Journal of Heredity, 2023, **114**, 612–624
https://doi.org/10.1093/jhered/esad049
Advance access publication 30 August 2023
Original Article

American Genetic Association

Original Article

Guidelines and quantitative standards for improved cetacean taxonomy using full mitochondrial genomes

Phillip A. Morin[1,*], Karen K. Martien[1], Aimee R. Lang[1], Brittany L. Hancock-Hanser[1], Victoria L. Pease[1], Kelly M. Robertson[1], Maya Sattler[1], Elizabeth Slikas[2], Patricia E. Rosel[3], C. Scott Baker[4], Barbara L. Taylor[1], Frederick I. Archer[1]

[1]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA, United States,
[2]School of Life Sciences, Center for Evolution and Medicine, Arizona State University, Tempe, AZ, United States,
[3]Marine Mammal and Turtle Division, Southeast Fisheries Science Center, National Marine Fisheries Service, NOAA, Lafayette, LA, United States,
[4]Marine Mammal Institute, Oregon State University, Newport, OR, United States

[*]Corresponding author: Southwest Fisheries Science Center, 8901 La Jolla Shores Dr., La Jolla, CA 92037, United States. Email: phillip.morin@noaa.gov

Corresponding Editor: Alfred Roca

Abstract

In many organisms, especially those of conservation concern, traditional lines of evidence for taxonomic delineation, such as morphological data, are often difficult to obtain. In these cases, genetic data are often the only source of information available for taxonomic studies. In particular, population surveys of mitochondrial genomes offer increased resolution and precision in support of taxonomic decisions relative to conventional use of the control region or other gene fragments of the mitochondrial genome. To improve quantitative guidelines for taxonomic decisions in cetaceans, we build on a previous effort targeting the control region and evaluate, for whole mitogenome sequences, a suite of divergence and diagnosability estimates for pairs of recognized cetacean populations, subspecies, and species. From this overview, we recommend new guidelines based on complete mitogenomes, combined with other types of evidence for isolation and divergence, which will improve resolution for taxonomic decisions, especially in the face of small sample sizes or low levels of genetic diversity. We further use simulated data to assist interpretations of divergence in the context of varying forms of historical demography, culture, and ecology.

Key words: divergence, fastsimcoal, mitogenome, simulated populations, subspecies delineation

Introduction

Taxonomic classification is the process of creating discrete categories within a continuous evolutionary process. As such, it is a difficult and often contentious process, with few fixed rules or constants across the breadth of biodiversity. It is, nevertheless, a critical component of understanding patterns of biodiversity and how they arise, and is essential for effective conservation management. The delimitation of populations, subspecies, and species is the basis for understanding when groups have reached certain levels of divergence or independence, and in the context of conservation, what level of protection might be warranted under certain frameworks (e.g. the IUCN Red List). For marine mammals, definitions by Taylor et al. (2017a) provided a useful framework for developing quantitative guidelines to evaluate taxonomic arguments for classification. Populations are understood to be demographically independent, such that within a population, demographic 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, 2017b). Subspecies, the lowest taxonomic level commonly recognized in vertebrates, are typically recognized to include some level of allopatry (Mayr 1969). They have recently been defined as "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" (Reeves et al. 2004; Taylor et al. 2017b). Whereas species, the next taxonomic level up and the basic unit of biology, are defined as "a separately evolving lineage comprised of a population or collection of populations" (Taylor et al. 2017b, following De Queiroz 2007).

Cetaceans constitute a somewhat special problem for taxonomists. Their large bodies, oceanic habitat, relatively recent radiation, and often ocean basin-wide or even global distributions make it difficult to apply traditional morphology-based methods, which historically have relied on collections of skulls. Taylor et al. (2017b) argued that 32% of currently recognized cetacean species and subspecies are under-classified, and that there could be twice as many subspecies. Molecular methods have played an increasing role in taxonomic descriptions within the Cetacea, sometimes providing the only available data for describing new

Received January 31, 2023; Accepted August 28, 2023

Published by Oxford University Press on behalf of The American Genetic Association 2023. This work is written by (a) US Government employee(s) and is in the public domain in the US.

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

species (Dalebout et al. 2002, 2007; Morin et al. 2017; Carroll et al. 2021). The inconsistent use of molecular markers and analytical methods (reviewed in Martien et al. 2017; Rosel et al. 2017b), however, has resulted in confusion about what level of molecular (and other) evidence is appropriate for taxonomic classification in cetaceans. A Special Issue of Marine Mammal Science (vol. 33, 2017) presented the reasons molecular methods are needed to advance cetacean taxonomy (Taylor et al. 2017b), a review of the genetic markers and analytical methods that have been used (Martien et al. 2017; Rosel et al. 2017b), and evaluations of how various divergence (Rosel et al. 2017a) and diagnosability (Archer et al. 2017b) metrics correlate with accepted taxonomic status. The Special Issue concluded with proposed guidelines and quantitative standards intended to "promote consistency, objectivity, and transparency in the classification of cetaceans" (Taylor et al. 2017a), including a decision flow diagram and checklist of considerations for presenting a persuasive argument for delimiting new subspecies or species, including the basis for taxonomic hypothesis, sampling considerations (sample size and distribution), review of relevant life history characteristics, data types and analytical methods, and synthesis of lines of evidence.

The reviews and guidelines in the Special Issue were necessarily limited to datasets that were widely available and comparable, specifically short sequences (~250 to 1000 basepairs) of the mitochondrial DNA (mtDNA) control region (Ross et al. 2003; Archer et al. 2017b; Martien et al. 2017; Rosel et al. 2017a,b). However, the level of variability in the control region can be too low for taxonomic resolution for some taxa due to population bottlenecks and social structure (e.g. Morin et al. 2010; Foote et al. 2016; Van Cise et al. 2019), and variation may predate the divergence of the taxa (incomplete lineage sorting) (Archer et al. 2017b).

A well-known principle of population genetics is that divergence due to genetic drift is correlated with effective population size (Wright 1931) such that species with high abundance, such as spotted and spinner dolphins, are expected to take so long to diverge at the subspecies level that genetic markers will likely not diagnose well-accepted subspecies (Rosel et al. 2017a; Leslie et al. 2019). In these taxa, selection on nuclear variation may act to drive divergence in the absence of control region differentiation (Andrews et al. 2021), or the rate of diversification may outpace lineage sorting in these highly variable mtDNA regions. Since the publication of the Special Issue of Marine Mammal Science in 2017, genomic datasets for marine mammals have become increasingly available, and complete mitochondrial genomes (mitogenomes) have been used for population, subspecies, and species-level studies (e.g. Morin et al. 2018; Archer et al. 2019; Leslie et al. 2019; Van Cise et al. 2019; Louis et al. 2020; Skovrind et al. 2021). In many cases, these longer mitogenome sequences (~16,400 bp) improved resolution of divergence among taxonomic groups (e.g. Morin et al. 2010, 2015, 2018; Archer et al. 2013; Van Cise et al. 2019; Albertson et al. 2022), but the impact of improved resolution on the use of genetics for taxonomic decisions has not yet been investigated.

Taylor et al. (2017a) proposed thresholds for two metrics, net nucleotide divergence (d_A) and percent diagnosability (PD), that can be used as guidelines for classification of cetacean strata as populations, subspecies, or species based on mtDNA control region sequences. To develop comparable thresholds for full mitogenome sequences, we have compiled published and unpublished cetacean mitogenomic sequence alignments and used them to evaluate the correlation of divergence metrics with taxonomic status. We compare the values of metrics calculated from full mitogenomes to those from hypervariable control region sequences to determine how the two sequence lengths are likely to perform in taxonomic assessments, including cases in which classification based on the guidelines would over- or under-classify recognized subspecies or species (i.e. genetic data place a taxon at a lower or higher level than currently recognized). We also conducted simulations of mitogenome and control region sequence divergence over continuous ranges of divergence time, effective population size (N_e), mutation rates, and migration rates, evaluating when divergence and diagnosability guidelines may be likely to under- or over-classify taxonomic groups.

Materials and methods

This study uses mitogenome sequence data from 18 cetacean species that have complete mitogenome sequence datasets previously published or generated for this study (Supplementary Table S1). Datasets were stratified into 30 different pairs of populations and subspecies within species, and sister- or closely related species based on published information associated with the data for accepted cetacean taxa (Committee on Taxonomy 2019). Populations within species were recognized based on independent data indicating genetic or geographic isolation (e.g. island-associated groups or population genetic analyses). Subspecies or species that have been suggested in published literature, but which remain uncertain due to lack of data or appropriate analyses, have been designated "unresolved" and analyzed only to suggest further taxonomic analyses that should be pursued (Supplementary Table S2). Only strata (populations, subspecies, or species) with at least 10 sequences available were used in the analyses described below. Capture of geographic variability within subspecies or species was limited by available data and/or samples, and datasets were selected to maximize diversity of pairs of subspecies and species regardless of geographic representation within each stratum.

