A review of molecular genetic markers and analytical approaches that have been used for delimiting marine mammal subspecies and species

This is the third of six papers forming a special issue of Marine Mammal Science (Vol. 33, Special Issue) on delimiting cetacean subspecies using 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.


1Corresponding author (e-mail: patricia.rosel@noaa.gov).
Abstract

Uncertainty in marine mammal taxonomy is increasingly being addressed using molecular genetic data. We examined 32 peer-reviewed articles published between 1994 and 2011 to review methodological practices, consistency of markers and analytical methods, and overall quality of arguments used when genetic data have been employed to delimit new species and subspecies of marine mammals. The mitochondrial DNA (mtDNA) control region was the primary genetic marker used in these studies, but analytical methods varied greatly across studies. Diagnosability, a common metric for delimiting subspecies with morphological data, was only used through citing of fixed differences in mtDNA sequences. Assignment tests based on microsatellite data were less common but were applied at both taxonomic levels. Nuclear DNA sequence data were rarely used. Basic background material needed to evaluate the strength of arguments, such as distribution and sampling maps, were often missing. For most studies, sample sizes were good, but adequate geographic sampling for broadly distributed taxa was often lacking, diminishing the strength of evidence for taxonomic distinctness. Examining these empirical cases revealed a mixture of sound and inadequate practices for genetic studies of cetacean taxonomy and suggested that improvements could be made to the field by developing standard guidelines.

Key words: subspecies definition, subspecies delimitation, cetacean taxonomy, mitochondrial DNA, control region, genetic data.

A central tenet of biology is that the species is the cohesive unit of taxonomy. Consistent delimitation of species (alpha taxonomy) is a necessary component of many biological inferences, for it allows appropriate comparisons of data sets. Accurate taxonomy and identification of species is critical to conservation and management of taxa and biodiversity and is often a consideration under legislation governing protection and recovery of species (Mace 2004, Reeves et al. 2004, Green 2005, Haig et al. 2006, Zhou et al. 2016, Taylor et al. 2017b). On the other hand, the need for and taxonomic worth of subspecies has been a matter of considerable debate (e.g., Wilson and Brown 1953, Starrett 1958, Zink 2004, Fitzpatrick 2010, Remsen 2010, Winker 2010). Subspecies classically represent geographically subdivided variants within a species (Mayr 1969) and lie along a continuum of variation somewhere between populations and recognized species. This variation likely represents adaptation, be it ecological, morphological, genetic, or other, to a local environment (Winker 2010) and as such is an integral component of understanding and conserving intraspecific biodiversity (Winston 1999, Haig et al. 2006, Taylor et al. 2017b).

Recognizing subspecies acknowledges such adaptive capacity by identifying pools of unique evolutionary potential and biodiversity. However, the number of subspecies identified within a taxonomic group is highly variable. In mammals, 12 of the 26 orders reported in the 2004 IUCN Red List had named subspecies (Gippoliti and Amori 2007). The number of recognized species and subspecies of cetaceans has varied widely over the past century. As of April 2016, there were 90 recognized cetacean species, 22 of which have named subspecies (Committee on Taxonomy 2015). Based on known disjunct populations and inferences from related cetacean species that do have recognized subspecies, it is likely that a large number of cetacean subspecies (and perhaps even some species) have been overlooked (Taylor et al. 2017b).

In recent years, genetic data have been increasingly used to delimit new taxa, and the number of studies that rely primarily on genetic data to make taxonomic
arguments for cetaceans is particularly high because morphology-based taxonomic
descriptions for this group are difficult. Broad and remote distributions of most taxa
make sampling challenging and expensive, and the generally large body size of the
animals severely limits osteological collections for morphological analysis (Taylor
et al. 2017b). The advent of molecular genetic techniques has allowed an increase in
sampling across many taxa resulting in a recent increase in delimitation of species
and subspecies in cetaceans (e.g., Dalebout et al. 2002, 2014; Beasley et al. 2005; Jef-
ferson and Wang 2011; Mendez et al. 2013). Though still challenging, it is often
easier to obtain tissue samples from a larger number of animals in the field than it is
to collect skeletal material or body dimension data for morphological comparison.
The advent of the polymerase chain reaction (PCR) greatly increased the types of tis-
sues that can provide viable DNA for study, e.g., bones and teeth from museum col-
lections (see Dalebout et al. 2002).

