Why we should develop guidelines and quantitative standards for using genetic data to delimit subspecies for data-poor organisms like cetaceans

This is the first 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 this paper. Together, these papers lead to a proposed set of guidelines that identify informational needs and quantitative standards (Taylor et al. 2017) 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.

BARBARA L. TAYLOR,[1] and WILLIAM F. PERRIN, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A.; RANDALL R. REEVES, Okapi Wildlife Associates, 27 Chandler Lane, Hudson, Quebec J0P 1H0, Canada; PATRICIA E. ROSEL, Southeast Fisheries Science Center, National Marine Fisheries Service, NOAA, 646 Cajundome Boulevard, Lafayette, Louisiana 70506, U.S.A.; JOHN Y. WANG, CetAsia Research Group, 310-7250 Yonge Street, Thornhill, Ontario L4J 7X1, Canada; FRANK CIPRIANO, Genomics/Transcriptomics Analysis Core, San Francisco State University, San Francisco, California 94132, U.S.A.; C. SCOTT BAKER, Marine Mammal Institute and Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Drive, Newport, Oregon 97365, U.S.A.; ROBERT L. BROWNELL, JR., Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A.

ABSTRACT

Obtaining the representative morphological data required for traditional taxonomy is difficult, and sometimes impossible, for cetaceans, especially large ones. As a result, three quarters of the 88 currently recognized extant species have no subspecies and 40 taxa likely have additional unnamed taxa. Conservation needs give urgency to improving taxonomy because unnamed taxa are unlikely to receive protection equivalent to that received by named taxa. Genetic data can improve efforts to delimit subspecies, but the markers and methods used have varied and the magnitude of genetic difference used to justify subspecies distinctions across studies has also varied. Here, we define the concepts of populations, subspecies, and species to establish a foundation for developing *guidelines* (data to include and analyses to conduct) and *quantitative standards* (the magnitude of differentiation expected at different taxonomic levels) for using genetic data to support taxonomic recognition. Our definition is particularly applicable to data-poor groups because it allows for naming a subspecies when there is uncertainty about whether lineages have diverged

[1]Corresponding author (e-mail: barbara.taylor@noaa.gov).

sufficiently for species-level recognition. This allows a species that lacks convincing data for lineage divergence to be recognized as a subspecies while sufficient data are accrued, which could take decades for some cetaceans.

"Whoever wants to hold firm rules, should give up taxonomic work. Nature is too disorderly for such a man."   Stresemann (1936)

"Evolutionary changes across space and time develop like snowflakes: no two are identical. The taxonomic consequences of this fact are diabolical. Because histories vary, and no two populations differentiate from one another in precisely the same way, any attempt to apply a single nomenclatural category to the process cannot help but encompass a diverse array of configurations, stages, and degrees of divergence."   Fitzpatrick (2010)

We agree with these sentiments. Yet, in the six articles that form this special issue on delimiting subspecies using primarily genetic data, we hope to convince readers of the benefit of developing guidelines and quantitative standards to be used in taxonomic arguments. This paper introduces the process of developing guidelines and standards by clarifying the rationale for such development and establishing the definitions for units to conserve, including our proposed definition of subspecies. The next two papers are reviews. Martien *et al.* 2017 review markers and analytical methods appropriate to assess different degrees of differentiation and recommend analytical methods and approaches that can be used to evaluate the key features distinguishing different taxonomic levels—diagnosability and apparent separate lineage evolution at the population/subspecies boundary, and separate lineage evolution at the subspecies/species boundary. The use of mitochondrial DNA (mtDNA) as a key marker appropriate for use in taxonomic arguments for cetaceans is considered in detail. Rosel *et al.* 2017*b* review recent papers that use genetic data to make taxonomic arguments for marine mammals. They find that basic information needed to evaluate taxonomic arguments, such as descriptions of the distributions of taxa and sample locations, were often lacking. Use of nuclear data in these studies was uncommon, but nearly all used mtDNA. Analytical methods varied widely, with diagnosability inferred only in cases with fixed differences. Together, these two reviews provide context and substantive input for developing guidelines for taxonomic arguments. Rosel *et al.* 2017*a* compare commonly used genetic metrics for mtDNA control region sequences across undisputed pairs of populations, subspecies, and species. For several metrics this study finds excellent separation of both populations and subspecies from species, however, subspecies with large effective population sizes are underclassified (classified as populations) using mtDNA alone. This empirical study reveals that underclassification is far more likely than overclassification (classifying accepted subspecies as species or populations as subspecies/species). Given that most studies examined by Rosel *et al.* 2017*b* relied on mtDNA control region sequences and the population/subspecies boundary is often described in terms of diagnosability (Patten and Unitt 2002), we needed a method to estimate diagnosability using sequence data. Here,

we use the definition of diagnosability as proposed by Archer *et al.* (2017): "a measure of the ability to correctly determine the taxon of a specimen of unknown origin based on a set of distinguishing characteristics." They refine the definition of diagnosability for sequence data, which are not described by Gaussian distributions like the morphological data used in past diagnosability methods and definitions. Their paper then demonstrates that the classification method Random Forests (Breiman 2001) can be used to estimate the percent diagnosable using DNA sequence data.

