ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 26/02/2024

DOI: https://doi.org/10.3897/arphapreprints.e121514

Hidden in the blow - a matrix to characterise cetaceans' respiratory microbiome: Short-finned pilot whale as model species

Beatriz Santos, Luís Afonso, Filipe Alves, Ana Dinis, Rita Ferreira, Ana Correia, Raul Valente, Ágatha Gil, Filipe Castro, Isabel Sousa-Pinto, Massimiliano Rosso, Cinzia Centelleghe, Sandro Mazzariol, Catarina Magalhães, Maria Paola Tomasino

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 26/02/2024* DOI: https://doi.org/10.3897/arphapreprints.e121514

Hidden in the blow - a matrix to characterise cetaceans' respiratory microbiome: Short-finned pilot whale as model species

Santos B.[1,2 *], Afonso L.[1 *], Alves F.[3], Dinis A.[3], Ferreira R.[3,4], Correia A. M.[1,2], Valente R.[1,2], Gil Á.[1,5,6], Castro L. Filipe C.[1,2], Sousa-Pinto I.[1,2], Rosso M.[7,8], Centelleghe C.[9], Mazzariol S.[9], Magalhães C[1,2 a], Tomasino M. P.[1 a]

[1]CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto 4450-208, Matosinhos, Portugal.
[2]FCUP – Department of Biology, Faculty of Sciences, University of Porto, 4169-007, Porto, Portugal.
[3]MARE - Marine and Environmental Sciences Centre/ARNET - Aquatic Research Network, Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI), 9020-105 Funchal, Portugal.
[4]Marine Biology Station of Funchal, Faculty of Life Sciences, University of Madeira, 9000-003, Funchal, Portugal
[5]CITAB – Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Department of Biology and Environment, University of Trás-os-Montes and Alto Douro, 5000-801, Vila Real, Portugal.
[6]IIM-CSIC – Institute of Marine Research of the Spanish National Research Council, 36208, Vigo, Pontevedra, Spain.
[7]CIMA Research Foundation - Centro Internazionale di Ricerca in Monitoraggio Ambientale, 17100, Savona, Italy
[8]NBFC - National Biodiversity Future Center, 90133, Palermo, Italy
[9]UNIPD - Department of Comparative Biomedicine and Food Science, University of Padua 35131, Padova, Italy.

* Authors contributed equally to the work.
[a] Authors contributed equally to the work. Corresponding authors: cmagalhaes@ciimar.up.pt, mtomasino@ciimar.up.pt.

Abstract

Cetaceans are key sentinel species of marine ecosystems and ocean health, being a strategic taxonomic group to evaluate the well-being of aquatic habitats and to detect harmful environmental trends. Respiratory diseases are among the main causes of death in these animals, so the identification of the microbiome community existent in their exhaled breath condensates (EBCs), i.e. blows, has been proposed as a key biomarker for assessing respiratory health. Yet, to characterise microbiomes related to these animals' respiratory tract and use them as a proxy for health status, it is necessary to develop baseline data on the microorganisms associated with cetaceans. Here, the short-finned pilot whale (SFPW, *Globicephala macrorhynchus*) was used as a model species to validate the most suitable primer set to explore the prokaryotic diversity of the cetaceans' respiratory tract. DNA extracted from blow samples (n = 12) of island-associated animals off Madeira Island was sequenced to amplify both V3-V4 and V4-V5 hypervariable regions of the 16S rRNA gene, using the same sequencing platform (Illumina MiSeq). Independently of the primer set used, all blows shared Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria phyla in their composition. V3-V4 resulted in higher diversity of taxa with relative abundance above 1%, whereas the V4-V5 primers captured a higher number of microbial Amplicon Sequence Variants, detecting the microbial rare biosphere with pathogen potential.

Additionally, it captured more efficiently the core microbiome. Thus, this study provides a detailed characterization of SFPW respiratory-associated microbial communities, also strengthening the idea of sociality influencing microbiome composition in the respiratory tract. Moreover, it supports the use of EBCs as a relevant biomarker for the physiological state of the airways in free-ranging cetaceans.

Keywords: blow sampling, exhale breath condensate, health status, metabarcoding, respiratory disease, short-finned pilot whale.

Introduction

As keystone species, cetaceans play vital ecological functions, considering their role as nutrient vectors, position in the food chains and use as bioindicators of environmental health (Ballance, 2018). Currently, multiple global threats pose a serious concern to the conservation of these aquatic mammals (Evans, 2018), affecting individual health and ultimately compromising population viability (Nicol et al., 2020). Addressing their impacts on cetacean populations is therefore crucial yet challenging, as many species are elusive and inhabit remote habitats with wide distributions (Parsons et al., 2015).

Among the primary causes of death observed in these marine mammals, respiratory tract infections make a substantial contribution (Díaz-Delgado et al., 2018; Cuvertoret-Sanz et al., 2020). Therefore, the identification of the microbiome community existent in their exhaled breath condensates (EBCs), i.e. blows, has been proposed as a key biomarker for assessing respiratory health and inform conservation management (Acevedo-Whitehouse et al., 2010). Cetacean blow is a relatively unexplored biological matrix (de Mello & De Oliveira, 2016). It is mostly composed of hormones, as well as respiratory microbes, a range of metabolites, and substances related to inflammation and the immune system (Hunt et al., 2013). Several studies have reported important differences in the microbial communities detected in the blow of cetaceans when compared to those occurring in the external environment, thus being considered specific of cetaceans' respiratory tract (Pirotta et al., 2017; Geoghegan et al., 2018; Vendl et al., 2021). Within this material, potential specific microbial pathogens have been identified (Acevedo-Whitehouse et al., 2010), highlighting the value of using the EBC collection and characterization as a method for monitoring respiratory microbial communities in cetaceans (Lima et al., 2012).

Previous research on the airway microbiota of cetaceans resorted to the analysis of blow samples, with different sampling methodologies, DNA extraction and amplification, targeting different hypervariable gene regions, some of these using a metabarcoding approach. Nevertheless, the majority of data accessible on cetacean-associated microbiome, namely pathogens, diseases, and parasites, come from captive, stranded, sick, or injured individuals, which cannot be considered representative of the free-ranging populations (Johnson et al., 2009; Acevedo-Whitehouse et al., 2010; Lima et al., 2012). Therefore, knowledge of the cetaceans' respiratory microbiome from free-ranging individuals is limited. Here, we focus on short-finned pilot whales (*Globicephala macrorhynchus*; SFPW), a thoroughly studied species due to its global distribution, abundance and propensity to mass strandings (Betty et al., 2023). With a strong social structure (Alves et al., 2013), SFPW represents an interesting case-study of how social complexity might shape microbiome composition.

To the best of our knowledge, there are no published studies on the SFPW respiratory microbiome. In addition, there are also no previous works on the topic for other cetacean species that have compared the amplification targeting different hypervariable gene regions using the same methodology on the same samples. In light of this, the present study addresses the respiratory microbiome of free-ranging cetaceans, using SFPW as a model species. The main goal is to provide a network of consistent microbiome core taxa of the SFPW blow, by comparing V3-V4/V4-V5 hypervariable regions of 16S rRNA gene. The expected outputs will serve as a baseline to set a working and optimised methodology for cetacean health monitoring, from sample collection to laboratory analysis.

Methods

Blow Sampling

Blow sampling was conducted in September and October 2018, during at-sea campaigns in the southern waters of Madeira Island, Portugal, targeting SFPW (Fig. 1). Besides blow collection, the natural behaviour (travelling, resting, feeding or socialising), age class (following Betty et al., 2023; Aguilar de Soto & Alves, 2023), and the number of individuals sampled in the group (Supplementary Table 1) were recorded.

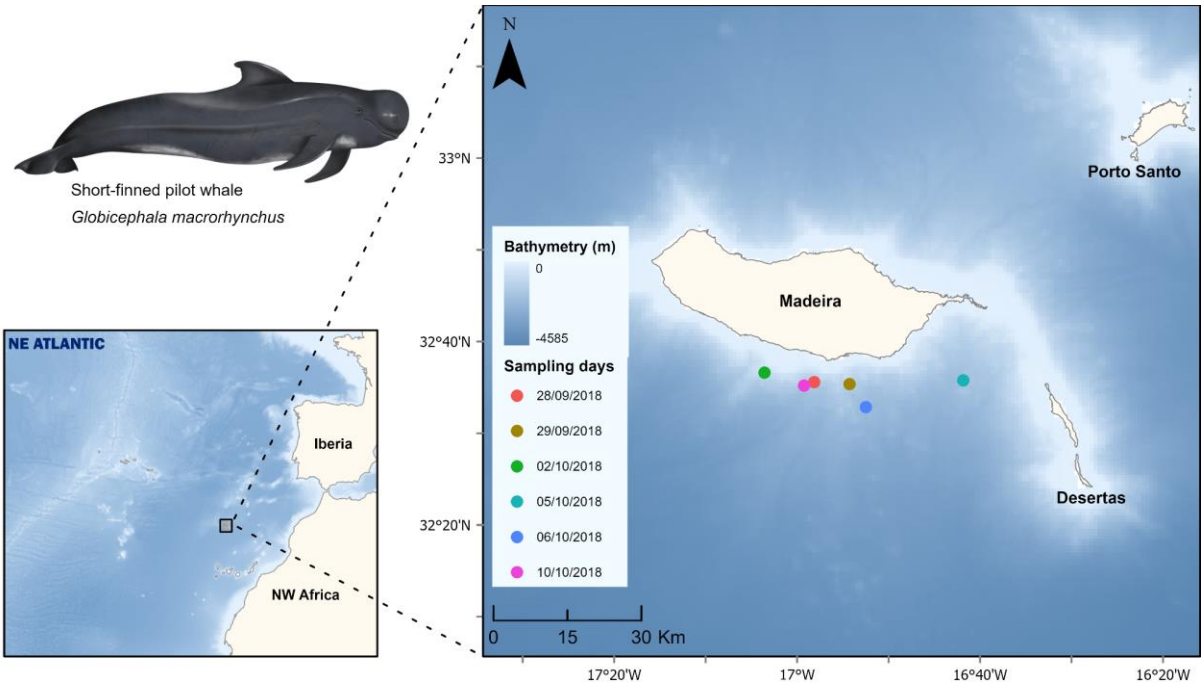


Figure 1 - Blow sampling locations and dates in the Madeira Archipelago, with illustration of the short-finned pilot whale (*Globicephala macrorhynchus*).