Samples and sequencing

We assembled mitogenomes from one genus, *Neophocaena*, from the genomic short-read archive (SRA) data after standard quality trimming with BBtools (sourceforge.net/projects/bbmap/), using the BWA mem algorithm (v. 0.7.15, Li and Durbin 2009) to align reads to a published mitogenome for *Neophocaena phocaenoides* (accession KC777291). We generated new data for four species (*Pepenocephala electra*, *Pseudorca crassidens*, *Phocoena phocoena*, and *Balaenoptera musculus*; Table 1) using previously described methods for capture enrichment of mtDNA prior to Illumina short-read sequencing (Hancock-Hanser et al. 2013). DNA extraction and sequencing library preparation for these species were conducted at SWFSC from samples stored in the Marine Mammal and Sea Turtle Research (MMaSTR) Collection. Consensus mitochondrial sequences were obtained from the newly generated data using reference-guided assembly based on the custom pipeline described by Hancock-Hanser et al. (2013). Previously unpublished mitogenome sequences of right whales (*Eubalaena* sp.) were generated by conventional (Sanger) and next-generation sequencing at Oregon

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

Table 1. Mitochondrial genome sequence datasets used for pairwise divergence analysis.

Strata type	Genus	Species	Strata pairs (sample sizes)	Source
Population	*Globicephala*	*macrorhynchus*	Mariana Islands (20) v. Hawaii (32)	Van Cise et al. (2019)
Population	*Orcinus*	*orca*	EAL-TRI (16) v. GOA (17)[a]	Morin et al. (2010, 2015)
Population	*Orcinus*	*orca*	CAL-WAL (23) v. EAL-TRI-GOA (33)[a]	Morin et al. (2010, 2015)
Population	*Orcinus*	*orca*	CAL (15) v. GOA (17)[a]	Morin et al. (2010, 2015)
Population	*Orcinus*	*orca*	CAL (15) v. EAL-TRI (16)[a]	Morin et al. (2010, 2015)
Population	*Peponocephala*	*electra*	Main Hawaiian Is. (10) v. Palmyra (11)	This study
Population	*Peponocephala*	*electra*	Bahamas (10) v. Main Hawaiian Is. (10)	This study
Population	*Peponocephala*	*electra*	Atlantic (10) v. Pacific (21)	This study
Population	*Peponocephala*	*electra*	Bahamas (10) v. Palmyra (11)	This study
Population	*Phocoena*	*phocoena*	N. California-BC (11) v. S. California (10)	Ben Chehida et al. (2020)
Population	*Physeter*	*macrocephalus*	Atlantic (35) v. Pacific (124)	Morin et al. (2018)
Population	*Pseudorca*	*crassidens*	E. Pacific (29) v. Hawaii (27)	This study
Population	*Stenella*	*attenuata*	Northern (14) v. WesternSouthern (17)	Leslie et al. (2019)
Species	*Balaenoptera*		B. musculus (101) v. B. physalus (123)	This study, Archer et al. (2013, 2019)
Species	*Eubalaena*		E. australis (15) vs. E. japonica (27)	This study
Species	*Eubalaena*		E. australis (15) vs. E. glacialis (11)	This study
Species	*Eubalaena*		E. glacialis (11) vs. E. japonica (27)	This study
Species	*Neophocaena*		N. asiaeorientalis (31) v. phocaenoides (17)	Zhou et al. (2018)
Species	*Phocoena*		P. phocoena (49) v. P. sinus (24)	Ben Chehida et al. (2020)
Species	*Stenella*		S. attenuata (59) v. S. longirostris (98)	Leslie et al. (2019)
Species	*Tursiops*		T. aduncus (21) v. T. truncatus (48)	Moura et al. (2013)
Subspecies	*Balaenoptera*	*musculus*	B. m. brevicauda (20) v. B. m. musculus (29)	This study
Subspecies	*Balaenoptera*	*musculus*	B. m. intermedia (65) v. B. m. musculus (29)	This study
Subspecies	*Balaenoptera*	*musculus*	B. m. brevicauda (20) v. B. m. intermedia (65)	This study
Subspecies	*Balaenoptera*	*physalus*	B. p. physalus (14) vs. B. p. quoyi (43)	Archer et al. (2013, 2019)
Subspecies	*Balaenoptera*	*physalus*	B. p. velifera (96) vs. B. p. quoyi (43)	Archer et al. (2013, 2019)
Subspecies	*Balaenoptera*	*physalus*	B. p. physalus (14) vs. B. p. velifera (96)	Archer et al. (2013, 2019)
Subspecies	*Neophocaena*	*asiaeorientalis*	N. a. asiaeorientalis (13) v. N. a. sunameri (18)	Zhou et al. (2018)[c]
Subspecies	*Stenella*	*attenuata*	S. a. graffmani (21) v. S. a. attenuata (38)	Leslie et al. (2019)
Subspecies	*Stenella*	*longirostris*	S. l. orientalis (30) v. S. l. longirostris (27)[b]	Leslie et al. (2019)
Unresolved	*Globicephala*	*macrorhynchus*	Atlantic (30) v. Naisa (72)	Van Cise et al. (2019)
Unresolved	*Globicephala*	*macrorhynchus*	Atlantic (30) v. Clade 3 (36)	Van Cise et al. (2019)
Unresolved	*Globicephala*	*macrorhynchus*	Clade 3 (36) v. Naisa (72)	Van Cise et al. (2019)
Unresolved	*Globicephala*	*macrorhynchus*	Atlantic (30) v. Shiho (43)	Van Cise et al. (2019)
Unresolved	*Globicephala*	*macrorhynchus*	Clade 3 (36) v. Shiho (43)	Van Cise et al. (2019)
Unresolved	*Globicephala*	*macrorhynchus*	Naisa (72) v. Shiho (43)	Van Cise et al. (2019)
Unresolved	*Orcinus*	*orca*	Antarctic_B1 (21) v. Antarctic_B2 (10)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_B1 (21) v. Antarctic_C (38)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_B2 (10) v. Antarctic_C (38)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_B1 (21) v. Resident (107)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_B2 (10) v. Resident (107)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_C (38) v. Resident (107)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Resident (107) v. Transient (97)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_B1 (21) v. Transient (97)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_B2 (10) v. Transient (97)	Morin et al. (2010, 2015)
Unresolved	*Orcinus*	*orca*	Antarctic_C (38) v. Transient (97)	Morin et al. (2010, 2015)
Unresolved	*Phocoena*	*phocoena*	Atlantic (13) v. Pacific (32)	Ben Chehida et al. (2020)

Accession numbers for mitogenomes sequenced for this study are in Supplementary Table S3.
[a]Strata based on Parsons et al. (2013): EAL = E. Aleutians; TRI = Trinity Islands; GOA = Gulf of Alaska; CAL = Central Aleutians; WAL = Western Aleutians.
[b]*S. l. longirostris* samples from "white belly" population in the Eastern Tropical Pacific.
[c]Mitogenomes assembled from short read archive data. Accession ID's in Supplementary Table S3.

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

State University, following methods described by Alexander et al. (2013) (Supplementary Methods). All previously unpublished mitogenome sequences generated for this study have been submitted to NCBI GenBank (Supplementary Table S3). The remainder of the data used in this analysis were previously published (Table 1).

Sequence alignment

All sequences were aligned first within species, then by combining species alignments within each genus. When available in the literature, full alignments were downloaded or requested from the authors. Otherwise, mitogenome sequences were downloaded from NCBI GenBank and alignments were created using MUSCLE (Edgar 2004) with default parameters as implemented in Geneious Prime (BioMatters, Ltd). All indels were visually inspected in the alignment files to ensure consistent placement. After complete mitochondrial genomes were aligned for each taxonomic group, coding sequences (CDS: defined here as all contiguous rRNA, tRNA, and protein coding sequence, excluding the full control region), and the first 500 bp of the control region were extracted as separate sequence alignments from each mitogenome alignment. The three datasets were used for parallel analyses of the shorter sequences with the full mitogenome sequences.

Data analyses

Unique haplotypes within each taxonomic-group alignment were assigned using the labelHaplotypes function (ignoring indels) in the *strataG* package (Archer et al. 2017a) for R (v. 4.0.4, R Core Team 2021). Samples were assigned a priori to populations, subspecies, and species pairs for analysis of divergence (F_{ST} [Weir and Cockerham 1984], Φ_{ST} [Excoffier et al. 1992], d_A [eq. 10.21; Nei 1987]), fixed differences, and PD (Archer et al. 2017b) in R using custom scripts and the *strataG* package. PD is the smallest percent of correctly classified individuals in a stratum. In a pairwise analysis, this is the percent of individuals classified at >50% probability (= PD_{50} in Archer et al. 2017b). We used the modelTest function in the Phangorn package (v.2.2.0, Schliep 2011) to determine the optimal nucleotide substitution model, which was the HKY (74%; Hasegawa et al. 1985) or GTR (26%; Tavaré 1986), for all mitogenome alignments. For divergence metrics involving genetic distance (d_A, Φ_{ST}) calculated in *strataG*, the most similar available model was the TN93 (Tamura and Nei 1993) model in the dist.dna function in the *ape* package (Paradis and Schliep 2019).

