A survey of the sperm whale (*Physeter catodon*) commensal microbiome

Chang Li [Equal first author, 1, 2], **Xiaoxuan Tan** [Equal first author, 2], **Jie Bai** [2], **Qiwu Xu** [2], **Shanshan Liu** [2, 3], **Wenjie Guo** [2], **Cong Yu** [2], **Guangyi Fan** [2, 4, 5], **Yishan Lu** [6], **He Zhang** [2], **Huanming Yang** [1, 5, 7], **Jianwei Chen** [Corresp., 2], **Xin Liu** [Corresp. 2, 5, 7, 8]

[1] BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China

[2] BGI-Qingdao, BGI-Shenzhen, Qingdao, China

[3] Shandong Technology Innovation Center of Synthetic Biology, Qingdao, China

[4] State Key Laboratory of Quality Research of Chinese Medicine and Institute of Chinese Medical Sciences, University of Macau, Macau, China

[5] BGI-Shenzhen, Shenzhen, China

[6] Guangdong Ocean University, Shenzhen, China

[7] China National GeneBank, BGI-Shenzhen, Shenzhen, China

[8] State Key Laboratory of Agricultural Genomics, Shenzhen, China

Corresponding Authors: Jianwei Chen, Xin Liu
Email address: chenjianwei@genomics.cn, liuxin@genomics.cn

Background. Mammalian commensal microbiota play important roles in the health of its host. In comparison to terrestrial mammals, commensal microbiota of marine mammals is mainly focused on the composition and function of skin and gut microbiota, with less attention paid to the health impact of bacteria and viruses. Previous studies on sperm whales (*Physeter catodon*) have affirmed their important phylogenetic position; however, studies on their commensal microbiota have not been published, due to difficulty in sample collection.

Methods. Here, we sequenced the metagenomes of blood, muscle and fecal samples from a stranded sperm whale using the BGISEQ-500 platform. We compared the diversity and abundance of microbiomes from three different tissues and tried to search pathogenic bacterial and virulence genes probably related to the health of the sperm whale. We also performed 16S rDNA sequencing of the fecal sample to compare to published gut metagenome data from other marine mammals.

Results. Our results demonstrated notable differences in species richness and abundance in the three samples. Extensive bacteria, including *Enterococcus faecium*, *Fusobacterium nucleatum*, *Pseudomonas aeruginosa*, *Streptococcus anginosus*, *Streptococcus pneumoniae*, and *Streptococcus suis*, and five toxigenic *Clostridium* species usually associated with infection, were found in the three samples. We also found the taxa composition of sperm whale gut microbiota was similar to that of other whales, suggesting co-evolution with its host. This study is the first report of the sperm whale gut microbiome, and provides a foundation for the pathogen detection and health assessment of the sperm whale.

A survey of the sperm whale (*Physeter catodon*) commensal microbiome

Chang Li[1,2,*], Xiaoxuan Tan[2,*], Jie Bai[2], Qiwu Xu[2], Shanshan Liu[2,3], Wenjie Guo[2], Cong Yu[2], Guangyi Fan[2,4,5], Yishan Lu[6], He Zhang[2], Huanming Yang[1,4,7], Jianwei Chen[2], Xin Liu[2,4,7,8]

[1] BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China.

[2] BGI-Qingdao, BGI-Shenzhen, Qingdao, 266555, China.

[3] Shandong Technology Innovation Center of Synthetic Biology, Qingdao, 266101, China.

[4] BGI-Shenzhen, Shenzhen, 518083, China.

[5] State Key Laboratory of Quality Research of Chinese Medicine and Institute of Chinese Medical Sciences, University of Macau, Macau, 999078, China.

[6] Guangdong Ocean University, Shenzhen, 518120, China.

[7] China National GeneBank, BGI-Shenzhen, Shenzhen, 518120, China.

[8] State Key Laboratory of Agricultural Genomics, Shenzhen, 518120, China.