Given that studies proposing new marine mammal taxa are rapidly emerging, the
timing is appropriate to assess whether the taxonomic arguments being made are
robust and whether the genetic markers and analytical methods are being used consist-
tently. Furthermore, despite the increase in molecular genetic analyses applied to
questions of marine mammal taxonomy, both the number of independent lines of evi-
dence necessary as well as the magnitude of differentiation required to delimit sub-
species and species have been applied differently in published marine mammal
studies (Caballero et al. 2007, Charlton-Robb et al. 2011), and this has, in some
cases, impacted conservation efforts (Reeves et al. 2004). To address the issue of the
number of lines of evidence necessary to delimit a subspecies, Reeves et al. (2004)
concluded that for marine mammals “...the subspecies concept should be understood
to embrace groups of organisms that appear to have been on independent evolution-
ary trajectories (with minor continuing gene flow), as demonstrated by morphological
evidence or at least one line of appropriate genetic evidence.” To what extent is this
recommendation being followed? To gain a better understanding of the methodolo-
gies in use, we reviewed publications that utilized molecular genetic data to address
questions in marine mammal taxonomy (primarily cetaceans, but also pinnipeds and
manatees). The goal of this review was to examine the taxonomic arguments relying
primarily on genetic data and to evaluate whether (1) similar markers were utilized
across studies, (2) analytical methods were consistent across studies, and (3) sufficient
background context was provided to allow readers to evaluate the quality of the argu-
ment. The results from this compilation were used to identify areas where molecular
taxonomy might be improved for marine mammals (see Taylor et al. 2017a for rec-
ommendations and guidelines).

For the literature search we used the keywords cetacean, subspecies, pinnipeds, genetics,
molecular, and sirenian. The time frame was limited to 1994–2011. We chose to start
in 1994 because of the advent of PCR and improved DNA sequencing technologies
in the 1990s. The literature search returned 70 publications, including some of the
authors’ previously published works that used molecular genetic data to examine
some aspect of genetic differentiation among groups of marine mammals (Appendix S1). Of those, 38 focused on population level questions or had a phyloge-
netic focus above the species level and were removed from consideration. In the end,
we evaluated 32 peer-reviewed publications (Table 1).

Upon completion of our review, we identified four topics deemed critical to a con-
vincing argument for subspecies or species delimitation and used these topics to eval-
uate the publications: (1) clear articulation of the taxonomic question being
investigated and some background taxonomic context for the research, (2) adequacy
Table 1. Summary of the 35 taxonomic comparisons that were evaluated and the taxa and data type(s) used in each. For full evaluation of each publication’s categories see Table S1. Question code refers to the taxonomic level each paper addressed: 1 = subspecies/species boundary; 2 = boundary uncertain; 3 = population/subspecies boundary. mtDNA = mitochondrial DNA; nuDNA = nuclear DNA; msat = microsatellites.

<table>
<thead>
<tr>
<th>Species involved</th>
<th>Reference</th>
<th>mtDNA sequence data</th>
<th>Fixed differences in mtDNA</th>
<th>Monophyly at mtDNA</th>
<th>Shared haplotypes</th>
<th>Estimate of percent sequence divergence</th>
<th>Fixed differences in nuDNA</th>
<th>Monophyly at nuDNA</th>
<th>Shared haplotypes</th>
<th>Microsatellites used</th>
<th>Private must alleles</th>
<th>Assignment test performed</th>
<th>Morphology used in species/subspecies argument?</th>
<th>Tree</th>
<th>Divergence</th>
<th>Diagnosability</th>
<th>Gene flow</th>
<th>Is proposed species designation accepted?</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Eubalaena australis</em>, <em>E. glacialis</em>, <em>E. japonica</em></td>
<td>Gaines et al. 2005</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>U</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td><em>Eubalaena australis</em>, <em>E. glacialis</em>, <em>E. japonica</em></td>
<td>Rosenbaum et al. 2000</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Balaenoptera acutorostrata</em></td>
<td>Paetene et al. 2007</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>Balaenoptera omurai</em></td>
<td>Sasaki et al. 2006</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td><em>Kogia sima</em>, <em>K. breviceps</em></td>
<td>Chivers et al. 2005</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Mesoplodon peruvianus</em></td>
<td>Dalebout et al. 2002</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td><em>Sousa spp.</em></td>
<td>Frese et al. 2008</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td><em>Sotalia flavirostris</em>, <em>Sotalia guianensis</em></td>
<td>Caballero et al. 2007</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Tursiops aduncus</em></td>
<td>Wang et al. 1999</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Tursiops aduncus</em></td>
<td>Natoli et al. 2004</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Tursiops spp.</em></td>
<td>Miller et al. 2008</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Tursiops australis</em></td>
<td>Charleton-Robb et al. 2011</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Delphinus sajensis</em></td>
<td>Rosel et al. 1994</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Orcasella heinsohni</em></td>
<td>Beasley et al. 2005</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td><em>Orcinus urvius</em></td>
<td>Morris et al. 2010</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td><em>Neophocaena australiensis</em>, <em>Neophocaena phocaenoides</em></td>
<td>Wang et al. 2008</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>—</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>—</td>
<td>—</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Zalophus californianus</em>, <em>Zalophus wollebaeki</em></td>
<td>Wolf et al. 2007</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><em>Globiphora melas</em>, <em>G. macrorhynchus</em></td>
<td>Oremus et al. 2009</td>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td><em>Mesoplodon sp. in South Pacific (unnamed)</em></td>
<td>Dalebout et al. 2007</td>
<td>2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Species involved</th>
<th>Reference</th>
<th>Methods used</th>
<th>Species involved</th>
<th>Reference</th>
<th>Methods used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delphinus spp.</td>
<td>Natoli et al. 2006</td>
<td>mtDNA sequence data?</td>
<td>Orcinus orca</td>
<td>LeDuc et al. 2008</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2009</td>
<td>mtDNA sequence data?</td>
<td>Balaenoptera musculus intermedia, B. m. brevicauda</td>
<td>LeDuc et al. 2007</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Orcinus orca</td>
<td>Garcia-Rodríguez et al. 1998</td>
<td>mtDNA sequence data?</td>
<td>Inia boliviensis, I. geoffrensis</td>
<td>Banguera-Hinestroza et al. 2002</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2007</td>
<td>mtDNA sequence data?</td>
<td>Cephalorhynchus c. commersoni, C. c. kerguelenensis</td>
<td>Pichler 2002</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2007</td>
<td>mtDNA sequence data?</td>
<td>Cephalorhynchus h. hectori, C. h. maui</td>
<td>Robineau et al. 2007</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2007</td>
<td>mtDNA sequence data?</td>
<td>Tursiops truncatus ponticus</td>
<td>Viaud-Martinez et al. 2008</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2007</td>
<td>mtDNA sequence data?</td>
<td>Globicephala melas melas, G. m. edwardsii</td>
<td>Oremus et al. 2009</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2007</td>
<td>mtDNA sequence data?</td>
<td>Phocoena phocoena relicta</td>
<td>Viaud-Martinez et al. 2007</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2007</td>
<td>mtDNA sequence data?</td>
<td>Arctocephalus australis</td>
<td>Rosel et al. 1995</td>
<td>mtDNA sequence data?</td>
</tr>
<tr>
<td>Delphinus spp.</td>
<td>Amaral et al. 2007</td>
<td>mtDNA sequence data?</td>
<td>Kogia sima</td>
<td>Chivers et al. 2005</td>
<td>mtDNA sequence data?</td>
</tr>
</tbody>
</table>