Sample collection was carried out using a PERFORMAgene™ PG-100 swab collection kit (DNA Genotek®). This kit was attached to an extendable 5 metre aluminium pole and used as a sampling device to collect the blow. When the animals surfaced to exhale near the boat, the blow collector device was positioned about 40 to 50 cm above the blowhole into the exhaled plume to collect the droplets in the blow. Simultaneously, the sampled animals were photographed to be compared with the (OOM/MARE-ARDITI) Madeira's photographic-ID catalogue of the species following Alves et al. (2020) to confirm if they belong to the island-associated population of short-

finned pilot whales in Madeira (i.e., regularly captured in different seasons throughout the years; detailed in Alves et al., 2013, 2020) or if they were transient.

DNA Extraction

DNA extraction of blow samples was performed using the QIAmp® DNA Mini Kit (QIAGEN), following the manufacturer's instructions. An additional sample concentration step in an Eppendorf Concentrator Plus™ was added to increase the concentration of the extracted DNA from blow samples. A Qubit™ 3 Fluorometer with a Qubit™ dsDNA High Sensitivity (HS) assay kit (Invitrogen™) was used for DNA quantification.

PCR Amplification, Library preparation and Sequencing

Samples were prepared for the amplification of the 16S rRNA gene hypervariable V4-V5 region (≈412 bp), using specific primers 515F-Y/926R (Parada et al., 2016). The V3-V4 region of the 16S rRNA bacterial and archaeal gene was amplified using the 341F and 806R primers (Takahashi et al., 2014).

Amplification of both regions was performed through high-throughput sequencing. Detailed description of the protocol is reported in Ribeiro et al. (2018). Briefly, KAPA HiFi HotStart PCR Kit was used in the first PCR reaction following manufacturer suggestions. Indexes and sequencing adapters were added to the target amplicon in the second PCR reaction. The amplified products obtained were purified and normalised with SequalPrep 96-well plate kit (ThermoFisher Scientific, USA). Pair-end sequencing was carried out in the Illumina MiSeq® sequencer with the V3 chemistry (Illumina, Inc., San Diego, CA, USA) at Genoinseq laboratories (Cantanhede, Portugal).

Bioinformatic Analysis

Upstream Analysis

The "*DADA2*" software package (v.1.20) (Callahan et al., 2016) on R studio (v.4.1.1) was used to process the FASTQ files obtained after Illumina MiSeq sequencing. Initially, the raw sequences were quality-filtered. Raw reads were truncated at 275 bp and 215 bp for forward and reverse sequences, respectively – the position where the final sequence numbers were higher. Afterward, the trimmed sequences were dereplicated, denoised, and merged. Sequences were then processed to obtain the Amplicon Sequence Variants (ASVs) table for taxonomic resolution, increased precision, and reproducibility. In the ASVs, reads represent differences of one nucleotide (Callahan et al., 2017). Additionally, chimeric sequences were identified and excluded. To determine the taxonomic classification of each ASV, the Naïve Bayes classifier was used against the SILVA database v132 (Quast et al., 2013). The upstream analysis performed for the primer set comparison pipeline followed Fadeev et al. (2021).

Downstream Analysis

The ASV counts and taxonomy tables from the upstream analysis, together with the metadata table containing the sample information (type of sample, species, primer set, sample day), were used as input for the "*phyloseq*" R package (v.1.36) for downstream analysis (McMurdie & Holmes, 2013). Summarily, "*phyloseq*" analyses and graphically

displays complex phylogenetic sequencing data that has already been clustered into ASVs. The ASVs that were taxonomically unclassified at phylum rank or were not assigned to bacterial or archaeal lineages, and the undesirable lineages, such as “Chloroplast”, “Eukaryota” and “Mitochondria”, were excluded from further analysis.

The distribution and diversity of the prokaryotic community across the different used primer sets were investigated. Alpha diversity was analysed for different subsets of samples by calculating four different indexes, also using “*phyloseq*” package: Observed and Chao 1 for the estimation of unique ASVs abundance, and Shannon and Inverse Simpson as species diversity measures. Differences in alpha diversity were tested using the Wilcoxon signed rank test, considering a level of significance of 0.05. Moreover, the beta diversity was analysed using the “*vegan*” package and a Non-metric MultiDimensional Scaling (NMDS) plot, based on Bray-Curtis dissimilarity (*phyloseq*” package). This measure is a statistical index used to quantify the compositional dissimilarity between two different sites, depending on the two communities' counts of shared and non-shared specimen (Bray and Curtis, 1957). To test if the set of primers used had a significant effect on the prokaryotic communities existent in the different blow samples, a PERMANOVA statistical test was implemented. Therefore, two hypotheses were generated: H_0: The use of different primer sets does not influence the distribution of prokaryotic communities; H_1: The use of different primer sets influences the distribution of prokaryotic communities. The level of significance was set to 0.05.

The microbial communities´ taxonomic composition was evaluated by creating taxonomy bar plots. The distribution of prokaryotes taxa across the samples was analysed at four taxonomic levels: phylum, class, family, and genus. Taxonomic distribution plots were performed using several R-packages, namely, “*phyloseq*” (v.1.36) (McMurdie & Holmes, 2013), “*ggplot2*” (v.3.3.6) (Wickham, 2009), “*tidyverse*” (v.1.32) (Wickham, 2009), and “*scales*” (v.1.20) (Wickham et al., 2019).

Core Microbiome

With the aim of identifying the common set of microbial taxa originated for each of the datasets, the core microbiome was calculated using the “*phyloseq*” package “*microbiome*” (Leo Lahti, Sudarshan Shetty et al., 2017). The core microbiome was determined using the “*core_members*” function, considering a detection threshold of 0.001 and a prevalence of 50% in all blow samples. The comparative analysis of the primers used to identify this core microbiota was then carried out using a NMDS plot, based on Bray-Curtis dissimilarity.

Results

Blow Sampling Characterization

A total of 12 blow samples were collected from six sampling events (Fig. 1), of which nine were from individual animals and three from a pool of animals (Supplementary Table 1). The photographic-ID comparison showed that the nine individual samples corresponded to seven individuals, i.e., two samples were obtained from the same animals in two separate sampling events (Supplementary Table 1). The photographic-ID also showed that all sampled individuals were island-associated animals.

Sequencing Output

After the sequencing process, a total of 685 692 reads (with a mean of 57 141 ± 7 920.8 reads per sample) was obtained for the V3-V4 dataset; and 862 763 for the V4-V5 (with a mean of 71 896,9 ± 12 990.5 reads per sample). After the quality filter steps, the V3-V4 dataset showed a lower decrease in the number of sequences throughout the workflow compared to V4-V5, with 60.4 ± 5.7% and 48.5 ± 7.7% of sequences retained per sample, respectively. Therefore, the non-target sequences were eliminated. Consequently, the final output was 415 340 sequences in the V3-V4 dataset that were assigned to 2 764 ASVs (with a mean of 34 611.7 ± 6 228.9 reads per sample) and 417 733 sequences in the V4-V5 dataset that were assigned to 3 665 ASVs (with a mean of 34 811.1 ± 7 554.7 reads per sample). The raw sequence data from this work is deposited to the European Nucleotide Archive (ENA) (Study accession number PRJEB72700).

Alpha and Beta diversity

Alpha diversity metrics varied in relation to the set of primers used (Fig. 2). The Observed ASVs and Chao1 had lower minimum/maximum and mean values in the V3-V4 dataset relative to values of the V4-V5 dataset. On the other hand, alpha indices (Shannon and Inverted Simpson) values were higher for the V3-V4, although with an evident high standard deviation. For the V4-V5 dataset, despite lower values for these aforementioned diversity indices, there was less variation within samples (lower standard deviations). In the V4-V5 dataset, sample Gma_03 was an outlier, presenting highest values in Observed and Chao1 indexes. The Shannon and Inverse Simpson indexes demonstrated a more considerable ASVs diversity in the V3-V4 region. The Wilcoxon Signed-Rank Test - used to compare the blow community richness for the Observed, Chao1, and Inverse Simpson measures - was significantly different between the two primers sets (see Supplementary Table 2), with ca. 32% more bacterial ASVs in the V4-V5. Statistically significant differences were not found for the Shannon index.