*These authors contributed equally to this work.

Correspondence authors:

Xin Liu[2,4,7,8], Jianwei Chen[2]

Hengyunshan Road, Qingdao, Shandong, 266555, China.

Email address: liuxin@genomics.cn, chenjianwei@genomics.cn.

Abstract

Background. Mammalian commensal microbiota play important roles in the health of its host. In comparison to terrestrial mammals, commensal microbiota of marine mammals is mainly focused on the composition and function of skin and gut microbiota, with less attention paid to the health

impact of bacteria and viruses. Previous studies on sperm whales (*Physeter catodon*) have affirmed their important phylogenetic position; however, studies on their commensal microbiota have not been published, due to difficulty in sample collection.

Methods. Here, we sequenced the metagenomes of blood, muscle and fecal samples from a stranded sperm whale using the BGISEQ-500 platform. We compared the diversity and abundance of microbiomes from three different tissues and tried to search pathogenic bacterial and virulence genes probably related to the health of the sperm whale. We also performed 16S rDNA sequencing of the fecal sample to compare to published gut metagenome data from other marine mammals.

Results. Our results demonstrated notable differences in species richness and abundance in the three samples. Extensive bacteria, including *Enterococcus faecium*, *Fusobacterium nucleatum*, *Pseudomonas aeruginosa*, *Streptococcus anginosus*, *Streptococcus pneumoniae*, and *Streptococcus suis*, and five toxigenic *Clostridium* species usually associated with infection, were found in the three samples. We also found the taxa composition of sperm whale gut microbiota was similar to that of other whales, suggesting co-evolution with its host. This study is the first report of the sperm whale gut microbiome, and provides a foundation for the pathogen detection and health assessment of the sperm whale.

Introduction

Microorganisms exist widely in mammal bodies and the environment, and they are indispensable to the health of mammals (Cho & Blaser 2012; Human Microbiome Project 2012; McFall-Ngai 2013). Marine mammals are emblematic members of aquatic ecosystems, but marine mammal populations are declining gradually, and may be progressively impacted by climate change, environmental pollution, and direct (e.g. hunting) and indirect (e.g. habitat exploitation) anthropogenic activities (Moore 2008; Pompa 2011). There is growing concern about the current health status and life habitat of marine mammals. The study of marine mammalian symbiotic microbes is important to understand the health of marine mammals and their conservation.

Until now, investigations of marine mammalian symbiotic microbes were mainly focused on

the composition and function of skin microbiota communities in order to examine the potential for a core bacterial community and its variability with specific host or environmental factors (Apprill 2014; Hooper 2019; K. C. Bierlich 2018). The composition of the gut microbiomes of marine mammals have also been extensively studied, focusing on the putative functionality of these symbiotic communities, differences between marine mammals, and correlationship between marine mammals' diet and evolution (Bik 2016; Erwin 2017; Merson 2014; Nelson 2013; Sanders 2015). As far as diseases in marine mammals have been concerned, few studies have been conducted on the health impact of bacteria and viruses on marine mammals, which might be limited due to the difficulties in sample collection and the opportunistic nature of stranding events (Godoy-Vitorino 2017; Van Bressem 2014). More marine mammals' samples are required to confirm these studies and apply this knowledge to conservation and rehabilitation.

The sperm whale (*Physeter catodon*) is the largest toothed whale and the only living member of genus *Physeter* (Marino 2004). As a representative species for *Physeteridae*, *Odontoceti* and *Cetacea*, they have unique adaptations and a phylogenetically important position (Wesley C. Warren 2018). They feed on several species, most notably the squid, but also octopuses, and fish such as demersal rays (Smith & Whitehead 2000). Recently, the chromosome level genome sequence of the sperm whale has been published (Fan 2019), providing a good genetic foundation for studying its conservation and population structure. However, the commensal microbiome of sperm whale based on metagenomic sequencing has not been reported yet. The diversity, composition and structure of sperm whale symbiotic microbiota are not known.