Collectively, these papers lead to a proposed set of guidelines and quantitative standards[2] (Taylor *et al.* 2017) intended to promote consistency, objectivity, and transparency in the classification of cetaceans. The guidelines will correct many errors of omission noted in arguments reviewed by Rosel *et al.* 2017*b* and promote use of the most appropriate analyses as revealed in Martien *et al.* 2017. The standards use the magnitudes of differentiation suggested by Rosel *et al.* 2017*a* and the newly developed diagnosability metric from Archer *et al.* 2017. These standards rely on use of mtDNA control region data, and were developed based on these data because such data are the most commonly available for marine mammals. The standards proposed here are not intended to be prescriptive, but rather to serve as a starting point for making taxonomic arguments. The standards are also intended as a "living" set of tools that can be revised as new types of data (particularly nuclear data) become available.

Here, we provide evidence that there is a problem requiring the solutions given in these papers: cetaceans are "underclassified" at the subspecies level, *i.e.*, a larger number of subspecies exist than have been described and are currently recognized, illustrating the widespread need for improved taxonomy in cetaceans. Taxonomic progress using traditional skull morphology has stalled due to the slow accrual of the necessary data to make the use of molecular genetic data[3] to characterize the subspecific diversity of cetaceans an obvious alternative (Milinkovich *et al.* 2001). Use of genetic data could lead to a long overdue increase in proposals for new taxa. However, inappropriate interpretation of molecular genetic data could also lead to classification errors. The time is ripe to develop guidelines and quantitative standards to facilitate rapid, well-founded subspecies classifications.

We focus on cetaceans because of their limited potential for further morphology-based taxonomic work and the urgency of advancing cetacean taxonomy for the purposes of conservation. The guidelines, and potentially the quantitative standards, should nevertheless provide a useful foundation for all marine mammals and all marine mammals were considered in the review by Rosel *et al.* (2017*b*).

Why Are Subspecies Important, Especially for Cetaceans?

It is important to advance the pace of cetacean taxonomic descriptions because management and conservation policy assume that the current taxonomy is correct and authoritative. Named taxonomic units, particularly when they occur entirely within the waters of single nations, are almost certain to be accorded a higher management priority than groups of individuals that are regarded simply as part of a widely distributed taxon. For example, when a group of killer whales (*Orcinus orca*)

---

[2]The quantitative standards set forth in Taylor *et al.* (2017*a*) are numerical values for the magnitude of a metric that corresponds to the degree of differentiation sufficient to justify a given taxonomic category.

[3]For brevity we use "genetics" and "genetic data" to refer to molecular genetics and molecular genetic data, although we recognize that molecular genetics is a subset of the field of genetics.

called "southern residents" was petitioned for listing under the U.S. Endangered Species Act, this raised the scientific question of whether the taxon to which it belonged was a widely distributed, globally abundant species of killer whale or instead a much less abundant taxon found only in the North Pacific. Taxonomic uncertainty was a key factor in the initial decision not to list this group (U.S. Federal Register 2005) and the later reversal of this decision based on expert opinion, which suggested there were likely at least multiple subspecies within the currently recognized species (Krahn *et al.* 2004).

At the most simplistic level, subspecies are somewhat distinct subdivisions (often geographic) of a species, that are diagnosably distinct from other such subdivisions (http://www.merriam-webster.com/dictionary/subspecies). If a subspecies eventually diverges to the point it becomes recognizable as a separate lineage, it will be described as a distinct species. Accordingly, we recognize that one important role for subspecies designation is to deal with taxonomic uncertainty in a way that improves cetacean taxonomy both to allow for conservation needs in a timely manner and to avoid adding new species prematurely, *i.e.*, before there is convincing evidence that the lineage has diverged. Thus, for some new subspecies further research may indicate that their lineages have diverged and therefore they are actually species. Classifying these cases as subspecies can be critical for taxa like cetaceans that are widely distributed and thus difficult to sample adequately. It may take decades for sufficient evidence to accrue to support a decision that the lineage has diverged and for someone to undertake the comprehensive review needed for a formal species proposal.

Why Are There So Few Cetacean Subspecies?

There are very few recognized cetacean subspecies and, in fact, generally less is known about regional variation in cetaceans than in most other groups of mammals. The reasons for this deficiency can be sorted into several classes: logistical, historical, and biological/ecological.

An important *logistical* problem is that cetaceans occur mainly in the oceans, where they are relatively elusive, inaccessible, and difficult to observe and sample. Collection of cetacean specimens and data is also complicated by the extensive legal protection given to marine mammals in recent decades. Collection and transfer of specimens from strandings and bycatch are further complicated by national laws (*e.g.*, the U.S. Marine Mammal Protection Act) and international agreements (*e.g.*, Convention on International Trade in Endangered Species, CITES). Because so many cetaceans warrant special protection, even gathering data nonlethally in the field is constrained in some countries by the need for permits to approach cetaceans closely for observation, photography, or biopsy collection. Another logistical difficulty pertains to the large size of some cetaceans that makes preparation and storage of skulls difficult.