Simulated data and modeling the effect of simulation parameters on classification accuracy

Simulated datasets were generated to examine the effects of a set of evolutionary and demographic parameters on values of d_A and PD that meet our empirically derived subspecies and species thresholds (see Results). Our simulations used the same structure as the model from Archer et al. (2017b). Specifically, the coalescent-based model implemented in *fastsimcoal2 (v. 2.6)* (Excoffier and Foll 2011) was used to simulate a single population with effective size $2{*}N_e$ that diverged into two populations, each with effective size N_e. The split occurred T generations in the past, after which the populations exchanged migrants at the rate m per generation. Mutation occurred at the rate μ substitutions/basepair/generation. Twenty individuals were sampled per stratum.

Parameter values for each simulation were drawn from the following distributions:

$N_e \sim 10^{\mathrm{Uniform}(2,\,5)}$
$T \sim 10^{\mathrm{Uniform}(1.398,\,5.398)}$
$m \sim 10^{\mathrm{Uniform}(-9,\,-2)}$
$\mu \sim 10^{\mathrm{Uniform}(-8,\,-5.5)}$

These parameter ranges were chosen to include all values likely to apply to cetacean species (see Archer et al. 2017b for details). Empirical estimates for mutation rates in the cetacean mitogenome and control region range from 1.6 × 10^{-8} (=$10^{-7.8}$) to 1.86 × 10^{-6} (=$10^{-5.73}$) (summarized in Table 4 of Alexander et al. 2013). As in Archer et al. (2017b), a large number (100,000) of parameter sets were generated, from which 3,000 sets were randomly chosen that were evenly split between three categories: 1) $N_e m \geq 1$, to represent populations, 2) $0 < N_e m < 1$, to represent subspecies, and 3) m set to zero, to represent species.

The *fastsimcoal* simulation was run once each with sequence lengths of 16,400 bp (full mitogenome simulations) and 500 bp (control region simulations). For each simulated dataset, d_A and PD were calculated. Logistic Generalized Additive Models (GAM; Wood 2006) were fit to the resulting estimates as described in Archer et al. (2017b) and used to visualize the relationships between the simulation parameters and d_A and PD. The simulations and GAM analyses were performed using functions in the R package *strataG* and custom scripts.

Results

We collected published mitogenome datasets for 13 cetacean species and generated new mitogenome sequences for 6 cetacean species (Table 1). We generated pairwise divergence and diagnosability metrics for 13 populations, 9 subspecies, and 8 species pairs (Table 1). These included 11 of the 14 families of cetaceans. Similar to results based on short (~340 bp) control region sequences (Rosel et al. 2017a), analysis of mitogenomes showed that fixed differences, net nucleotide divergence (d_A), Φ_{ST}, and diagnosability all showed large differences in comparisons between subspecies and species pairs (Fig. 1). The frequency-based measure F_{ST} performed poorly for differentiating taxonomic groups. Diagnosability (PD) was high for all strata, with increasing diagnosability of subspecies relative to populations, and near-complete diagnosability for species, as expected (e.g. Morin et al. 2010, 2018; Archer et al. 2013; Van Cise et al. 2019). Differences between complete mitogenomes and only the coding sequences in pairwise statistics were minimal (Supplementary Figs. S1 and S2), so primarily full mitogenomes and control region datasets are presented and discussed.

Empirical control region data from cetacean pairwise comparisons were previously used to establish guidelines that minimized over- and under-classification of populations, subspecies, and species based on genetic data (Taylor et al. 2017a). Those guidelines were based on ~340 bp of the control region and would be expected to change for longer sequences and lower mutation rates, such as for the complete mitogenomes. Based on our empirical comparisons, we propose new guidelines for mitogenome datasets, based on similar minimization of over- and under-classification errors. For

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

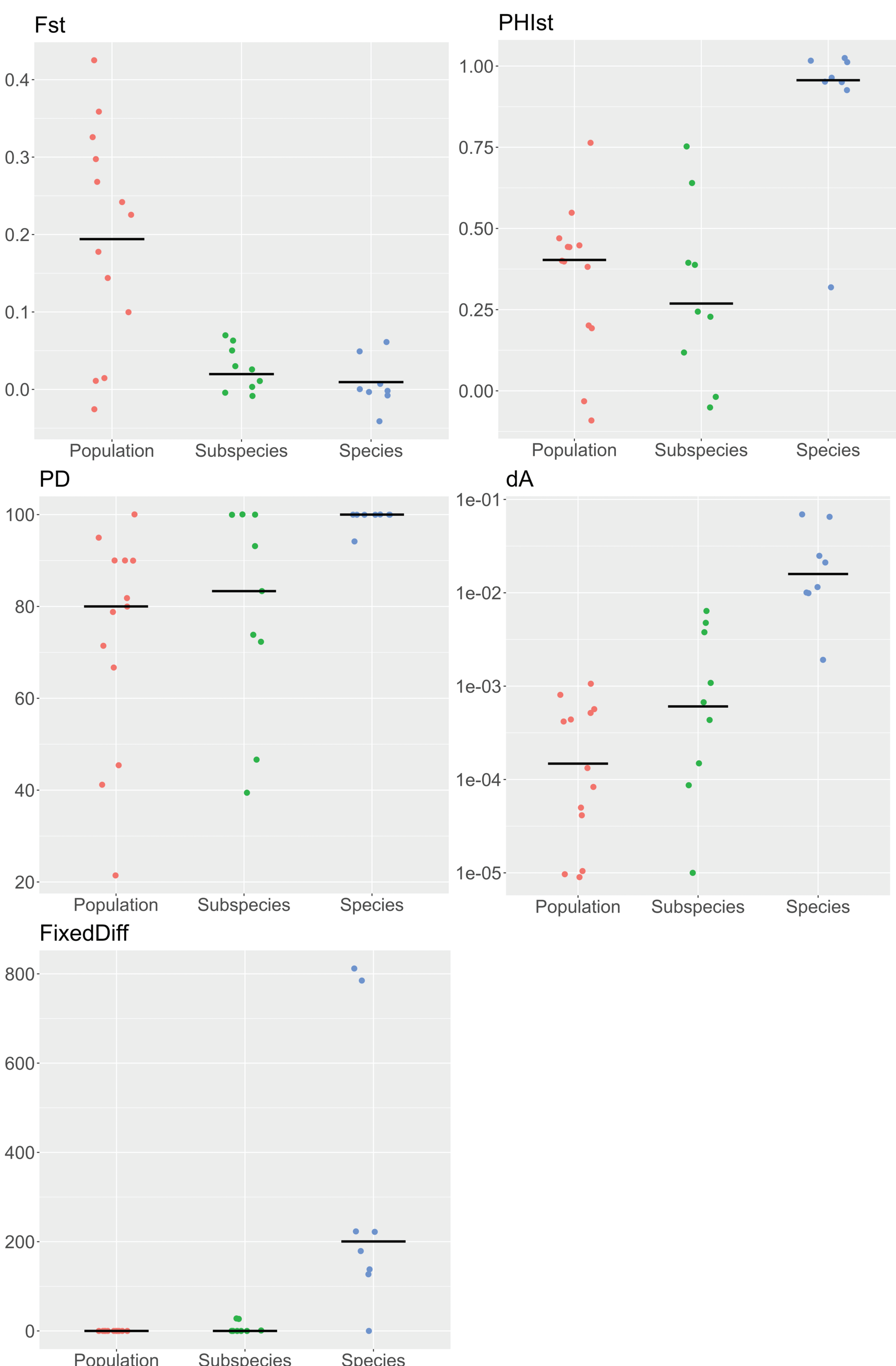

Fig. 1. Plots of mitogenome divergence metrics (F_{ST}, Φ_{ST} (PHIst), PD, net nucleotide divergence (d_A), and fixed differences) between accepted pairs of populations, subspecies, and species of cetaceans. The median values for each dataset are indicated by a horizontal bar.