Hence, a survey of the sperm whale commensal microbiome is very important for understanding its specific microorganisms and finding some possible pathogenic microorganisms related to its health. In this study, we collected a female stranded sperm whale and characterized the blood, muscle and fecal microbiome using metagenomic sequencing. We also screened for pathogenic microorganisms and virulence genes which might be related to its health in these samples. Additionally, we investigated its gut microbiomes using 16S community sequence analysis to detect the unique microbiome taxa composition by comparing it to published cetacean

81 microbiomes. Our reference metagenome data and the potential pathogenic microbes could be used
82 to further monitor and evaluate the health of sperm whales and other marine mammals.
83

84 Materials and Methods

85 Sample collection and DNA extraction

86 The present study sampled a living female sperm whale stranded near the bay area of Huizhou
87 City in southern China on 12th March 2017, and we collected a blood sample of about 5 ml from
88 the vein via syringe (Supplementary Figure S1). The whale died on 15th March, approximately 79
89 h after rescue. The dead animal was salvaged and post-mortem analysis was performed according
90 to standard protocols (Kuiken T 1991). We collected fecal and muscle samples during the
91 necropsy. We dissected the intestine with sterile scissors and forceps, then we used a swab to
92 collect about 2g of fecal matter from inside the incision to a sterile plastic tube. To collect a muscle
93 sample, we cut the sperm whale cortex tissue and about 5g of muscle tissue was collected with
94 scissors. Each sample was immediately stored in -20 °C freezers before transport to the laboratory
95 within 12 h for further investigation.
96 The blood sample was combined with RNAlater to dissolve eukaryotic cells, and then
97 centrifuged to obtain bacteria cells to extract DNA. Fecal and muscle sample DNA extractions
98 were performed using the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA, USA). To
99 eliminate RNA contamination, DNase-free RNase was used to treat extracts. Finally, DNA
100 quantity was determined using a Qubit 3.0 fluorometer, and DNA integrity was evaluated by gel
101 electrophoresis (Pan 2018). All lab work in this study was conducted in a sterile flow hood to
102 prevent contamination. Blood and other samples were dealt in different laboratories to prevent
103 same microbiota contamination. This study was approved by the Institutional Review Board of
104 BGI (NO.BGI-R052-3, NO.FT 17160 and NO.SIDSSE-SYLL-MMMBL-01).
105

106 Library construction and sequencing

107 We totally constructed three metagenome pair-end libraries for the three samples and one 16S

rDNA amplicon library for the fecal sample. To construct metagenome sequencing libraries, the extracted DNA was sheared into fragments between 500 bp and ~800 bp in size. Fragments between 150 bp and 250 bp were selected using AMPure XP beads (Agencourt, Beverly, U.S.) and then repaired using T4 DNA polymerase, (ENZYMATICS, Beverly, U.S.) to obtain blunt ends which were then 3-adenlyated to create sticky ends. These DNA fragments were ligated at both ends to T-tailed adapters and amplified for eight cycles. These amplification products were subjected to a single-strand circularization process using T4 DNA Ligase to generate a single-stranded circular DNA library. For constructing 16S V4 rDNA sequencing library, binding sites for sequencing primers of BGISEQ-500 were added at the 5' end of V4 region common primers (515F: GTGCCAGCMGCCGCGGTAA, 806R: GGACTACHVGGGTWTCTAAT). The PCR products with target bands were mixed in equal amounts, and 2% agarose gel was used for electrophoresis and gel cutting, and then were subjected to generate a single-stranded circular DNA library.

All libraries were sequenced on the BGISEQ-500 platform in the paired-end model with 100 bp length reads for metagenome in BGI-Shenzhen and 150 bp for 16S rDNA library in BGI-Qingdao. Different sequencing machines were used for different libraries to prevent same microbiota contamination.