History is another factor partly responsible for the small number of cetacean subspecies currently recognized. Describing subspecies is not as popular an activity as it was early in the last century, and the few classically trained marine mammal taxonomists have also required more evidence prior to recognizing subspecies than in earlier times. So, while there are many nominal cetacean subspecies (Hershkovitz 1966), a recent review (Rice 1998) resulted in the relatively small number recognized today (Committee on Taxonomy 2016). Application of even more rigorous criteria for morphological data than were used for the current subspecies

classifications (by Rice 1998) would likely reduce the number even further (Perrin *et al.* 2009).

Finally, cetaceans' basic *biology* and *ecology* also impede subspecies description. Marine cetaceans inhabit seemingly continuous habitat, which offers few clues to the existence of regional forms. The breeding season is often in winter when poor weather coupled with remote distributions results in poor information at this critical time of the year for understanding gene flow.

Why Do We Classify?

To many biologists, taxonomy, the science of classification, is assumed to be a rather arcane subject, and few outside the field of evolutionary studies have given much thought to why categorizing a continuous process, like evolution, is important. To many managers, taxonomy is assumed to be known and static and the rationale for further classification efforts is obscure. Therefore, before delving into the subspecies concept, it is useful to consider why the classification process is useful to both biology and management. The search to understand evolution forces biologists to assign categories within an ongoing process that results in observing different lineages at different levels of differentiation from their closest relatives. Given all these issues, taxonomy is fated to be contentious and iterative.

The challenge of communicating effectively about a continuous process is certainly not unique to evolution. For example, descriptions of sea state, like the Beaufort scale, require mariners to make practical decisions and assign conditions to categories along a continuum. Thus, the transition from a light breeze (Beaufort 2) to a gentle breeze (Beaufort 3) is considered to occur at the point when waves just begin to break (*i.e.*, wind speed = 11 km/h). Somewhat arbitrary increments in perception and language such as these enable us to communicate different states, and this can be useful for enhanced understanding and, sometimes, as a basis for actions. The Beaufort scale, for example, eased communication among sailors in the British Navy, giving them a standard protocol for when to put up or take down their sails. Despite the contentiousness of taxonomy, classifying the levels of differentiation among biological groupings or units (such as social units, subspecies, and even phyla) is useful both to biologists trying to understand biological processes and to conservation scientists trying to communicate to managers the changes in human behavior needed to meet specific conservation objectives. There is ample literature attempting to resolve whether subspecies and even species are simply useful categories or instead actual biological entities (Mayr 1982, Hey *et al.* 2003). Like the example of the Beaufort scale, where Beaufort 3 is both a convenient category and really does differ from the sea state at Beaufort 2, we argue that subspecies is both a useful category and a biological entity that differs from both populations and species. However, in most cases subspecies delimitation is primarily related to practical considerations and not theoretical concepts, and consequently we are primarily concerned with the operational criteria used to apply the concept in practice (see also de Queiroz 2011 for a discussion of how explicitly making this distinction has allowed a better understanding of the actual differences between competing species concepts and the development of a unified species concept).

What Is a Subspecies and How Is it Currently Delimited?

The subspecies concept is best explained in relation to the classification units that bound it from below (populations) and from above (species). We refer to the unit below a subspecies as a "population." For our purposes, "population" refers collectively to a series of Units to Conserve ranging from a Demographically Independent Population (DIP) to an Evolutionarily Significant Unit (ESU) (Waples 1995), both of which are defined below. At the lowest level of the hierarchical biological organization is the DIP, defined as "a sympatric group of individuals whose dynamics are more a consequence of births and deaths within the group (internal dynamics) than of immigration or emigration (external dynamics)" (Taylor 2005, Taylor *et al.* 2010). The DIP is an appropriate level of population structure for management objectives related to ecosystem function, like those of the U.S. Marine Mammal Protection Act (Taylor 1997).

The ESU is defined as "a DIP or a collection of DIPs that is substantially reproductively isolated from other conspecific population units and represents an important component in the evolutionary potential of the species" (Waples 1995). Diagnostic differentiation (defined below and expected of subspecies) is not expected of ESUs.

De Queiroz (2007) proposed a unified species concept as the common ground underlying all other taxonomic concepts: "a species is a separately evolving metapopulation lineage." To make our definition consistent with our population and subspecies definition and because several cetacean species are comprised of a single population our definition is: "A *species* is a separately evolving lineage composed of a population or collection of populations." De Queiroz (1998) points out that much of the debate concerning species "concepts" revolves around different opinions on how to delimit species, *i.e.*, how to operationalize the species concept. Properties such as differing phenotypes, reproductive systems, and mate recognition systems, are the evidence used to delimit species.

Our definition of subspecies is: "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[4] distinct." Our definition is consistent with the subspecies concept discussed in Reeves *et al.* (2004), but is more explicit in requiring diagnosability. The subspecies is delimited by a lower population/subspecies boundary and an upper subspecies/species boundary. Our concept of subspecies is captured in Figure 1. The main difference between species and subspecies is that species *are* a separately evolving lineage and subspecies *appear to be* headed in that direction. Note that, in agreement with Patten and Unitt (2002), subspecies cannot be clinal[5] and subspecies, unlike species, do not have to be reproductively isolated from other subspecies. We also emphasize that diagnosability must be based on a heritable character.

---

[4]Diagnosability implies a high probability (but not necessarily a 100% probability) of identifying an individual as belonging to the taxon.