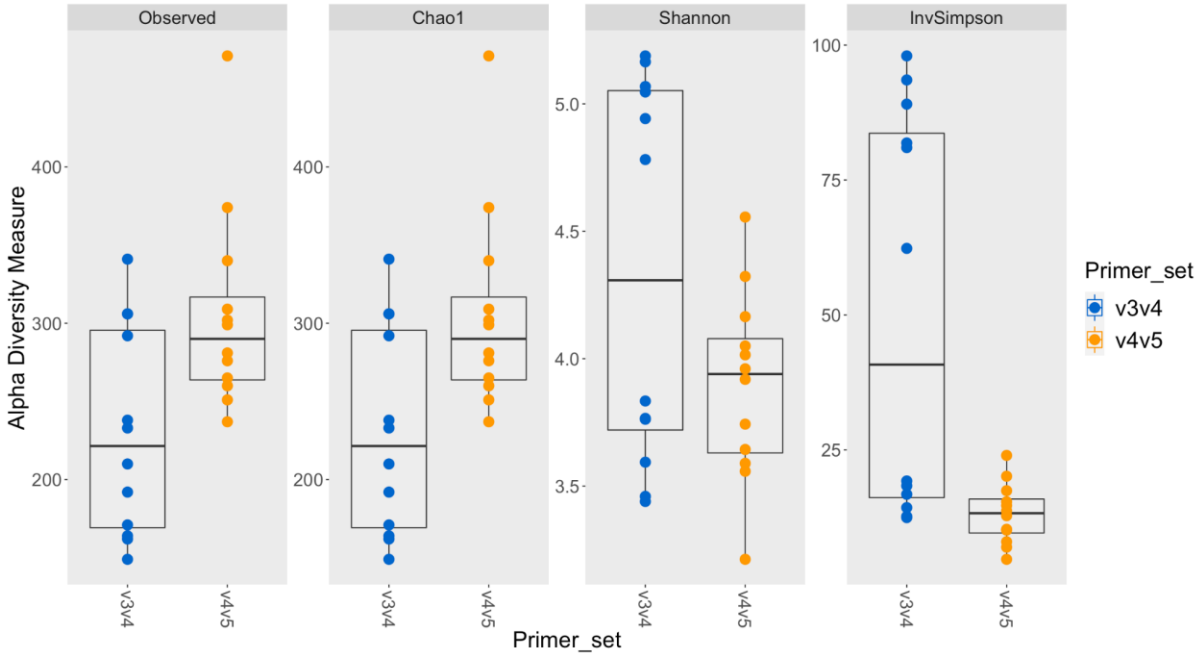


Figure 2 - Observed richness, Chao1, Shannon, and Inverse Simpson indices of alpha diversity among blow microbiome. Different primer sets were represented by different colours (blue for V3-V4 and orange for V4-V5).

The NMDS plot for beta diversity of microbial communities depicted a separation of the samples based on the used primer set, with 4 distinct main clusters, two for each primer set (Fig. 3). One of the clusters of the V3-V4 dataset was composed of samples Gma_01a, 02, 03, 04, 05, and 06, and the other cluster with samples Gma_01b, 07a, 07b, P1, P2, and P3. In the V4-V5 dataset, one of the clusters was composed of Gma_01a, 01b, 02, 03, 04, and 05 samples, and the other one of Gma_06, 07a, 07b, P1, P2, and P3. Blow samples from individuals that were travelling together clustered in the same group for the V4-V5 dataset, which was not the case for V3-V4 dataset. The most similar values were obtained for the samples of the individuals Gma_04 and 05 (sampled in the same day and in the same group of animals), within the V4-V5 dataset.

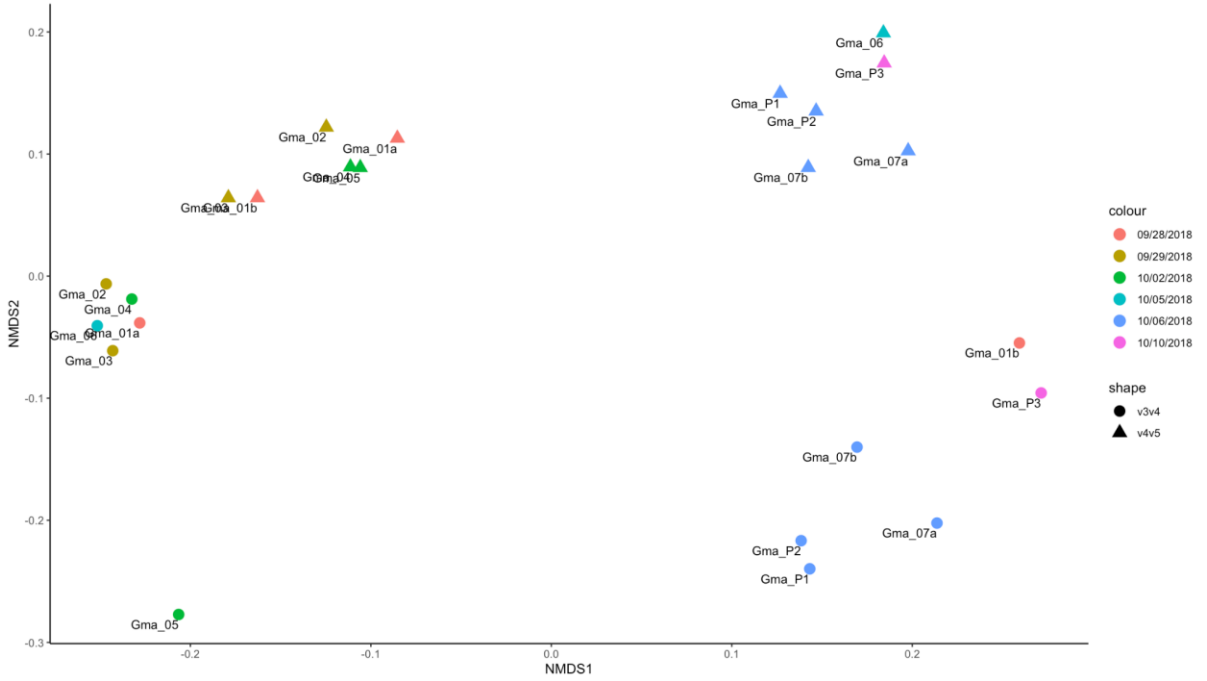


Figure 3 - Non-metric multidimensional scaling (NMDS) of the microbiota found in short-finned pilot whale blow samples, in merged V3-V4 and V4-V5 datasets, based on Bray-Curtis dissimilarity. The primer sets used are represented by different shapes (circles for V3-V4 and triangles for V4-V5) and the days of sampling by different colours.

Regarding the PERMANOVA test results, the primer set used influenced the distribution of prokaryotic communities (p=0.006). Specifically, approximately 25% (R^2=0,249) of the variation of the prokaryotic community distribution was explained by the use of different primers (Supplementary Table 3).

Blow taxonomic characterization

The composition of the blow prokaryotic communities was investigated on the two datasets (V3-V4 and V4-V5) to analyse the differences and commonalities between the use of the two primer sets. Considering the richness, a total of 1 890 bacterial and 38 archaeal taxa were recorded in the V3-V4 dataset, whereas V4-V5 recorded 2 327 bacterial and 30 archaeal taxa. Taxonomic analysis at phylum level (Fig. 4A) revealed a core of 4 most abundant phyla identified by the two datasets: Actinobacteria,

Bacteroidetes, Firmicutes, and Proteobacteria. In the V3-V4 dataset, the prokaryotic community was distributed across 15 different phyla, 7 of them had more than 1% of the total proportion of all the ASVs sequences of this dataset. This dataset was dominated by sequences of the phylum Proteobacteria (47%). On the other hand, V4-V5 was distributed within 21 phyla, with only 5 phyla with more than 1% of the total proportion of the V4-V5 dataset. This dataset was dominated by Proteobacteria (37%) and Actinobacteria (34%). Cyanobacteria was present in more samples of the V4-V5 dataset.

At the Class level (Supplementary Fig. 1), the V3-V4 dataset included 37 different classes (11 of them were highly abundant, with >1% of the sequence reads), and it was dominated by Gammaproteobacteria (31%), followed by the same proportion of Actinobacteria and Alphaproteobacteria (15%). The V4-V5 dataset was dominated by the Actinobacteria class (34%), with low abundance of the majority of the sequence reads (representing <1%) (30 out of 36). The V4-V5 dataset was mainly composed of 6 classes (Actinobacteria, Gammaproteobacteria, Alphaproteobacteria, Bacilli, Bacteroidia, and Oxyphoyobacteria), whereas the V3-V4 dataset had in its composition the same taxa plus Gracilibacteria, Campylobacteria, Saccharimonadia, and Clostridia. Within the classes Actinobacteria, Alphaproteobacteria, Bacteroidia, Clostridia, Gammaproteobacteria, Gracilibacteria, Parcubacteria, and Saccharimonadia, there were many differences between datasets in the number of observed ASVs. These classes belong to the top 10 of both datasets, except Clostridia (position 11 of V3-V4 dataset), Gracilibacteria (position 13 of V4-V5 dataset), and Saccharimonadia (position 27 of V4-V5 dataset).

Regarding the family level (Supplementary Fig. 2), the prokaryotic community was distributed within 108 different families, and 24 families had more than 1% of the total proportion. This dataset was dominated by sequences of the family *Pseudomonadaceae* (18%), whereas V4-V5 was distributed by 153 families, and only 17 families with more than 1% of the total proportion. Despite the fact that the V4-V5 dataset presented more families, most of them had a relative abundance <1%. *Propionibacteriaceae* (25%) was the predominant family of this dataset. A comparison between the different datasets regarding the sequence proportion of the major taxonomic families identified can be found in the supplementary material (Supplementary Fig. 3).