Metagenome assembly and taxonomic assignment.

Adaptor contamination, low quality reads and duplication for metagenomes were filtered from the raw reads by SOAPnuke (1.5.6) (Chen 2018), and the metagenome remaining reads were removed to eliminate host DNA on the basis of sperm whale reference (Fan 2019) using SOAP2 (Version 2.21) (Gu 2013) with the parameters “-v 8 -m 4”. To account for the possibility of contamination with human-associated microorganisms, clean data were mapped to the human genome, and samtools (Version 1.9) with the parameters “view -q 30 -F 4 -F 256” was used to account mapping reads. After removing the host data, the unmapped metagenome data of 16 Gb, 53 Gb and 15 Gb data for the blood, fecal and muscle samples, respectively, were subjected to further analysis. We

135 then used metaSPAdes (Version 3.10.1) (Nurk 2013) to assemble each sample separately.
136 MetaGeneMark (Version 3.38) (Zhu 2010) was used to identify the coding sequences (CDSs) of
137 the assembled sequences.

138 Taxonomic assignment of the predicted genes was carried out using BLAST+ (Version 2.2.26,
139 best match according to BlastN, evalue <1e-5, coverage >50%) by aligning against the prokaryotic
140 representative genomes (Version 201608) and non-redundant nucleotide (NT) database (Version
141 20170616) (Zhao 2018). The reads were then aligned to the predicted genes using Bowtie2
142 (Version 2.2.5) (Langmead & Salzberg 2012), thus the sequence-based gene abundance profiling
143 was calculated using PathoScope (Version 2.0.6) (Francis 2013). The relative abundances of
144 species were calculated from the relative abundance of their respective genes, and the microbes
145 with a relative abundance below 1e-8 were removed to reduce contamination (Guo 2018).

146 The blood, muscle and fecal clean reads were aligned to the VFDB (Virulence Factors of
147 Bacterial Pathogens) database using Bowtie2 (Version 2.2.5) and the blood assembly contigs were
148 aligned using blat (Version 3.2.1). The Virulence Factors identified in the two results that were
149 annotated to *Clostridium novyi* and *Clostridium botulinum* were selected. The core bacterial
150 virulent genes were aligned by MUSCLE (Version 3.8.31) with the same bacteria genome from
151 NCBI and the trees were built by FastTree (Price 2009).

152

153 **16S rDNA community sequence analysis**

154 In order to compare the gut microbiota with that of other marine mammals, we first collected the
155 16S V4 rDNA tags of the fecal samples of six species from the NCBI and EMBI database (see
156 more data details in Supplementary Table S1). Then the tags were clustered to the OTUs
157 (Operational Taxonomic Units) using USEARCH (Version 7.0.1090) (Rognes 2016) with a 97%
158 threshold. Taxonomic assignment of OTU sequences were classified using Ribosomal Database
159 Project (RDP) Classifier (v.2.2) trained on the RDP database (version trainset14_032015) (Cole
160 2014), using 0.8 confidence values as a cutoff. The data of sperm whale fecal sample 16S rDNA
161 amplicon was discarded to eliminate the low quality and adapter pollution, then paired-end reads

with overlap were merged to tags using FLASH (v1.2.11). Finally, all tags were mapped into the OTU representative sequences using USEARCH GLOBAL to obtain the OTUs and species abundance profiles. Bray-Curtis distance was used to estimate β-diversity (between-sample diversity).

Results

Community diversity and structure of the samples' microbiome.

To obtain a reference metagenomic sequences of the three tissues, we performed a strict data quality control and obtained ~16 Gb, ~53 Gb and ~15 Gb high-quality data of blood, fecal and muscle samples, respectively (Supplementary Table S2). Subsequently, we assembled the metagenomic reads of the three tissues, with an assembly size of ~37 Mb, ~470 Mb and ~151 Mb for blood sample, fecal sample and muscle sample, respectively (Table 1). A total of 1,715 microorganisms were found in the three tissues, among which the microbial species in the fecal samples were the most abundant. There were notable differences in species richness and abundance in the different samples (Figure 1A).