[5]In social species like cetaceans, clinal refers to a series of populations that differ from one another (for example, frequency differences in mtDNA) but where no strong discontinuities are apparent. Many coastal species, like porpoises, are found in a stepping-stone pattern with positive correlation between genetic and geographic distance (*i.e.*, isolation by distance but no clear discontinuities resulting from restrictions to gene flow).

By considering that subspecies *appear to be* separately evolving lineages, we are capturing within the definition several different types of uncertainty encountered by practitioners. When the degree of divergence is small, an apparently diverging lineage may reconverge as conditions change. For example, a barrier to gene flow, such as water of an unsuitable temperature, may be removed, allowing gene flow to resume and divergence to cease. There may be cases where there is only partial divergence with some ongoing gene flow, as seems to be the case for spinner dolphin (*Stenella longirostris*) subspecies in the Eastern Tropical Pacific. Finally, there may be cases where the evidence for divergence is too weak to make a case for a new species and more evidence is needed. An example would be the case of fin whales (*Balaenoptera physalus*) in the North Pacific and North Atlantic. Mitochondrial DNA suggests divergence between these ocean basins (Archer *et al.* 2013). Although allopatric distributions in the North Pacific, North Atlantic, and Southern Hemisphere suggest that these lineages have diverged, the case for lineage divergence would be much stronger with nuclear DNA data to support it. Hey *et al.* (2003) would probably consider this case
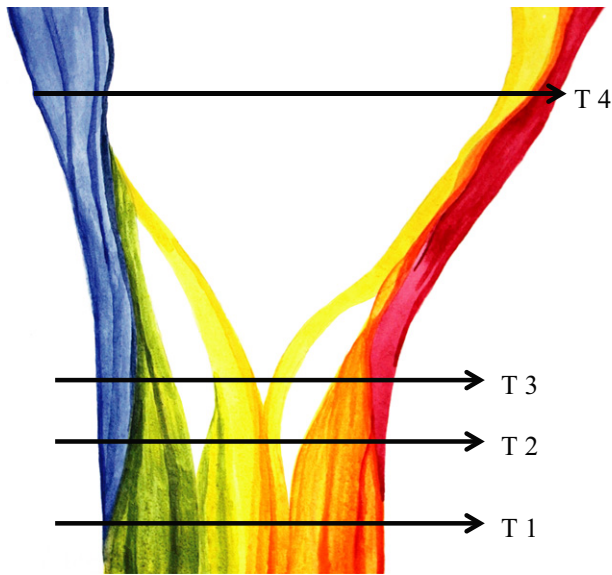

Figure 1. Depiction of the divergence of lineages with four times (T1–T4) chosen to illustrate different levels of biological organization. At T1 the yellow lineage is found across the distribution and although there are likely Demographically Independent Populations (DIPs) that differ in frequencies of the blue, yellow, and red lineages, there are no discontinuities. At T2 some lineages may be diagnosable but likely do not yet appear to be separate lineages. At T3 three groups (the blue/green, yellow, and orange/red lineages) meet the subspecies definition (they are diagnosable and appear to be diverging separately). The divergence level is not sufficient that reconvergence can be ruled out. Between T3 and T4, barriers to gene flow change such that the yellow lineage comes into contact with the blue/green and red-dominated lineages. Blue has diverged in a manner by which gene flow does not resume and the green/yellow lineage dies out. The yellow lineage reconverges and persists alongside the red lineage with a small level of gene flow (orange). At T4 the blue lineage is a species evolving separately from the yellow/red species. The yellow/red species has two subspecies that are both diagnosable and partially diverged.

as an hypothesis for a species where sufficient evidence has not yet been acquired for testing.

Rosel *et al.* (2017*b*) found no consistent pattern to how subspecies were delimited in papers that relied primarily on genetic data. Many papers cite Reeves *et al.* (2004), which states: "In addition to the use of morphology to define subspecies, the subspecies concept should be understood to embrace groups of organisms that appear to have been on independent evolutionary trajectories (with minor continuing gene flow), as demonstrated by morphological evidence or at least one line of appropriate genetic evidence. Geographical or behavioral differences can complement morphological and genetic evidence for establishing subspecies." The Society for Marine Mammalogy formed the Committee on Taxonomy that maintains a list of species and subspecies and considers the merits of new taxonomic proposals. Several subspecies that are delimited on the basis of morphology and on mtDNA control region data have been recently included in the list (see for example *Cephalohynchus hectori maui*; Baker *et al.* 2002).

Why Is the Subspecies Classification Important to Conservation Science?