For the sake of documenting how often monophyly is presented in cetacean taxonomic studies, we only considered nodes with bootstrap values greater than 70% (Hillis and Bull 1995) or Bayesian posterior probabilities greater than 0.95 (Huelsenbeck and Ronquist 2004) as demonstrating monophyly.
of sampling, (3) choice of genetic marker, and (4) analytical methods and strength of evidence. This paper summarizes how the reviewed papers addressed each of these topics and provides rationale for the aspects of particular topics we identified as important for taxonomic arguments. We did not evaluate the studies against specific criteria for delimiting subspecies or species, but rather documented the analyses and results the authors used to support their conclusions. We organized the lessons learned from our review into the above four topics, and the results and discussion of each are presented in sequence below.

Articulation of the Taxonomic Question

Deciding whether a group of organisms deserves subspecies or species status is a classification problem, and evaluating whether the data collected and analyses performed are sufficient to address the classification problem first requires clear articulation of the taxonomic question under consideration. We therefore sought within each publication a description of the taxonomic question or hypothesis being addressed. Many of the studies we examined did not clearly articulate the species or subspecies concept against which they were testing their taxa. In fact, many studies did not clearly specify whether they were examining a subspecies-level or a species-level question. This omission sometimes interfered with the reader’s ability to evaluate the other topics, i.e., sample adequacy, marker choice, analytical rigor, and the final outcome of the paper. Because we ultimately wanted to be able to compare papers at each taxonomic level (subspecies or species) to summarize findings regarding the current state of practice, we categorized each paper we reviewed into one of three taxonomic-question types: (1) concerned with subspecies delimitation, (2) uncertain whether unit is subspecies or species, and (3) concerned with species delimitation. Each study was categorized based on either direct expression by the authors or our inference of the taxonomic level the paper considered, based on the accepted taxonomy at the time of our review of the literature, in cases where the authors did not specifically declare it. Several papers addressed both boundaries and were therefore evaluated independently for each boundary. In cases where it was not clear whether a subspecies- or species-level question was being addressed, we used the subspecies definition from Reeves et al. (2004) and Taylor et al. (2017b) to categorize the study: “A species is a separately evolving lineage comprised of a population or collection of populations; 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.” Finally, the definition of population encompasses 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) through to Evolutionarily Significant Units (ESUs) as defined in Waples (1995).

Across the 32 publications, we reviewed a total of 35 taxonomic comparisons, 12 of which examined the population/subspecies boundary, 18 that examined the subspecies/species boundary, and 5 that could not be assigned to either of these two categories (Table 1). One paper focused on manatees, two on pinnipeds, and the rest on cetaceans. In most cases the molecular genetic studies were designed to address

2Diagnosability implies a high probability (but not necessarily a 100% probability) of identifying an individual as belonging to the taxon.
questions of taxonomic standing originating from morphological studies, and all drew conclusions congruent with conclusions drawn from the morphological studies. There was at least one instance where genetic data sparked the search for and identification of morphological differentiation between two groups (eventually identified as two species; Dalebout et al. 2002). However, a successful argument for subspecies designation using genetic data as the sole line of reasoning has yet to be made, despite the fact that the 2004 Taxonomy Workshop (Reeves et al. 2004) indicated that a single line of such evidence would be sufficient at this lower boundary. In addition to providing a clear description of the taxonomic question under investigation, a few additional pieces of information vastly improved the general taxonomic context for reviewing a publication. This information included a description of the current state of taxonomy, what previous studies (morphological, genetic, or other) have been undertaken, and, importantly, what new information or insight the present study contributes.