In all three tissues, only 113 microbial species were detectable, and they were those belonging to seven phyla Firmicutes, Bacteroidetes, Deferribacteres, Spirochaetes, Fusobacteria, Proteobacteria and Tenericutes (Supplementary Table S3). A total of 21 phyla and 74 genera were identified in at least two tissues (Figure 1B). Each sample had a dominance of different taxa; for instance, the blood sample was dominated by Firmicutes. In comparison with blood and muscle samples, the fecal sample was more abundant in Bacteroidetes and Firmicutes. The muscle sample shared fewer genera when compared to the other two samples.

Detection of potentially pathogenic microorganisms

To detect potentially pathogenic microbes of the sperm whale, we screened the pathogenic bacterial species in the metagenome sequence data of three tissues. We found pathogenic microbes including *Enterococcus faecium*, *Fusobacterium nucleatum*, *Pseudomonas aeruginosa*,

189 *Streptococcus anginosus*, *Streptococcus pneumoniae*, *Streptococcus suis*, and five toxigenic
190 *Clostridium* in the bloodstream and intestine, and six of them appeared in muscle tissue (Table 2)
191 (Chevalier 2017; Engholm 2017; Fortier 2017; Gao 2018; Han 2015; Okwumabua 2017). Then we
192 aligned the blood, muscle and fecal assembly results and sequencing data against the VFDB
193 databases to identify bacterial virulence genes (Chen LH 2016). Twelve virulence genes were
194 found in *Clostridium novyi* and *Clostridium botulinum*. By aligning the pathogenic genes to all
195 strain sequences of *C. novyi* and *C. botulinum*, we found the nearest bacteria to them were *C. novyi*
196 A str. 4540 (GCA_000724445) and *C. botulinum* C str. Stockholm (GCA_000219255) from
197 Sweden (Figure 2A and 2B).
198
199 **Comparison of gut microbiome in different marine mammals**
200 The composition of mammalian gut microbiota could be shaped by host diet, age and phylogenetic
201 position (Hullar & Fu 2014; Muegge 2011). To profile the unique gut microbiota of sperm whales,
202 we compared the composition of gut microbiome between sperm whales and other marine
203 mammals. A total of 122 bacterial Operational Taxonomic Units (OTUs) were identified in the
204 sperm whale's gut microbiota (Supplementary Table S4). The majority of microbes were members
205 of the phyla Bacteroidetes (51.49% of the total) and Firmicutes (31.18% of the total), as well as
206 the phyla Euryarchaeota, Spirochaetes and Verrucomicrobia (Figure 3A). The sperm whale's
207 sample had either very few or no reads assigned to Proteobacteria, which was comparatively
208 common among six other marine mammals.
209 We also calculated these compositional dissimilarities using β-diversity (Whittaker 1960), and
210 built cluster trees using β-diversity to disentangle the effect of different factors shaping community
211 assembly at different phylogenetic scales. The Bray-Curtis distance in mammalian gut
212 microbiomes showed that the sperm whale was closest to the Beluga Whale (*Delphinapterus*
213 *leucas*) and furthest from the manatee (*Florida manatee*), which is consistent with the biological
214 evolutionary tree of these marine species (Figure 3B) (Groussin 2017).
215

