Evolutionary history of the porpoises (*Phocoenidae*) across the speciation continuum: a mitogenome phylogeographic perspective

Yacine Ben Chehida[1], Julie Thumloup[1], Cassie Schumacher[2], Timothy Harkins[2], Alex Aguilar[3], Asunción Borrell[3], Marisa Ferreira[4], Lorenzo Rojas-Bracho[5], Kelly M. Roberston[6], Barbara L. Taylor[6], Gísli A. Víkingsson[7], Arthur Weyna[8], Jonathan Romiguier[8], Phillip A. Morin[6], Michael C. Fontaine[1,9]*

[1] Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, PO Box 11103 CC, Groningen, The Netherlands
[2] Swift Biosciences, 674 S. Wagner Rd., Suite 100, Ann Arbor, MI 48103, USA
[3] IRBIO and Department of Evolutive Biology, Ecology and Environmental Sciences, Faculty of Biology, University of Barcelona, Diagonal 643, 08071 Barcelona, Spain
[4] MATB-Sociedade Portuguesa de Vida Selvagem, Estação de Campo de Quiaios, Apartado EC Quiaios, 3080-530 Figueira da Foz, Portugal & CPRAM-Ecomare, Estrada do Porto de Pesca Costeira, 3830-565 Gafanha da Nazaré, Portugal
[5] Instituto Nacional de Ecología, Centro de Investigación Científica y de Educación Superior de Ensenada, Carretera Ensenada-Tijuana 3918, Fraccionamiento Zona Playitas, Ensenada, BC 22860, Mexico
[6] Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Dr., La Jolla, California 92037, USA
[7] Marine and Freshwater Research Institute, PO Box 1390, 121 Reykjavik, Iceland
[8] Institut des Sciences de l'Évolution (Université de Montpellier, CNRS UMR 5554), Montpellier, France
[9] Laboratoire MIVEGEC (Université de Montpellier, UMR CNRS 5290, IRD 229), Centre IRD de Montpellier, Montpellier, France

Corresponding author: Michael C. Fontaine (michael.fontaine@cnrs.fr). MIVEGEC (IRD 224-CNRS 5290-UM). Institut de Recherche pour le Développement (IRD), 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France.

ORCID ID:
Yacine Ben Chehida: https://orcid.org/0000-0001-7269-9082
Alex Aguilar: https://orcid.org/0000-0002-5751-2512
Asunción Borrell: https://orcid.org/0000-0002-6714-0724
Marisa Ferreira : https://orcid.org/0000-0002-0733-3452
Gísli A. Víkingsson: https://orcid.org/0000-0002-4501-193X
Phillip A. Morin: https://orcid.org/0000-0002-3279-1519
Michael C. Fontaine: https://orcid.org/0000-0003-1156-4154

Abstract

Historical changes affecting food resources are a major driver of cetacean evolution. Small cetaceans like porpoises (Phocoenidae) are among the most metabolically challenged marine mammals and are particularly sensitive to changes in their food resources. The seven species of this family inhabit mostly temperate waters and constitute a textbook example of antitropical distribution. Yet, their evolutionary history remains poorly known despite major conservation issues threatening the survival of some porpoises (e.g., vaquita and Yangzte finless porpoises). Here, we reconstructed their evolutionary history across the speciation continuum, from intraspecific subdivisions to species divergence. Phylogenetic analyses of 63 mitochondrial genomes suggest that, like other toothed whales, porpoises radiated during the Pliocene in response to deep environmental changes. However, all intra-specific phylogeographic patterns were shaped during the Quaternary Glaciations. We observed analogous evolutionary patterns in both hemispheres associated with convergent adaptations to coastal *versus* oceanic environments. This result suggests that the mechanism(s) driving species diversification in the relatively well-known species from the northern hemisphere may apply also to the poorly-known southern species. In contrast to previous studies, we showed that the spectacled and Burmeister's porpoises share a more recent common ancestor than with the vaquita that diverged from southern species during the Pliocene. The low genetic diversity observed in the vaquita carried signatures of a very low population size throughout at least the last 5,000 years, leaving one single relict mitochondrial lineage. Finally, we observed unreported subspecies level divergence within Dall's, spectacled and Pacific harbor porpoises, suggesting a richer evolutionary history than previously suspected. These results provide a new perspective on the mechanism driving the adaptation and speciation processes involved in the diversification of cetacean species. This knowledge can illuminate their demographic trends and provide an evolutionary framework for their conservation.

1. Introduction

Most cetaceans possess a tremendous potential for dispersal in an environment that is relatively unobstructed by geographical barriers. This observation begs the question of how do populations of such highly mobile pelagic species in such an open environment split and become reproductively isolated from each other to evolve into new species. Recent micro- and macro-evolutionary studies showed that changes in environmental conditions (Steeman et al., 2009), development of matrilinearly transmitted cultures (Whitehead, 1998), and resource specialization (Fontaine et al., 2014; Foote et al., 2016; Louis et al., 2014) are major drivers of population differentiation and speciation in cetacean species. Yet, the processes that link these two evolutionary timescales are still not fully understood and empirical examples are limited (Foote et al., 2016; Steeman et al., 2009).

Several cetacean taxa display an antitropical distribution where the distribution of closely related taxa occurs on either side of the equator but are absent from the tropics (Barnes, 1985; Hare et al., 2002; Pastene et al., 2007). Multiple mechanisms have been proposed to explain such a peculiar distribution (Burridge, 2002). In cetaceans, the predominant hypothesis implies dispersal and vicariance of temperate species enabled by climatic fluctuations, especially during the Quaternary glacial-interglacial cycles (Davies, 1963). During glacial periods, cold adapted taxa extended their range into the tropics and possibly crossed the equator. In the subsequent warmer interglacial periods, these taxa would shift their range to higher latitudes. This geographic isolation in both hemispheres promoted allopatric divergence of conspecific taxa, resulting in their antitropical distribution. A closely related scenario suggests that the rise of the overall sea temperature during interglacial periods would have altered the winds direction and upwelling strength, leading to a redistribution of feeding areas for cetaceans toward higher latitudes, which in turn promoted their antitropical distribution (Pastene et al., 2007). Another plausible hypothesis implies that broadly distributed species, such as several cetacean species, were outperformed in the tropics by more competitive species (Burridge, 2002). A combination of these different mechanism is also possible.

The porpoises family (Phocoenidae) displays one of the best known example of antitropical distribution (Barnes, 1985). Porpoises are among the smallest cetaceans and represent an interesting evolutionary lineage within the Delphinoidea, splitting from the Monodontidae during the Miocene (~15 Million years ago) (McGowen et al., 2009; Steeman et al., 2009). Gaskin (1982) suggested that porpoises originated from a tropical environment and then radiated into temperate zones in both hemispheres. In contrast, based on the location

1   of the oldest fossils, Barnes (1985) suggested that they arose in a more temperate environment

2   of the North Pacific Ocean and subsequently colonized the southern hemisphere, the Indian and

3   Atlantic Oceans. Porpoises consist of seven extant species that occur in both hemispheres in

4   pelagic, coastal, and riverine habitats (Fig. 1a). The family includes primarily cold-water

5   tolerant species, but two sister species of finless porpoises (*Neophocoena phocaenoides, N.*

6   *asiaeorientalis*) (Zhou et al., 2018) inhabit the tropical waters of the Indo-Pacific preferring

7   mangrove zones. They are also found in the Yellow Sea, East China Sea, Sea of Japan and in

8   estuaries and large river systems of the Indus, Ganges, and Yangtze rivers. The remaining

9   species are considered cold-water tolerant and are antitropically distributed. In the Northern

10   Hemisphere, the harbor porpoise (*Phocoena phocoena*) inhabits the coastal waters of the North

11   Pacific and North Atlantic, while the Dall's porpoise (*Phocoenoides dalli*) occupies the oceanic

12   waters of the North Pacific. This neritic-oceanic habitat segregation is mirrored in the southern

13   hemisphere with the Burmeister's porpoise (*Phocoena spinipinnis*), occupying the coastal

14   waters of South America and the poorly known spectacled porpoise (*Phocoena dioptrica*)

15   occupying the circum-Antarctic oceanic waters. The vaquita depart from the other species with

16   an extremely localized geographical distribution in the upper Gulf of California and is now

17   critically endangered (Morell, 2017).

18   With the exception of vaquitas, all species of porpoises exhibit a relatively broad

19   distribution range that appear fairly continuous at a global scale. Nevertheless, despite having

20   the ability to disperse over vast distance in an open environment, many include distinct

21   subspecies, ecotypes, or morphs. For example, the finless porpoises not only include two

22   recognized species, but also an incipient species within the Yangtze River (Chen et al., 2017;

23   Zhou et al., 2018); the harbor porpoise also displays a disjunct distribution with at least four

24   sub-species reported (Fontaine, 2016); at least two forms of Dall's porpoise have been described

25   (Hayano et al., 2003); and the Burmeister's porpoise also shows evidence of deep population

26   structure (Rosa et al., 2005). Such intraspecific subdivisions, also observed in killer whales

27   (Foote et al., 2016) and dolphins (Louis et al., 2014), illustrate the evolutionary processes in

28   action, and can, in some cases, eventually lead to new species. Porpoises are thus an interesting

29   model to investigate the evolutionary processes at both micro- and macroevolutionary time

30   scale to better understand present and historical mechanisms governing population structure,

31   adaptation to different niches, and speciation.

32   From a conservation perspective, the coastal habitat of many porpoise species largely

33   overlaps with areas where human activities are the most intense. These can have dramatic

34   impacts on natural populations. For example, the vaquita lost 90% of its population between

1 2011 and 2016 leaving about 30 individuals in 2017 (Thomas et al., 2017), and less than 19 in
2 2019 (Jaramillo Legorreta et al., 2019). This species is on the brink of extinction and currently
3 represents the most endangered marine mammal. Finless porpoise also faces major
4 conservation issues, especially the lineage within the Yangtze River (*N. a. asiaeorientalis*) in
5 China, also on the brink of extinction due to human activities (Wang et al., 2013; Wang and
6 Reeves, 2017). Similarly, several populations of harbor porpoises are highly threatened (Birkun
7 and Frantzis, 2008; Read et al., 2012). Little information about the spectacled and Burmeister's
8 porpoises is available. While anthropogenic activities are an undeniable driver of the current
9 threats to biodiversity, the evolutionary context should also be considered when assessing their
10 vulnerability (Dufresnes et al., 2013). Such knowledge is useful to characterize population
11 dynamics, identify evolutionary significant units relevant for conservation, recent or historical
12 split related to environmental variation, evolutionary or demographic trends, and evolutionary
13 processes that could explain or enhance or mitigate the current threats experienced by species
14 (Malaney and Cook, 2013; Moritz and Potter, 2013).

15   Porpoise evolutionary history and biogeography remains contentious and superficial
16 (Fajardo-Mellor et al., 2006). Previous phylogenetic studies led to incongruent results, as there
17 are disagreements regarding some of their relationships, in particular about the position of the
18 vaquita, Dall's, Burmeister's and spectacled porpoises (Barnes, 1985; Rosel et al., 1995b;
19 Fajardo-Mellor et al., 2006; McGowen et al., 2009). So far, molecular phylogenetic
20 relationships among porpoises have been estimated using short sequences of the D-loop and
21 cytochrome *b* (McGowen et al., 2009; Rosel et al., 1995b). However, the D-loop can be
22 impacted by high levels of homoplasy that blurs the resolution of the tree (Torroni et al., 2006)
23 and the cytochrome *b* may have limited power to differentiate closely related taxa (Viricel and
24 Rosel, 2011).

25   In this study, we sequenced and assembled the whole mitogenome from all extant
26 porpoise species, including most of the known lineages within species to resolve their
27 phylogenetic relationships and reconstruct their evolutionary history. More specifically, (1) we
28 assessed the phylogenetic and phylogeographic history of the porpoise family based on the
29 whole mitogenomes including the timing and tempo of evolution among lineages; (2) we
30 assessed the genetic diversity among species and lineages and (3) reconstructed the
31 demographic history for some lineages for which the sample size was suitable. (4) We placed
32 the evolutionary profile drawn for each lineage and species into the framework of past
33 environmental changes to extend our understanding of porpoise biogeography. Finally, (5) we
34 interpreted the IUCN conservation status of each taxa in the light of their evolutionary history.

2. Material & methods

2.1. Data collection

Porpoise tissue samples from 56 live-biopsy or dead stranded animals (Table S1) were stored in salt-saturated 20% DMSO or 95% Ethanol and stored at -20°C until analyses. Genomic DNA was extracted from tissues using the *PureGene* or *DNeasy* Tissue kits (Qiagen), following the manufacturer's recommendations. DNA quality and quantity were assessed on an agarose gel stained with ethidium bromide, as well as using a Qubit-v3 fluorometer. Genomic libraries for 44 porpoise samples including 3 spectacled porpoises, 3 Burmeister's porpoises, 12 vaquita, 6 Dall's porpoises, 3 East Asian finless porpoises, 2 Yangtze finless, 1 Indo-Pacific finless and 14 North Pacific harbor porpoises. Libraries were prepared by Swift Biosciences Inc. using either the Swift Biosciences Accel-NGS double-strand 2S (harbor porpoises) or single-strand 1S Plus DNA Library Kit (all other species), following the user protocol and targeting an insert-size of ~350 base-pairs (bps). The libraries were indexed and pooled in groups of 2 to 12 for paired-end 100 bps sequencing in five batches on an Illumina MiSeq sequencer at Swift Biosciences. Additional libraries for 12 samples of North Atlantic harbor porpoises were prepared at BGI Inc. using their proprietary protocol, indexed and pooled for paired-end 150 bps sequencing on one lane of HiSeq-4000 at BGI Inc. The total sequencing effort produced reads for 56 individuals (Table S2). Previously published reads from one additional finless porpoise sequenced with a Hiseq-2000 (Yim et al., 2014) were retrieved from NCBI (Table S2). For this individual, we down-sampled the raw FASTQ files to extract 0.5% of the total reads and used the 5,214,672 first reads to assemble the mitogenomes.