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

d_A, the values that best delineated subspecies from populations and species from subspecies are $d_A > 0.0006$ and $d_A > 0.008$, respectively (Fig. 2). As argued by Taylor et al. (2017a), we propose to use 95% as the diagnosability (PD) threshold for subspecies and species in keeping with standard practices for diagnosability based on morphology, but include a reduced threshold (80%) in Fig. 2 that would decrease the number of under-classification errors for recognized subspecies, but

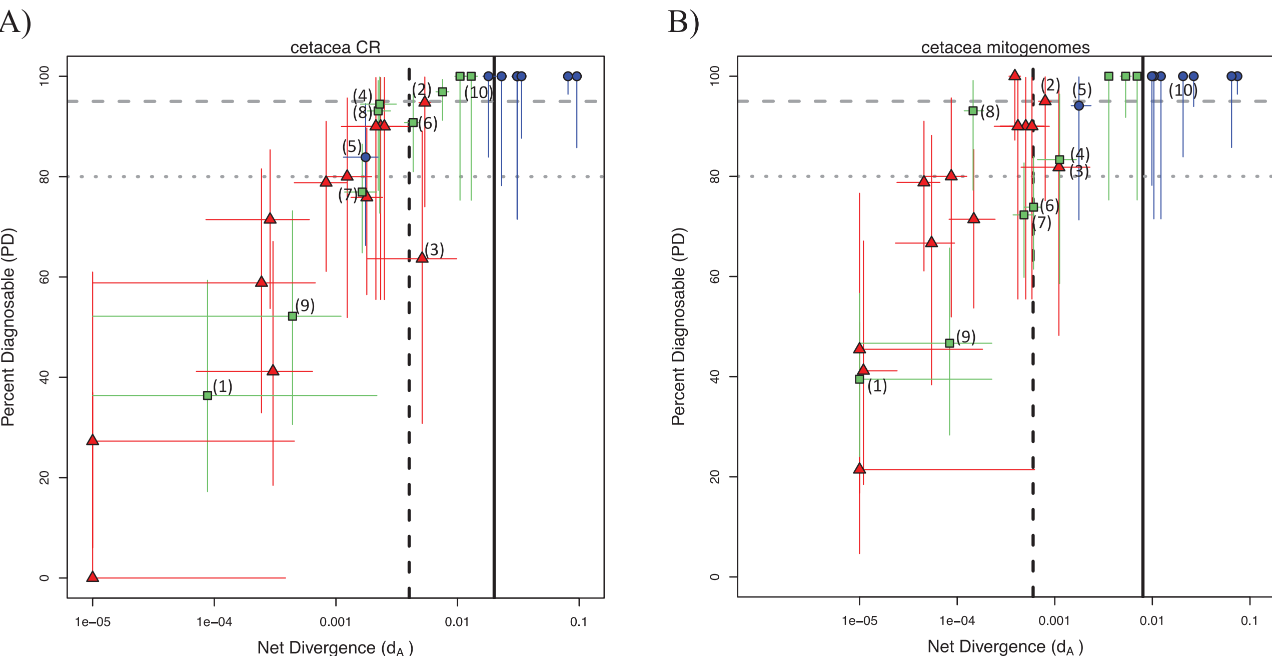

Fig. 2. Plot of net divergence (d_A) (on a natural log scale) vs. PD for accepted populations (triangle), subspecies (square), and species (circle) of cetaceans, based on A) the first 500 bp of the control regions extracted from the mitogenomes and B) complete mitochondrial genomes. Thin horizontal and vertical lines show the central 95th percentile of the estimate distributions. The dotted and solid vertical lines are the minimum cutoffs for subspecies and species, respectively, corresponding to d_A = 0.004, 0.02 for control regions (Taylor et al. 2017a) and d_A = 0.0006, 0.008 for mitogenomes. The gray horizontal dotted lines demark 80% and 95% diagnosability. The numbers highlight over- and under-classified pairs in one or both plots based on the taxonomic guidelines. 1. Spotted dolphin (Coastal/Pantropical), 2. Short-finned pilot whale (Hawaii/Mariana Islands), 3. Harbor porpoise (S. California/British Columbia – N. California), 4. Narrow-ridged finless porpoise (Yangtze/East Asian), 5. Finless porpoise (Narrow-ridged/Indo-Pacific), 6. Blue whale (Pygmy/Antarctic), 7. Blue whale (Antarctic/Northern), 8. Blue whale (Pygmy/Northern), 9. Spinner dolphin (Eastern/Gray's [white belly]), 10. Common and Indo-Pacific bottlenose dolphins. Divergence data in Supplementary Table S2.

increase the potential for misidentification of populations as subspecies. This lower threshold might be appropriate, for example, when sample sizes are small (see Archer et al. 2017b for additional discussion). Additional information can be useful to further justify taxonomic decisions (see Discussion).

We evaluated molecular classification accuracy using our proposed guidelines (i.e. using d_A and PD) for the full mitogenome sequences and a 500 bp portion of the control region from the same datasets (Fig. 2, Supplementary Table S1). Only one (short-finned pilot whale (Hawaii/Mariana Islands)) population pair was over-classified by molecular data as subspecies for both the control region and complete mitogenome datasets. Rates of under-classification based on molecular data were also similar for both control regions and mitogenomes, with six recognized subspecies pairs falling below the subspecies thresholds for each dataset, and two and one recognized species pairs for control regions and mitogenomes, respectively, falling below the species threshold. In the latter cases, one of the species pairs was the recently described species of finless porpoises, which are believed to have diverged within the past ~20 kyr (Jefferson and Wang 2011); this pair was also under-classified with the shorter control region dataset in Rosel et al. (2017a).

We similarly plotted net divergence and PD for additional pairs of intraspecific cetacean groups based on geographically and/or ecologically divergent pairs with unresolved taxonomic status in order to provide evidence in support of further taxonomic evaluation (Fig. 3, Supplementary Table S2). All short-finned pilot whale (*Globicephala macrorhynchus*) pairs were classified as subspecies based on our mitogenome guidelines. This is not surprising, as most pairs were previously proposed as subspecies based on nuclear single nucleotide polymorphism (SNP) divergence, mitogenome phylogeography, and control region d_A and PD (Van Cise et al. 2019). Killer whales (*Orcinus orca*), known for a global distribution with genetically and ecologically divergent ecotypes, especially in high latitudes (Pitman and Ensor 2003; Leduc et al. 2008; Morin et al. 2010, 2015; Foote et al. 2016, 2019), clustered into the population and subspecies levels of divergence based on d_A, though even the populations (d_A < 0.0006) all had high diagnosability. Interestingly, the ecotype pairs classified as populations were comparisons of all of the ice-associated Antarctic types (B1, B2, and C), which are among the most morphologically diverse of all killer whale ecotypes (Pitman and Ensor 2003; Durban et al. 2017). The partially sympatric North Pacific resident and transient ecotypes were classified as subspecies, as were allopatric comparisons between those two ecotypes and the Antarctic types. It has been previously suggested based on geographic isolation, control region data, and morphological differences (Rosel et al. 1995; Reeves et al. 2004) that the two harbor porpoise subspecies in the North Pacific and North Atlantic ocean basins, which form reciprocally monophyletic mitogenome clades separated for an approximately 0.86 Myr (Ben Chehida et al. 2020), could be considered separate species. Based on our analyses, they fall firmly within the "species" part of the d_A/PD plot, with 196 fixed differences across the mitogenome, supporting an elevation from subspecies to species status.

Similar to previous analyses of the number of fixed differences in control region sequences between subspecies

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

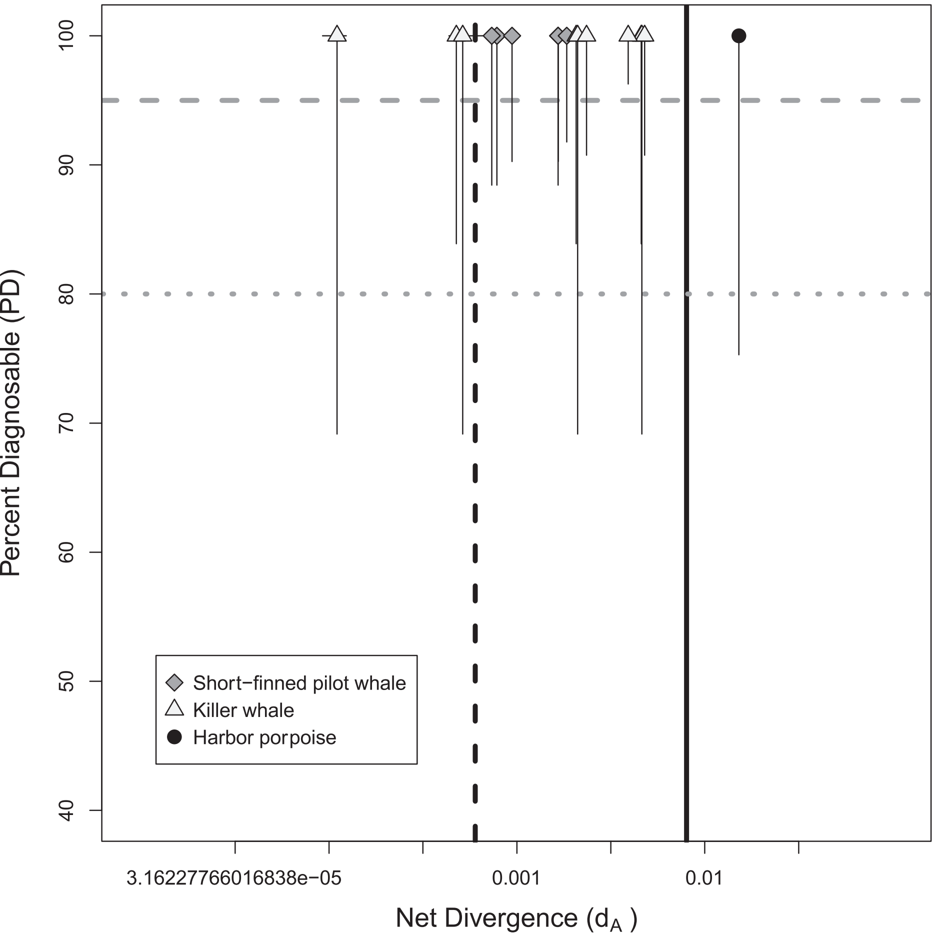

Fig. 3. Plot of net divergence (d_A) vs. PD values based on complete mitochondrial genomes (see Fig. 2B for plot details) for geographically and/or ecologically divergent pairs of unresolved taxonomic status. See Supplementary Table S2 for group pairs and divergence metrics.