Sampling Considerations

Sample size—The sample size necessary to robustly characterize structure among the groups examined is a critical consideration for any study. The number of genetic samples needed to test a taxonomic hypothesis depends on the effective population size ($N_e$) of each taxon, the genetic diversity and inheritance mode of the genetic markers being used, the taxonomic level under investigation, and the degree of divergence among taxa being examined. Many of these issues are discussed in depth elsewhere (Taylor and Dizon 1996, Ryman et al. 2006, Martien et al. 2017, Taylor et al. 2017). Since subspecies are expected to have diverged relatively recently and may continue to experience low levels of genetic exchange, we generally expect lower levels of detectable genetic differentiation between them, thereby requiring larger sample sizes to provide sufficient power to detect differences. Furthermore, as discussed below, appropriate geographic sampling for studies at the subspecies level is also important. Studies of putative species, which should exhibit considerably higher levels of genetic divergence (but see Wang et al. (2008) for counterexample) should generally require fewer sampled individuals per taxon. For phylogenetic studies that examine relationships among species or higher taxonomic levels, the required numbers of individuals per taxon may be lower but thorough taxon sampling, i.e., sampling of all the potential species in the group, is important for accurate phylogenetic reconstructions (Heath et al. 2008).

In reviewing the marine mammal literature, we found that sample size varied widely across studies. Approximately half the studies had total sample sizes of 50–200 individuals for the focal taxon (Fig. 1a). Appropriately, more studies at the species level had lower sample sizes than those at the population/subspecies boundary, although several had sample sizes greater than 200. However, total sample size can be somewhat misleading, and it is important to examine the number of samples per locality, especially for studies at the population/subspecies boundary, where frequency-based analyses may be employed. After re-examining sample sizes for each study and recording, where possible, the sampling locality with the largest sample size, the most common sample size class dropped to the 20 to 50 category (Fig. 1b). We also found wide variation in the clarity of presentation of sample size. For some studies, it was difficult to determine the total sample size for the focal group, and in others, while a total sample size was clearly stated, how those samples were partitioned across sampling locations was unclear. In addition, although few papers
Figure 1. Sample sizes used in publications of molecular genetic studies of marine mammals at different taxonomic levels. Graphs present the proportion of studies at each taxonomic level that fall into each sample size category. (A) minimum total sample size per focal taxon; (B) maximum sample size per single sampling locality. Papers were categorized as examining taxonomic questions at: species = subspecies/species boundary; subspecies = population/subspecies boundary; uncertain = taxonomic boundary uncertain (see text).

provided it, information on life history characteristics of the species further helped the reader interpret the adequacy of the sampling for the hypotheses being addressed. For example, knowing whether the focal taxa are resident or migratory, exhibit strong
matrilineal social structure, have a distribution that fluctuates interannually, or have habitat constraints that contribute to allopatry could all influence a reader's assessment of the adequacy of sampling. Interestingly, very few papers presented information on how many biopsies were collected per pod/school or details sufficient to infer whether biopsy sampling of multiple individuals from a pod/school of animals could lead to bias.

**Geographic coverage of sampling**—Accurate subspecies and species delimitation relies on obtaining a sample set that captures the genetic variation within the putative taxa in an unbiased manner. Sampling the full range of a species is a difficult goal to reach for most marine mammals, but a lack of adequate sampling can lead to a variety of biases (Martien *et al.* 2017). If samples come only from the extremes of the range, the genetic or morphological variation seen could be clinal rather than discrete, and a lack of intermediate sampling diminishes the strength of the evidence. For example, Frère *et al.* (2008) sampled *Sousa* populations at the ends of their range in the Indian Ocean and suggested two species were present. A lack of samples between Africa and Australia, where the species is/are continuously distributed in coastal waters, means that clinal variation could not be rejected, nor was it possible, as the authors point out, to determine the location of the break between the two putative species in the Indian Ocean. Further sampling in the intermediate regions by Mendez *et al.* (2013) provided additional support for the initial hypothesis of two species in the Indian Ocean put forth by Frère *et al.* (2008) and also helped identify the location of the geographic break between the two *Sousa* taxa.

From a taxonomic standpoint, inadequately sampling the range could result in the application of an incorrect species name to a group. For example, long-beaked and short-beaked common dolphins in the eastern North Pacific (ENP) exhibit significant genetic (Rosel *et al.* 1994) and morphological (Heyning and Perrin 1994) differentiation consistent with a species-level difference. The Latin name, *Delphinus capensis,* first attributed to long-beaked animals from southern Africa, was applied to the long-beaked animals from the ENP (Heyning and Perrin 1994). Subsequent to these studies, however, genetic analysis revealed that long-beaked common dolphins from the ENP and southern Africa are also quite differentiated (Natoli *et al.* 2006), suggesting that *D. capensis* was probably not the correct name to apply to long-beaked animals in the ENP (now recognized as *Delphinus delphis bairdii*), and resulting in complications in taxonomic nomenclature.