The taxonomic categories defined above are useful in, first, determining how, and how much, human activities need to be modified to meet conservation objectives and, second, identifying the evolutionary consequences if needed actions are not taken. For example, an unsustainable rate of removal of individuals from a cetacean "group" by a fishery might warrant different actions or timing of actions depending on what that "group" is: if the "group" in question constitutes an entire species, immediate closure of that fishery is warranted, but if the "group" is one of several populations of a species, actions short of immediate closure may suffice, *e.g.*, convening a team to seek ways to reduce human-caused mortality. When considered from the perspective of losses that would result from unregulated or inadequately or inappropriately regulated human activities, the hierarchy of taxonomic categories corresponds to increasing risk of the irreversible loss of biological function. Loss of a DIP could mean it takes decades for the area of local extirpation to be recolonized. Recolonization of an area where an ESU was extirpated may never occur, but (given that entire ecosystems have recovered following ice ages) recolonization is not improbable over periods of hundreds of years. Recolonization of an area following the loss of a subspecies becomes more questionable as considerable local adaptation is expected, and the conditions engendering such adaptation may never be quite the same. Nevertheless, recolonization of the North Island of New Zealand following the loss of the subspecies currently found there (*Cephalorhynchus hectori maui*) by the subspecies from the South Island (*Cephalorhynchus hectori hectori*) is possible because the latter lineage remains extant. In contrast, loss of a species is a permanent loss with unknown consequences to the ecosystem of which it was a part. There is no potential for recolonization of the Yangtze River by baiji (*Lipotes vexillifer*) because the lineage no longer exists anywhere.

A host of laws and treaties assume that taxonomists have their house in good order (for reviews on the importance of subspecies see Haig *et al.* 2006 and Haig and D'Elia 2010, and on taxonomy in general see Mace 2004). Unfortunately, the house is often untidy, and the degree of disorder depends on the taxon involved. Importantly (as discussed in detail for cetaceans below), it is not just the degree of disorder but the direction of disorder that differs among taxonomic groups. Underclassification and overclassification errors confuse both resource managers and the general public, and compromise the credibility of taxonomists and conservation scientists alike.

How Much Underclassification Is Likely for Cetaceans?

For the 88 species of extant cetaceans currently recognized, there is a total of only 57 recognized extant subspecies (Table 1 data from Committee on Taxonomy 2016) (details in Table S1). Around three quarters of cetacean species have no recognized subspecies even though most have extensive distributions and many occur in multiple ocean basins with little or no connecting suitable habitat. To give a rough idea of the magnitude of the problem, we counted the taxa considered to have a high probability of underclassificstion errors. To have a high probability, at least one of the following conditions had to apply: (1) multiple allopatric populations during the breeding season, (2) multiple morphotypes from different localities described in publications or guidebooks, (3) multiple described ecotypes (sympatric or parapatric), (4) marked genetic differences published. Although this suffices to indicate that the number of unrecognized unit errors is likely high, the actual number of undescribed species and subspecies is probably much greater. We made no attempt to quantify the likely number of such errors within a taxon, and in many cases that number could be large. For example, only two subspecies of killer whales are currently recognized in the official Committee on Taxonomy (2016) list. Numerous publications have suggested more subspecies or even species (Pitman and Ensor 2003, Pitman *et al.* 2007, LeDuc *et al.* 2008, Morin *et al.* 2010, Foote *et al.* 2013), and it would not be surprising if at least seven new taxa of killer whales were eventually described and recognized. In Table 1, this case with possibly multiple underclassification errors counts as just one taxon (species) for which the probability of underclassification error is high.

Several important points arise from an examination of these data. First, biodiversity (the number of species and subspecies) in cetaceans is certainly much higher than is currently recognized. Of the 124 currently recognized taxa (species and subspecies combined) 32% have a high chance of an underclassification error. It would not be at all surprising if the number of subspecies is at least double the number currently recognized. Some underclassification errors may have no consequences for conservation. For example, populations of some widely distributed pelagic species, such as the sei

Table 1. Number of recognized extant subspecies and species with probable underclassification errors, sorted by family (see Table S1 for details).

Family	No. of species	No. of subspecies[a]	No. of taxa with a high probability of an underclassification error
Balaenidae	4	0	0
Balaenopteridae	8	18	5
Delphinidae	37	27	18
Eschrichtiidae	1	0	0
Iniidae	1	2	2
Kogiidae	2	0	2
Monodontidae	2	0	0
Neobalaenidae	1	0	0
Phocoenidae	7	8	6
Physeteridae	1	0	0
Platanistidae	1	2	2
Pontoporiidae	1	0	1
Ziphiidae	22	0	4
Total	88	57	40

[a]Including recognized but unnamed subspecies.

whale (*Balaenoptera borealis*), are unlikely to be under imminent threat unless there is a return to commercial whaling. Other underclassification errors, however, may have more obviously serious consequences. For example, all populations of the South Asian river dolphin (*Platanista gangetica*) are at high risk of extinction. A workshop on cetacean taxonomy held in 2004 developed a method to prioritize taxonomic research according to both taxonomic uncertainty and the level of risk (table 1 in Reeves *et al.* 2004) and gave these river dolphins the highest rank. Two subspecies are currently recognized—the Indus dolphin (*Platanista gangetica minor*), which lives in the Indus River system that discharges into the Arabian Sea, and the Ganges dolphin (*P. g. gangetica*), which lives in the Ganges-Megna-Brahmaputra and Karnaphuli River systems that all discharge into the Bay of Bengal. The Indus has been separate from the other river systems for several million years, and the estuaries are separated by several thousand kilometers of marine waters and peninsular India, meaning that movement between the "subspecies" almost certainly has not occurred for a considerable time. There have been only very limited taxonomic studies to date but both genetic (Braulik *et al.* 2014) and skull morphology data (Braulik 2012) raise the possibility that these subspecies are full species. This change could have considerable consequences in terms of elevating their conservation priority and making funding available for more conservation action and research.