2.2. Data cleaning

The quality of the reads was first evaluated for each sample using *FastQC* v.0.11.5 (Andrews, 2010). *Trimmomatic* v.0.36 (Bolger et al., 2014) was used to remove low quality regions, overrepresented sequences and Illumina adapters. Different filters were applied according to the type of Illumina platform used (see Text S1 for details). Only mated reads that passed *Trimmomatic* quality filters were used for the subsequent analyses.

2.3. Mitogenome assembly

Porpoises mitogenomes assemblies were reconstructed using two different approaches. First, we used *Geneious* v.8.1.8 (Kearse et al., 2012) to perform a direct read mapping to the reference

1   mitogenome of the harbor porpoise (accession number AJ554063; Arnason et al., 2004), with

2   a minimum mapping quality of 20 (see Fig. S1 for details). This step was followed by a

3   reconstruction of the consensus sequences (Table S2). The second approach implemented in

4   *MITOBIM* (Hahn et al., 2013) is an hybrid approach combining a baiting and iterative

5   elongation procedure to perform a *de-novo* assembly of the mitogenome (see details in Text

6   S2). We visually compared the assemblies provided by the two methods in *Geneious* to assess

7   and resolve inconsistencies (Text S2 and Table S2).

8

9   In addition to the 57 assembled individuals, we retrieved six porpoise mitogenomes from

10 Genbank (Table S2). We also added eight complete mitogenomes from four outgroup species:

11 one narwhal (*Monodon monoceros*) (Arnason et al., 2004)*,* three bottlenose dolphins *(Tursiops*

12 *truncatus*) (Moura et al., 2013)*,* one Burrunan dolphin (*Tursiops australis*) (Moura et al., 2013)

13 and three orcas (*Orcinus orca*) (Morin et al., 2010)*.*

14

2.4. Sequences alignments

16 We performed the alignment of the 71 mitogenomes with *MUSCLE* (Edgar, 2004) using default

17 settings in *Geneious*. A highly repetitive region of 226 bps in the D-loop was excluded from the

18 final alignment (from position 15,508 to 15,734) because it was poorly assembled, and included

19 many gaps and ambiguities. We manually annotated the protein-coding genes (CDS), tRNA and

20 rRNA of the final alignment based on a published harbor porpoise mitogenome (Arnason et al.,

21 2004). Contrary to the remaining CDSs, ND6 is transcribed from the opposite strand (Clayton,

22 2000). Therefore, to assign the codon positions in this gene, we wrote a custom script to reverse

23 complement ND6 in the inputs of all the analyses that separates coding and non-coding regions

24 of the mitogenomes. This led to a 17 bps longer alignment due to the overlapping position of

25 ND5 and ND6.

26

2.5. Phylogenetic relationships

28 We estimated the phylogenetic relationships among the assembled mtDNA sequences using 3

29 approaches: a maximum-likelihood method in *PHYML* v3.0 (Guindon et al., 2010) (ML); a

30 distance based method using the Neighbour-Joining algorithm (NJ) in *Geneious*; and an

31 unconstrained branch length Bayesian phylogenetic tree (BI) in *MrBayes* v3.2.6 (Ronquist et

32 al., 2012). We used the Bayesian information criterion (BIC) to select the substitution model

33 that best fits the data in *jModelTest2* v2.1.10 (Darriba et al., 2012). The best substitution model

1   and parameters were used in the ML, NJ and BI approaches. For the ML approach, we fixed

2   the proportion of invariable sites and the gamma distribution parameters to the values estimated

3   by *jModelTest2*. The robustness of the ML and NJ tree at each node was assessed using 10,000

4   bootstrap replicates. For the Bayesian inference, a total of 1×10^6 MCMC iterations was run after

5   a burn-in of 1×10^5 steps, recording values with a thinning of 2,000. We performed 10

6   independent replicates and checked the consistency of the results. Runs were considered to have

7   properly explored the parameter space if the effective sample sizes (ESS) for each parameters

8   was greater than 200 and by visually inspecting the trace plot for each parameter using *Tracer*

9   v1.6 (Rambaut et al., 2014). We assessed the statistical robustness and the reliability of the

10   Bayesian tree topology using the posterior probability at each node.

11   Phylogenetic inference can be influenced by the way the data is partitioned and by the

12   substitution model used (Kainer and Lanfear, 2015). Therefore, in the ML analysis, besides

13   trying to find the best substitution model for the complete mitogenome using *jModelTest2,* we

14   also tested a partitioned scheme of the mitogenomes into 42 subsets using *PARTITIONFINDER*

15   v1.1.0 (Lanfear et al., 2012) and search for the optimal substitution model for each partition

16   defined. The partition scheme included (1) one partition per codon position for each CDS

17   (3x13=39 partitions in total); (2) one partition for the non-coding regions; (3) one partition for

18   all rRNAs genes; and (4) one partition for all tRNAs genes. We used the greedy algorithm of

19   *PARTITIONFINDER*, the *RAxML* set of substitution models and the BIC to compare the fit of

20   the different models. ML tree was reconstructed using the results of *PARTITIONFINDER* in

21   *RAxML* v8 (Stamatakis, 2014). The robustness of partitioned ML tree was assessed using

22   10,000 rapid bootstraps (Stamatakis et al., 2008).

23   Finally, the four phylogenetic trees were rooted with the eight outgroup sequences and

24   plotted using the R package *ggtree* v1.4 (Yu et al., 2018).

25

26   **2.6. Divergence time estimate**

27   We estimated the divergence time of the different lineages using a time-calibrated Bayesian

28   phylogenetic analysis implemented in *BEAST* v2.4.3 (Bouckaert et al., 2014). We assumed two

29   calibration points in the tree: (1) the split between the Monodontidae and Phocoenidae, node

30   calibrated with a lognormal mean age at 2.74 Myr (McGowen et al., 2009) (sd=0.15) as a prior

31   and (2) the split between the Pacific and Atlantic harbor porpoise lineages, node calibrated with

32   a Uniform distribution between 0.7 and 0.9 Myr as a prior (Tolley and Rosel, 2006).

33   Divergence times were estimated using a relaxed log-normal molecular clock model for

34   which we set the parameters *ucldMean* and *ucldStdev* to exponential distributions with a mean

9

1   of 1 and 0.3337, respectively. We used a Yule speciation tree model and fixed the mean of the

2   speciation rate parameter to 1. The *BIC* was used in *jModelTest2* to identify the substitution

3   model best fitting to the data, using the empirical base frequencies. We assumed a substitution

4   rate of $5x10^{-8}$ substitutions per-site and per-year (Nabholz et al., 2007). A total of $1.2x10^9$

5   MCMC iterations were run after a burn-in length of $1.2x10^8$ iterations, with a thinning of 5,000

6   iterations. We performed eight independent replicates and checked for the consistency of the

7   results among replicates. A run was considered as having converged if the ESS values were

8   greater than 200, and if they produced consistent trace plots using *Tracer* v1.6. Subsequently,

9   we combined all runs together after a burn-in of 98% using *LogCombiner* (Bouckaert et al.,

10   2014). The best supported tree topology was the one with the highest product of clade posterior

11   probabilities across all nodes (maximum clade credibility tree), estimated using *TreeAnnotator*

12   (Bouckaert et al., 2014). We also calculated the 95% highest posterior density for the node ages

13   using *TreeAnnotator*. The final chronogram was rooted with the eight outgroups sequences and

14   plotted using *FigTree* v.1.4.3 (Rambaut and Drummond, 2012).

15

2.7. Genetic diversity within species and subspecies

17   We subdivided each species into their distinct lineages in order to compare genetic diversity at

18   the different taxonomic level. Specifically, we divided the harbor porpoise into five subgroups,

19   North Pacific (*P. p. vomerina*), Black Sea (*P. p. relicta*), Mauritanian, Iberian (*P.p.*

20   *meridionalis*) and North Atlantic (*P. p. phocoena*) in accordance with the subdivisions proposed

21   for this species in the literature (Fontaine, 2016). Finless porpoise was split into Indo-Pacific

22   finless (*N. phocaenoides;* IPF), East Asian finless (*N. a. sunameri*; EAF) and Yangtze finless

23   porpoises (*N. a. asiaeorientalis*; YFP). Additionally, we subdivided the other groups into

24   lineages that were as divergent or more divergent than the sub-species that were described in

25   the literature. This includes splitting the North Pacific harbor porpoises into NP1 and NP2,

26   Dall's porpoises into DP1 and DP2 and spectacled porpoises into SP1 and SP2 to reflect the

27   deep divergence observed in the phylogenetic tree within these three lineages (Fig. 1b and S2).

28   For each species and subgroup, several summary statistics capturing different aspects of

29   the genetic diversity were calculated for different partition of the mitogenome, including the

30   whole mitogenomes, the non-coding regions (i.e. inter-gene regions and D-loop) and the CDSs

31   (i.e. 13 protein coding genes excluding tRNAs and rRNA). The number of polymorphic sites,

32   nucleotide diversity (π), number of singletons, number of shared polymorphisms, number of

33   haplotypes, haplotype diversity and Watterson estimator of θ were calculated. For CDSs, we

34   also estimated the number of synonymous (*#Syn*) and non-synonymous mutations (*#NSyn*), π

based on synonymous (π_S) and non-synonymous mutations (π_N), and the ratio π_N/π_S. All these statistics were computed in *DnaSP* v.5.10.01 (Librado and Rozas, 2009). Since we only have a unique sample for IPF, DP1 and SP1 we did not estimate these statistics for these lineages.

Differences in sample sizes can potentially influence some of these statistics. As our sample size ranged from three to 26 individuals per group, we used a rarefaction technique (Sanders, 1968) to account for the differences in sample size. We assumed a sample size of three individuals in order to compare the genetic diversity among lineages that have different sample sizes. For each lineage, we randomly subsampled 2,500 times a pool of three sequences and estimated the median, mean and 95% confidence interval for π.

2.8. Test for selective neutrality

We tested for evidence of natural selection acting on the mitogenomes using a McDonald-Kreitman test (MK-tests) (McDonald and Kreitman, 1991). This test infers deviation from the neutral theory by comparing the ratio of divergence between species (d_N/d_S) *versus* polymorphism within species (π_N/π_S) at synonymous (silent) and non-synonymous (amino acid-changing) sites in protein coding regions using the so-called neutrality index (*NI*). *NI* is 1 when evolution is neutral, greater than 1 under purifying selection, and less than 1 in the case of adaptation. MK-tests were conducted on the 13 CDSs regions of the mitogenome using *DnaSP*. We conducted this test in two different ways: first comparing all the interspecific lineages to a same outgroup, the killer whale, and second comparing all interspecific lineages to each other in order to assess how the MK-tests were affected by the outgroup choice. The significance of the *NI* values was evaluated using a G-tests in *DnaSP*. Furthermore, the distribution of *NI* values for each lineage were compared among each other using a PERMANOVA with the R package *RVAideMemoire* (Hervé, 2019). Pairwise comparisons were assessed with a permutation tests and were adjusted for multiple comparisons using the false rate discovery method (Benjamini and Hochberg, 1995). The PERMANOVA and pairwise comparisons were conducted using 9999 permutations. The neutral theory predicts that the efficacy of purifying selection increases with *Ne* (Kimura, 1983). Under these assumptions, *Ne* is expected to be proportional to *NI* (Hughes, 2008; Phifer-Rixey et al., 2012). To test this hypothesis, we assessed the correlation between *NI*s and π derived by rarefaction as a proxy of *Ne*.

In addition to the MK-tests, we quantified the branch-specific non-synonymous to synonymous substitution ratios (d_N/d_S) to infer direction and magnitude of natural selection along porpoise phylogenetic tree. To estimate the branch-specific ratio we first counted the number of synonymous (*#S*) and non-synonymous *(#NS)* substitutions for the 13 CDSs. Then

1 *#S* and *#NS* were mapped onto a tree using the mapping procedure of Romiguier et al. (2012).

2 Next, we divided *#S* and *#N* by the number of synonymous and nonsynonymous sites to obtain

3 an approximation of d_S and d_N, respectively. More specifically, we first fitted the YN98 codon

4 model using the *BPPML* program (Dutheil and Boussau, 2008), then we mapped the estimated

5 d_N/d_S values onto the branches of the ML tree using the program *MAPNH* of the *TESTNH*

6 package v1.3.0 (Dutheil et al., 2012). Extreme d_N/d_S ratio (> 3) are often due to branches with

7 very few substitution (d_N or d_S) (Romiguier et al., 2012) and were discarded. We compared the

8 distribution of d_N/d_S among species (*i.e.*, across all the branches) using a PERMANOVA.