and species pairs (Rosel et al. 2017a), the number of fixed differences in mitogenomes was much better for identifying species than subspecies (Supplementary Fig. S1c). Species were more strongly differentiated by fixed differences across the full mitogenome than just the control region. There were an average of 311 fixed differences for six of seven species pairs (range 127 to 812, excluding the species pair, *N. phocaenoides*, *N. asiaeorientalis*, with no fixed differences). The 500 bp control region sequences for the same pairs contained an average of 10 fixed differences (range 0 to 28). One species pair (*Tursiops truncatus*, *T. aduncus*) had no fixed differences in the control region, but 223 fixed differences in the complete mitogenome. For subspecies pairs, none of the 10 pairs had fixed differences in the control region, while they had zero to 28 fixed differences in the complete mitogenomes. The frequency of fixed differences between species pairs was nearly the same in coding and noncoding regions (average control region = 0.020/bp, CDS = 0.019/bp) across species pairs, so the benefit of using complete mitogenomes is due to the increased sequence length rather than different rates of fixation.

Simulated data were used to investigate the effect of a range of divergence times, mutation rates, and migration rates on nucleotide divergence and diagnosability. Simulations showed patterns that followed population genetic predictions, with larger effects of divergence time and effective population size on accumulation of diagnosable differences between groups, and smaller effects of mutation rate (Fig. 4). This was especially true for full mitogenomes, where mutation rate had little effect on d_A through most of the empirically observed range ($\log_{10}\mu$ = −7.8: −5.7). The simulated mitogenome data also replicated the divergence time results based on control region sequences in Archer et al. (2017b), but with a ≥10-fold decrease in divergence times required to reach >85% diagnosability (Fig. 4) across most population sizes and mutation rates.

Discussion

When genetic data are the only or primary source of information available, quantitative guidelines and decision criteria can help to make consistent, transparent taxonomic decisions. We have used empirical comparisons of well-accepted pairs of cetacean populations, subspecies, and species to establish a new divergence (d_A) threshold based on whole mitogenome sequences to use as quantitative standards to guide taxonomic decisions. Correct classification of accepted taxa was maximized when the population/subspecies boundary was d_A = 0.0006 with PD > 95%, and the subspecies/species boundary was d_A = 0.008, PD = 100%. While control region sequences will continue to be valuable for population and taxonomy studies, especially because they can be generated quickly and

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

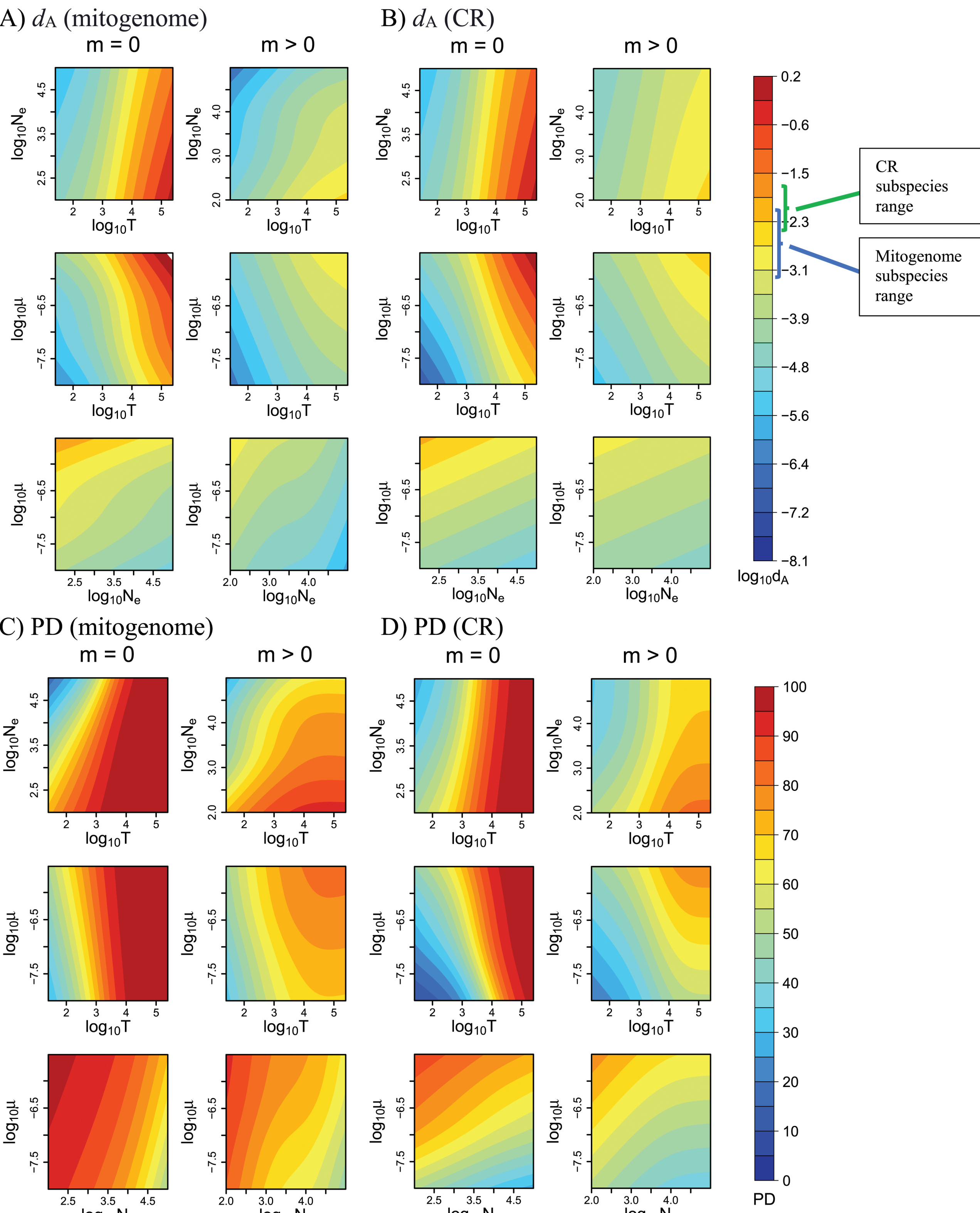

Fig. 4. Two-dimensional GAM fits of effective population size (N_e), divergence time in generations (T), mutation rate (μ), and migrants per generation (m) for d_A and PD from simulated data. A and C) 16.4 kb sequences (mitogenomes); B and D) 500 bp sequences (control regions [CR]). Each column contains the three possible pairwise comparisons for each simulation, with and without migration (see methods). Guideline ranges of d_A (see Fig. 2) for subspecies delineation are indicated on the $\log_{10}d_A$ legend.

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

cheaply from a wide variety of sample types (e.g. biopsies, teeth, and eDNA; Parsons et al. 2018; Costa et al. 2022), complete mitogenome sequences offer some advantages. The primary benefit of using whole mitogenome sequences instead of short control region (or other mitochondrial locus) sequences (Taylor et al. 2017a) are 1) a 10-fold increase in the rate of diagnosability relative to divergence time, as mutations across ~16.4 kb of sequence and drift lead to divergent sequences and fixation of differences between demographically isolated lineages, and 2) a decreased effect of differences in substitution rates on d_A over time, reducing variability in estimates of divergence due to stochastic sampling of variant sites from short sequences.

Complete mitogenome sequence comparisons across taxonomic groups also showed strong differences in the number of fixed differences in coding and noncoding (control region) sequence between subspecies and species categories (Supplementary Fig. S1). Among species, all pairs examined except for one, the recently divergent narrow- and wide-ridged porpoises (*Neophocaena* sp.) had at least 127 fixed differences across the mitogenomes. Thus, accumulation of fixed differences in the coding region reflects the longer isolation and divergence of lineages, and provides another mitogenome metric for delimiting species from subspecies. Although our sample size is small, we suggest a conservative guideline of >50 fixed differences for delineating species based on full mitogenomes, roughly half way between the maximum number of fixed differences found between subspecies pairs and the minimum for species pairs (excluding the *Neophocaena* species pair). Since the observed number of fixed differences may be influenced by the sample size and geographic distribution, the probability of observing a given number of fixed differences in a given sample size when they are not truly fixed can be calculated to support the strength of results (Braulik et al. 2021).