Discussion

Here, the first microbiome inventory of a stranded sperm whale was presented. Among previous metagenomic studies, few have focused on the commensal microbiome of stranded marine mammals, possibly due to the high contamination risk during sample collection (Godoy-Vitorino 2017). In this study, we significantly focused on sperm whale salvage in order to ensure that in the collection of fecal and muscle samples, there was no contamination by the external environment. To monitor DNA extraction kits and other laboratory reagent contamination (Lusk 2014; Susannah J Salter 2014), we set up a negative control for the meta 16S rDNA library (Supplementary Figure S2) and removed microbes with relative abundance below 1e-8 for metagenomic data (Guo 2018). To reduce possibility contamination with human-associated microorganisms, we removed the reads which can map to human genome with high quality (Hooper 2019). The commensal microbiome of the stranded sperm whale in this study contained five main phyla (>1% relative abundance), which were also reported in pygmy (*Kogia breviceps*) and dwarf (*K. sima*) sperm whales, corroborating the validity of contamination control in this study (Erwin 2017; Godoy-Vitorino 2017).

The pathogens detected in this study, including *E. faecium*, *P. aeruginosa, S. anginosus*, *S. pneumoniae,* S. *suis*, and *F. nucleatum*, are opportunistic pathogens and can cause blood infection and immunity damage (Asam & Spellerberg 2014; Engholm 2017; Gao 2018; Mulcahy 2014; Prithiviraj 2005). *F. nucleatum* and five toxigenic *Clostridium* were found in all three samples, and *F. nucleatum* is associated with a wide spectrum of human diseases (Bashir 2016; Han 2015; Signat 2011). Twelve virulence genes were found in *C. novyi* and *C. botulinum* which were the nearest two highly virulent bacteria of *C. novyi* A str. 4540 (GCA_000724445) and *C. botulinum* C str. Stockholm (GCA_000219255) from Sweden (Figure 2A and 2B). *C. novyi A* can cause Black disease and is characterized as lethal and necrotizing (Kahn 2005; Skarin & Segerman 2014), and *C. botulinum C* (group III) can cause diseases in animals. This suggests that *C. novyi* and *C. botulinum* found in this study may also affect host health (Skarin H 2011).

Many pathogenic bacteria have been found in humans and other economic animals (Baumler

& Sperandio 2016), although the reports of them in marine mammals were limited. Live stranding causes a number of Cetacean deaths (CM Kemper 2005; Karen Evans 2002), and bacterial infection has been suggested to be an important and underestimated factor leading to stranding of marine mammals (Cools 2013; Cowan 2001). The pathogens we detected also can be used to monitor the health of marine mammals. Additionally, we found the pathogenic nematodes *Elaeophora elaphi* in the metagenome data of blood that was supported by our dissection result (Table 2, Supplementary Figure S3) (Hernandez Rodriguez 1986). This also proved that this is a comprehensive method to understand the overall health of the marine mammals.

Additionally, the order of pathogenic species found in fecal samples were also identified in meta 16S rDNA data analysis results. Genomic DNA was extracted from the blood sample and other samples from different laboratories, and metagenomic library construction and meta 16S library construction were also independent. Thus, the pathogenic microorganisms in the three samples were real components of the sperm whale commercial microbiome (Hooper 2019).

Moreover, when comparing the gut microbes of other marine mammals, we performed 16S V4 sequencing for the sperm whale fecal sample instead of using metagenomics data. Since most of the published gut microbe data of marine mammals are 16S data, we maintained consistency with them to ensure the accuracy of the analysis results (Sanders 2015). High abundance of Bacteroidetes and Firmicutes identified in the gut of the sperm whale were similar with pygmy and dwarf sperm whales, indicating similar richness and evenness of gut microbiomes in these closely related species (Erwin 2017; LANGER 2001; Sanders 2015; Zhu 2011). Interestingly, the β -diversity phylogenetic tree is consistent with the evolution of host, suggesting that sperm whales and their gut microbiota have a coevolutionary relationship that was also consistent with reports from previous research (Groussin 2017; Sanders. 2015). In the future, more gut microbiome data of sperm whales and other marine mammals should be studied to identify the structural determinants of gut microbiome composition in marine mammals.