Conversely, if samples are concentrated in areas where gene flow between two putative subspecies is highest (*e.g.*, in areas of parapatry), then the degree of overlap in potentially diagnostic characters could be overestimated, resulting in failure to confer subspecies status when it is warranted (Martien *et al.* 2017, Remsen 2010). Finally, consideration should be given to the historical range of taxa, particularly for those taxa that have been heavily exploited by human activities. For example, recent fragmentation of geographic distributions may have led to relic populations at geographic extremes that appear to be highly divergent (see also Taylor *et al.* 2017a). Designing field sampling to address the taxonomic question appropriately can help minimize these types of errors, but in many cases, opportunistic sampling is a common practice for many marine mammal studies. Recognizing and describing the potential problems with sampling, however, provides a reader with a better means to evaluate the taxonomic argument being presented. Suggestions for future sampling to solve outstanding questions, when appropriate, would also be helpful.

We examined the geographic coverage of sampling in the papers we reviewed by estimating the proportion of the taxon’s geographic range covered by the samples and
by categorizing the sampling efforts into poor (25% or less of the range sampled), fair
(25%–75% of the range sampled), and good (more than 75% of the range sampled),
recognizing that for some species the full range is currently unknown. Fifty percent of
the studies examining the population/subspecies boundary were categorized as having
good coverage, but only 22% of the subspecies/species boundary studies could be so
classified. Most studies fell in the “fair coverage” category, admittedly also our broad-
est category. We further categorized papers into two species-range types: small (for
taxa that inhabit rivers, seas, or have contiguous coastal distributions) and large (for
taxa that inhabit a large range or multiple ocean basins). We found that the propor-
tions of poor, fair, and good sampling coverage were quite similar between taxa with
large geographic ranges and those with small ranges, and were not significantly
different ($\chi^2 = 0.985$, 2 degrees of freedom, $P > 0.5$). This result is somewhat sur-
prising since one might expect small-range species to have had better sampling cov-
erage and it highlights that even marine mammal species with “small” distributional
ranges may still cover thousands of linear kilometers and be difficult to sample
adequately.

Evaluating sampling adequacy was much easier when maps of both the approxi-
mate range of the species and the exact sampling locations and sample sizes were pro-
vided. Only 23% of the reviewed studies provided information on both sampling
location and the range of the taxon under study. Wang et al. (2008) provided an
excellent map of the range of the taxa under consideration and highlighted the area of
sympathy, but did not illustrate the exact locations of the samples used in the study,
instead providing a detailed verbal description of the sampling localities. Frère et al.
(2008) combined both information on the range and the distribution of the taxa and
general sampling locations into a single map that allowed the reader to identify
potential gaps in sampling. A map or detailed description of both the distribution
of the species and the samples helps the reader place the extent of the sampling into the
context of the species range and helps the reader determine whether sampling covered
important areas, for example to address the possibility of clinal variation in a linearly
dispersed taxon, or areas of sympatry between potential species. Unfortunately, for
many marine mammal taxa, distributional ranges are poorly defined, and in these
cases, it is equally important to include uncertainties in the range, particularly if they
occur in an area that affects the argument being presented in the study.

**Marker Choice**

Mitochondrial DNA sequences were by far the most commonly used marker in the
studies we reviewed (Table 1, Fig. 2). Ninety-seven percent of the 32 studies used
mtDNA, and all but one of these used mtDNA control region sequence data. Studies
focused on the upstream (5′) portion of the control region, except for 1 that used the
entire control region; 56% used control region sequence lengths ≤400 bp, while
88% used sequences ≤500 bp in length. Twelve studies also presented cytochrome b
data, while one paper used only cytochrome b sequences. The one study that did not
use mtDNA used only microsatellite data. The prominence of mtDNA sequence data
in the papers we reviewed is consistent with the pattern seen in taxonomic and phylo-
geographic studies in general, where mtDNA has played a major role for the past sev-
eral decades (Simon et al. 2006, Zink and Barrowclough 2008). The popularity of
mtDNA sequence data reflects the fact that this marker exhibits many attributes that
make it particularly well suited to phylogeographic studies (Avise et al. 1987, Avise
1992, Rubinoff and Holland 2005, Zink and Barrowclough 2008, Martien et al.
However, the use of mtDNA data in phylogeographic and taxonomic studies has engendered considerable argument in the past decade (Ballard and Whitlock 2004, Edwards et al. 2005, Rubinoff and Holland 2005, Zink and Barrowclough 2008, Barrowclough and Zink 2009, Edwards and Bensch 2009, Dupuis et al. 2012). MtDNA is a single locus and as such may not accurately reflect the true evolutionary history of the study group. Introgressive hybridization can obscure relationships among taxa at deeper divergences (e.g., McCracken and Sorenson 2005). A final concern inherent in the use of a single-locus, maternally inherited marker is the possibility of falsely concluding that a unit is a subspecies when instead it is a social unit with female philopatry and male-mediated gene flow, demographic characteristics that are not uncommon among cetacean taxa. A thorough description of life history traits and a sampling design accounting for those life history traits helps a reader evaluate whether such a false positive error was possible and whether mtDNA may reliably be interpreted in the absence of nuclear DNA (nuDNA) evidence. Despite these potential weaknesses in mtDNA, Rosel et al. (2017), in an examination of cetacean mtDNA control region sequence data from pairs of closely-related cetacean populations, subspecies, and species, found that although a false designation (e.g., confusing social structure for taxonomic structure is possible), the opposite is more likely (i.e., not finding subspecies-level differences for actual subspecies pairs). Nevertheless, inclusion of nuclear data is a critical consideration, particularly at the subspecies-species boundary, where strong evidence for separately evolving lineages is key and nuclear data address the level of both female and male-mediated gene flow.