We divided the mitogenome alignments into coding and noncoding regions to investigate the differences between parameters for portions of the mitogenome presumably under different selective constraints, and where there might be advantages of using short noncoding sequences (control region) vs. complete mitogenomes. Comparison of d_A between coding regions and the control region indicate that divergence between pairs is greater in the control region in populations (Supplementary Fig. S2A), as would be expected due to higher mutation rate in the control region (Duchene et al. 2011). As taxonomic differentiation increases, net divergence remains greater in the control region, but the rate of change becomes more similar to that of the coding regions. This suggests that there is no advantage in using just coding sequences or the complete mitogenome for comparisons of divergence (d_A) at lower levels (populations and subspecies). However, diagnosability is higher for subspecies and species when the complete mitogenome (or complete coding sequence) is used, as fixed or nearly fixed differences will accumulate faster across the longer sequences (Supplementary Fig. S2B). Thus, when possible, use of complete mitogenome sequences in conjunction with the divergence and diagnosability guidelines is recommended for inference of taxonomic status.

The guideline for using d_A and PD to evaluate taxonomic status should serve as a starting point for making an argument about taxonomic status, rather than an inflexible rule. We acknowledge that mtDNA, even when the whole mitogenome is sequenced, still represents only a single locus and a single gene genealogy. Rapid speciation can lead to incomplete lineage sorting resulting in mito-nuclear discordance in the absence of gene flow (e.g. Duran et al. 2020; Deraad et al. 2023), and mito-nuclear discordance can arise from historical gene flow among diverging lineages, as has recently been suggested to have been common in cetaceans (Arnason et al. 2018; Foote et al. 2019; Westbury et al. 2023). Indeed, there is a suggestion of historical mitochondrial introgression in *Steno bredanensis* (Vilstrup et al. 2011; McGowen et al. 2020), and a putative hybrid origin of *Stenella clymene*, resulting in discordant mitochondrial and nuclear phylogenetic relationships (Kingston et al. 2009; Amaral et al. 2014; McGowen et al. 2020). However, there have been multiple studies comparing mitochondrial and nuclear phylogenetics and taxonomy across cetacea and within subgroups (Dalebout et al. 2008; McGowen et al. 2009; Amaral et al. 2012; Banguera-Hinestroza et al. 2014; Leslie et al. 2019; McGowen et al. 2020), and while patterns of inferred phylogenetic relationships have varied somewhat among studies and marker sets, changes to taxonomic level (as opposed to changing nomenclature or inferred phylogenetic relationships) due to mito-nuclear discordance to date have been limited to a recommendation to synonymize two congeneric species (*Delphinus delphis*, *D. capensis*) until more detailed analyses can be completed (Mcgowen et al. 2020; Committee on Taxonomy 2023).

There is a rich history of analyses of mtDNA sequences to inform systematics and taxonomy, as well as methods that do not rely on phylogenetic arguments (e.g. reciprocal monophyly) for taxonomic inference. The "flow diagram for species delineation" provided by Taylor et al. (2017a) (Fig. 5) provides the context for interpretation (see Taylor et al. 2017a for additional discussion). The genetic criteria provide PD and an independent evolutionary trajectory (d_A) as initial thresholds for taxonomic delimitation. Additional considerations for arguments about whether a group of animals merits subspecies or species status include effects of demography, gene flow, and mutation rate on the metrics considered and other factors that could lead to a different taxonomic outcome than using mtDNA metrics alone. In recognition of the needed increase in evidence to make an argument for a species classification, the guidelines flow diagram (Fig. 5) requires evidence beyond mtDNA sequence data, including ruling out male-mediated gene flow (often provided through analysis of nuDNA) and more traditional lines of evidence like geographic distribution of samples, acoustic signals, color patterns, morphology, distribution, and life history differences (Martien et al. 2019a).

There are many reasons why a taxonomic unit might be classified differently from what is suggested by the guidelines. The simulated data provide some guidance to assist researchers in interpreting divergence and diagnostic data when comparing groups in the context of their historical demography, culture, and ecology. For instance, the simulation results clearly demonstrate that very low values of N_e can result in high values of PD, even in the face of ongoing gene flow (Fig. 4C and D, $m > 0$ panels). Although low N_e can result from low abundance, it can also result from strong social structure (e.g. false killer whales, Martien et al. 2019b) or historical events such as bottlenecks and founder effects (Weber et al. 2000; Foote et al. 2019). In such cases, it might be appropriate to classify a unit at a lower taxonomic level than suggested by the guidelines. On the other hand, under-classification (not recognizing a taxon when it

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

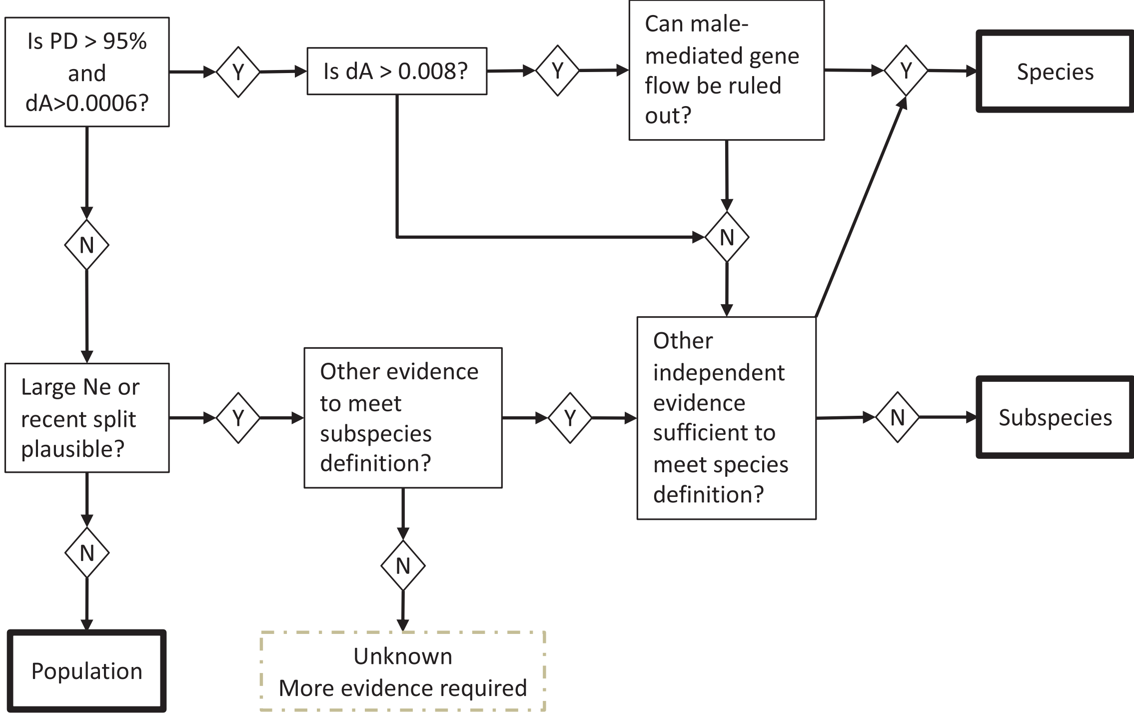

Fig. 5. Flow diagram for subspecies delineation using combined quantitative and qualitative standards, reprinted from Taylor et al. (2017a) (with permission from John Wiley & Sons, license 5453770777609) and modified for use with complete mitogenome sequence data. The threshold values assume the user is evaluating a case relying on complete mitochondrial genome data. PD is the smallest strata-specific correct classification score in a given comparison (e.g. PD_{50} in two-strata comparisons in Archer et al. 2017b). The text box "other evidence to meet subspecies definition" allows for subspecies delineation when both conditions based on mitogenome standards are not met. This box could be used either for the case when one condition is met and one unmet or when both just barely miss meeting the standards. For example, consider the case with PD < 95% and d_A > 0.0006. Diagnosability could be achieved with morphological data or nuclear data that are sufficient for subspecies but not for full species.

is warranted) of empirical, recognized subspecies pairs was more common for both mitogenome and control region datasets. Under-classification may be due to demographic effects, but the speciation process also may proceed at a faster pace than the evolution of mtDNA (e.g. resulting in mtDNA polyphyly; Funk and Omland 2003; Wang et al. 2018), especially for highly social species that show evidence of niche partitioning or sexual selection (e.g. killer whales [Foote et al. 2016], spinner dolphins [Andrews et al. 2021]). In the empirical mitogenome data presented here, under-classification based on the d_A threshold involved subspecies of only three species: blue whales, spinner dolphins, and spotted dolphins. Blue whales are globally distributed with high historical abundance in at least some oceans, and spinner and spotted dolphins fall at the highest demographic extreme, with global distributions and pelagic populations numbering in the hundreds of thousands, in the range where both d_A and PD would be expected to be lowest (Fig. 2).