Concerning the genus level (Fig. 4B), the V3-V4 dataset detected 24 genera (with more than 1% of the total proportion of the sequences) associated with 128 different ASVs. This dataset was dominated by sequences of the genus *Pseudomonas* (18%) with only 3 ASVs associated with this genus. On the other hand, the V4-V5 dataset had only 14 genera (with more than 1% of the total proportion) associated with 79 different ASVs and dominated by *Cutibacterium* (25%). The top-10 genera included some taxa that are common to both datasets: *Cutibacterium*, *Dyella*, *Flavobacterium*, *Prochlorococcus_MIT9313*, *Pseudomonas*, *Rhodococcus*, and *Shpingomonas*. In the V3–V4 dataset, the ASVs were merged into 104 different genera and 53 lineages that were affiliated to higher taxonomic ranks. In the V4–V5 dataset, the ASVs were merged into 184 different genera and 68 lineages that were affiliated with higher taxonomic ranks. Overall, at this level, 110 (36.8% of the total) lineages were observed in both datasets. Considering all the taxa, in the V3–V4 dataset, 47 (15.7% of the total)

lineages were absent from the V4–V5 dataset. Moreover, in the V4–V5 dataset there were 142 lineages (47.5% of the total) that were absent from the V3–V4 dataset.

At the family and genus levels, it is possible to observe two clusters of samples in terms of composition: Gma_01a, 02, 03, 04, 05, 06 and Gma_01b, 07a, 07b, P1, P2, P3 in V3-V4; Gma_01 to 05 and Gma_06 to P3 in V4-V5 (Supplementary Fig. 2 and 4B, respectively).

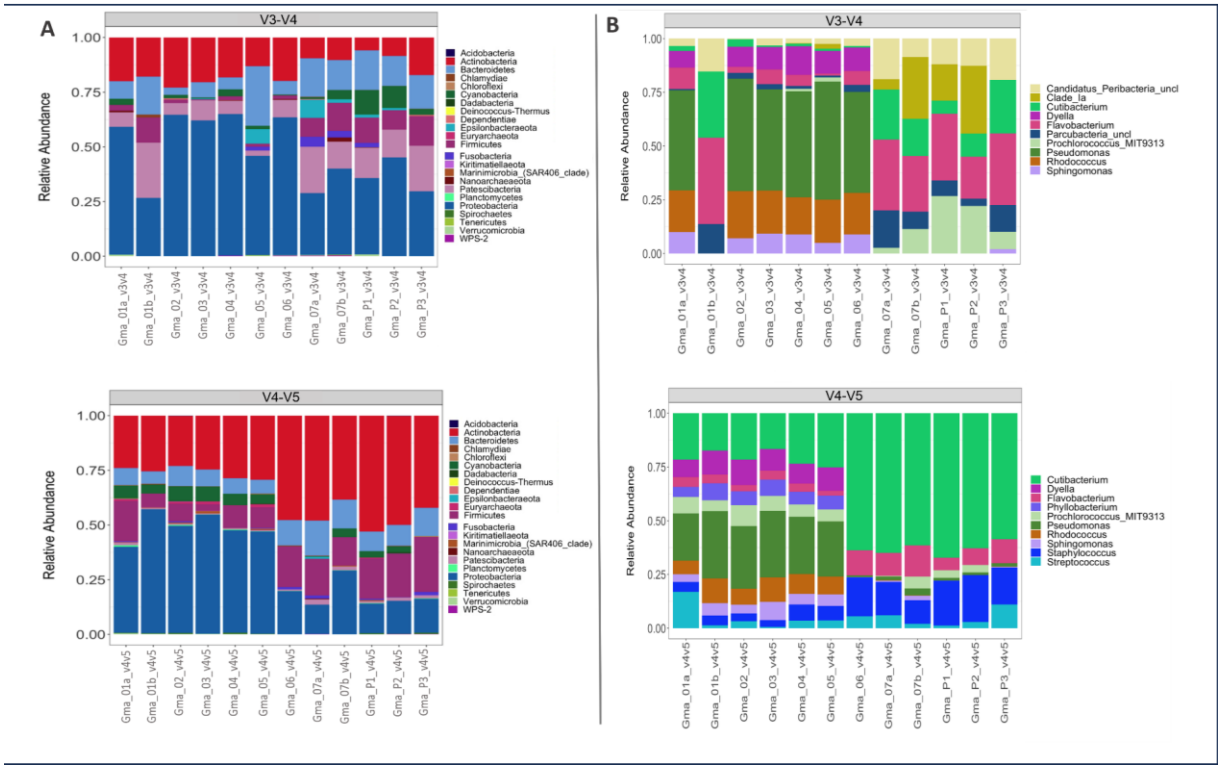


Figure 4 - Relative abundance of prokaryotic phyla (A) and top 10 genera (B) identified across the SFPW blow samples, by V3-V4 (on the left) and V4-V5 (on the right) datasets.

Core microbiome

The NMDS for the core microbiome analysis demonstrated an obvious separation of the samples based on the primer set used, being possible to differentiate 3 distinct main clusters, 2 originated from the V3-V4 dataset and a unique cluster for the V4-V5 dataset (Fig.5).

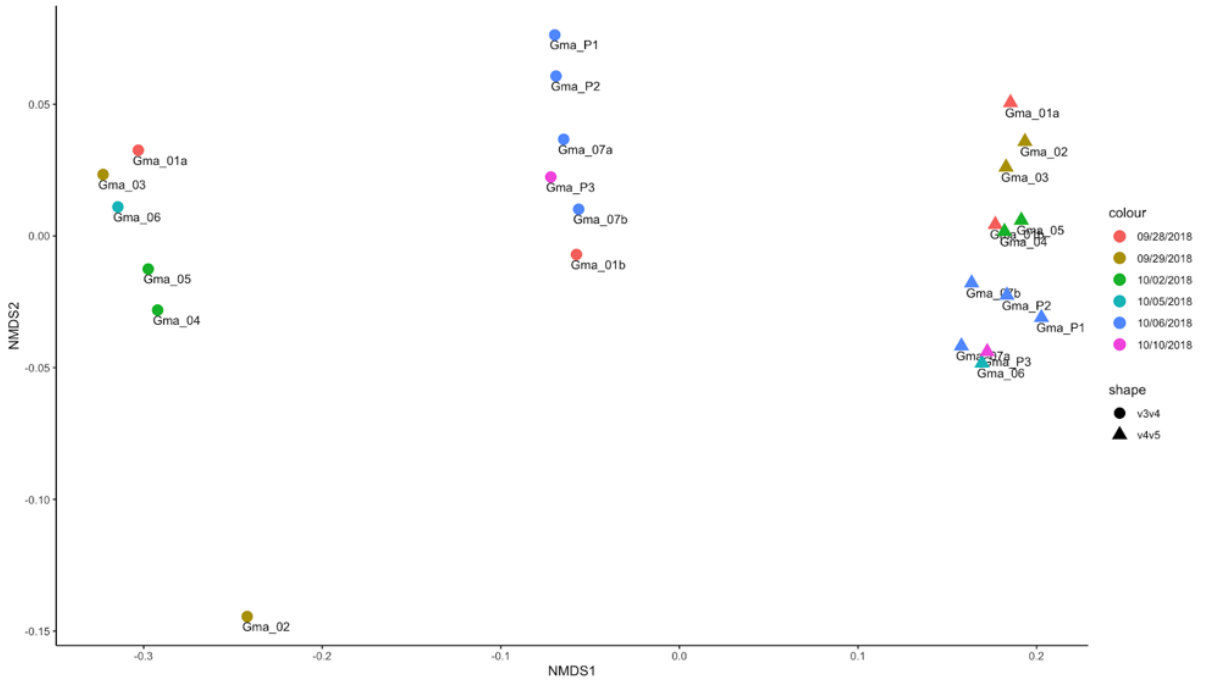


Figure 5 - Non-metric multidimensional scaling (NMDS) of the core microbiota found in short-finned pilot whale blow samples (Genus Level), in merged V3-V4 and V4-V5 datasets, based on Bray-Curtis dissimilarity. The primer sets used are represented by different shapes (circles for V3-V4 and triangles for V4-V5) and the days of sampling by different colours.

Discussion

Sampling methodology

The main goal of non-invasive sampling techniques for cetaceans is to avoid disturbing, injuring, or negatively influencing the sampled individual during sample collection (Robinson & Nuuttila, 2020). Apparently, none of the SFPW sampled here were negatively impacted by the sampling procedure used. This is based on no evident changes in the animals' swimming patterns or anomalous behaviour during sampling, suggesting that the blow-sampling method causes minimal distress to the individuals.

The sampling kit used (PERFORMAgene) has been employed in previous studies, applied to different animals such as livestock (Neary et al., 2014) and companion animals (Sacco et al., 2017; Colpitts et al., 2022). This kit is suitable for genotyping, sequencing, parent-child appraisal, and biobanking. Until now, no studies in the literature have reported the use of the PERFORMAgene kit for sampling the blow of free-ranging cetaceans. In the present study, this kit was tested for the first time for the collection of blow samples from SFPW. DNA was extracted from the swab of the PERFORMAgene kit, and although low concentrations were obtained, the method allowed for the employment of metabarcoding analysis.

Primers set comparison

In this study, the prokaryotic community harboured in the respiratory tract of SFPW was described for the first time, and a comprehensive comparison of the performance of two different 16S primer sets' (V3-V4 and V4-V5) was conducted. Both hypervariable 16S regions are recommended in the literature for assessing marine microbial diversity (Klindworth et al., 2013; Walters et al., 2015).