Of the studies we examined, fewer than 30% included nuDNA sequences when examining the subspecies/species boundary, while none of the studies examining the population/subspecies boundary included nuDNA sequences (Table 1, Fig. 2).
Utility of nuDNA sequences has been limited historically in part because of the overall low level of genetic variability in the nuclear genome of cetaceans (Schlotterer et al. 1991, Martin and Palumbi 1993, Bininda-Emonds 2007), even at higher taxonomic levels. However, next-generation sequencing (NGS) technologies should improve the ability to survey larger amounts of nuclear sequence data that can then be used to directly estimate the proportion of individuals diagnosable to their unit of origin. NGS data may also improve evaluation of the degree of male-mediated gene flow for cases where high diagnosability in mtDNA result from strong matrilineal social structure.

Microsatellite data were used equally in the marine mammal studies we reviewed at both the population/subspecies level and the subspecies/species level (Table 1, Fig. 2). They played an important role in many of the papers focused on the population/subspecies boundary. For instance, de Oliveira et al. (2008) used microsatellites as their primary genetic data set to evaluate the taxonomic status of South American fur seals (Arctocephalus australis). They demonstrated high diagnosability based on microsatellite genotypes and a high proportion of fixed and/or private microsatellite alleles, providing strong genetic evidence that the two units represent separately evolving lineages. The utility of microsatellite data at the subspecies/species boundary is more limited. Microsatellite loci are susceptible to homoplasy, i.e., shared alleles not identical by descent (Estoup et al. 2002). Therefore, observed similarities may not accurately represent the evolutionary distinctiveness of two groups (e.g., Martien et al. 2017) and, similarly, observed differences are difficult to interpret as population, subspecies, or species–level differences. Recognizing the difficulties inherent in interpreting microsatellite data for taxonomic decisions, when differences are found they may provide evidence of evolutionary independence and certainly identify cases worthy of further investigation using additional lines of evidence. For example, Wang et al. (2008) found a very high proportion of fixed or private alleles at microsatellite loci between two subspecies of finless porpoises. They used this result together with strong morphological differences to make a compelling argument for elevating the subspecies to full species status. An added advantage of microsatellite over mtDNA data is that the multiple microsatellite loci are independent and essentially provide multiple lines of evidence for divergence rather than just one.

Analytical Methods and Strength of Evidence

A variety of analytical methods and metrics can be used to evaluate taxonomic hypotheses. Four major lines of evidence were commonly used in the studies we reviewed: (1) number of shared haplotypes between focal groups, (2) whether the group of interest was monophyletic, (3) the presence of fixed differences (i.e., nucleotide positions in the DNA sequence at which all of the sequences in one sample are different from all of the sequences in a second sample) between focal groups, and (4) percent divergence calculations. Martien et al. (2017) provide a review of analytical methods and their strengths and weakness in genetic taxonomic studies, particularly for marine mammals, while Rosel et al. (2017) conduct comparative analyses aimed at evaluating the utility of a variety of divergence metrics for delimiting subspecies. Here, we focused on documenting the prevalence of different types of methods in published studies and the variability in their application.

A lack of shared haplotypes (but with no fixed differences between groups) was commonly used at the subspecies/species boundary (89% of studies) and even at the population/subspecies boundary (67% of studies). However, it should be noted that
the usefulness of this metric varies across the two taxonomic levels. At the subspecies/species boundary, a lack of shared haplotypes alone (with no fixed differences between them) does not necessarily translate into a strong signal of divergence along unique evolutionary pathways. Brower (1999) provided several examples of data sets in which haplotypes are not shared, but analysis of the data in an evolutionary framework illustrates the limited phylogenetic/phylogeographic signal these haplotypes contain. Conversely, the presence of shared haplotypes does not necessarily refute species status, as demonstrated for cetaceans in the recognition of two *Neophocaena* species by Jefferson and Wang (2011), which show diagnostic morphological traits and strong differentiation using microsatellites yet share one mtDNA control region haplotype. Although the presence or absence of shared haplotypes is not conclusive at the subspecies/species boundary, at the population/subspecies boundary a lack of shared haplotypes may be sufficient for diagnosability, particularly for taxa with a large *N*, assuming that sampling has been adequate and appropriately distributed across the geographic range, *i.e.*, that the two putative units are both well characterized for the markers involved (see Archer et al. 2017, Martien et al. 2017, Taylor et al. 2017a).