As genomic methods continue to expand the quantity and quality of data available to investigate evolutionary and demographic history of species, genomic tools will continue to expand and refine taxonomic delineation in the absence of morphological data. Our analysis of whole mitogenome datasets of cetacean populations, subspecies, and species provides a benchmark for understanding the patterns of mtDNA diversity, divergence, and diagnosability through the speciation process, with simulated data to further examine the ranges of parameters that will affect these processes across a wide range of possibilities.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Acknowledgments

We are grateful to the many people who provided samples for mitogenome sequences generated for all of the included studies, and to facilitate this comparative analysis: Diane Claridge, Bahamas Marine Mammal Research Organisation; Emma Carroll, University of Auckland; Jon Seger and Luciano Valenzuela, University of Utah; Timothy Frasier, Saint Mary's University; Erin Oleson, Pacific Islands Fisheries Science Center; Robin Baird and John Calambokidis, Cascadia Research Collective; Kim Parsons, Phil Clapham, and Pat Gearin, National Marine Mammal Lab, Alaska Fisheries Science Center; Barbara Mahoney, Alaska Regional Office, NOAA Fisheries; Jim Dines, Los Angeles County Museum of Natural History; Jay Barlow, Robert Pitman, and Karin Forney, Southwest Fisheries Science Center; and Suzanne Yin and the Hawai'i Marine Mammal Consortium. We are grateful to the Associate Editor and three external reviewers for helpful comments.

Funding

Funding for the world-wide survey of right whale mitogenomes was provided to CSB by the Pacific Life Foundation. Samples

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

for mitogenome sequencing of additional species for this study were provided by the SWFSC Marine Mammal and Sea Turtle Research (MMASTR) collection, with funding from the SWFSC Marine Mammal and Turtle Division.

Conflict of interest

The authors declare no conflicts of interest.

Data availability

All mitogenome sequences generated for this study are available in the NCBI nucleotide database (see Supplementary Table S3 for accession ID's for sequences generated for this study). Custom R scripts for divergence metrics and simulations are available at https://github.com/PAMorin/Mitogenome_taxonomy.

References

Albertson GR, Alexander A, Archer FI, Caballero S, Martien KK, Hemery LG, Baird RW, Oremus M, Poole MM, Duffield DA, et al. Worldwide phylogeography of rough-toothed dolphins (*Steno bredanensis*) provides evidence for subspecies delimitation. *Mar Mamm Sci*. 2022:38:1371–1397.

Alexander A, Steel D, Slikas B, Hoekzema K, Carraher C, Parks M, Cronn R, Baker CS. Low diversity in the mitogenome of sperm whales revealed by next-generation sequencing. *Genome Biol Evol*. 2013:5:113–129.

Amaral AR, Jackson JA, Moller LM, Beheregaray LB, Manuela Coelho M. Species tree of a recent radiation: the subfamily Delphininae (Cetacea, Mammalia). *Mol Phylogenet Evol*. 2012:64:243–253.

Amaral AR, Lovewell G, Coelho MM, Amato G, Rosenbaum HC. Hybrid speciation in a marine mammal: the clymene dolphin (*Stenella clymene*). *PLoS One*. 2014:9:e83645.

Andrews KR, Epstein B, Leslie M, Fiedler P, Morin PA, Hoelzel AR. Genomic signatures of divergent selection are associated with social behavior for spinner dolphin ecotypes. *Mol Ecol*. 2021:30:1993–2008.

Archer FI, Adams PE, Schneiders BB. strataG: an R package for manipulating, summarizing and analysing population genetic data. *Mol Ecol Resour*. 2017a:17:5–11.

Archer FI, Brownell RL Jr, Hancock-Hanser BL, Morin PA, Robertson KM, Sherman KK, Calambokidis J, Urbán JR, Rosel PE, Mizroch SA, et al. Revision of fin whale *Balaenoptera physalus* (Linnaeus, 1758) subspecies using genetics. *J Mammal*. 2019:100:1653–1670.

Archer FI, Martien KK, Taylor BL. Diagnosability of mtDNA with Random Forests: using sequence data to delimit subspecies. *Mar Mamm Sci*. 2017b:33:101–131.

Archer FI, Morin PA, Hancock-Hanser BL, Robertson KM, Leslie MS, Berube M, Panigada S, Taylor BL. Mitogenomic phylogenetics of fin whales (*Balaenoptera physalus* spp.): genetic evidence for revision of subspecies. *PLoS One*. 2013:8:e63396.

Arnason U, Lammers F, Kumar V, Nilsson MA, Janke A. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Sci Adv*. 2018:4:eaap9873.

Banguera-Hinestroza E, Hayano A, Crespo E, Hoelzel AR. Delphinid systematics and biogeography with a focus on the current genus *Lagenorhynchus*: multiple pathways for antitropical and trans-oceanic radiation. *Mol Phylogenet Evol*. 2014:80:217–230.

Ben Chehida Y, Thumloup J, Schumacher C, Harkins T, Aguilar A, Borrel A, Ferreira M, Rojas-Bracho L, Robertson KM, Taylor BL, et al. Mitochondrial genomics reveals evolutionary history of the porpoises (Phocoenidae) across the speciation continuum. *Sci Rep*. 2020:10:15190.

Braulik GT, Archer FI, Khan U, Imran M, Sinha RK, Jefferson TA, Donovan C, Graves JA. Taxonomic revision of the South Asian River dolphins (*Platanista*): Indus and Ganges River dolphins are separate species. *Mar Mamm Sci*. 2021:37:1022–1059.

Carroll EL, Mcgowen MR, Mccarthy ML, Marx FG, Aguilar N, Dalebout ML, Dryer S, Gaggiotti OE, Hansen SS, Van Helden A, et al. Speciation in the deep: genomics and morphology reveal a new species of beaked whale *Mesoplodon eueu*. *Proc R Soc Lond Biol*. 2021:288:20211213.

Committee on Taxonomy. *List of marine mammal species and subspecies*. Society for Marine Mammalogy; 2019. www.marinemammalscience.org, consulted 28 June, 2019.

Committee on Taxonomy. *List of marine mammal species and subspecies*. Society for Marine Mammalogy; 2023. www.marinemammalscience.org, consulted 9 September 2023.

Costa APB, Mcfee W, Wilcox LA, Archer FI, Rosel PE. The common bottlenose dolphin (*Tursiops truncatus*) ecotypes of the western North Atlantic revisited: an integrative taxonomic investigation supports the presence of distinct species. *Zool J Linn Soc*. 2022:196:1608–1636.

Dalebout ML, Baker CS, Steel D, Robertson KM, Chivers SJ, Perrin WF, Mead JG, Grace RV, Schofield TD. A divergent mtDNA lineage among *Mesoplodon* beaked whales: molecular evidence for a new species in the tropical Pacific? *Mar Mamm Sci*. 2007:23:954–966.

Dalebout ML, Mead JG, Baker CS, Baker AN, Van Helden AL. A new species of beaked whale *Mesoplodon perrini* sp N. (Cetacea: Ziphiidae) discovered through phylogenetic analyses of mitochondrial DNA sequences. *Mar Mamm Sci*. 2002:18:577–608.

Dalebout ML, Steel D, Baker CS. Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: implications for the evolution of male tusks. *Syst Biol*. 2008:57:857–875.

De Queiroz K. Species concepts and species delimitation. *Syst Biol*. 2007:56:879–886.

Deraad DA, Mccullough JM, Decicco LH, Hime PM, Joseph L, Andersen MJ, Moyle RG. Mitonuclear discordance results from incomplete lineage sorting, with no detectable evidence for gene flow, in a rapid radiation of *Todiramphus* kingfishers. *Mol Ecol*. 2023:32:4844–4862.

Duchene S, Archer FI, Vilstrup J, Caballero S, Morin PA. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One*. 2011:6:e27138.

Duran DP, Laroche RA, Gough HM, Gwiazdowski RA, Knisley CB, Herrmann DP, Roman SJ, Egan SP. Geographic life history differences predict genomic divergence better than mitochondrial barcodes or phenotype. *Genes (Basel)*. 2020:11:265.

Durban JW, Fearnbach H, Burrows DG, Ylitalo GM, Pitman RL. Morphological and ecological evidence for two sympatric forms of Type B killer whale around the Antarctic Peninsula. *Polar Biol*. 2017:40:231–236.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004:32:1792–1797.

Excoffier L, Foll M. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*. 2011:27:1332–1334.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 1992:131:479–491.

Foote AD, Martin MD, Louis M, Pacheco G, Robertson KM, Sinding M-HS, Amaral AR, Baird RW, Baker CS, Ballance L, et al. Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Mol Ecol*. 2019:28:3427–3444.

Foote AD, Vijay N, Ávila-Arcos MC, Baird RW, Durban JW, Fumagalli M, Gibbs RA, Hanson MB, Korneliussen TS, Martin MD, et al. Genome-culture coevolution promotes rapid divergence in the killer whale. *Nat Commun*. 2016:7:Article No. 11693.

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst*. 2003:34:397–423.

Hancock-Hanser B, Frey A, Leslie M, Dutton PH, Archer EI, Morin PA. Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Mol Ecol Resour*. 2013:13:254–268.

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

Jefferson TA, Wang JY. Revision of the taxonomy of finless porpoises (genus *Neophocaena*): the existence of two species. *J Mar Anim Ecol*. 2011:4:3–16.

Kingston SE, Adams LD, Rosel PE. Testing mitochondrial sequences and anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group: molecular systematics of the Delphininae (Cetacea: Odontoceti: Delphinidae). *BMC Evol Biol*. 2009:9:245.