Another important point made clear by Table 1 is that the tradition of relying primarily on morphological data to classify cetaceans has likely resulted in a large number of underclassification errors. Perrin *et al.* (2009) reviewed the subspecies listed by Rice (1998) together with those described since 1998 (a total of 49 subspecies in 19 species) and assessed them against the criteria recommended by the Workshop on Shortcomings of Cetacean Taxonomy in Relation to Needs of Conservation and Management (Reeves *et al.* 2004). They concluded that only 38 of the 49 subspecies would qualify as such if one relied solely on morphological data, but that an additional six would qualify if geographic isolation were allowed as a strong line of evidence. Thus, using morphological data alone and a criterion requiring only one strong line of evidence (but disallowing allopatry) would almost certainly result in increasing the number of taxonomic errors.

Why Rely Heavily on Genetic Data for Cetaceans?

In this series of special issue papers we focus on the use of molecular genetic data to delimit subspecies for two reasons: (1) it enables the advancement of taxonomic research where morphological data are unavailable or unattainable for the reasons discussed earlier,[6] and (2) the review of recent papers relying on genetic data (Rosel *et al.* 2017*a*) suggests that guidelines and standards would promote consistency of taxonomic arguments.

Fortunately, at least some of the logistical issues facing cetologists can be obviated by using DNA extracted from tissue samples for genetic analyses. For example, the tissue collection at the laboratory of the Southwest Fisheries Science Center of NOAA covers 82 cetacean species and contains over 41,000 tissue vials[7] and the New

---

[6]Morphological data are very valuable for taxonomic study and should be used whenever possible. Situations in which cetaceans are killed deliberately or die from fishing gear entanglement, ship strike, stranding, or other causes present opportunities to pursue traditional taxonomic studies based on morphology, and such opportunities should be capitalized upon.

[7]Some individuals are represented in more than one vial.

Zealand cetacean tissue archive at the University of Auckland holds samples from 35 species, 11 of which are from the poorly described beaked whales (Thompson *et al.* 2012). The logistics of specimen storage are greatly reduced, as these extensive collections are housed in just a few freezers. DNA can also be extracted from dried museum specimens. Genetic data are publicly available through GenBank and specialized programs for species identification of cetaceans (Ross *et al.* 2003).

Genetic data, when properly analyzed and interpreted, can also provide direct evidence for the evolution of lineages (see Milinkovich *et al.* 2001 for a cetacean-specific review). In theory, both neutral markers and those under selection are available for making inferences about evolutionary processes. Although markers under selection hold great promise for future analyses, we focus our attention here on markers that are generally agreed to be neutral (see Martien *et al.* 2017 for details) and that allow estimates of both diagnosability and divergence, which are key to our subspecies concept.

In spite of the encouraging progress, the use of genetic data in classification of subspecies remains challenging. Rosel *et al.* (2017*b*) reviewed recent taxonomic publications to document the markers, analytical methods, and rationales used to construct arguments concerning classification of marine mammals. They found that a range of markers and analytical methods have been used, and in this case, such variability is not necessarily a good thing. In the absence of an operational definition for subspecies, guidelines, or quantitative standards, the use of such varied metrics and analyses has often meant that it is very difficult to evaluate the strength of the taxonomic arguments. Many of the publications reviewed by Rosel *et al.* (2017*b*) do not provide sufficient detail for the reader to adjudicate the case with a high degree of confidence. For example, few contain descriptions of locations where samples were collected even though assessing representativeness of the sampling design is a fundamental aspect of judging the strength of a taxonomic argument. This type of shortcoming could be remedied with a simple list of important data and analyses to include in genetics papers that propose taxonomic changes.

Why Should We Develop Guidelines and Quantitative Standards?

Our suggested guidelines include reference to the need to provide information on the distributions of the taxa or taxon and on sample locations, descriptions of life history, and comments on choice of genetic markers and analytical methods (Taylor *et al.* 2017). Such guidelines for best practices are an easy means to improve consistency in the field of taxonomy. This is especially true as taxonomists continue to decline in number. With this decline, changes in taxonomy are now being proposed increasingly by scientists who do not specialize in taxonomy and are unfamiliar and/ or inexperienced with how to construct compelling supportive arguments. Further, much of the taxonomic work most in need of progress is in less-developed countries where there are very few taxonomists. "Living" guidelines developed in journals like this one and improved upon by professional societies or in future publications should help to broaden participation in the field without loss of quality.

Guidelines are also useful for reviewers and readers of taxonomic literature who, for example, can be reminded that sufficiently detailed distribution descriptions are essential to making taxonomic judgments. Many reviewers of taxonomic proposals based on genetic data may be unaware that subspecies distinctions, which in the past have primarily been based on morphology, require a high level of diagnosability, and that therefore if a manuscript does not include such analyses, the authors need to explain and justify this omission.

"Quantitative standards" refers to the magnitude of differentiation that warrants subspecies classification. Standards differ somewhat from criteria or decision rules, however, in that the user is encouraged to make an argument to support his/her case even if it does not meet the standards and he/she still feels the proposed taxon deserves subspecies classification. Therefore, although our quantitative standards are similar to the rules used by bird taxonomists (Tobias *et al.* 2010) and the 75% rule for subspecies used by some others (Patten and Unitt 2002, Amadon 1949), we do not intend for them to be applied rigidly. Rather, we would hope that practitioners use the standards to hone their arguments and clarify the inevitable strengths and weaknesses of any given proposal. We reiterate that although the focus of this special issue is on the use of genetic data, we recognize that other lines of evidence can be used in taxonomic arguments and genetic analyses do not take precedence over lines like morphology when such evidence is available.

