because of an inability to detect male-mediated gametic gene flow. Strongly disjunct distributions (for example tropical species in the Atlantic and Pacific) are not a

concern. Similarly, species with known fission-fusion social structure and high abundance are unlikely cases for such false positives. In contrast, caution should be used in cases with low effective population size and known or suspected male-biased dispersal or maternally inherited site or social group fidelity. In these cases, the structure identified using mtDNA analyses must be corroborated with nuclear loci.

Fortunately, technological advances associated with Next Generation Sequencing (NGS) are likely to result in a substantial increase in the number and types of loci brought to bear in future taxonomic studies (*e.g.*, Emerson *et al.* 2012, Funk *et al.* 2012, Peterson *et al.* 2012, Puritz *et al.* 2012, Hancock-Hanser *et al.* 2013). New analytical methods being developed for use with mitochondrial genome sequences and multi-locus nuclear sequence data from neutral markers hold the potential to resolve even very recent speciation events that cannot be resolved with control region or cytochrome *b* data (Morin *et al.* 2010, McGowen 2011, Amaral *et al.* 2012, Fujita *et al.* 2012, Andrews *et al.* 2013). Another promising avenue of development is the use of sequence data to identify selected regions of the nuclear genome. Genome-wide scans of polymorphisms (see Davey *et al.* 2011 for a review of laboratory methods) can provide thousands of SNPs from which loci under selection may be detected using outlier tests or neutrality tests (Baird *et al.* 2008, Hohenlohe *et al.* 2010). Combined with annotated whole-genome reference sequences, these methods may facilitate the discovery of "speciation genes" (Nosil and Schulter 2011) and gene regions that underlie the diagnosable morphological characters that were the basis of initial subspecies hypotheses (for example within the Delphininae; Perrin 1990, Perrin *et al.* 1994).

Appendix 2. Evaluating divergence: Comparing apples to apples.

A comparative approach to identifying quantitative thresholds for different metrics of divergence is problematic because all of these metrics are affected by effective population size (N_e), mutation rate, and whether the groups are diverging due to neutral drift or different selection pressures. Large populations experience drift much more slowly than small populations. Consequently, it will take new mutations much longer to reach fixation in large populations, slowing the accumulation of fixed differences. Similarly, large populations maintain higher levels of genetic diversity than small populations, again due to their slower rate of drift. Thus, two large populations must be separated much longer in order to develop the same level of net nucleotide divergence (corrected for within-population variability) as two small populations. The slow rate of genetic drift in large populations is easily overcome by even low levels of ongoing or periodic gene flow. As a result, two very small populations can attain a very high level of divergence, as measured either by percent divergence or an analog of Wright's (1931) *F*-statistics, very quickly, even in the face of ongoing gene flow, while two very large populations may exhibit very low divergence estimates despite having been isolated long enough to develop strong morphological differences.

If two groups diverge solely due to random drift, then the dependence of divergence metrics on N_e would not be problematic. After all, it really would take longer for two large populations to diverge by random drift alone than it would for two small populations. However, two geographically disjunct groups will almost always be subject to different selective pressures. Certain types of selection can cause two groups to diverge adaptively much more rapidly than expected by drift alone, but that divergence might not be reflected in the neutral markers typically used in

taxonomic studies. Thus, detecting species-level divergence of large populations that are diverging under strong selection can be very challenging (*e.g.*, Galver 2002, Andrews *et al.* 2013).

Estimates of divergence between two groups will also depend on the mutation rate of the loci used and, in the case of sequence data, the length of sequence obtained. Thus a comparison made using sequences of the full mtDNA control region (~1,000 base pairs in cetaceans; Hoelzel *et al.* 1991) would be expected to reveal more fixed differences than a comparison using only the ~400 base pairs of control region containing hypervariable region 1 (HV1), which exhibits a higher substitution rate than the rest of the control region. Although the difference in sequence length can be compensated for by calculating the percent of base pairs that are fixed (number of fixed differences divided by sequence length; Rosel *et al.* 2017*a*), the high substitution rate and consequent homoplasy in HV1 may result in a lower proportion of fixed sites than in the rest of the control region. The net percent divergence for full control region sequence could be either higher or lower than for HV1 depending on whether the between group diversity (the signal) or within group variability (the noise) increases faster when the slower-mutating portion of the sequence is included. Most analogs of Wright's (1931) *F*-statistics are inversely correlated with mutation rate (Meirmans and Hedrick 2011). Tree-based methods will produce greater resolution and divergence time estimates will be more precise when applied to sequences from loci that have mutation rates optimized to the question being addressed—mutation rates high enough for informative variation to have developed but not so high that there is substantial homoplasy (Duchêne *et al.* 2011).

Received: 19 December 2016
Accepted: 4 January 2017