Leduc RG, Robertson KM, Pitman RL. Mitochondrial sequence divergence among Antarctic killer whale ecotypes is consistent with multiple species. *Biol Lett*. 2008:4:426–429.

Leslie MS, Archer EI, Morin PA. Mitogenomic differentiation in spinner (*Stenella longirostris*) and pantropical spotted dolphins (*S. attenuata*) from the eastern tropical Pacific Ocean. *Mar Mamm Sci*. 2019:35:522–551.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009:25:1754–1760.

Louis M, Skovrind M, Samaniego Castruita JA, Garilao C, Kaschner K, Gopalakrishnan S, Haile JS, Lydersen C, Kovacs KM, Garde E, et al. Influence of past climate change on phylogeography and demographic history of narwhals, *Monodon monoceros*. *Proc R Soc B*. 2020:287:20192964.

Martien KK, Lang AR, Taylor BL, Rosel PE, Simmons SE, Oleson EM, Boveng PL, Hanson MB. *The DIP delineation handbook: a guide to using multiple lines of evidence to delineate demographically independent populations of marine mammals*. NOAA Technical Memorandum NMFS; 2019a. NOAA-TM-NMFS-SWFSC-622.

Martien K, Leslie M, Taylor BL, Morin PA, Archer FI, Hancock-Hanser B, Rosel P, Vollmer N, Viricel A, Cipriano F. Analytical approaches to subspecies delimitation with genetic data. *Mar Mamm Sci*. 2017:33:27–55.

Martien KK, Taylor BL, Chivers SJ, Mahaffy SD, Gorgone AM, Baird RW. Fidelity to natal social groups and mating within and between social groups in an endangered false killer whale population. *Endanger Species Res*. 2019b:40:219–230.

Mayr E. *Principles of systematic zoology*. New York (NY): McGraw-Hill; 1969.

Mcgowen MR, Spaulding M, Gatesy J. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol*. 2009:53:891–906.

Mcgowen MR, Tsagkogeorga G, Álvarez-Carretero S, Dos Reis M, Struebig M, Deaville R, Jepson PD, Jarman S, Polanowski A, Morin PA, et al. Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Syst Biol*. 2020:69:479–501.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, et al. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res*. 2010:20:908–916.

Morin PA, Baker CS, Brewer RS, Burdin AM, Dalebout ML, Dines JP, Fedutin I, Filatova O, Hoyt E, Jung J-L, et al. Genetic structure of the beaked whale genus *Berardius* in the North Pacific, with genetic evidence for a new species. *Mar Mamm Sci*. 2017:33:96–111.

Morin PA, Foote AD, Baker CS, Hancock-Hanser BL, Kaschner K, Mate BR, Mesnick SL, Pease VL, Rosel PE, Alexander A. Demography or selection on linked cultural traits or genes? Investigating the driver of low mtDNA diversity in the sperm whale using complementary mitochondrial and nuclear genome analyses. *Mol Ecol*. 2018:27:2604–2619.

Morin PA, Parsons KM, Archer FI, Ávila-Arcos MC, Barrett-Lennard LG, Dalla Rosa L, Duchêne S, Durban JW, Ellis G, Ferguson SH, et al. Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Mol Ecol*. 2015:24:3964–3979.

Moura AE, Nielsen SC, Vilstrup JT, Moreno-Mayar JV, Gilbert MT, Gray HW, Natoli A, Moller L, Hoelzel AR. Recent diversification of a marine genus (*Tursiops* spp.) tracks habitat preference and environmental change. *Syst Biol*. 2013:62:865–877.

Nei M. *Molecular evolutionary genetics*. New York (NY): Columbia University Press; 1987.

Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019:35:526–528.

Parsons KM, Durban JW, Burdin AM, Burkanov VN, Pitman RL, Barlow J, Barrett-Lennard LG, Leduc RG, Robertson KM, Matkin CO, et al. Geographic patterns of genetic differentiation among killer whales in the northern North Pacific. *J Hered*. 2013;104(6):737–754.

Parsons KM, Everett M, Dahlheim M, Park L. Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. *R Soc Open Sci*. 2018:5:180537.

Pitman RL, Ensor P. Three forms of killer whales (*Orcinus orca*) in Antarctic waters. *J Cetacean Res Manag*. 2003:5:131–139.

R Core Team. *R: a language and environment for statistical computing*. Vienna (Austria): R Foundation for Statistical Computing; 2021.

Reeves RR, Perrin WF, Taylor BL, Baker CS, Mesnick SL. *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*. Sponsored by the U.S. Department of Commerce, NOAA Technical Memorandum; 2004. Contract No. NOAA-TM-NMFS-SWFSC-363.

Rosel PE, Dizon AE, Haygood MG. Variability of the mitochondrial control region in populations of the harbor porpoise, *Phocoena phocoena*, on interoceanic and regional scales. *Can J Fish Aquat Sci*. 1995:52:1210–1219.

Rosel PE, Hancock-Hanser BL, Archer FI, Robertson KM, Martien KK, Leslie MS, Berta A, Cipriano F, Viricel A, Viaud-Martinez KA, et al. Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences. *Mar Mamm Sci*. 2017a:33:76–100.

Rosel PE, Taylor BL, Hancock-Hanser BL, Morin PA, Archer FI, Lang AR, Mesnick SL, Pease VL, Perrin WF, Robertson KM, et al. A review of molecular genetic markers and analytical approaches that have been used for delimiting marine mammal subspecies and species. *Mar Mamm Sci*. 2017b:33:56–75.

Ross HA, Lento GM, Dalebout ML, Goode M, Ewing G, Mclaren P, Rodrigo AG, Lavery S, Baker CS. DNA surveillance: web-based molecular identification of whales, dolphins, and porpoises. *J Hered*. 2003:94:111–114.

Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011:27:592–593.

Skovrind M, Louis M, Westbury MV, Garilao C, Kaschner K, Castruita JS, Gopalakrishnan S, Knudsen SW, Haile JS, Dalen L, et al. Circumpolar phylogeography and demographic history of beluga whales reflect past climatic fluctuations. *Mol Ecol*. 2021:30:2543–2559.

Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 1993:10:512–526.

Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Live Sci*. 1986:17:57–86.

Taylor BL. Identifying units to conserve. In: Reynolds JE III, Perrin WF, Reeves RR, Montgomery S, Ragen TJ, editors. *Marine mammal research: conservation beyond crisis*. Baltimore (MD): The John Hopkins University Press; 2005.

Taylor BL, Archer FI, Martien KK, Rosel PE, Hancock-Hanser BL, Lang AR, Leslie MS, Mesnick SL, Morin PA, Perrin WF, et al. Guidelines and quantitative standards to improve consistency in cetacean subspecies and species delimitation relying on molecular genetic data. *Mar Mamm Sci*. 2017a:33:132–155.

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024

Taylor BL, Martien KK, Morin PA. Identifying units to conserve using genetic data. In: Boyd I, Bowen D, Iverson S, editors. *Marine Mammal Ecology and Conservation: a handbook of techniques. Techniques in Ecology and Conservation*. Oxford: Oxford University Press; 2010. p. 306–324.

Taylor BL, Perrin WF, Reeves RR, Rosel PE, Wang JY, Cipriano F, Scott Baker C, Brownell RL. Why we should develop guidelines and quantitative standards for using genetic data to delimit subspecies for data-poor organisms like cetaceans. *Mar Mamm Sci*. 2017b:33:12–26.

Van Cise AM, Baird RW, Baker CS, Cerchio S, Claridge D, Fielding R, Hancock-Hanser B, Marrero J, Martien KK, Mignucci-Giannoni AA, et al. Oceanographic barriers, divergence, and admixture: phylogeography and taxonomy of two putative subspecies of short-finned pilot whale. *Mol Ecol*. 2019:28: 2886–2902.

Vilstrup JT, Ho SYW, Foote AD, Morin PA, Kreb D, Krützen M, Parra GJ, Robertson KM, De Stephanis R, Verborgh P, et al. Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evol Biol*. 2011:11:65.

Wang K, Lenstra JA, Liu L, Hu QJ, Ma T, Qiu Q, Liu JQ. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Commun Biol*. 2018:1:169.

Weber DS, Stewart BS, Garza JC, Lehman N. An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Curr Biol*. 2000:10:1287–1290.

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984:38:1358–1370.

Westbury MV, Cabrera AA, Rey-Iglesia A, De Cahsan B, Duchene DA, Hartmann S, Lorenzen ED. A genomic assessment of the marine-speciation paradox within the toothed whale superfamily Delphinoidea. *Mol Ecol*. 2023:32:4829–4843. doi:10.1111/mec.17069

Wood S. *Generalized additive models: an introduction with R*. Boca Raton (FL): CRC Press; 2006.

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

Zhou X, Guang X, Sun D, Xu S, Li M, Seim I, Jie W, Yang L, Zhu Q, Xu J, et al. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nat Commun*. 2018:9:1276.

Downloaded from https://academic.oup.com/jhered/article/114/6/612/7255533 by NOAA Central Library user on 11 January 2024