The use of different hypervariable regions, as well as different types of samples, storage, methods of DNA extraction, and 16S databases, can influence the obtained data and the interpretation of the results. In McNichol et al. (2021), that compared commonly-used primers, with >300 million rRNA gene sequences retrieved from marine metagenomes around the world, the best-performing primers, when comparing predicted median coverage to Bacteria and Archaea of 16S rRNA, were 515Y/926R (regions V4-V5) and 515Y/806RB (region V4). The V4–V5 regions are currently recommended to target marine microbes (including both bacteria and archaea) (Walters et al., 2015). On the other hand, prior studies analysing the blow microbiota of cetaceans have amplified the regions V3-V4 (Bik et al., 2016; Robles-Malagamba et al., 2020). Specifically, Centelleghe et al. (2020) used the primers 341F and 806R (as in this study) to amplify the V3-V4 regions of the 16S rRNA bacterial and archaeal genes. One of the aims of the present research was to assess the advantages and limitations in the amplification of both V3-V4 and V4-V5 regions, regarding prokaryotic diversity and richness, without the influence of technical biases (same experimental and bioinformatic processes on the same samples). Moreover, with this work we also intended to select an optimal PCR primer set (or different applications of each primer set), that can be applied to the study of the microbiome of cetacean blow samples. The results denote that the different hypervariable regions tested provide different degrees of resolution in taxonomic identification, resulting in different estimates of microbial community composition.

Alpha diversity measures captured by each primer set differed significantly. Our results showed that the V4-V5 dataset captured more abundance in unique ASVs (higher values in the Observed ASVs and Chao1 measures) and subsequently, more identified taxa. On the other hand, V3-V4 resulted in higher values for the applied alpha diversity indexes (higher values in Shannon and Inverse Simpson measures). This is probably explained by the fact that these indexes take into account since the used indexes only take into account unique ASVs in relative abundance above 1%. Therefore, although V4-V5 had better results in capturing higher prokaryotic taxa richness. Thus, data generated with this primer set reflects better the prokaryotic community structure of the blow. Beta diversity revealed that samples were grouped according to the different primer sets used. Despite the separation of the samples in different clusters in relation to the primer set used, the V4-V5 dataset appeared to better represent the distribution of prokaryotic communities. In this dataset, individuals that were travelling together when they were sampled (namely Gma_2/Gma_3, Gma_4/Gma_5 and Gma_7/Gma_P1/Gma_P2) appeared within the same cluster and had more approximate values in the NMDS when compared to the V3-V4 dataset. There is some evidence that sociality affects microbes in the respiratory tract, which seems to be the case of the SFPW sampled in the present work. Cetaceans exhibit behaviours, such as surfacing and breathing near each other, or feeding cooperatively, which could facilitate the transfer of microbes and the spread of pathogens between individuals (Bogomolni et al., 2008; Apprill et al., 2017). Such behaviours are common in highly social and matrilineal species, such as the SFPW (Olson, 2018; Boran & Heimlich, 2019), as demonstrated in the target population (Alves et al., 2013; Esteban et al., 2022) to which the sampled animals belong. This process has been recognized as an exclusive and important aspect of social living, providing health benefits to animals. Access to associated microbes is hypothesised to be a driving force in the evolution of sociality (Lombardo, 2008). Nevertheless, this hypothesis needs further

evidence, since the V3-V4 dataset does not corroborate these results, possibly not reflecting the true blow prokaryotic community composition. Vendl et al. (2020) targeted solely the V4 region to study the microbiome in the blow of different whale species and observed a species-specific clustering in the microbiome beta-diversity, also detecting a positive correlation between sociality and microbial diversity. Additionally, as a future challenge, it may be relevant to analyse the microbial composition in the breath of animals with different levels of residence at the site studied.

The analysis of the blow core microbiome could provide useful features for the health monitoring of cetaceans worldwide. All samples from both datasets shared a main core microbiota in their blow, composed of Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria phyla. Nevertheless, the dominant ASVs were not the same between the results obtained from the amplification of V3-V4 and V4-V5 regions. Lima et al. (2012) described a temporal analysis of the blow in captive dolphins, suggesting that microbial community composition in healthy animals is quite stable and that individual dolphins harbour consistently unique microbial communities. Several studies provided preliminary evidence that cetaceans host a core group of bacteria associated with the respiratory system (Johnson et al., 2009; Lima et al., 2012; Bik et al., 2016; Apprill et al., 2017). Apprill et al. (2017) detected, on average, 25 sequence clusters that were found in 100% of humpback whales, accounting for 36% of the microbiota, being one of the most extensive core microbiotas found in any mammal to date. This large core microbiome was shared across individual whales from populations separated geographically into two ocean basins. Out of these core microbes, 20 closest phylogenetic relatives have been previously found in other cetaceans (normally in the mouths and blowholes of bottlenose dolphins), indicating that this core microbiome assemblage is unique to cetaceans and may denote a healthy, non-infected, pulmonary system. Conversely, Vendl et al. (2019) found low microbiota richness and a small core microbiome in humpback whales off the coast of Australia. Those sampled animals were migrating for four months, which is the period of fasting; whereas the animals sampled in the study by Apprill et al. (2017) were at their foraging sites and early stages of migration. Vendl et al. (2019) hypothesised that the lack of a core microbiome might be related to the animals' physiological state at the sampling time. Also, none of the five core genera detected matched with any core found in other studies that investigated cetacean blow (Johnson et al., 2009; Lima et al., 2012; Apprill et al., 2017; Pirotta et al., 2017).

In this study, at a phylum level, the most dominant taxa recovered from both datasets was Proteobacteria. This aligns with previous studies of other species of cetaceans (Apprill et al., 2017; Pirotta et al., 2017; Centelleghe et al., 2018; Nelson et al., 2019; Vendl et al., 2019; Atkinson, 2021) where this taxon is highly abundant. Lima et al. (2012) showed that the aforementioned common phyla, here present in both datasets (Proteobacteria, Actinobacteria, Bacteroidetes, and Firmicutes), plus the Fusobacteria and unclassified bacteria, represented 98% of the total community composition in bottlenose dolphins. In the present work, Fusobacteria was also detected, but in lower abundance (<1% of the total sequence reads) in both datasets, contrasting with the results from Nelson et al. (2019) where this taxon accounted for 7,2% of the total community. The presence of Cyanobacteria, a seawater phylum, in blow samples is not surprising, having been detected in previous studies of the cetacean blow microbiota (Pirotta et al., 2017). The occurrence of prokaryotes common to the

microbiome present in the surface layers of seawater in blow samples is something that has already been reported, namely by Geoghegan et al. (2018) and Centelleghe et al. (2020). In both studies, the authors demonstrate a lower abundance of common prokaryotic taxa between the microbial composition of the blow and seawater, also showing a clear distinction between them. Epsilonbacteraeota, a phylum with a marked presence in the V3-V4 dataset, has been already identified as part of the core microbiota of bottlenose dolphins' respiratory system (Lima et al., 2012). The phyla Euryarchaeota, Dadabacteria, Kiritimatiellaeota, Acidobacteria, and Deinococcus-Thermus, in the present work only found in the V4-V5, were not reported in previous studies investigating the blow microbiota of different cetacean species. The same occurs with Dependentiae in the V3-V4 dataset. Contrary, Planctomycetes, Chloroflexi (Bik et al., 2016), and Spirochaetes (Lima et al., 2012; Bik et al., 2016; Vendl et al., 2021), here only found in the V4-V5 dataset, were described in other studies concerning blow microbiota.

Regarding the genera level, *Cutibacterium* (dominant in the V4-V5 dataset) is a typical dominant microbial community in the nasal microbiota (Kumpitsch et al., 2019). The genus *Prochlorococcus*, detected in both datasets, is one of the most abundant bacteria in the ocean (Zehr et al., 2017), and it is probably a seawater-associated bacteria within the SFPW blow. Four of the most abundant genera detected in both datasets (*Pseudomonas*, *Flavobacterium*, *Rhodococcus*, and *Sphingomonas*) are potential pathogenic genera within the blow microbiome, commonly related to different diseases such as various pulmonary infections (Higgins, 2000; Venn-Watson et al., 2008; Venn-Watson et al., 2012; Apprill et al., 2017). The V4-V5 dataset detected more diversity of the less abundant taxa in the microbiome composition of the blows, therefore highlighting the value of this primer set to detect the “rare biosphere” in the blow. This can be especially relevant, considering the low represented taxa that can potentially have a pathogenic role in causing respiratory diseases. In Nelson et al. (2019), different genera with a potential pathogenic role were a less abundant taxa in the blow samples (*Staphylococcus* with 0.01% relative abundance of identified genera in blow samples; and *Streptococcus* with 0.09%). Identification of taxa with pathogenic potential may therefore be of special relevance in the assessment of the health status of cetacean populations, such as the SFPW targeted in this case study. Here, the identification of various pathogenic potential genera across all samples provides an important insight of the vulnerability of this particular population to respiratory diseases. Other health assessment methodologies such as nucleic acid-derived indices have also recently been tested to study the ecophysiological traits of these animals that commonly occur in the surroundings of the island of Madeira (Alves et al., 2020). In this cited study, the authors concluded that this SFPW population showed good ecophysiological conditions, although significantly lower when compared to other species from the same study area, which could be due to interspecific variations and not environmental conditions. Thus, the present blow microbiome work then provides complementary information to evaluate health status of this marine mammal population.

Viral diseases, such as cetacean morbillivirus (CeMV) infection (Paramyxoviridae family), and algal toxins, have been identified as the main causes of large-scale mass mortality in cetaceans across different geographic areas (Van Bressem et al., 2014). Previous studies reported that CeMV circulates in SFPW, sometimes without relevant pathological changes acting as viral vectors for other other marine mammals in which

CeMV shows higher morbidity and mortality (Di Guardo & Mazzariol, 2016; Sierra et al., 2016). Morbillivirus is thought to be transmitted through exhaled air (Groch et al., 2021). In some areas, short-finned pilot whales associate frequently with bottlenose dolphins, acting as a vector for inter-specific morbilliviruses infections (Olson, 2009).