9 Finally, the estimated ratios were correlated with π obtained by rarefaction using a Pearson's

10 correlation tests in R (R Core Team, 2018). To do so, we pooled the signal from each lineage

11 as a single data point as suggest by Figuet et al. (2014). We considered the intraspecific and

12 interspecific lineages, except for those where no non-synonymous substitutions were observed

13 (ex. NP2). Within a lineage, π was summarized as the mean of the $\log_{10}$-transformed value of

14 its representatives and the d_N/d_S was obtained by summing the non-synonymous and

15 synonymous substitution counts across its branches and calculating the ratio (Figuet et al.,

16 2014).

17

2.9. Inference of demographic changes

19 To conduct reliable demographic inferences, we investigated changes in *Ne* through time for

20 the lineages that included a sample size ≥ 10 to conduct reliable demographic inferences. This

21 includes the vaquitas and North Pacific harbor porpoise lineage NP1. We first tested for

22 deviations from neutral model expectations using three statistics indicative of historical

23 population size changes: Tajima's *D* (Tajima, 1989), Fu and Li's *D** and *F** (Fu and Li, 1993)

24 in *DnaSP*. The *p*-values were assessed using coalescent simulations in *DnaSP* using a two tailed

25 test as described in Tajima (1989). We then reconstructed the mismatch distributions indicative

26 of population size changes using *Arlequin* v.3.5.2.2 (Excoffier and Lischer, 2010). Mismatch

27 distribution were generated under a constant population size model and a sudden

28 growth/decline population model (Schneider and Excoffier, 1999). This later model is based on

29 three parameters: the population genetic diversity before the population change (θ_i); the

30 population genetic diversity after the population change (θ_f), and the date of the change of

31 population size in mutational units ($\tau = 2.\mu.t$, where μ is the mutation rate per sequence and

32 generation and *t* is the time in generations). These parameters were estimated in *Arlequin* using

33 a generalized non-linear least-square approach. The 95% confidence interval was obtained

1   using 10,000 parametric bootstraps (Schneider and Excoffier, 1999). Finally, we used the

2   coalescence based Bayesian Skyline Plot (BSP) (Drummond et al., 2005) to estimates

3   demographic changes in *Ne* back to the T_{MRCA}. BSP analysis was performed in *BEAST* v2.4.3

4   using the empirical bases frequencies and a strict molecular clock. We applied *jModelTest2*

5   separately to both lineages to evaluate the best substitution models. We assumed a substitution

6   rate of 5×10^{-8} substitutions per site and per year (Nabholz et al., 2007) in order to obtain the

7   time estimates in years. We conducted a total of 10 independent runs of 10^8 MCMC iterations

8   following a burn-in of 1×10^7 iterations, logging every 3000 iterations. We constrained *Ne*

9   between 0 and 150,000 individuals and between 0 and 380,000 individuals for the vaquita and

10   the NP1 harbor porpoise lineage, respectively. This upper boundary on *Ne* was empirically set

11   to encompass the entire marginal posterior distribution. All other parameters were kept at

12   default values. The convergence of the analysis was assessed by checking the consistency of

13   the results over 10 independent runs. For each run, we also used *Tracer* to inspect the trace plots

14   for each parameter to assess the mixing properties of the MCMCs and estimate the ESS value

15   for each parameter. Run were considered as having converged if they displayed good mixing

16   properties and if the ESS values for all parameters was greater than 200. We discarded the first

17   10% of the MCMC steps as a burn-in and obtained the marginal posterior parameter

18   distributions from the remaining steps using *Tracer*. *Ne* values were obtained by assuming a

19   generation time of 10 years (Birkun and Frantzis, 2008).

20

3. Results

22

3.1. Porpoise mitogenomes assemblies

24   A total of 57 mitogenomes of the seven species of porpoise (Table S1) were newly sequenced

25   and assembled using Illumina sequencing. After read quality check, trimming, and filtering

26   (Table 1 and S2), between 1,726,422 and 67,442,586 cleaned reads per sample were used to

27   assemble the mitogenomes. The two methods used to assemble the mitogenomes delivered

28   consistent assemblies with an average depth coverage ranging from 15 to 4,371X for *Geneious*

29   and 17 to 4,360X for *MITOBIM* (Table 1 and S2). In total, 35 of the 57 mitogenome assemblies

30   were strictly identical with both methods. The 22 remaining assemblies included between 1 and

31   4 inconsistencies which were visually inspected and resolved (Text S2 and Table S2).

32   Augmented with the 14 previously published mitogenome sequences, the final alignment

33   counted 71 mitogenome sequences of 16,302 bps and included less than 0.2% of missing data.

34   The alignment was composed of 627 singletons and 3,937 parsimony informative sites defining

1   68 haplotypes (including the outgroup sequences). Within the 63 ingroup porpoise sequences,

2   we observed 2,947 segregating sites, including 242 singletons and 2,705 shared polymorphisms

3   defining 58 distinct haplotypes with a haplotype diversity of 99.6%.

4

3.2. Phylogenetic history of the porpoises

6   A Hasegawa-Kishino-Yano (HKY+G, Gamma=0.19) model was selected as the best-fitting

7   nucleotide substitution model. Phylogenetic inferences using a maximum-likelihood (ML)

8   approach based on the whole mitogenome (Fig. 1b and Fig. S2a) or on a partitioned data set

9   (Fig. S2b), a distance-based neighbor-joining method (Fig. S2c), and a Bayesian approach (Fig.

10   S2d) all produced concordant phylogenies (i.e., similar topologies and statistical supports). All

11   phylogenies were fully resolved with high statistical support at the inter- and intra-specific

12   levels (bootstrap supports ≥93% or node posterior probability of one). One exception was the

13   node 5 grouping the Burmeister's and spectacled porpoises in the neighbor-joining tree (Fig.

14   S2b) as it displayed a slightly lower bootstrap support of 61% (red star in Fig. 1b and S2b), but

15   it was fully supported by the other approaches (Fig. 1b and S2).

16   The resulting phylogenetic reconstruction (Fig. 1b) showed that all porpoises formed a

17   monophyletic group (node 1). The most basal divergence in the tree split the more tropical

18   finless porpoises from the other temperate to subpolar porpoises. Then, the temperate species

19   split into two clades (node 2) composed of two reciprocally monophyletic groups. The first is

20   composed of the southern hemisphere species (spectacled and Burmeister's porpoises) and

21   vaquitas (node 3). The second is composed of the porpoises from the northern hemisphere

22   (harbor and Dall's porpoises, node 4). In contrast with a previous phylogenetic study based on

23   control region sequences (Rosel et al., 1995b), the phylogenetic tree based on the whole

24   mitogenome suggested that vaquitas split from a common ancestor to the spectacled and

25   Burmeister's porpoises (node 3), and thus that the two species from the southern hemisphere

26   are more closely related to each other (node 5) than they are to vaquitas. Finally, the

27   mitogenome tree supported the monophyly of each recognized species (nodes 6-11).

28   Intraspecific subdivisions were also evident from the mitogenome phylogeny in some

29   species, such as in the harbor and finless porpoises (Fig 1b). In the harbor porpoises, the split

30   between the North Atlantic and North Pacific subspecies constituted the deepest divergence of

31   all intraspecific subdivisions across all species. Within the North Atlantic harbor porpoises,

32   further subdivisions were also observed and corresponded to the three previously described

33   ecotypes or sub-species (Fontaine, 2016). These included the relict population in the Black Sea,

34   the harbor porpoises from the upwelling waters with two closely related but distinct lineages in

14

1   Iberia and Mauritania, and the continental shelf porpoises further north in the North Atlantic.

2   Finally, within in the North Pacific, two cryptic subgroups were also observed (NP1 and NP2;

3   Fig. 1b). Among the finless porpoises, the three taxonomic groups currently recognized (Zhou

4   et al., 2018), including IPF and the two closely related species of narrow-ridged finless

5   porpoises, were clearly distinct from each other on the mitogenome phylogenetic tree (node

6   11). Finally, despite a limited sampling, Dall's (node 7) and spectacled porpoises (node 10) each

7   also displayed distinct lineages (DP1/DP2 and SP1/SP2, respectively; Fig. 1b and S2) as

8   divergent as those observed in the harbor and finless porpoises.

9   The time-calibrated Bayesian mitochondrial phylogeny (Fig. 2 and Table S5) suggested

10   that all extant porpoises find their common ancestor ~5.42M years ago (95% Highest Posterior

11   Density, HPD, 4.24-6.89; node 1). This time corresponds to the split between the finless and

12   the other porpoise species. Spectacled, vaquita and Burmeister's porpoises diverged from

13   harbor and Dall's porpoise ~4.06 Myr ago (95% HPD, 3.15-5.12; node 2). The split between

14   vaquitas, spectacled and Burmeister's porpoises was estimated at ~2.39 Myr (95% HPD, 1.74-

15   3.19; node 3) and between spectacled and Burmeister's at ~2.14 Myr (95% HPD, 1.51-2.86;

16   node 4). The Dall's and harbor porpoises split from each other ~3.12 Myr ago (95% HPD, 2.31-

17   3.98; node 5). Finally, the common ancestor of the subdivisions observed within each species

18   was dated within the last million years (nodes 6-11).

19

20   **Genetic diversity of the mitogenome.** Mitochondrial genetic diversity varied greatly among

21   species (Table 2 and Fig. 3). The highest values of π were observed in the harbor porpoises

22   (π=1.15%), followed by the spectacled (π=0.60%), Dall's (π=0.50%), finless (π=0.35%), and

23   Burmeister's porpoises (π=0.13%), while vaquitas displayed the lowest values (π=0.02%). The

24   variation among species was strongly related to the occurrence of distinct mitochondrial

25   lineages within species that corresponds to ecologically and genetically distinct sub-species or

26   ecotypes (Table 2 and Fig. 3). Once the lineages that included more than three sequences were

27   compared to each other while accounting for the difference in sample size using a rarefaction

28   procedure (Sanders, 1968) (Fig. 3), π values were more homogeneous among lineages and

29   species, with however some variation. The most diversified lineages included DP2 in Dall's

30   porpoise (π=0.32%) and the North Atlantic (NAT) lineage in harbor porpoise (π=0.33%). In

31   contrast, the vaquita lineage (π=0.02%), harbor porpoises North Pacific lineages NP2

32   (π=0.02%) and Black Sea (BS) (π=0.07%), and the Yangtze finless porpoise YFP lineage

33   (π=0.06%) displayed the lowest nucleotide diversity. The other lineages displayed intermediate

1   π values.

2

3   **3.3. Molecular evolution of the mitogenome**

4   The nucleotide diversity also varied greatly along the mitogenome. It was lowest in the origin

5   of replication, tRNA and rRNAs, intermediate in the coding regions and highest in non-coding

6   regions (Fig. S3). This result indicates different levels of molecular constraints along the

7   mitogenome.

8   The π_N/π_S ratio in the 13 CDSs, displaying the relative proportion of non-synonymous

9   *versus* synonymous nucleotide diversity, was lower than one in all the lineages. This is

10   consistent with purifying selection acting on the coding regions. At the species level, the ratio

11   π_N/π_S ranged from 0.04 in Dall's and spectacled to 0.10 in finless porpoises (Table S3). The

12   vaquita displayed an intermediary value of 0.06. Within each species, π_N/π_S also varied

13   markedly across lineages: in the harbor porpoise, π_N/π_S ratios ranged from 0 in the North Pacific

14   NP2 lineage to 0.21 in the Black Sea BS lineage; in the finless porpoises from 0.14 in EAF to

15   0.17 in YFP; 0.05 in the DP2 Dall's porpoise lineage; and 0.06 in SP2 spectacled porpoise

16   lineage (Table S3).

17   The branch-specific non-synonymous to synonymous substitution rate (d_N/d_S, Fig. S4a)

18   were fairly conserved across the phylogenetic tree and ranged from 0 in the finless porpoise to

19   1.08 in the harbor porpoise, with a median value at 0.12. A d_N/d_S lower than one implies

20   purifying selection. Thus, similar to π_N/π_S, the branch-specific d_N/d_S suggested that the porpoise

21   mitogenomes were mostly influenced by purifying selection. Furthermore, the d_N/d_S ratios did

22   not differ significantly among species (PERMANOVA, *p*-value=0.49). Interestingly, the d_N/d_S

23   ratio was negatively correlated with the nucleotide diversity (Fig. S4b; Pearson's r = -0.64, *p*-

24   value = 0.01) suggesting that purifying selection removes deleterious mutations more

25   effectively in genetically more diversified lineages.

26   The Mcdonald-Kreitman (MK) tests using first the orca as an outgroup showed that all

27   the lineages for each species had neutrality index (*NI)* values greater or equal to one (Fig. 4a).

28   In particular, some lineages displayed *NI* values significantly higher than one (G-tests, *p*-

29   value<0.05), consistent with a signal of purifying selection. These included the EAF and YFP

30   lineages in the finless porpoises and the MA, IB and BS in the harbor porpoises (Fig. 4a).

31   Vaquitas and NP2 harbor porpoise lineages also displayed marginally significant *NI* values

32   (*NI* > 2, *p*-value≤0.10; Fig. 4a). The remaining lineages showed *NI*s close to one suggestive of

33   selective neutrality. The MK tests applied to all pairs of interspecific lineages showed *NI* values

34   often higher than one (Fig. 4b and Fig. S5a). The values were especially high (Fig. S5a) and

16

1   significant (Fig. S5b) when comparing the harbor porpoise lineages (MA, IB, and BS) with the

2   finless porpoise linages (YFP and EAF). The variation in the distribution of *NI* among

3   interspecific lineages (Fig. 4b) showed that these same lineages displayed significantly larger

4   *NI* values compared to spectacled SP2 and Dall's DP2 porpoise lineages (PERMANOVA, *p*-

5   value<0.001 and all pairwise comparisons have a *p*-value<0.04 after False Rate discovery

6   adjustment). We observed a significant negative correlation between π and *NI* (Pearson's *r* = -

7   0.28, *p*-value=0.003), suggesting that purifying selection is stronger in lineages with small *Ne*

8   or that demographic events impacted the polymorphism of these lineages.