These standards may be difficult for many taxonomists to accept because they set specific numerical values for the desired magnitude of differentiation. However, the analysis by Rosel *et al.* 2017*a* basically follows the philosophy of using the comparative method for allopatric species. Allopatric populations have often proven difficult for taxonomists, particularly taxonomists who adhere strictly to the Biological Species Concept, because there is no potential to test the criterion of reproductive isolation. As a proxy, some taxonomists advocate placing the degree of differentiation between allopatric populations in a comparative framework, namely comparing the degree of morphological or genetic differentiation in the allopatric forms to that seen for closely related sympatric or parapatric species (*e.g.,* Remsen 2010). In this way a reasonable and testable hypothesis can be stated concerning whether the allopatric forms have or have not differentiated to the degree shown by related forms that do or do not freely interbreed (Miller 1955; Mayr 1969, 1996; Mayr and Ashlock 1991; Helbig *et al.* 2002; Futuyma 2005). In essence, this comparative method forms the basis of recently proposed quantitative criteria for species delimitation in birds (Tobias *et al.* 2010) and Taylor *et al.* (2017) follow this philosophy using genetic metrics developed by Rosel *et al.* (2017*a*).

The continuous nature of divergence and the unique pathways of evolutionary change noted in the quotations at the beginning of this paper would seem to argue against the use of quantitative standards to categorize biological units. We agree that taxonomic units are, by their nature, somewhat arbitrary constructs, particularly in borderline cases where different lines of evidence suggest quite different levels of differentiation (Hey 2001, de Queiroz 2007). Indeed, Hey and Pinho (2012) found considerable overlap in several genetic metrics when comparing species to "populations." We note, however, that their usage of the term "populations" included everything below the species level, which differs from our use of that term. Rosel *et al.* 2017*a* made empirical comparisons and found considerable overlap in some metrics but little overlap in others. In many cases, levels of genetic differentiation did not fall in the overlap zone, and hence the units being analyzed fell clearly into the population, subspecies, or species category. For genetic data, we anticipate that some of the markers and metrics used today will be supplanted by more powerful markers and metrics in the near future. We therefore advocate that the standards set forth by Taylor *et al.* (2017) be used to facilitate improvement of the quality and transparency of arguments advanced on behalf of taxonomic proposals and that they be viewed as "living standards" that can evolve with experience and as knowledge grows.

LITERATURE CITED

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

Archer, F. I., P. A. Morin, B. L. Hancock-Hanser, *et al.* 2013. Mitogenomic phylogenetics of fin whales (*Balaenoptera physalus* spp.): Genetic evidence for revision of subspecies. PLoS ONE 8:e63396.

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

Baker, A. N., A. N. H. Smith and F. B. Pichler. 2002. Geographical variation in Hector's dolphin: Recognition of a new subspecies of Cephalorhynchus hectori. Journal of the Royal Society of New Zealand 32:713–727.

Braulik, G. T. 2012. Conservation ecoloty and phylogenetics of the Indus River dolphin (*Platanista gangetica minor*). Ph.D. thesis, University of St. Andrews, St. Andrews, U.K. 267 pp.

Braulik, G. T., R. Barnett, V. Odon, V. Islas-Villanueva, A. R. Hoelzel and J. A. Graves. 2014. One species or two? Vicariance, lineage divergence and low mtDNA diversity in geographically isolated populations of South Asian river dolphin. Journal of Mammal Evolution 22:111–120.

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

Committee on Taxonomy. 2016. List of marine mammal species and subspecies. Society for Marine Mammalogy. Available at http://www.marinemammalscience.org.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. Pages 57–75 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: Species and speciation. Oxford: Oxford University Press, Oxford, U.K.

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

de Queiroz, K. 2011. Branches in the lines of descent: Charles Darwin and the evolution of the species concept. Biological Journal of the Linnean Society 103:19–35.

Fitzpatrick, J. W. 2010. Subspecies are for convenience. Ornithological Monographs 67:54–61.

Foote, A. D., P. A. Morin, R. L. Pitman, *et al*. 2013. Mitogenomic insights into a recently described and rarely observed killer whale morphotype. Polar Biology 36:1519–1523.

Futuyma, D. J. 2005. Evolution. Sinauer Associates, Sunderland, MA.

Haig, S. M., and J. D'Elia. 2010. Avian subspecies and the U.S. endangered species act. Ornithological Monographs 67:170–173.

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

Helbig, A. J., A. G. Know, D. T. Parkin, G. Sangster and M. Collinson. 2002. Guidelines for assigning species rank. Ibis 144:518–525.

Hershkovitz, P. 1966. Catalog of living whales. Bulletin of the United States National Museum 246. i–vii + 259 pp.

Hey, J. 2001. Genes, categories and species. Oxford University Press, Oxford, U.K.