Regarding the comparison of the core microbiome between the different datasets used, our results show that V4-V5 provides less variation in the data obtained from all the samples, with all the EBC microbiomes showing similarity between them. Furthermore, it is clear that within the same differentiated cluster for this dataset, the NMDS values for the sampled individuals who travelled together (Gma_2/Gma_3, Gma_4/Gma_5 and Gma_7/Gma_P1/Gma_P2) are tendentiously close. This reinforces the idea of the probable and important influence of sociality in the microbiome composition, similarly to what was inferred in the beta-diversity analysis. Therefore, our results suggest that the V4-V5 dataset could be more effective in determining the core microbiome present in the respiratory tract of free-ranging SFPW.

Conclusion

The comparison between the primers sets showed that all samples from both datasets (V3-V4 and V4-V5) shared the main taxa, being composed of Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria phyla. In this study, it was also provided a detailed characterization of the microbial richness present in the blow of SFPW across multiple taxonomic levels. Following this work's results, it is concluded that the selection of the primer set to use for the assessment of microbiomes in cetacean blow samples should depend mainly on the goal of the analyses. If the main goal is to capture more diversity present in higher relative abundance, the V3-V4 primer set is demonstrated to have a better performance; whereas if the purpose is to gather more information in form of unique ASVs and to identify the microbial rare biosphere, we propose the use of the primer set targeting the hypervariable regions V4-V5. Indeed, despite the V4-V5 dataset detected a higher number of unique ASVs and taxa, the majority of them had a relative abundance <1%. The V4-V5 dataset showed better results regarding the determination of the core microbiome in the blow samples used. Additionally, this work provides further proof of sociality being impactful for the microbiome composition of the respiratory tract of cetaceans' species.

Nevertheless, several other aspects require consideration and future development to further advance the blow microbiome as a health monitoring tool for cetaceans. Besides optimisation of the sampling and processing protocols, it is also relevant to test different methodologies in order to enhance sequencing efficiency and downstream procedures. Moreover, crossing this type of data with photogrammetry datasets for assessment of body condition is of relevance to properly infer about the pathogenic potential of these microbial communities in cetacean species.

In conclusion, this work acts as an important first step for a proper understanding of the microbiome existent in the respiratory tract of free-ranging cetaceans. Finally, this study underpins the utility of the EBC microbiome as a future biomarker of health status and physiological state of the airways in free-ranging cetaceans.

References

Acevedo-Whitehouse, K., Rocha-Gosselin, A., & Gendron, D. (2010). A novel non-invasive tool for disease surveillance of free-ranging whales and its relevance to conservation programs. Animal conservation, 13(2): 217-225.

Aguilar de Soto, N., & Alves, F. (2023). Short-Finned Pilot Whale *Globicephala macrorhynchus* Gray, 1846. Handbook of the Mammals of Europe: 1-32. Cham: Springer International Publishing.

Alves, F., Alessandrini, A., Servidio, A., Mendonça, A. S., Hartman, K. L., Prieto, R., ... & Aguilar de Soto, N. (2019). Complex biogeographical patterns support an ecological connectivity network of a large marine predator in the north-east Atlantic. Diversity and Distributions, 25(2): 269-284.

Alves, F., Dromby, M., Baptista, V., Ferreira, R., Correia, A. M., Weyn, M., ... & Teodósio, M. A. (2020). Ecophysiological traits of highly mobile large marine predators inferred from nucleic acid derived indices. Scientific Reports, 10(1): 4752.

Alves, F., Quérouil, S., Dinis, A., Nicolau, C., Ribeiro, C., Freitas, L., ... & Fortuna, C. (2013). Population structure of short-finned pilot whales in the oceanic archipelago of Madeira based on photo-identification and genetic analyses: implications for conservation. *A*quatic Conservation: Marine and freshwater ecosystems, 23(5): 758-776.

Apprill, A., Miller, C. A., Moore, M. J., Durban, J. W., Fearnbach, H., & Barrett-Lennard, L. G. (2017). Extensive Core Microbiome in Drone-Captured Whale Blow Supports a Framework for Health Monitoring. MSystems, 2(5): e00119-17.

Atkinson, S., Rogan, A., Baker, C. S., Dagdag, R., Redlinger, M., Polinski, J., ... & Kerr, I. (2021). Genetic, endocrine, and microbiological assessments of blue, humpback and killer whale health using unoccupied aerial systems. Wildlife Society Bulletin, 45(4): 654-669.

Ballance, L. T. (2018). Cetacean ecology. Encyclopedia of marine mammals (Third Edition): 172-180. Academic Press.

Betty, E.L., Zwamborn, E.M.J., Weyn, M., Luck, E., Alves, F. (2023). Life History Parameters, Sociobiology, and Reproductive Strategies of Pilot Whales. In: Würsig, B., Orbach, D.N. (eds) Sex in Cetaceans. Springer, Cham.

Bik, E. M., Costello, E. K., Switzer, A. D., Callahan, B. J., Holmes, S. P., Wells, R. S., Carlin, K. P., Jensen, E. D., Venn-Watson, S., & Relman, D. A. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nature Communications, 7: 1–13.

Bogomolni, A. L., Gast, R. J., Ellis, J. C., Dennett, M., Pugliares, K. R., Lentell, B. J., & Moore, M. J. (2008). Victims or vectors: A survey of marine vertebrate zoonoses from coastal waters of the Northwest Atlantic. Diseases of Aquatic Organisms, 81(1): 13–38.

Boran, J., & Heimlich, S. (2019). Pilot whales: delphinid matriarchies in deep seas. Ethology and behavioral ecology of odontocetes: 281-304.

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. Ecological monographs, 27(4): 326-349.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME Journal, 11(12): 2639–2643.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods, 13(7): 581-583.

Centelleghe, C., Carraro, L., Gonzalvo, J., Rosso, M., Esposti, E., Gili, C., ... & Mazzariol, S. (2020). The use of Unmanned Aerial Vehicles (UAVs) to sample the blow microbiome of small cetaceans. PLoS One, 15(7): e0235537.

Colpitts, J., McLoughlin, P. D., & Poissant, J. (2022). Runs of homozygosity in Sable Island feral horses reveal the genomic consequences of inbreeding and divergence from domestic breeds. BMC genomics, 23(1): 1-17.

Cuvertoret-Sanz, M., López-Figueroa, C., Byrne, A. O., Canturri, A., Martí-Garcia, B., Pintado, E., ... & Domingo, M. (2020). Causes of cetacean stranding and death on the Catalonian coast (western Mediterranean Sea), 2012-2019. Diseases of aquatic organisms, 142: 239-253.

De Mello, D. M. D., & De Oliveira, C. A. (2016). Biological matrices for sampling free-ranging cetaceans and the implications of their use for reproductive endocrine monitoring. Mammal Review, 46(2): 77-91.

Di Guardo, G., & Mazzariol, S. (2016). Cetacean Morbillivirus-associated pathology: Knowns and unknowns. Frontiers in Microbiology, 7: 1–5.

Díaz-Delgado, J., Fernández, A., Sierra, E., Sacchini, S., Andrada, M., Vela, A. I., ... & Arbelo, M. (2018). Pathologic findings and causes of death of stranded cetaceans in the Canary Islands (2006-2012). PloS One, 13(10): e0204444.

Evans, P. G. (2018). Habitat pressures. Encyclopedia of marine mammals (Third Edition): 441-446. Academic Press.

Fadeev, E., Cardozo-Mino, M. G., Rapp, J. Z., Bienhold, C., Salter, I., Salman-Carvalho, V., ... & Boetius, A. (2021). Comparison of two 16S rRNA primers (V3–V4 and V4–V5) for studies of arctic microbial communities. Frontiers in microbiology, 12: 637526.

Geoghegan, J. L., Pirotta, V., Harvey, E., Smith, A., Buchmann, J. P., Ostrowski, M., ... & Holmes, E. C. (2018). Virological sampling of inaccessible wildlife with drones. Viruses, 10(6): 300.

Groch, K. R., Blazquez, D. N., Marcondes, M. C., Santos, J., Colosio, A., Díaz Delgado, J., & Catão-Dias, J. L. (2021). Cetacean morbillivirus in Humpback whales' exhaled breath. Transboundary and Emerging Diseases, 68(4): 1736-1743.

Higgins, R. (2000). Bacteria and fungi of marine mammals: a review. *T*he Canadian veterinary journal, 41(2): 105.

Hunt, K. E., Moore, M. J., Rolland, R. M., Kellar, N. M., Hall, A. J., Kershaw, J., ... & Kraus, S. D. (2013). Overcoming the challenges of studying conservation physiology in large whales: a review of available methods. Conservation Physiology, 1(1).

Johnson, W. R., Torralba, M., Fair, P. A., Bossart, G. D., Nelson, K. E., & Morris, P. J. (2009). Novel diversity of bacterial communities associated with bottlenose dolphin upper respiratory tracts. Environmental Microbiology Reports, 1(6): 555–562.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research, 41(1): 1–11.

Kumpitsch, C., Koskinen, K., Schöpf, V., & Moissl-Eichinger, C. (2019). The microbiome of the upper respiratory tract in health and disease. BMC Biology, 17(1): 1–20.

Lima, N., Rogers, T., Acevedo-Whitehouse, K., & Brown, M. V. (2012). Temporal stability and species specificity in bacteria associated with the bottlenose dolphins respiratory system. Environmental microbiology reports, 4(1): 89-96.

Lombardo, M. P. (2008). Access to mutualistic endosymbiotic microbes: An underappreciated benefit of group living. Behavioral Ecology and Sociobiology, 62(4): 479–497.

McNichol, J., Berube, P. M., Biller, S. J., & Fuhrman, J. A. (2021). Evaluating and Improving SSU rRNA PCR Primer Coverage for Bacteria, Archaea, and Eukaryotes Using Metagenomes from Global Ocean Surveys. MSystems, 3(6).