9

10 **3.4. Demographic history**

11 The vaquita displayed significant departure from neutral constant population size expectations

12 with significant negative values for Fu and Li's *D** and *F**, but not Tajima's *D* (Table 3). This

13 result indicates a significant excess of singleton mutations compare to a neutral expectation,

14 consistent with a bottleneck or a selective sweep. In contrast, the harbor porpoise NP1 lineage

15 did not show any such deviation, even though all the statistics displayed negative values.

16

17 The mismatch distribution and the coalescent-based BSP also captured this contrast (Fig. 5).

18 For the North Pacific harbor porpoise NP1 lineage, the mismatch distribution was consistent

19 with an ancient population expansion (Fig. 5a) with a modal value close to 20 differences on

20 average between pairs of sequences. Despite the ragged distribution and large 95% CI, the best

21 fitted model suggested an ancient increase in genetic diversity ($\theta=2.Ne.\mu$) by a factor of 40

22 after a period ($\tau=2.t.\mu$) of 18 units of mutational time. This old expansion was also detected by

23 the BSP analysis (Fig. 5c). Indeed, NP1 displayed an old steady increase in *Ne* with time since

24 the most recent common ancestor (T_{MRCA}) 16,166 years before present (yrs. BP) (95%CI:

25 12,080–20,743), with a median *Ne* increasing from 1,660 to 6,436 (Fig. 5c). For the vaquita

26 lineage, the mismatch distribution and the BSP analyses both supported a much more recent

27 expansion than in NP1. The mismatch distribution (Fig. 5b) showed an increase of θ by a factor

28 of 1,000 after a τ of four units of mutational time. The mode of the bell shape distribution for

29 the best fitted model was around three differences among pairs of sequences, which is consistent

30 with a recent population expansion. The BSP analysis (Fig. 5d) detected this expansion and

31 dated it back to 3,079 yrs. BP (95%CI: 1,409–5,225), with median *Ne* increasing from 613

32 (95%CI: 45-5,319) to 1,665 (95%CI: 276-9,611). Thus, the estimated current *Ne* was three to

33 six times lower than in NP1 (Fig. 5d).

4. Discussion

The phylogeny of the Phocoenidae has been debated for decades, in part due to the lack of polymorphism and statistical power that came from the analyses of short fragments of the mitochondrial genome (McGowen et al., 2009; Rosel et al., 1995b). Using massive parallel sequences technologies, the analyses of newly sequenced and assembled whole mitogenomes from all the species and subspecies of porpoises provide the first robust comprehensive picture of the evolutionary history of the porpoises. The phylogenetic relationships estimated here delivered a fully resolved evolutionary tree (Fig 1b and S2). While most of the phylogenetic relationships were suggested previously (McGowen et al., 2009; Rosel et al., 1995b), the resolution and statistical support recovered here was maximal. Our results support the monophyly and branching of each species and subspecies. Moreover, the comparative view of the mitochondrial polymorphism within and among species provides the first attempt to our knowledge to bridge macro- to micro-evolutionary processes in a cetacean group. This perspective across evolutionary time-scales can shed light on the isolation dynamics and their drivers across the speciation continuum of the Phocoenidae.

4.1. New insights into the biogeography of the Phocoenidae

The biogeographical history of cetacean species has been hypothesized to results of a succession of vicariant and dispersal events influenced by geological, oceanic, and climatic reorganization during the Late Miocene, Pliocene and early Pleistocene (McGowen et al., 2009; Steeman et al., 2009). Changes in climate, ocean structure, circulation, and marine productivity opened new ecological niches, enhanced individual dispersal and isolation, and fostered specialization to different food resources. All these factors promoted the adaptive radiation in cetaceans which led to the extant species diversity in the odontocete families (Slater et al., 2010). For example, the Monodontidae and Delphinidae are the closest relatives to the Phocenidae. They originated during the Miocene and displayed an accelerated evolution marked with the succession of speciation events during 3 Myr, leading to the extant species diversity in these groups (McGowen et al., 2009; Steeman et al., 2009). The time calibrated phylogeny of the Phocoenidae (Fig. 2) suggests that porpoises also diversified following similar processes during the late Miocene until the early-Pleistocene (between 6 and 2 Myr). This timing is about 2 to 3 Myr more recent than those proposed by McGowen et al. (2009) and Steeman et al. (2009). The increase of the genetic information, node calibrations and number

1 of sequences per species are known to influence phylogenetic inferences and the divergence

2 time estimates (Duchêne et al., 2011; Zheng and Wiens, 2015). The use of complete

3 mitogenomes, two node calibrations (instead of one) and several sequences per species in our

4 study likely explain the difference compare to previous studies.

5   Consistent with previous findings (Fajardo-Mellor et al., 2006; McGowen et al., 2009;

6 Rosel et al., 1995b), the finless porpoises were the first species to split among the Phocoenidae.

7 As the vast majority of the porpoise fossils found so far come from tropical or subtropical

8 regions (Fajardo-Mellor et al., 2006), and considering their current affinity for warm waters,

9 the finless porpoises seem to be the last members of a group of porpoise species that adapted

10 primarily to tropical waters. Interestingly, finless porpoises further diversified and colonized

11 more temperate waters of the Yellow Sea and Sea of Japan (Fig. 1a). The five remaining

12 porpoise species diverged ~4.0 Myr ago and all but the vaquita occupy temperate regions with

13 an antitropical distribution (Fig. 1a). Harbor and Dall's porpoises inhabit the cold water of the

14 Northern hemisphere whereas spectacled and Burmeister's porpoises are found in the Southern

15 hemisphere. This result is consistent with the hypothesis that antitropically distributed

16 cetaceans have evolved with the deep environmental changes that occurred during the late

17 Pleistocene and as a response to the fluctuations in surface water temperatures in the tropics,

18 concomitant with the changes in oceanographic currents, marine productivity, and feeding areas

19 (Barnes, 1985; Davies, 1963; Pastene et al., 2007). About 3.2 Myr ago, the formation of the

20 Panama Isthmus altered the tropical currents and water temperatures in coastal regions of the

21 Pacific. It promoted the dispersion of numerous taxa from the Northern to the Southern

22 Hemisphere (Lindberg, 1991). This process is also a plausible driver that led to the antitropical

23 distribution of the modern porpoises (Fajardo-Mellor et al., 2006).

24   During the porpoises' evolutionary history, a symmetric evolution took place

25 approximatively at the same time in both hemispheres resulting in analogous ecological

26 adaptations. In the Northern hemisphere, the split between Dall's and harbor porpoises ~3.1

27 Myr ago led to offshore *versus* coastal specialized species, respectively. This pattern was

28 mirrored in the Southern hemisphere with the split between the spectacled and Burmeister's

29 porpoises ~2.1 Myr ago which led also to the divergence between offshore *versus* coastal

30 specialized species. Such a symmetric habitat specialization likely reflects similar ecological

31 opportunities that opened in both hemispheres and triggered a convergent adaptation in the

32 porpoises. Interestingly, this parallel offshore evolution observed in Dall's and spectacled

33 porpoises was accompanied by a convergent highly contrasted countershading coloration

19

1 pattern with a white ventral side and black dorsal back side in both species. This color pattern

2 is thought to be an adaptation to the offshore environment serving as camouflage for prey and

3 predators (Perrin, 2009).

4 Resources and diet specializations are known to be a major driver of cetacean evolution

5 as their radiation is linked to the colonization of new vacant ecological niches in response to

6 past changes (Slater et al., 2010). As small endothermic predators with elevated energetic needs,

7 limited energy storage capacity and a rapid reproductive cycle, porpoises are known for their

8 strong dependency on food availability (Hoekendijk et al., 2017; Koopman et al., 2002). These

9 characteristics reinforce the hypothesis that their adaptive radiation has been strongly shaped

10 by historical variation in food resources and should also be visible at the intraspecific level.

11

4.2. Porpoises phylogeography and microevolutionary processes

13 The processes shaping the evolution of porpoises at the macro-evolutionary time scale

14 find their origins at the intraspecific level (micro-evolutionary scale), with the split of multiple

15 lineages within species that may or may not evolve into fully distinct and reproductively

16 isolated species. We showed that all lineages forming the intraspecific subdivisions (sub-

17 species or ecotypes) were all monophyletic groups. All these distinct lineages found their most

18 recent common ancestor within the last million years. These results corroborate previous

19 phylogeographic studies suggesting that intraspecific subdivisions observed in many porpoise

20 species were mediated by environmental changes during the last glacial cycles of the

21 Quaternary (Chen et al., 2017; Fontaine, 2016; Hayano et al., 2003; Zhou et al., 2018). The Last

22 Glacial Maximum (LGM) and the subsequent ice retreat have profoundly shaped the

23 phylogeographic patterns of many organisms, leaving behind multiple divergent lineages in

24 many cetacean taxa that are vestiges of past environmental variations (Chen et al., 2017;

25 Fontaine, 2016; Foote et al., 2016; Louis et al., 2014). Adaptive evolution to different niches in

26 response to past changes associated with variation in marine sea ice, primary production, and

27 redistribution of feeding areas led to intraspecific divergence in many cetaceans in terms of

28 genetics, behavior, morphology, and geographical distribution. For example, the specialization

29 between coastal *versus* offshore ecotypes in bottlenose dolphins has been dated back to the

30 LGM, and explain the observed patterns of genetic and morphological differentiation (Louis et

31 al., 2014). Likewise, the behavior, size, color patterns and genetic divergence among some

32 different types of killer whales were attributed to specialization onto different food resources

33 since the LGM (Foote et al., 2016). The present study shows that analogous processes occurred

34 in each porpoise species too.

1   The finless porpoises represent probably one of the best documented case of incipient
2   speciation related to habitat specializations among the porpoises. Consistently with the results
3   of Zhou et al. (2018) based on whole genome sequencing, our mitogenome results dated the
4   radiation of the finless porpoise species within the last ~0.5 Mya, which coincide with profound
5   environmental changes associated with the Quaternary glaciations. In particular, our results are
6   congruent with the hypothesis suggesting that the diversification of the finless porpoises has
7   been driven by the elimination of the Taiwan Strait associated with the sea-level drop during
8   glacial periods (Wang et al., 2008). Indeed, at least three land bridges connected Taiwan to the
9   mainland China since the last 0.5 Myr and could have enhanced the separation between the
10   Indo-Pacific and the narrow ridged finless porpoises (Wang et al., 2008). Likewise, we dated
11   the emergence of the Yangtze finless porpoise to the last ~0.1 Myr, which is consistent with
12   previous studies suggesting that the last glacial event have strongly determined the evolutionary
13   history of this species (Chen et al., 2017; Zhou et al., 2018).

14   Similar to the finless porpoises, the harbor porpoises are also divided into several
15   lineages previously recognized as distinct sub-species. The deepest split is observed between
16   the North Pacific (*P. p. vomerina*) and North Atlantic lineages, and is deeper than the genetic
17   divergence observed between the two species of finless porpoises. The lack of shared
18   haplotypes between Pacific and Atlantic porpoises confirm previous results supporting their
19   total isolation (Rosel et al., 1995a). Their splitting time was estimated at ~0.86 Mya, which is
20   consistent with the presumed time when the North Pacific porpoises colonized the North
21   Atlantic basin (Tolley and Rosel, 2006). The two ocean basins were last in contact across the
22   Arctic approximately 0.7 to 0.9 Myr. ago, with estimated sea surface temperatures of *ca.* 0.5°C
23   (Harris, 2005), which corresponds to the lowest temperature at which harbor porpoises are
24   currently observed. Within the North Atlantic, the three known subspecies (Fontaine, 2016)
25   (i.e. *P. p. phocoena*; *P. p. meridionalis* and *P. p. relicta*) were also detected as distinct
26   monophyletic groups based on the mitogenome (Fig. 1b and S2). Their evolutionary history has
27   been strongly influenced by recent environmental changes during the LGM (reviewed in
28   Fontaine, 2016). North Pacific porpoises also showed unreported cryptic subdivisions (i.e. NP1
29   and NP2 in Fig. 1b and S2), displaying a level of divergence deeper than the one observed
30   between the Iberian and Mauritanian harbor porpoises or between the Yangtze and East Asian
31   finless porpoises. These two clades (NP1 and NP2) may also represent lineages that split during
32   the LGM. The steady increase in genetic diversity observed in the NP1 lineage since 12 kyrs
33   ago (Fig. 5c) is consistent with the end of the LGM.

1   Compared to the finless and harbor porpoises, little is known about the Dall's,
2   Burmeister's and spectacled porpoises. This is in part due to the limited number of observations
3   and access to biological data (but see Hayano et al., 2003; Pimper et al., 2012; Rosa et al.,
4   2005), especially for the spectacled porpoise. Despite these limitations, our study revealed that
5   the patterns and processes described for the finless porpoise and harbor porpoise may apply
6   also to the majority of the other porpoise species. Previous studies identified multiple
7   intraspecific subdivisions within the Dall's (Hayano et al., 2003) and Burmeister's (Rosa et al.,
8   2005) porpoises. The long branches in the phylogenetic tree (Fig. 1b) for Dall's porpoise
9   (DP1/DP2 lineages) and spectacled porpoises (SP1/SP2) imply that distinct evolutionary units
10   may also exist in these species. Furthermore, their vast distribution (Fig 1a) and divergence
11   times among lineages (Tables S5) suggest that the different lineages in these species also split
12   during the Quaternary glaciations.