Although monophyly can suggest subspecies/species level divergence, its use in taxonomic studies has been strongly criticized (see review in Funk and Omland 2003). Incomplete lineage sorting and introgressive hybridization can result in a lack of reciprocal monophyly among true species (*e.g.*, Bernatchez et al. 1995), while matrifocal social structure and rapid drift in recent isolates can produce monophyly in groups that do not warrant specific or even subspecific status. Despite these limitations, evidence of monophyly was cited in nearly all studies at the subspecies/species boundary, as well as in ~33% at the population/subspecies boundary (Table S1). However, we found that use of the term monophyletic varied across studies. In discussing evidence for monophyly of a particular group, many papers used standard lower limits for bootstrap values or posterior probabilities in Bayesian analyses that were derived from simulation studies (*i.e.*, Hillis et al. 1993, Huelsenbeck and Rannala 2004). A few studies concluded groups were monophyletic in the absence of any tests for nodal support, while others qualified bootstrap values below the standard lower limits as providing “weak” or “moderate” support for a given clade.

Evidence of fixed differences between putative taxa was presented in 64% of subspecies studies and 89% of species studies (Table 1). Assuming sample sizes were adequate to ensure that putative fixed differences were genuine, such differences could be used to address the diagnosability criterion, the primary metric by which subspecies are delimited using morphological data (Amadon 1949, Patten and Unitt 2002). Only six of the papers we reviewed (two at the subspecies level and four at the species level) included explicit estimates of diagnosability. This result likely reflects the fact that most of the studies we reviewed relied primarily or entirely on mtDNA sequence data, yet few methods to directly estimate diagnosability from sequence data exist (but see Austerlitz et al. 2009, Archer et al. 2017). An assignment test based on multilocus microsatellite genotypes is a common genetic method for estimating diagnosability (Martien et al. 2017), and is the approach that was used in the studies we reviewed that did include diagnosability estimates. However, assignment tests are typically not useful when applied to a single marker like mtDNA. Even if two putative taxa do not share any haplotypes, any haplotypes represented by a single individual cannot be meaningfully assigned because when they are treated as being of unknown origin for assignment purposes, their frequency of occurrence in either population cannot be estimated. Thus, without using a sequence classification method...
like those reviewed in Austerlitz et al. (2009) and Archer et al. (2017), overall diagnosability cannot be accurately estimated.

Finally, estimates of genetic divergence between putative taxa were presented in 45% and 67% of subspecies and species studies, respectively. Interpreting these divergence estimates requires the use of “threshold” values against which potential taxonomic units can be evaluated. Establishing such thresholds is challenging. “DNA barcoding” analysis (Hebert et al. 2003), initially proposed as a means to identify species through use of cytochrome oxidase I sequences, is based on the establishment of thresholds (Valentini et al. 2009), but the method has not been universally successful for cetaceans (Amaral et al. 2007, Viricel and Rosel 2012). In the absence of established thresholds for divergence estimates, researchers often compare the divergence estimates from their study to a yardstick of divergence values estimated for “accepted” species or subspecies pairs (e.g., Beasley et al. 2005, Caballero et al. 2007). For instance, Caballero et al. (2007) provided a useful comparison of percent divergence values and number of fixed differences across appropriately consistent gene regions and taxa in their analysis of the genus Sotalia. However, in many studies such comparisons were either lacking entirely or were uninformative due to variation across studies in the methods used to estimate and interpret levels of divergence with respect to taxonomic questions. For example, some publications corrected for within-group variation while others did not. Some estimates of percent divergence accounted for recognized differences in mutation rates and mutation patterns that may impact the accrual of nucleotide substitutions, while others did not. Although none of this variation in methodology creates “wrong” divergence estimates, it does make comparisons across studies difficult or even inappropriate, and may have contributed to the broad range of published sequence divergence values seen for both the population/subspecies and subspecies/species boundaries (Fig. 3).

We also note that 83% of the studies we reviewed at the subspecies level and 28% at the species level applied $F$-statistics (Wright 1943) (or analogues) to microsatellite and sequence data. While these are appropriate population-level analytical methods, they are of limited utility in taxonomic studies (Martien et al. 2017, Rosel et al. 2017, Hey and Pinho 2012). We therefore did not judge them to be very informative in most of the studies we reviewed. If $F_{ST}$ values are high among putative subspecies, this may provide evidence that gene flow is limited, but care must be taken when interpreting these high values, as it has been shown that, under certain circumstances, they may be anomalously high for reasons other than taxonomic distinctiveness (see Rosel et al. 2017). The exceptions were a few studies aimed at the lower boundary, where some gene flow can be expected and the groups may represent populations. In such cases, $F_{ST}$ estimates may be helpful for interpreting the degree of interconnectedness of two groups. For instance, when based on nuclear data, they can be used to evaluate low levels of male-mediated gene flow and thereby support stronger results based on mtDNA data.