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one, 8(4).

Neary, M. T., Neary, J. M., Lund, G. K., Garry, F. B., Holt, T. N., Mohun, T. J., & Breckenridge, R. A. (2014). A comparison of DNA collection methods in cattle and yaks. Journal of Animal Science, 92(9): 3811-3815.

Nelson, T. M., Wallen, M. M., Bunce, M., Oskam, C. L., Lima, N., Clayton, L., & Mann, J. (2019). Detecting respiratory bacterial communities of wild dolphins: Implications for animal health. Marine Ecology Progress Series, 622: 203–217.

Nicol, C., Bejder, L., Green, L., Johnson, C., Keeling, L., Noren, D., ... & Simmonds, M. (2020). Anthropogenic threats to wild cetacean welfare and a tool to inform policy in this area. Frontiers in Veterinary Science, 7: 57.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... & Wagner, H. (2018). Community ecology package. R package version, 2: 5-2.

Olson, P. A. (2009). Pilot Whales: *Globicephala melas* and *G.macrorhynchus*. Encyclopedia of marine mammals (Second edition): 847–852. Academic Press.

Parsons, E. C. M., Baulch, S., Bechshoft, T., Bellazzi, G., Bouchet, P., Cosentino, A. M., ... & Sutherland, W. J. (2015). Key research questions of global importance for cetacean conservation. Endangered Species Research, 27(2): 113-118.

Pirotta, V., Smith, A., Ostrowski, M., Russell, D., Jonsen, I. D., Grech, A., & Harcourt, R. (2017). An economical custom-built drone for assessing whale health. Frontiers in Marine Science, 425.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research, 41(D1): 590–596.

Robinson, C. V., & Nuuttila, H. K. (2020). Don't hold your breath: Limited DNA capture using non-invasive blow sampling for small cetaceans. Aquatic Mammals, 46(1): 32–41.

Robles-Malagamba, M. J., Walsh, M. T., Ahasan, M. S., Thompson, P., Wells, R. S., Jobin, C., Fodor, A. A., Winglee, K., & Waltzek, T. B. (2020). Characterization of the bacterial microbiome among free-ranging bottlenose dolphins (*Tursiops truncatus*). Heliyon, 6(6): e03944.

Sacco, J., Ruplin, A., Skonieczny, P., & Ohman, M. (2017). Polymorphisms in the canine monoamine oxidase a (MAOA) gene: identification and variation among five broad dog breed groups. Canine genetics and epidemiology, 4(1): 1-8.

Sambolino, A., Alves, F., Fernandez, M., Krakauer, A. B., Ferreira, R., & Dinis, A. (2022). Spatial and temporal characterization of the exposure of island-associated cetacean populations to whale-watching in Madeira Island (NE Atlantic). Regional Studies in Marine Science, 49: 102084.

Sierra, E., Fernández, A., Suárez-Santana, C., Xuriach, A., Zucca, D., Bernaldo De Quirós, Y., García-Álvarez, N., De La Fuente, J., Sacchini, S., Andrada, M., Díaz-Delgado, J., & Arbelo, M. (2016). Morbillivirus and pilot whale deaths, Canary Islands, Spain, 2015. Emerging Infectious Diseases, 22(4): 740–742.

Van Bressem, M. F., Duignan, P. J., Banyard, A., Barbieri, M., Colegrove, K. M., de Guise, S., di Guardo, G., Dobson, A., Domingo, M., Fauquier, D., Fernandez, A., Goldstein, T., Grenfell, B., Groch, K. R., Gulland, F., Jensen, B. A., Jepson, P. D., Hall, A., Kuiken, T., … Wellehan, J. F. X. (2014). Cetacean morbillivirus: Current knowledge and future directions. Viruses, 6(12): 5145–5181.

Vendl, C., Ferrari, B. C., Thomas, T., Slavich, E., Zhang, E., Nelson, T., & Rogers, T. (2019). Interannual comparison of core taxa and community composition of the blow

microbiota from East Australian humpback whales. FEMS Microbiology Ecology, 95(8): 102.

Vendl, C., Nelson, T., Ferrari, B., Thomas, T., & Rogers, T. (2021). Highly abundant core taxa in the blow within and across captive bottlenose dolphins provide evidence for a temporally stable airway microbiota. *BMC microbiology, 21*(1), 1-15.

Vendl, C., Slavich, E., Nelson, T., Acevedo-Whitehouse, K., Montgomery, K., Ferrari, B., ... & Rogers, T. (2020). Does sociality drive diversity and composition of airway microbiota in cetaceans?. *Environmental Microbiology Reports, 12*(3), 324-333.

Venn-Watson, S., Daniels, R., & Smith, C. (2012). Thirty year retrospective evaluation of pneumonia in a bottlenose dolphin *Tursiops truncatus* population. Diseases of aquatic organisms, 99(3): 237-242.

Venn-Watson, S., Smith, C. R., & Jensen, E. D. (2008). Primary bacterial pathogens in bottlenose dolphins *Tursiops truncatus*: needles in haystacks of commensal and environmental microbes. Diseases of aquatic organisms, 79(2): 87-93.

Walters, W., Hyde, E. R., Berg-lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. a, & Jansson, J. K. (2015). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. MSystems, 1(1): e0009-15.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43): 1686.

Zehr, J. P., Weitz, J. S., & Joint, I. (2017). How microbes survive in the open ocean. Science, 357(6352): 646–647.

Supplementary material

Supplementary Table 1 - Metadata of the blow samples collected and analysed: Sample ID, sampling date, geographical coordinates (DD), behaviour, age class, number of individuals sampled (1/pool) and residency pattern. Residency pattern was confirmed via Photo-ID, comparing with (OOM/MARE-ARDITI) Madeira's photographic-ID catalogue of the species following Alves et al. (2020): Resident - individuals that exhibited multiyear and year-round site fidelity, i.e. photographed in at least 4 years and in all 4 seasons; Visitor - individuals that exhibited multi-year but seasonal specific presence between July and December (i.e., covering 2 following seasons); Extended Visitor - individuals that exhibited multi-year but seasonal specific presence between July and March (i.e., covering 3 following seasons); Transient - individuals photographed once (i.e. not previously catalogued) and not mixed with catalogued.

Sample ID	Sampling date	Latitude (DD)	Longitude (DD)	Natural behaviour	Class age	Number of individuals sampled	Residency pattern
gma_01a *	28/9/2018	32.594066667	-16.96951667	Travelling	Adult	1	Visitor
gma_01b *	28/9/2018	32.594066667	-16.96951667	Travelling	Adult	1	Visitor
gma_02	29/9/2018	32.590233333	-16.90541667	Travelling	Adult	1	Resident
gma_03	29/9/2018	32.590233333	-16.90541667	Travelling	Adult	1	Resident
gma_04	2/10/2018	32.610883333	-17.06	Travelling	Adult	1	Extended Visitor
gma_05	2/10/2018	32.610883333	-17.06	Travelling	Adult	1	Extended Visitor
gma_06	5/10/2018	32.597533333	-16.69801667	Travelling	Adult	1	Resident
gma_07a *	6/10/2018	32.548483333	-16.87531667	Travelling	Adult	1	Transient
gma_07b *	6/10/2018	32.548483333	-16.87531667	Travelling	Adult	1	Transient
gma_P1 **	6/10/2018	32.548483333	-16.87531667	Travelling	Adult	Pool	Visitor
gma_P2 **	6/10/2018	32.548483333	-16.87531667	Travelling	Adult	Pool	Visitor
gma_P3 **	10/10/2018	32.587533333	-16.98825	Travelling	Adult	Pool	Resident

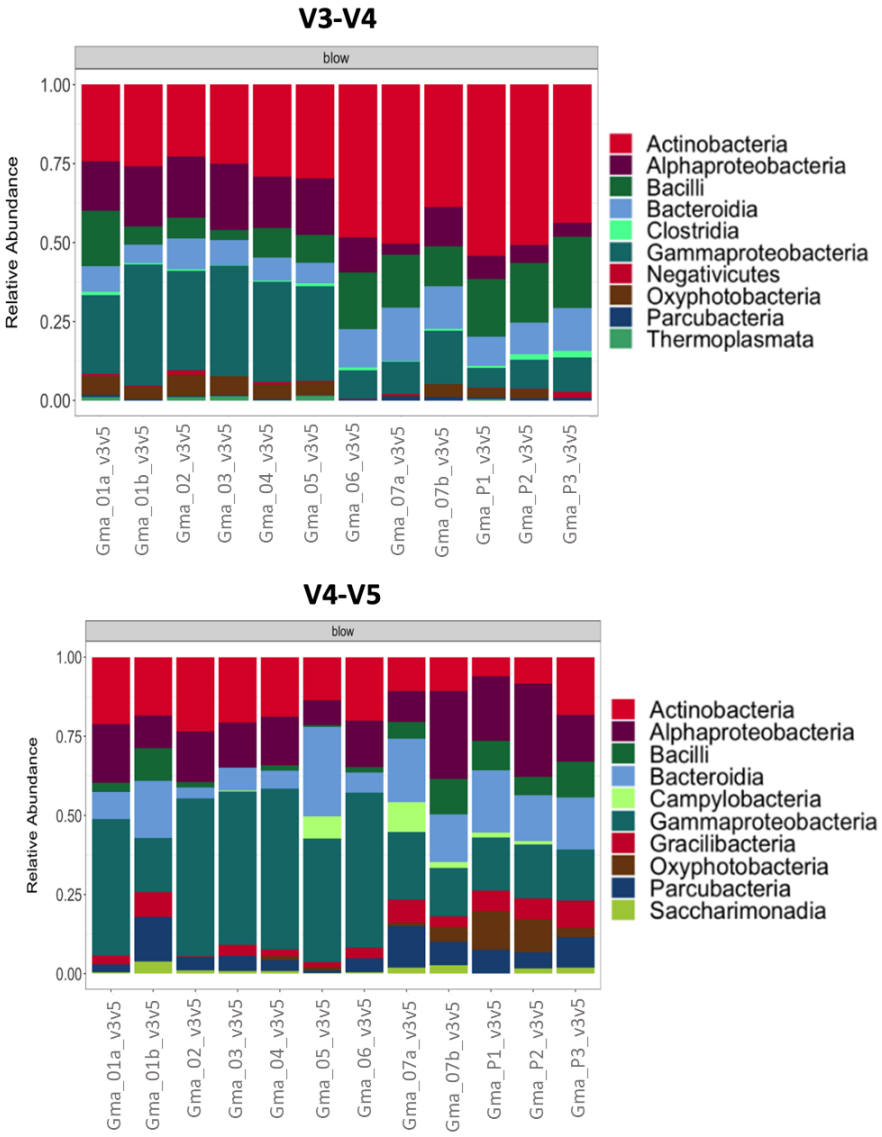
* Samples gma_01a and gma_01b correspond to the same individual, as samples gma_07a and gma_07b; ** Samples gma_P1, gma_P2, gma_P3 were collected from a pool of individuals within the same group