13   The vaquita contrasts strikingly with the other species with its narrow distribution, the
14   smallest of all marine mammals (see Fig. 1a and Lundmark, 2007). Previous studies based on
15   a short fragment of the mitochondrial *D-loop* and *Cyt-b* identified the Burmeister's porpoise as
16   the closest relative to the vaquitas. However, the phylogenetic results reported here using the
17   whole mitogenome support that the vaquita coalesce with the ancestor of the Burmeister's and
18   spectacled porpoise with maximal support. The estimated split time of ~2.4 Myr ago between
19   vaquita and the southern porpoises is consistent with the onset of the Quaternary glaciations,
20   2.6 Myr ago (Ehlers and Gibbard, 2014). The fact that the vaquita is found in the northern
21   hemisphere, while the Burmeister's and spectacled porpoises are in the southern hemisphere
22   implies that some ancestors with cold-affinities from the southern species crossed the equator
23   in the past and became isolated long enough to become the ancestors of the extant vaquita. The
24   most parsimonious hypothesis is that the decrease in water temperature associated with a glacial
25   maximum likely allowed vaquita's ancestor to cross the equator and disperse to the Northern
26   Hemisphere (Norris and McFarland, 1958). The current vaquita representatives form thus a
27   relic population of the temperate southern species' ancestor that crossed the intertropical zone.
28   In contrast with previous mitochondrial studies that found no variation at the D-loop (Rosel and
29   Rojas-Bracho, 1999) we observed 16 sites segregating along the entire mitogenome. Among
30   them, 13 were singleton mutations. This extremely low nucleotide diversity was the lowest of
31   all porpoise lineages, as illustrated by the extremely short branches in the phylogenetic tree
32   (Fig. 1b and 2). The origin of the present mitochondrial diversity is also relatively recent, with
33   a T_{MRCA} estimated at ~20,000 to 70,000 years with the phylogenetic approach and ~1,500 to

1   5,000 years with coalescent approach. The main reason behind this discrepancy is called "time

2   dependency of molecular rates" (Ho et al., 2011). Population genetics coalescent-based

3   estimates reflect the spontaneous mutation rates, whereas phylogeny-based estimates reflect the

4   substitution rates (i.e. mutations fixed among taxa). Contrasting with these recent estimates, the

5   branch connecting to the ancestor of vaquitas and southern species dates back to ~2.4Mya. This

6   suggest that either additional lineages may have existed in the past and went extinct or that only

7   a single lineage crossed the intertropical zone. Whole genome analyses may help enlightening

8   the evolutionary history of this peculiar species.

9

10   *4.3. Genetic diversity and conservation*

11   Maintenance of genetic diversity has been considered as an important factor in conservation

12   biology. Genetic factors can contribute to the "extinction vortex" by a mutual reinforcement

13   with demographic processes speeding-up population decline and increasing their extinction risk

14   (Allendorf et al., 2012). As a consequence, ideal conservation measures should be designed to

15   maximize genetic diversity, especially through the management of evolutionary significant

16   units (ESU) (Moritz, 1994). However, conservation status does not explicitly take this

17   parameter into account since the relationship between IUCN status and genetic diversity is not

18   always straightforward (Nabholz et al., 2008). In this study, the genetic diversity of each

19   porpoise species correlates well with its IUCN status, especially when we account for

20   intraspecific subdivision. The *Critically Endangered* taxa, such as the vaquitas or the YPF

21   porpoises displayed extremely low π suggesting a low *Ne*. The *Endangered* (EAF) and

22   *Vulnerable* (Black Sea harbor porpoise) taxa displayed low to average π. *Least Concern* taxa

23   (e.g. North Atlantic harbor porpoise and DP2) exhibited higher π, suggesting larger *Ne* and/or

24   the presence of internal subdivision. This link between π and the IUCN conservation status may

25   thus provide a useful proxy to assess the conservation status of taxa for which an IUCN status

26   has not been yet established, due to data deficiency. For example, the Iberian harbor porpoise

27   population is among the marine mammals displaying the highest stranding and by-catch rates

28   reported (Read et al., 2012). The low genetic diversity reported in the present study represents

29   thus an additional signal indicating how possibly vulnerable are these Iberian porpoises. On the

30   other hand, spectacled porpoises represent currently one of the least known cetacean species.

31   Their genetic diversity is comparable to the one observed in the Dall's porpoise DP2 lineage or

32   the North Atlantic (NAT) harbor porpoise linage, suggesting these populations display large

33   *Ne*.

34   Mitochondrial diversity may not always be a good proxy of population abundance. Other

1   evolutionary processes than just demography may impact genetic diversity (i.e., such as natural

2   selection) (Bazin et al., 2006). The d_N/d_S and π_N/π_S lower than 1 highlighted in this study would

3   be usually indicative of purifying selection acting on the mitochondrial genetic diversity. The

4   negative relationships observed between π and d_N/d_S (Fig. S4b) lend further support to this

5   hypothesis, suggesting that purifying selection is more effective in large populations as

6   predicted by the neutral theory (Hughes, 2008; Phifer-Rixey et al., 2012). Surprisingly, the MK-

7   tests suggested that purifying selection was prevailing on the mitochondrial genomes of the

8   endangered porpoises with *NI* values larger than 1. In contrast, selective neutrality could not be

9   rejected for less threatened species. At first glance, this result for the endangered porpoises

10   seems counter intuitive. Purifying selection is expected to be less effective in lineages where

11   population size is very small, since genetic drift is expected to outperform selective forces

12   (Kimura, 1983). However, when a lineage harbors a low *Ne*, slightly deleterious variants are

13   expected to increase in frequency and segregate for a longer time without being fixed. Parsch

14   et al. (2009) showed that this effect will increase the number of polymorphic nonsynonymous

15   mutations compared to the divergent nonsynonymous mutations ($\pi_N \gg d_N$). Hence, despite the

16   reported increase of d_N/d_S in endangered taxa, it induces a bias in the neutrality index toward

17   positive values creating an artefactual signal of purifying selection. The higher values of π_N/π_S

18   observed in the endangered porpoise taxa (Table S3) and the negative relationships reported

19   between π and *NI* support this hypothesis. All these elements suggest that demographic

20   processes rather than selective forces drive the genetic diversity of the mitochondrial genome,

21   and led to high values of NI or d_N/d_S in the endangered taxa.

22

5. Conclusions

24

25   Using complete mitochondrial genomes, we reconstructed the most comprehensive picture of

26   the evolutionary history of the Phocoenidae to date. Besides clarifying the debated phylogenetic

27   relationships among porpoises, our results provided new insights into the process driving

28   species diversification in the porpoises across the speciation continuum. Similar to the

29   Delphinidae, the Phocoenidae recently radiated in response to past environmental variation,

30   adapting to different environments, ecological niches, and food resources. Furthermore, our

31   results suggested that the processes governing their divergence at the macro-evolutionary scale

32   find their origins at the micro-evolutionary scale. We revealed unreported genetic subdivision

33   for several taxa suggesting that our knowledge about many species, especially the data deficient

34   southern species, is still scarce. Finally, the level of mitochondrial genetic diversity within a

species seems to be primarily driven by demographic processes, rather than natural selection and turned out to be a good proxy for the conservation issues reported in these groups (i.e. Yangtze finless porpoises or vaquita). In the future, the analyses of genome-wide data and a better description of porpoise distinctive environments hold great promise to shed light on the processes underpinning their adaptation and divergence. Such knowledge can facilitate the identification of units following distinct evolutionary trajectories, which in turn can greatly improve our ability to set conservation plans to mitigate the decline of several porpoise groups.

Acknowledgments

Author's contributions

MCF designed the study with the contribution of PAM; YBC analyzed the data under the supervision of MCF, with help from AW and JR; AA, AB, MF, BLT, LRB, KMR, GAV, PAM contributed with the biological materials; JT, KMR performed the DNA extractions; CS and TH constructed the genomic libraries at Swift Bioscience and performed part of the sequencing; YBC and MCF wrote the manuscript with input and feedbacks from all the co-authors.

Declaration of interest

None.

Data accessibility

1   Mitochondrial genomes were deposited on NCBI under the accession numbers To Be

2   Announced (TBA). Data and scripts have also been deposited on Dryad doi: TBA

3

Reference

Allendorf, F.W., Luikart, G.H., Aitken, S.N., 2012. Conservation and the Genetics of Populations. John Wiley & Sons.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Arnason, U., Gullberg, A., Janke, A., 2004. Mitogenomic analyses provide new insights into cetacean origin and evolution. Gene 333, 27–34. doi:10.1016/j.gene.2004.02.010

Barnes, L.G., 1985. Evolution, taxonomy and antitropical distributions of the porpoises (Phocoenidae, Mammalia). Marine Mammal Science 1, 149–165.

Bazin, E., Glemin, S., Galtier, N., 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312, 570–572. doi:10.1126/science.1122033

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Berta, A., Sumich, J.L., Kovacs, K.M., 2015. Evolution and Geography, in: Marine Mammals: Evolutionary Biology. Elsevier, pp. 131–166. doi:10.1016/B978-0-12-397002-2.00006-5

Birkun, A.A., Jr, Frantzis, A., 2008. *Phocoena phocoena ssp. relicta*. The IUCN Red List of Threatened Species. doi:10.2305/IUCN.UK.2008.RLTS.T17030A6737111.en

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol 10, e1003537–6. doi:10.1371/journal.pcbi.1003537

Burridge, C.P., 2002. Antitropicality of Pacific fishes: molecular insights. Environmental Biology of Fishes.

Chen, M., Fontaine, M.C., Ben Chehida, Y., Zheng, J., Labbé, F., Mei, Z., Hao, Y., Wang, K., Wu, M., Zhao, Q., Wang, D., 2017. Genetic footprint of population fragmentation and contemporary collapse in a freshwater cetacean. Sci Rep 7, 14449. doi:10.1038/s41598-017-14812-z

Clayton, D.A., 2000. Transcription and replication of mitochondrial DNA. Hum. Reprod. 15 Suppl 2, 11–17.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9, 772–772. doi:10.1038/nmeth.2109

Davies, J.L., 1963. The antitropical factor in cetacean speciation. Evolution 17, 107–116. doi:10.1111/j.1558-5646.1963.tb03258.x

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192. doi:10.1093/molbev/msi103

Duchêne, S., Archer, F.I., Vilstrup, J., Caballero, S., Morin, P.A., 2011. Mitogenome Phylogenetics: The Impact of Using Single Regions and Partitioning Schemes on Topology, Substitution Rate and Divergence Time Estimation. PLoS ONE 6, e27138. doi:10.1371/journal.pone.0027138

Dufresnes, C., Wassef, J., Ghali, K., Brelsford, A., Stöck, M., Lymberakis, P., Crnobrnja-Isailovic, J., Perrin, N., 2013. Conservation phylogeography: does historical diversity contribute to regional vulnerability in European tree frogs (*Hyla arborea*)? Mol. Ecol. 22, 5669–5684. doi:10.1111/mec.12513

Dutheil, J., Boussau, B., 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. BMC Evol. Biol. 8, 255. doi:10.1186/1471-2148-8-255

Dutheil, J.Y., Galtier, N., Romiguier, J., Douzery, E.J.P., Ranwez, V., Boussau, B., 2012. Efficient selection of branch-specific models of sequence evolution. Mol. Biol. Evol. 29, 1861–1874. doi:10.1093/molbev/mss059

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797. doi:10.1093/nar/gkh340

Ehlers, J.R., Gibbard, P., 2014. Quaternary Glaciation, in: Encyclopedia of Snow, Ice and Glaciers, Encyclopedia of Earth Sciences Series. Springer Netherlands, Dordrecht, pp. 873–882. doi:10.1007/978-90-481-2642-2_423

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Fajardo-Mellor, L., Berta, A., Brownell, R.L., Boy, C.C., Natalie P Goodall, R., 2006. The Phylogenetic Relationships and Biogeography of True Porpoises (Mammalia: Phocoenidae) Based on Morphological Data. Marine Mammal Science 22, 910–932. doi:10.1111/j.1748-7692.2006.00080.x

Figuet, E., Romiguier, J., Dutheil, J.Y., Galtier, N., 2014. Mitochondrial DNA as a tool for reconstructing past life-history traits in mammals. Journal of Evolutionary Biology 27, 899–910. doi:10.1111/jeb.12361

Fontaine, M.C., 2016. Harbour Porpoises, *Phocoena phocoena*, in the Mediterranean Sea and Adjacent Regions: Biogeographic Relicts of the Last Glacial Period. Adv. Mar. Biol. 75, 333–358. doi:10.1016/bs.amb.2016.08.006

Fontaine, M.C., Roland, K., Calves, I., Austerlitz, F., Palstra, F.P., Tolley, K.A., Ryan, S., Ferreira, M., Jauniaux, T., Llavona, A., Oztürk, B., Öztürk, A.A., Ridoux, V., Rogan, E., Sequeira, M., Siebert, U., Vikingsson, G.A., Borrell, A., Michaux, J.R., Aguilar, A., 2014. Postglacial climate changes and rise of three ecotypes of harbour porpoises, *Phocoena phocoena*, in western Palearctic waters. Mol. Ecol. 23, 3306–3321. doi:10.1111/mec.12817

Foote, A.D., Vijay, N., Ávila-Arcos, M.C., Baird, R.W., Durban, J.W., Fumagalli, M., Gibbs, R.A., Hanson, M.B., Korneliussen, T.S., Martin, M.D., Robertson, K.M., Sousa, V.C., Vieira, F.G., Vinar, T., Wade, P., Worley, K.C., Excoffier, L., Morin, P.A., Gilbert, M.T.P., Wolf, J.B.W., 2016. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nat Commun 7, 11693. doi:10.1038/ncomms11693

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Gaskin, D.E., 1982. The Ecology of Whales and Dolphins. London: Heinemann.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. doi:10.1093/sysbio/syq010

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. Nucleic Acids Research 41, e129–e129. doi:10.1093/nar/gkt371

Hare, M.P., Cipriano, F., Palumbi, S.R., 2002. Genetic evidence on the demography of speciation in allopatric dolphin species. Evolution 56, 804–816.