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

Hey, J., R. S. Waples, M. L. Arnold, R. K. Butlin and R. G. Harrison. 2003. Understanding and confronting species uncertainty in biology and conservation. Trends in Ecology & Evolution. 18:597–603.

Krahn, M. M., M. J. Ford, W. F. Perrin, *et al*. 2004. 2004 status review of southern resident killer whales (*Orcinus orca*) under the Endangered Species Act. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-62. 73 pp.

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.

Mace, G. M. 2004. The role of taxonomy in species conservation. Philosophical Transactions of the Royal Society of London, Series B 359:711–719.

Martien, K. K., M. S. Leslie, B. L. Taylor, *et al*. 2017. Analytical approaches to subspecies delimitation with genetic data. Marine Mammal Science 33(Special Issue):27–55.

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

Mayr, E. 1982. Of what use are subspecies? The Auk 99:593–595.

Mayr, E. 1996. What is a species and what is not? Philosophy of Science 63:262–277.

Mayr, E., and P. D. Ashlock. 1991. Principles of systematic zoology. 2nd edition. McGraw-Hill, New York, NY.

Milinkovich, M. C., R. LeDuc, R. Tiedemann and A. Dizon. 2001. 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.

Miller, A. H. 1955. Concepts and problems of avian systematics in relation to the evolutionary processes. Pages 1–22 *in* A. Wolfson, ed. Recent studies in avian biology. University of Illinois Press, Urbana, IL.

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

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

Perrin, W. F., J. G. Mead and R. L. Brownell, Jr. 2009. Review of evidence used in the description of currently recognized cetacean subspecies. U.S. Department of Commerce, NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-450. 38 pp.

Pitman, R. L., and P. Ensor. 2003. Three forms of killer whales (*Orcinus orca*) in Antarctic waters. Journal of Cetacean Research and Management 5:131–139.

Pitman, R. L., W. L. Perryman, D. LeRoi and E. Eilers. 2007. A dwarf form of killer whale in the Antarctic. Journal of Mammalogy 88:43–48.

Reeves, R. R., W. F. Perrin, B. L. Taylor, C. S. Baker and S. L. Mesnick. 2004. Report of the workshop on the 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 NOAA-TM-NMFS-SWFSC-363. 94 pp.

Remsen, J. V., Jr. 2010. Subspecies as a meaningful taxonomic rank in avian calssification. Ornithological Monographs 67:62–78.

Rice, D. W. 1998. Marine mammals of the world. Systematics and distribution. Society for Marine Mammalogy Special Publication Number 4. 231 pp.

Rosel, P. E., B. L. Hancock-Hanser, F. I. Archer, *et al*. 2017*a*. Examining metrics and magnitudes of 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 genetic markers and analytical approaches that have been used for delimiting marine mammal subspecies and species. Marine Mammal Science 33(Special Issue):56–75.

Ross, H. A., G. M. Lento, M. L. Dalebout, *et al.* 2003. DNA surveillance: Web-based molecular identification of whales, dolphins and porpoises. Journal of Heredity 94:111–114.

Stresemann, E. 1936. The Formenkreis-theory. Auk 53:150–158.

Taylor, B. L. 1997. Defining "population" to meet management objectives for marine mammals. Pages 49–65 *in* A. E. Dizon, S. J. Chivers and W. F. Perrin, eds. Molecular genetics of marine mammals. Marine Mammal Science Special Publication Number 3.

Taylor, B. L. 2005. Identifying Units to Conserve. Pages 149–164 *in* Marine mammal research: Conservation beyond crisis. J. E. Reynolds III, W. F. Perrin, R. R. Reeves, S. Montgomery and T. J. Ragen, eds. The John Hopkins University Press, Baltimore, MD.

Taylor, B. L., K. Martien and P. Morin. 2010. Identifying units to conserve using genetic data. Pages 306–344 *in* Marine mammal ecology and conservation—A handbook of techniques. I. L. Boyd, W. D. Bowen and S. J. Iverson, eds. Oxford University Press, Oxford, U.K.

Taylor, B. L., F. I. Archer, K. K. Martien, *et al.* 2017. 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.

Thompson, K. F., C. D. Millar, C. S. Baker, M. Dalebout, D. Steel, A. L. van-Helden and R. Constantine. 2012. A novel conservation approach provides insights into the management of rare cetaceans. Biological Conservation 157:331–340.

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.

U.S. Federal Register. 2005. Endangered and threatened wildlife and plants: Endangered status for Southern Resident Killer Whales. FR 70(222):69903–69912 (18 November 2005) National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Department of Commerce, Washington, DC.

Waples, R. S. 1995. Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. American Fisheries Society Symposium 17:8–27.

SUPPORTING INFORMATION

The following supporting information is available for this article online at http://onlinelibrary.wiley.com/doi/10.1111/mms.12413/suppinfo.

Table S1. Assessment of high likelihood for additional taxa for current listing of cetaceans. Codes for the conditions considered to merit high likelihood are: (1) likely disjunct breeding distributions within taxa, (2) multiple described morphological forms, (3) marked different genetics, (4) multiple described ecotypes, (5) subspecies that could be full species. Meeting any one of these five conditions results in a "Yes" under "High likelihood of additional species or subspecies." Note that in some cases with apparent disjunct breeding distributions, the condition was not met because there were genetic data suggesting nontrivial levels of gene flow.