Combining Evidence: The Convincing Argument

There has been increasing support in recent years for a more integrative approach to taxonomy wherein multiple “disciplines” or lines of evidence are examined, evaluated, and integrated, preferably in an evolutionary framework, before proposing a new taxonomic classification (Padial et al. 2010). Not surprisingly, in our review of the literature, papers that cited multiple lines of evidence
were generally seen as providing stronger arguments for taxonomic distinctness. An example is the description of the new subspecies of Commerson’s dolphin (*Cephalorhynchus commersonii kerguelenensis*) by Robineau *et al.* (2007). The evidence cited that distinguished Commerson’s dolphins from South America and the Kerguelen Islands included morphological divergence, genetic divergence based on mtDNA control region sequences, and private microsatellite alleles, and behavioral divergence in the form of acoustic differences (Dziedzic and Buffrénil 1989) to complement the geographic distance between forms (a 8,500 km range disjunction). These authors also specifically addressed the “75% rule” for diagnosability (Amadon 1949, Patten and Unitt 2002) and showed that all individuals from the two regions are easily assignable to location of origin using total

---

**Figure 3.** Published values of percent divergence between cetacean subspecies (black bars), species (white bars), and taxa of uncertain taxonomic status (gray bars). Values are based on mtDNA control region sequence data. Not all values represent net sequence divergence. See Table 1 for list of papers corresponding to each value. Since completing this work, *Sousa* species have been supported ((Mendez *et al.* 2013) and *Inia* subspecies changed.)
length and all skull measurements (Robineau et al. 2007). However, the phylo-
genetic analysis of mtDNA control region sequences revealed that C. c. kerguele-
nensis haplotypes nested within C. c. commersonii haplotypes, indicating that the
time since separation has not been sufficient to establish reciprocal monophyly
between the two taxa. The authors’ conclusion was to designate two subspecies.
The nested nature of the phylogenetic analysis and the inability to know
whether dolphins from these two regions would interbreed if they did come into
contact makes designation at the subspecies level a conservative approach. Fur-
ther analyses with an increased sample size and additional nuclear data could
potentially support elevation to species status.

More recently, Jefferson and Wang (2011) presented a comprehensive analysis of
finless porpoise (Neophocaena spp.) populations using morphological and molecular
genetic data. The paper provides a clearly articulated hypothesis and basis for evaluat-
ing the question of whether the two Neophocaena morphotypes should be elevated to
species. The authors found concordance across all lines of evidence except mtDNA
sequence data, where the two forms shared one of seven haplotypes. Thus, this study
provides an example of where reliance on mtDNA data alone could lead to failing to
recognize two species as such. The authors suggested that the presence of a shared
haplotype results from very recent divergence (estimated at approximately 18,000 yr
ago) allowing insufficient time for ancestral haplotypes to differentiate. The authors
concluded that two subspecies should be considered full species (N. phocaenoides and
N. asiaeorientalis). Unlike the Commerson’s dolphin, the two taxa of finless porpoises
are partially sympatric, strengthening the argument for species-level status (Jefferson
and Wang 2011).

Conclusions

Our review of the published literature illustrates that good cases for taxo-
nomic revision of marine mammal taxa have been made relying on genetic data.
Most studies used the mtDNA control region as the primary marker, but meth-
ods to analyze the data varied widely. As a result, the type of information pro-
vided on how much genetic difference constituted species-level or subspecies-
level differentiation was inconsistent across studies. Thus, compilation of a com-
parative data set that could be comprehensively used in future arguments for
species or subspecies delimitation could not be made (but see Rosel et al. 2017).
Importantly, many papers suggested that the genetic data were consistent with
new taxa but could not resolve whether the observed differences suggested sub-
species or species. This suggests the field of cetacean taxonomy might be
improved with a set of guidelines to provide needed tenets for solid taxonomic
arguments. Such guidelines could address these weakness and inconsistencies,
thereby improving robustness of and consistency across future taxonomic studies
in cetaceans and other marine mammals (see Taylor et al. 2017a).

Acknowledgments

This work is the outcome of discussions in a Journal Club with participants from multiple
institutions held via conference call. The content of the manuscript was also influenced by
comments and discussion provided by attendees of a workshop on defining subspecies held in
October, 2009 at the 18th Biennial Conference on the Biology of Marine Mammals, and we
thank all involved for the valuable dialogue. We also thank Randy Reeves for serving as Rapporteur of the workshop discussions. The manuscript was improved by comments from three anonymous reviewers. Finally, we thank Daryl Boness, Editor of Marine Mammal Science, for his willingness to bring this set of six papers to light, and Michael Schwartz, who served as Associate Editor and provided valuable comments and input on each manuscript.

**Literature Cited**


Received: 20 March 2015
Accepted: 3 January 2017

**Supporting Information**

The following supporting information is available for this article online at http://onlinelibrary.wiley.com/doi/10.1111/mms.12412/suppinfo.

**Appendix S1.** List of publications using molecular genetic data to examine questions of marine mammal taxonomy along the continuum from populations to species that were reviewed for this study; * indicates those used in final evaluation.

**Table S1.** Results of evaluation of published papers that used molecular genetic data to examine questions of marine mammal taxonomy along the continuum from populations to species with a focus on subspecies. Question code refers to the taxonomic level each paper addressed: 1 = subspecies/species boundary; 2 = boundary uncertain; 3 = population/subspecies boundary. mtDNA = mitochondrial DNA; nuDNA = nuclear DNA; msat = microsatellites. Key to column headers follows below in Table S2.