Harris, S.A., 2005. Thermal history of the Arctic Ocean environs adjacent to North America during the last 3.5 Ma and a possible mechanism for the cause of the cold events (major glaciations and permafrost events). Progress in Physical Geography: Earth and Environment 29, 218–237. doi:10.1191/0309133305pp444ra

Hayano, A., Amano, M., Miyazaki, N., 2003. Phylogeography and population structure of the Dall's porpoise, *Phocoenoides dalli*, in Japanese waters revealed by mitochondrial DNA. Genes Genet. Syst. 78, 81–91.

Hervé, M., 2019. RVAideMemoire: Testing and Plotting Procedures for Biostatistics.

Ho, S.Y.W., Lanfear, R., Bromham, L., Phillips, M.J., Soubrier, J., Rodrigo, A.G., Cooper, A., 2011. Time-dependent rates of molecular evolution. Mol. Ecol. 20, 3087–3101. doi:10.1111/j.1365-294X.2011.05178.x

Hoekendijk, J.P.A., Spitz, J., Read, A.J., Leopold, M.F., Fontaine, M.C., 2017. Resilience of harbor porpoises to anthropogenic disturbance: Must they really feed continuously? Marine Mammal Science 34, 258–264. doi:10.1111/mms.12446

Hughes, A.L., 2008. Near neutrality: leading edge of the neutral theory of molecular evolution. Ann. N. Y. Acad. Sci. 1133, 162–179. doi:10.1196/annals.1438.001

Jaramillo Legorreta, A.M., Cardenas-Hinojosa, G., Nieto-Garcia, E., Rojas-Bracho, L., Thomas, L., Ver Hoef, J.M., Moore, J., Taylor, B., Barlow, J., Tregenza, N., 2019. Decline towards extinction of Mexico's vaquita porpoise (*Phocoena sinus*). R Soc Open Sci 6, 190598. doi:10.1098/rsos.190598

Kainer, D., Lanfear, R., 2015. The effects of partitioning on phylogenetic inference. Mol. Biol. Evol. 32, 1611–1627. doi:10.1093/molbev/msv026

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. doi:10.1093/bioinformatics/bts199

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge. doi:10.1017/CBO9780511623486

Koopman, H.N., Pabst, D.A., McLellan, W.A., Dillaman, R.M., Read, A.J., 2002. Changes in Blubber Distribution and Morphology Associated with Starvation in the Harbor Porpoise (*Phocoena phocoena*): Evidence for Regional Differences in Blubber Structure and Function. Physiological and Biochemical Zoology 75, 498–512. doi:10.1086/342799

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701. doi:10.1093/molbev/mss020

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452. doi:10.1093/bioinformatics/btp187

Lindberg, D.R., 1991. Marine biotic interchange between the northern and southern hemispheres. Paleobiology 17, 308–324. doi:10.1017/S0094837300010629

Louis, M., Fontaine, M.C., Spitz, J., Schlund, E., Dabin, W., Deaville, R., Caurant, F., Cherel, Y., Guinet, C., Simon-Bouhet, B., 2014. Ecological opportunities and specializations shaped genetic divergence in a highly mobile marine top predator. Proc. Biol. Sci. 281, 20141558–20141558. doi:10.1098/rspb.2014.1558

Lundmark, C., 2007. Science Sings the Blues: Other Words for Nothin' Left to Lose. BioScience 57, 208–208. doi:10.1641/B570218

Malaney, J.L., Cook, J.A., 2013. Using biogeographical history to inform conservation: the case of Preble's meadow jumping mouse. Mol. Ecol. 22, 6000–6017. doi:10.1111/mec.12476

McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351, 652–654. doi:10.1038/351652a0

McGowen, M.R., Spaulding, M., Gatesy, J., 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Mol. Phylogenet. Evol. 53, 891–906. doi:10.1016/j.ympev.2009.08.018

Morell, V., 2017. World's most endangered marine mammal down to 30. Science 355, 558–559. doi:10.1126/science.355.6325.558

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

Moritz, C., 1994. Defining "Evolutionarily Significant Units" for conservation. Trends in Ecology & Evolution 9, 373–375. doi:10.1016/0169-5347(94)90057-4

Moritz, C.C., Potter, S., 2013. The importance of an evolutionary perspective in conservation policy planning. Mol. Ecol. 22, 5969–5971. doi:10.1111/mec.12565

Moura, A.E., Nielsen, S.C.A., Vilstrup, J.T., Moreno-Mayar, J.V., Gilbert, M.T.P., Gray, H.W.I., Natoli, A., Moller, L., Hoelzel, A.R., 2013. Recent Diversification of a Marine Genus (*Tursiops* spp.) Tracks Habitat Preference and Environmental Change. Syst. Biol. 62, 865–877. doi:10.1093/sysbio/syt051

Nabholz, B., Glemin, S., Galtier, N., 2007. Strong variations of mitochondrial mutation rate across mammals--the longevity hypothesis. Mol. Biol. Evol. 25, 120–130. doi:10.1093/molbev/msm248

Nabholz, B., Mauffrey, J.-F., Bazin, E., Galtier, N., Glemin, S., 2008. Determination of mitochondrial genetic diversity in mammals. Genetics 178, 351–361. doi:10.1534/genetics.107.073346

Norris, K.S., McFarland, W.N., 1958. A New Harbor Porpoise of the Genus *Phocoena* from the Gulf of California. Journal of Mammalogy 39, 22. doi:10.2307/1376606

Parsch, J., Zhang, Z., Baines, J.F., 2009. The influence of demography and weak selection on the McDonald-Kreitman test: an empirical study in *Drosophila*. Mol. Biol. Evol. 26, 691–698. doi:10.1093/molbev/msn297

Pastene, L.A., Goto, M., Kanda, N., Zerbini, A.N., Kerem, D., Watanabe, K., Bessho, Y., Hasegawa,

M., Nielsen, R., Larsen, F., Palsbøll, P.J., 2007. Radiation and speciation of pelagic organisms during periods of global warming: the case of the common minke whale, *Balaenoptera acutorostrata*. Mol. Ecol. 16, 1481–1495. doi:10.1111/j.1365-294X.2007.03244.x

Perrin, W.F., 2009. Coloration, in: Würsig, B., Perrin, W., Thewissen, J.G.M. (Eds.), Encyclopedia of Marine Mammals. Elsevier, pp. 243–249. doi:10.1016/B978-0-12-373553-9.00061-4

Phifer-Rixey, M., Bonhomme, F., Boursot, P., Churchill, G.A., Piálek, J., Tucker, P.K., Nachman, M.W., 2012. Adaptive evolution and effective population size in wild house mice. Mol. Biol. Evol. 29, 2949–2955. doi:10.1093/molbev/mss105

Pimper, L.E., Goodall, R.N.P., Remis, M.I., 2012. First mitochondrial DNA analysis of the spectacled porpoise (*Phocoena dioptrica*) from Tierra del Fuego, Argentina. Mammalian Biology - Zeitschrift für Säugetierkunde 77, 459–462. doi:10.1016/j.mambio.2012.02.002

R Core Team, 2018. *R*: A language and environment for statistical computing.

Rambaut, A., Drummond, A.J., 2012. FigTree version 1.4.3. available at: tree.bio.ed.ac.uk/software/figtree.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v.1.6 [WWW Document]. URL http://tree.bio.ed.ac.uk/software/tracer.

Read, F.L., Santos, M.B., González, A.F., 2012. Understanding harbour porpoise (*Phocoena phocoena*) and fishery interactions in the north-west Iberian Peninsula. Final report to ASCOBANS.

Romiguier, J., Figuet, E., Galtier, N., Douzery, E.J.P., Boussau, B., Dutheil, J.Y., Ranwez, V., 2012. Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. PLoS ONE 7, e33852. doi:10.1371/journal.pone.0033852

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. 61, 539–542. doi:10.1093/sysbio/sys029

Rosa, S., Milinkovitch, M.C., Van Waerebeek, K., Berck, J., Oporto, J., Alfaro-Shigueto, J., Van Bressem, M.-F., Goodall, N., Cassens, I., 2005. Population structure of nuclear and mitochondrial DNA variation among South American Burmeister's porpoises (*Phocoena spinipinnis*). Conservation Genetics 6, 431–443. doi:10.1007/s10592-005-4988-9

Rosel, P.E., Dizon, A.E., Haygood, M.G., 1995a. Variability of the mitochondrial control region in populations of the harbour porpoise, *Phocoena*, on interoceanic and regional scales. Canadian Journal of Fisheries and Aquatic Sciences 52, 1210–1219. doi:10.1139/f95-118

Rosel, P.E., Haygood, M.G., Perrin, W.F., 1995b. Phylogenetic relationships among the true porpoises (Cetacea:Phocoenidae). Mol. Phylogenet. Evol. 4, 463–474. doi:10.1006/mpev.1995.1043

Rosel, P.E., Rojas-Bracho, L., 1999. Mitochondrial DNA variation in the Critically Endangered Vaquita *Phocoena Sinus* Norris and Macfarland, 1958. Marine Mammal Science 15, 990–1003. doi:10.1111/j.1748-7692.1999.tb00874.x

Sanders, H.L., 1968. Marine Benthic Diversity: A Comparative Study. The American Naturalist 102, 243–282. doi:10.1086/282541

Schneider, S., Excoffier, L., 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152, 1079–1089.

Slater, G.J., Price, S.A., Santini, F., Alfaro, M.E., 2010. Diversity versus disparity and the radiation of modern cetaceans. Proc. Biol. Sci. 277, 3097–3104. doi:10.1098/rspb.2010.0408

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57, 758–771. doi:10.1080/10635150802429642

Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sorensen, M.V., Willerslev, E., 2009. Radiation of Extant Cetaceans Driven by Restructuring of the Oceans. Syst. Biol. 58, 573–585. doi:10.1093/sysbio/syp060

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Thomas, L., Jaramillo-Legorreta, A., Cardenas-Hinojosa, G., Nieto-Garcia, E., Rojas-Bracho, L., Ver

1   Hoef, J.M., Moore, J., Taylor, B., Barlow, J., Tregenza, N., 2017. Last call: Passive acoustic
2   monitoring shows continued rapid decline of critically endangered vaquita. The Journal of the
3   Acoustical Society of America 142, EL512–EL517. doi:10.1121/1.5011673
4   Tolley, K.A., Rosel, P.E., 2006. Population structure and historical demography of eastern North
5   Atlantic harbour porpoises inferred through mtDNA sequences. Marine Ecology Progress Series
6   327, 297–308. doi:10.3354/meps327297
7   Torroni, A., Achilli, A., Macaulay, V., Richards, M., Bandelt, H.-J., 2006. Harvesting the fruit of the
8   human mtDNA tree. Trends Genet. 22, 339–345. doi:10.1016/j.tig.2006.04.001
9   Viricel, A., Rosel, P.E., 2011. Evaluating the utility of cox1 for cetacean species identification. Marine
10   Mammal Science 28, 37–62. doi:10.1111/j.1748-7692.2010.00460.x
11   Wang, D., Turvey, S.T., Zhao, X., Mei, Z., 2013. *Neophocaena asiaeorientalis* ssp. *asiaeorientalis*.
12   The IUCN Red List of Threatened Species. doi:10.2305/IUCN.UK.2013-
13   1.RLTS.T43205774A45893487.en
14   Wang, J.Y., Frasier, T.R., Yang, S.C., White, B.N., 2008. Detecting recent speciation events: the case
15   of the finless porpoise (genus *Neophocaena*). Heredity (Edinb) 101, 145–155.
16   doi:10.1038/hdy.2008.40
17   Wang, J.Y., Reeves, R.R., 2017. *Neophocaena phocaenoides*. The IUCN Red List of Threatened
18   Species. doi:10.2305/IUCN.UK.2017-3.RLTS.T198920A50386795.en
19   Whitehead, H., 1998. Cultural selection and genetic diversity in matrilineal whales. Science 282,
20   1708–1711. doi:10.1126/science.282.5394.1708
21   Yim, H.-S., Cho, Y.S., Guang, X., Kang, S.G., Jeong, J.-Y., Cha, S.-S., Oh, H.-M., Lee, J.-H., Yang,
22   E.C., Kwon, K.K., Kim, Y.J., Kim, T.W., Kim, W., Jeon, J.H., Kim, S.-J., Choi, D.H., Jho, S.,
23   Kim, H.-M., Ko, J., Kim, H., Shin, Y.-A., Jung, H.-J., Zheng, Y., Wang, Z., Chen, Y., Chen, M.,
24   Jiang, A., Li, E., Zhang, S., Hou, H., Kim, T.H., Yu, L., Liu, S., Ahn, K., Cooper, J., Park, S.-G.,
25   Hong, C.P., Jin, W., Kim, H.-S., Park, C., Lee, K., Chun, S., Morin, P.A., O'Brien, S.J., Lee, H.,
26   Kimura, J., Moon, D.Y., Manica, A., Edwards, J., Kim, B.C., Kim, S., Wang, J., Bhak, J., Lee,
27   H.S., Lee, J.-H., 2014. Minke whale genome and aquatic adaptation in cetaceans. Nat. Genet. 46,
28   88–92. doi:10.1038/ng.2835
29   Yu, G., Lam, T.T.-Y., Zhu, H., Guan, Y., 2018. Two Methods for Mapping and Visualizing
30   Associated Data on Phylogeny Using Ggtree. Mol. Biol. Evol. 35, 3041–3043.
31   doi:10.1093/molbev/msy194
32   Zheng, Y., Wiens, J.J., 2015. Do missing data influence the accuracy of divergence-time estimation
33   with BEAST? Mol. Phylogenet. Evol. 85, 41–49. doi:10.1016/j.ympev.2015.02.002
34   Zhou, X., Guang, X., Sun, D., Xu, S., Li, M., Seim, I., Jie, W., Yang, L., Zhu, Q., Xu, J., Gao, Q.,
35   Kaya, A., Dou, Q., Chen, B., Ren, W., Li, S., Zhou, K., Gladyshev, V.N., Nielsen, R., Fang, X.,
36   Yang, G., 2018. Population genomics of finless porpoises reveal an incipient cetacean species
37   adapted to freshwater. Nat Commun 9, 1276. doi:10.1038/s41467-018-03722-x
38
39