Supplementary Table 2 - Output of a Wilcoxon Signed-Rank Test, comparing the indexes of the short-finned pilot whale blow community richness obtained with two different primer sets (V3-V4 and V4-V5).

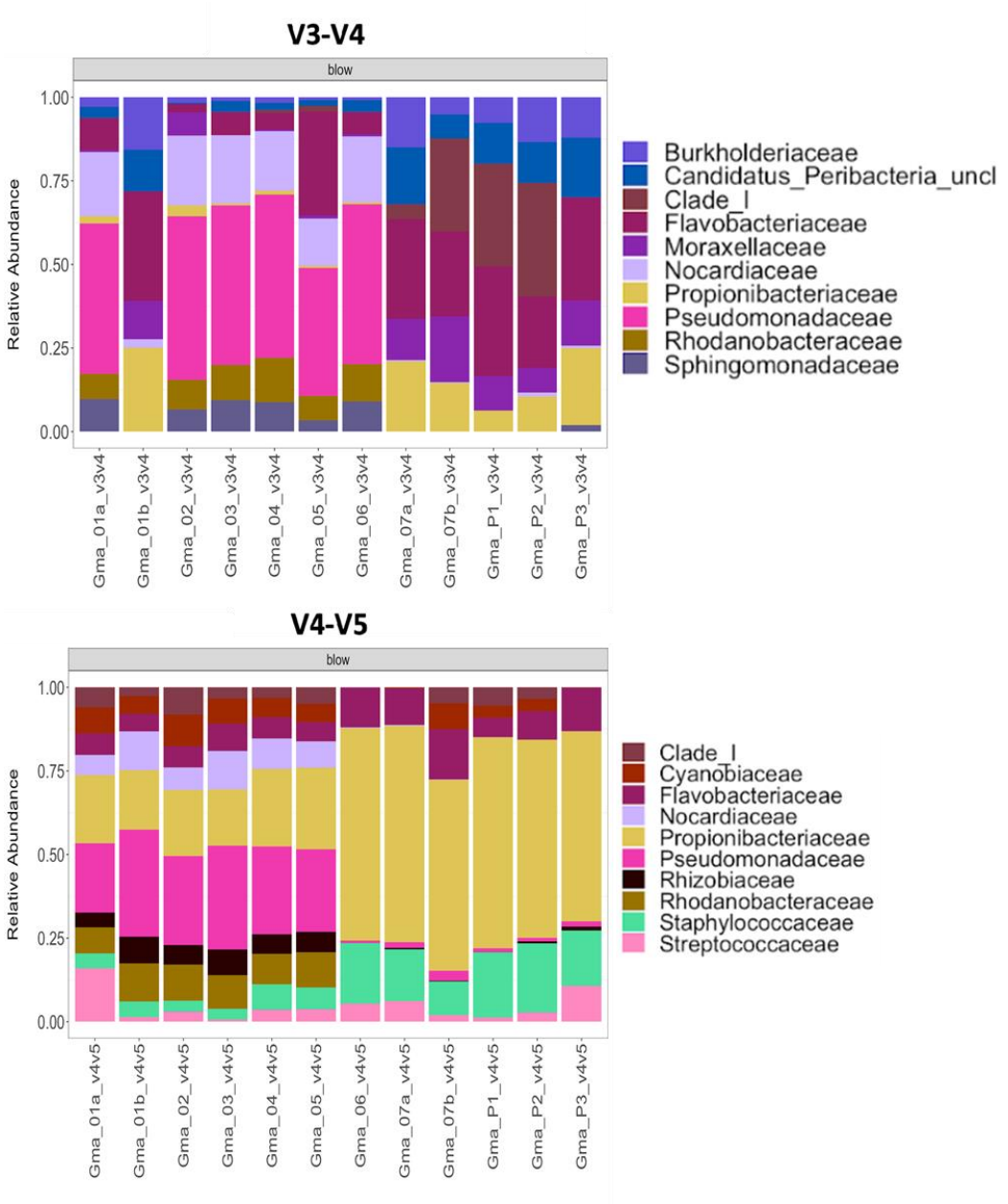
	Index	group1	group 2	n1	n2	statistic	p	p.signif
Blow	Observed	V3-V4	V4-V5	12	12	33	0.0262	*
	Chao1							
	Shannon					91.5	0.273	ns
	Inverse Simpson					118	0.00681	**

Supplementary Table 3 - Output of the PERMANOVA analysis (Df=degrees of freedom, SS=sum of squares, R2= coefficient of determination, PR(>F)=p-value) to test for the influence of the primer sets (V3-V4 and V4-V5) on the distribution of prokaryotic communities of the blows of short-finned pilot whales.

	Df	SS	R^2	F	Pr(>F)
Primer_set	1	0,38948	0,24864	7,2802	0,006
Residual	22	1,17697	0,75136		
Total	23	1,56645	1		

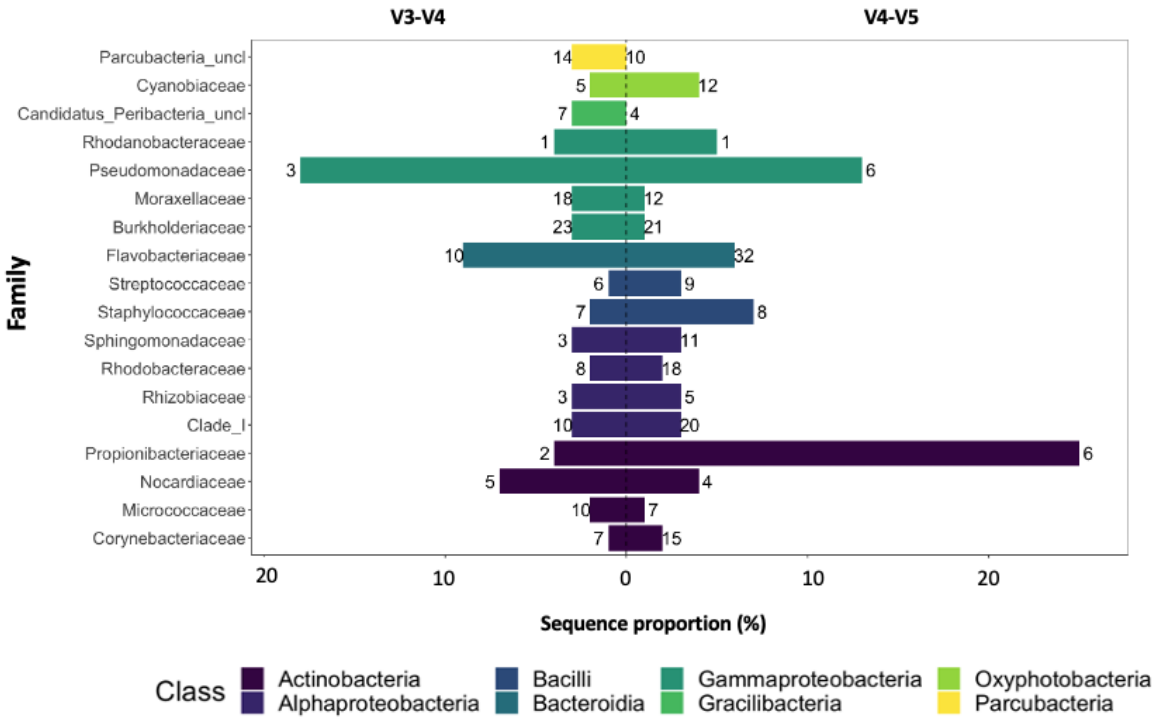


Supplementary Figure 1 - Relative abundance of prokaryotic top 10 classes identified across the SFPW blow samples, by V3-V4 (on the top) and V4-V5 (on the bottom) datasets.



859

860 Supplementary Figure 2 - Relative abundance of prokaryotic top 10 families identified
861 across the SFPW blow samples, by V3-V4 (on the top) and V4-V5 (on the bottom)
862 datasets.



Supplementary Figure 3 - Major taxonomic families obtained for the short-finned pilot whale blow samples (in V3–V4 and V4–V5 datasets). The total sequence proportion of each family in the V3–V4 (left side) and V4–V5 (right side) datasets is represented in the x-axis. The numbers in each column represent the number of observed ASVs affiliated with each taxonomic family. Colour coding represents the different taxonomic classes. Only families that comprised at least 1% of sequences, in at least one of the datasets, were included in the figure.