Tables

Table 1. Taxon sample size (n) and descriptive statistics for the shotgun sequencing and mitochondrial assembly per species and mitochondrial lineage. The statistics include the total number of reads before and after filtering (R_b and R_a), the sequencing coverage depth, the size of the mitochondrial assembly (in base-pairs), and the GC content in percent (%GC). The mean value and the standard deviation are shown.

Lineage	n	R_b	R_a	Depth	Assembly size	% GC
Harbor porpoise (HP)	27	30.8 ± 28.3	28.5 ± 25.9	1323.9 ± 1800.0	16383.8 ± 0.8	40.6 ± 0.1
NP1	10	5.1 ± 1.8	5.0 ± 1.7	99.6 ± 92.6	16384.3 ± 0.5	40.7 ± 0.0
NP2	4	5.9 ± 1.3	5.6 ± 1.2	97.2 ± 119.8	16384.0 ± 0.0	40.7 ± 0.0
NAT	4	56.9 ± 3.4	52.2 ± 2.9	1656.3 ± 1058.9	16383.3 ± 0.6	40.6 ± 0.0
MA	3	59.8 ± 5.4	54.9 ± 5.2	1431.0 ± 648.1	16384.0 ± 0.0	40.6 ± 0.0
IB	3	65.2 ± 8.2	60.5 ± 7.6	3169.7 ± 1882.6	16384.0 ± 0.0	40.5 ± 0.0
BS	3	60.0 ± 2.2	54.8 ± 1.3	4755.3 ± 1376.9	16382.0 ± 0.0	40.6 ± 0.0
Dall's porpoise (DP)	6	2.8 ± 0.9	2.7 ± 0.9	49.5 ± 14.5	16367.5 ± 0.8	40.5 ± 0.1
DP1	1	4.5	4.4	74.0	16369.0	40.6
DP2	5	2.5 ± 0.5	2.4 ± 0.5	44.6 ± 9	16367.2 ± 0.4	40.5 ± 0.0
Vaquita (V)	12	5.8 ± 5.7	5.7 ± 5.5	113.3 ± 108.6	16370.0 ± 0.0	39.8 ± 0.0
Burmeister's porpoise (BP)	3	2.3 ± 0.3	2.3 ± 0.3	125.0 ± 121.2	16378.7 ± 8.1	40.1 ± 0.0
Spectacled porpoise (SP)	3	2.4 ± 0.2	2.3 ± 0.2	55.0 ± 30.5	16371.0 ± 0.0	39.7 ± 0.1
SP1	1	2.2	2.1	41.0	16371.0	39.6
SP2	2	2.5 ± 0.0	2.4 ± 0.0	62.0 ± 39.6	16371.0 ± 0.0	39.7 ± 0.1
Finless porpoises (FP)	12	2.7 ± 1.1	2.6 ± 1.1	92.9 ± 74.9	16383.1 ± 5.8	40.8 ± 0.0
IPF	1	2.3	2.2	77.0	16385.0	40.8
EAF	5	3.0 ± 1.5	2.9 ± 1.4	53.2 ± 22.8	16381.2 ± 7.5	40.0 ± 0.0
YF	6	2.3 ± 0.3	2.2 ± 0.3	180.0 ± 101.8	16386.0 ± 0.0	40.8 ± 0.0

The lineage code includes the vaquita (*P. sinus*, V), Burmeister's porpoise (*P. spinipinnis,* BP), spectacled porpoise (*P. dioptrica,* SP), Dall's porpoise (*P. dalli,* DP), harbor porpoise (*P. phocoena,* HP) and finless porpoise (*N. phocaenoides* + *N. asiaeorientalis,* FP). Some species have been also further subdivided into distinct groups: harbor porpoises are divided into North Pacific (NP) and North Atlantic (NA), and within each of these groups, further subdivisions are recognized. Four groups are recognized within NA: NAT (North Atlantic), IB (Iberian), MA (Mauritanian), BS (Black Sea). NP is divided into two subgroups: NP1 and NP2. Three sub-species or species are also recognized among the finless porpoises: Indo-Pacific (IPF), Yangtze finless (YFP) and East Asia finless (EAF). Spectacled porpoise subgroups are designated as SP1 and SP2. Dall's porpoise subgroups as DP1 and DP2.

32

1   **Table 2.** Summary statistics describing the genetic diversity of the mitochondrial genomes
2   among porpoise species and their distinct lineages. The statistics includes the sample size (N),
3   number of sites with missing data (MD), segregating sites (S), singletons (*Singl.*), shared
4   polymorphism (*Shared P.*), theta from S (θ_W), average nucleotide diversity per site (π) and its
5   standard deviation (SD_π), number of haplotypes (H), haplotypic diversity (H_d) and its standard
6   deviation (SD_{HD}).
7

		N	MD	S	*Singl.*	*Shared P.*	θ_W (%)	π (%)	SD_π (%)	H	H_d (%)	SD_{HD} (%)
Species	All porpoises	63	51	2947	242	2705	4.28	5.35	0.23	58	99.6	0.04
	FP*	12	36	229	173	56	0.47	0.35	0.00	12	100.0	3.40
	BP[†]	3	31	33	33	0	0.13	0.13	0.04	3	100.0	27.20
	V[†]	12	32	16	13	3	0.03	0.02	0.00	8	89.4	7.80
	SP*	3	30	145	145	0	0.59	0.59	0.21	3	100.0	27.20
	DP*	6	34	208	152	56	0.56	0.49	0.13	5	93.3	14.80
	HP*	27	37	602	158	444	0.98	1.11	0.06	26	99.7	1.10
HP	NAT[†]	4	36	102	85	17	0.34	0.33	0.07	4	100.0	3.12
	IB[†]	3	32	31	31	0	0.12	0.12	0.04	3	100.0	27.20
	MA[†]	3	32	28	28	0	0.11	0.11	0.03	3	100.0	27.20
	BS[†]	3	34	18	18	0	0.07	0.07	0.02	3	100.0	27.20
	NP1[†]	10	31	76	53	23	0.16	0.12	0.01	10	100.0	4.50
	NP2[†]	4	31	6	4	2	0.02	0.02	0.00	3	83.3	4.94
FIN	YF[†]	6	34	33	30	3	0.09	0.07	0.01	6	100.0	0.93
	EAF[†]	5	33	58	55	3	0.17	0.15	0.04	5	100.0	16.00
DAL	DP2[†]	5	33	107	61	46	0.32	0.32	0.07	4	90.0	16.10
SPE	SP2	2	29	39	39	0	0.24	0.24	0.12	2	100.0	50.00

8   *Species including multiple distinct mitochondrial lineages; †Species with a single mitochondrial lineage.
9   The meaning of the group acronyms is provided in Table 1.
10

1   **Table 3.** Neutrality tests based on the site frequency spectrum. The neutrality tests were
2   only applied to lineages where at the sample size (*n*) was at least 10. The statistics
3   includes the Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F*.

4

Lineage	*D*	*D**	*F*
V *(n=12)*	-1.44^{NS}	-1.89^{*}	-1.84^{*}
NP1 *(n=10)*	-1.18^{NS}	-1.27^{NS}	-1.28^{NS}

5   NS, Not significant; *, p-value < 0.05. The acronyms are provided in Table 1.
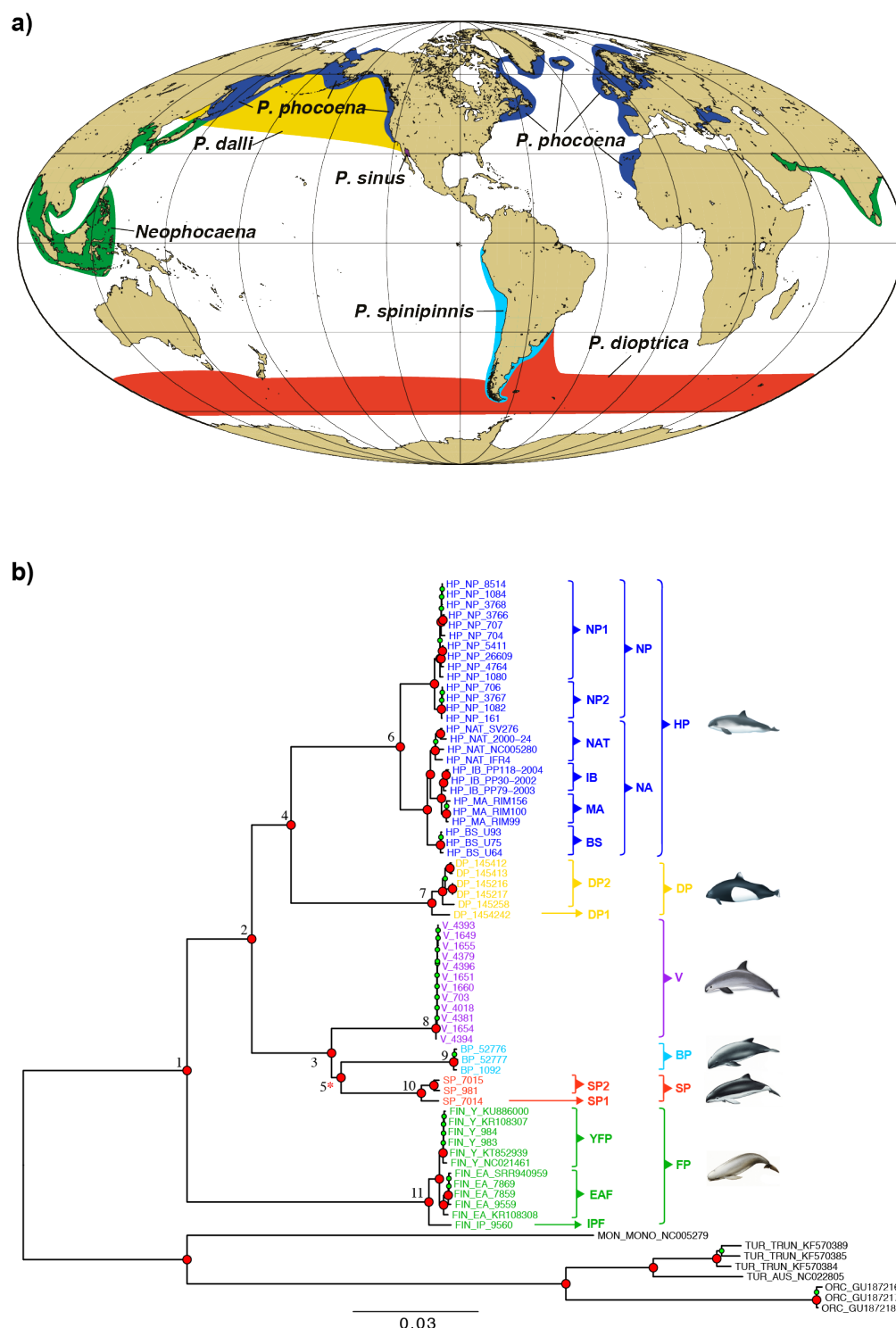
Fig. 1. (a) Geographical range of the seven porpoise species modified from Berta et al. (2015). (b) Maximum-likelihood mitochondrial phylogeny. The external branches and the tip labels are colored by species. The tree is rooted with eight sequences from four closely related *Odontoceti* species (in black). Numbers at the nodes are discussed in the text. The nodes indicated in red and green represent nodes with bootstrap support ≥ 90% and ≥70%, respectively. The red star indicates a node with a 63% bootstrap support in the Neighbor-joining tree (Fig. S2c). The acronyms are provided in Table 1. The scale bar unit is in substitution per site.
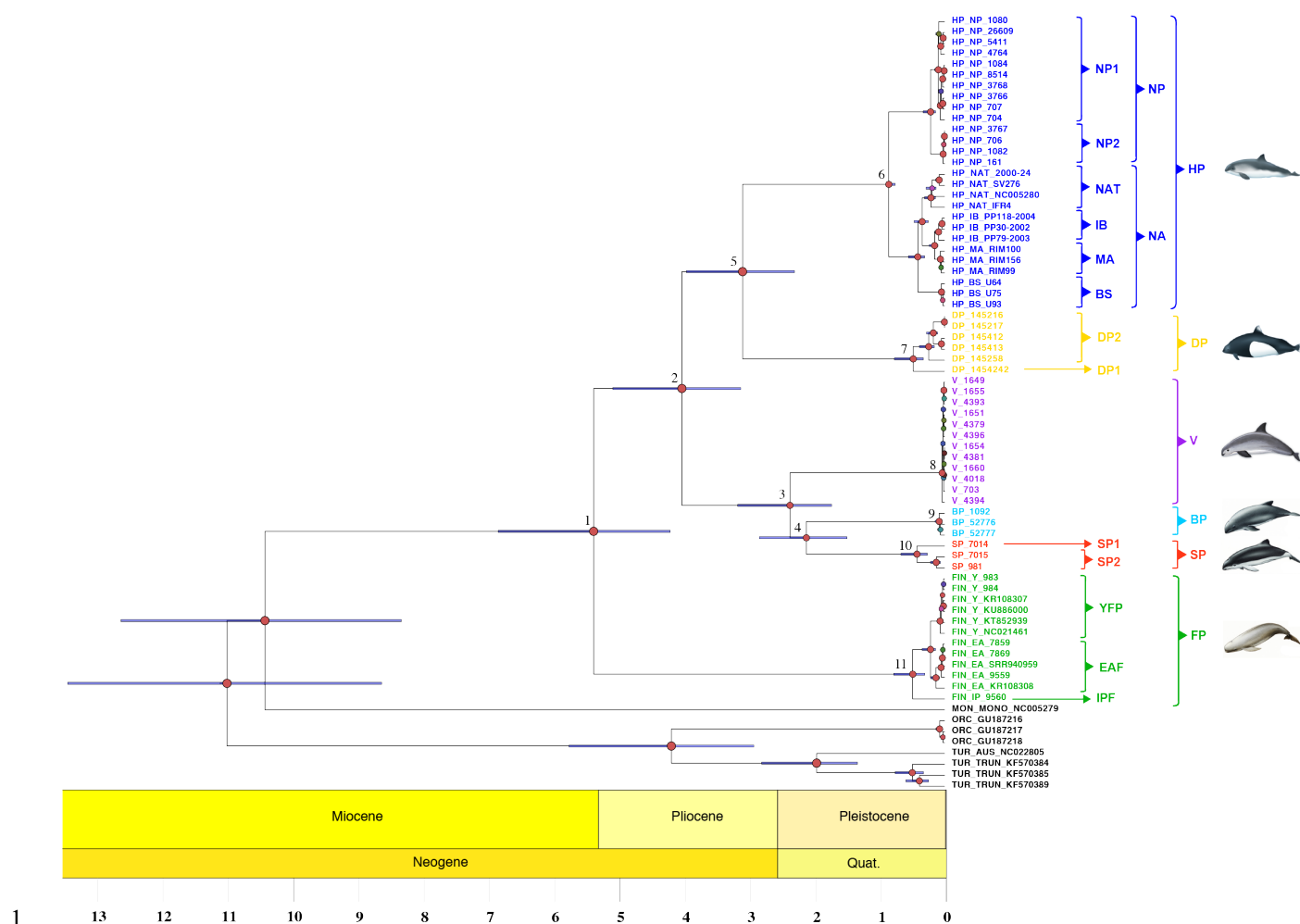
Fig. 2. Bayesian chronogram of the porpoise family. The tree represents the maximum clade credibility tree. Red node labels indicate posterior probabilities of 1; node position indicates median node age estimates and the error bars around the node indicate the 95% highest posterior density of the estimated node age. Time is in millions of years. Numbers at the nodes are discussed in the text. The acronyms are provided in Table 1.

1
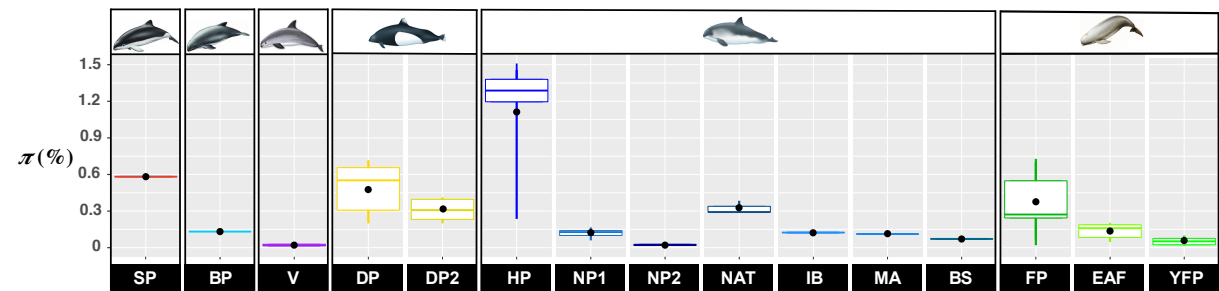
2
3   **Fig. 3.** Nucleotide diversity (π) among groups and species of porpoises. The median and mean
4   π values are represented respectively by the colored line in the boxplot and the black dot. The
5   whiskers represent the 95% confidence interval. The boxes represent the upper and lower
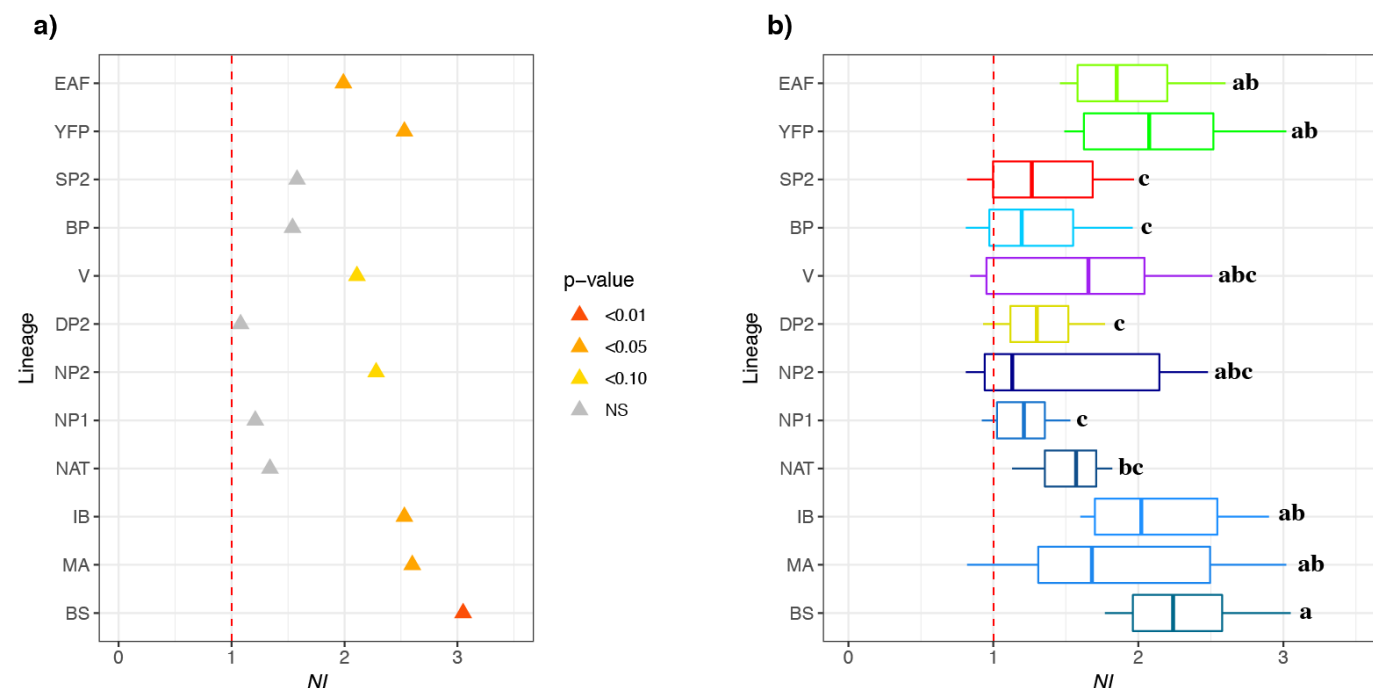6   quartile. The acronyms are provided in Table 1.

37

1
2   **Fig. 4.** Mcdonald-Kreitman tests on the 13 protein coding regions of the mitogenome among
3   porpoise subgroups. (a) Neutrality index (*NI*) estimated using the orca as outgroup. (b) NI
4   distributions per mitochondrial lineages, calculated using various outgroups, including orca
5   and all possible interspecific comparisons. The letters on the right of the boxplots indicate
6   significant differences in the mean *NI* between the different lineages (i.e. boxplot with
7   different letters are statistically different from one another). The red dashed lines represent the
8   limit at which *NI* reflects positive (*NI*<1), neutral (*NI* =1), or purifying (*NI*>1) selection. NS:
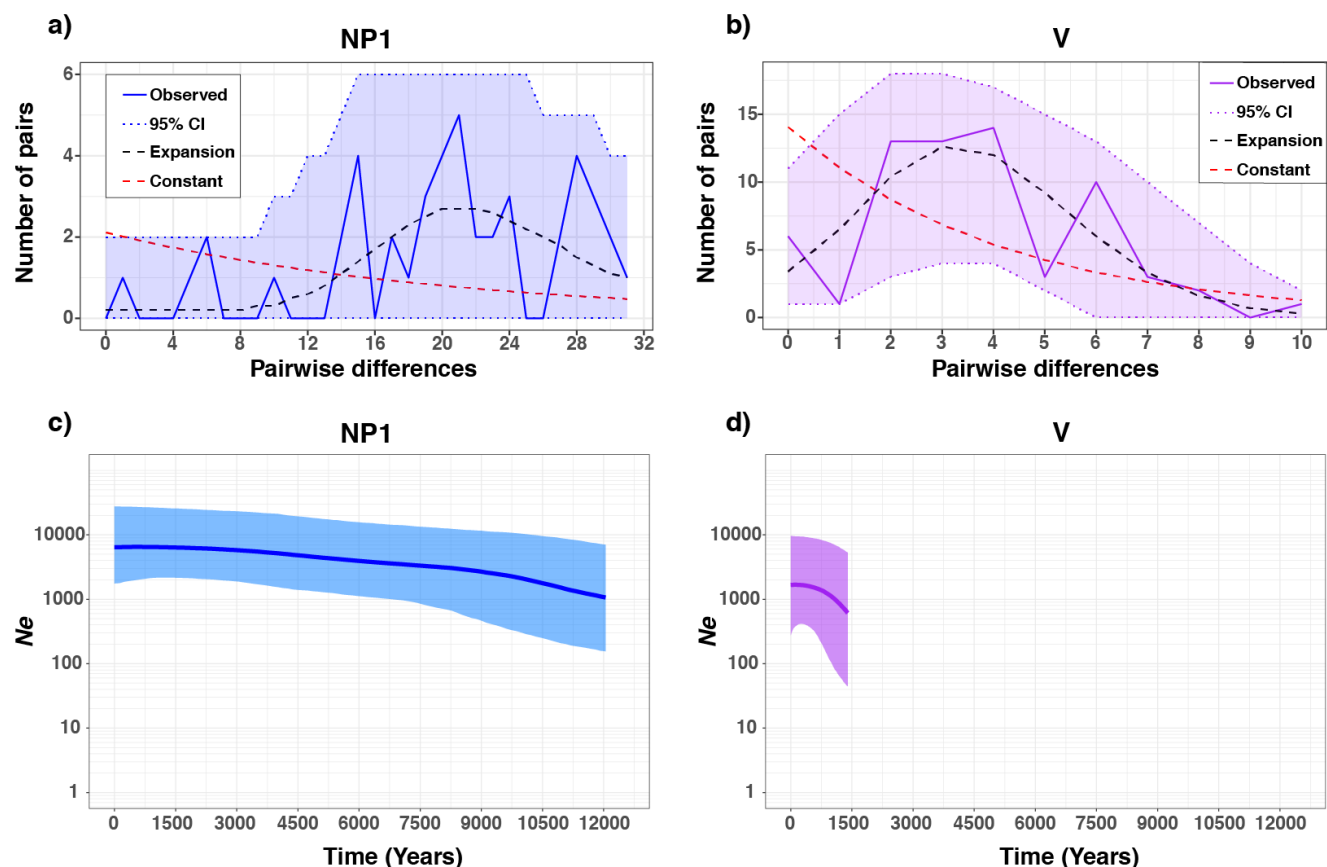9   Not Significant. The acronyms are provided in Table 1.

Fig. 5. Mismatch distributions (a and b) and Bayesian Skyline Plots (c and d) for the North Pacific harbor porpoise NP1 lineage and for the vaquita. The Skyline Plots (c and d) show the temporal changes in mitochondrial diversity. The y-axis of the Bayesian Skyline Plot shows the genetic diversity expressed as the effective female population size (*Ne*). The bold line inside each Bayesian Skyline Plot represents the median value and thin lines the 95% HPD intervals. The acronyms are provided in Table 1.