Conclusions

As the first metagenomic survey of sperm whale blood, fecal and muscle microbiome using next-generation sequencing, our results showed the differences in symbiotic microorganisms among the three tissues, and demonstrated the gut microbiota of sperm whales has a coevolutionary relationship with its host. The pathogens we detected in the blood, muscle and fecal samples may cause some health problems in the sperm whale, knowledge which will contribute to the monitoring of the health of marine mammals.

Acknowledgements

We thank Guangdong Ocean University and the Institute of Deep-sea Science and Engineering Chinese Academy of Sciences for collecting samples. We thank Xiangqun Chi and Santasree Banerjee from BGI-Qingdao for revising the manuscript. We thank Shandong Technology Innovation Center of Synthetic Biology for their advises to support this work.

Reference

Apprill A, Robbins J, Eren AM. 2014. Humpback whale populations share a core skin bacterial community: towards a health index for marine mammals? *PLoS One* 9:e90785. 10.1371/journal.pone.0090785

Asam D, and Spellerberg B. 2014. Molecular pathogenicity of Streptococcus anginosus. *Mol Oral Microbiol* 29:145-155. 10.1111/omi.12056

Bashir A, Miskeen AY, Hazari YM. 2016. Fusobacterium nucleatum, inflammation, and immunity: the fire within human gut. *Tumour Biol* 37:2805-2810. 10.1007/s13277-015-4724-0

Baumler AJ, and Sperandio V. 2016. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 535:85-93. 10.1038/nature18849

Bik EM, Costello EK, Switzer AD. 2016. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. *Nat Commun* 7:10516. 10.1038/ncomms10516

Chen LH ZD, Liu B, Yang J and Jin Q. 2016. VFDB 2016 hierarchical and refined dataset for big data analysis-10 years. *Nucleic Acids Res* 44(Database issue):D694-D697.

Chen Y, Chen Y, Shi C. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7:1-6. 10.1093/gigascience/gix120

Chevalier S, Bouffartigues E, Bodilis J. 2017. Structure, function and regulation of Pseudomonas aeruginosa porins. *FEMS Microbiol Rev* 41:698-722. 10.1093/femsre/fux020

Cho I, and Blaser MJ. 2012. The human microbiome: at the interface of health and disease. *Nat Rev Genet* 13:260-270. 10.1038/nrg3182

CM Kemper AF, SE Gibbs , M Hill , M Long. 2005. Cetacean captures, strandings and mortalities in South Australia

1881-2000, with special reference to human interactions. *Australian Mammalogy* 27(1) 37 - 47.

Cole JR, Wang Q, Fish JA, Chai B. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42:D633-642. 10.1093/nar/gkt1244

Cools P, Haelters J, Lopes dos Santos Santiago G. 2013. Edwardsiella tarda sepsis in a live-stranded sperm whale (Physeter macrocephalus). *Vet Microbiol* 166:311-315. 10.1016/j.vetmic.2013.05.020

Cowan DF, House, C., House, J.A. 2001. Handbook of Marine Mammal Medicine. *Public health*:767-778.

Engholm DH, Kilian M, Goodsell DS. 2017. A visual review of the human pathogen Streptococcus pneumoniae. *FEMS Microbiol Rev* 41:854-879. 10.1093/femsre/fux037

Erwin PM, Rhodes RG, Kiser KB, Keenan-Bateman TF. 2017. High diversity and unique composition of gut microbiomes in pygmy (Kogia breviceps) and dwarf (K. sima) sperm whales. *Sci Rep* 7:7205. 10.1038/s41598-017-07425-z

Fan G, Zhang Y, Liu X, Wang J, Sun Z, Sun S. 2019. The first chromosome-level genome for a marine mammal as a resource to study ecology and evolution. *Mol Ecol Resour*. 10.1111/1755-0998.13003

Fortier LC. 2017. The Contribution of Bacteriophages to the Biology and Virulence of Pathogenic Clostridia. *Adv Appl Microbiol* 101:169-200. 10.1016/bs.aambs.2017.05.002

Francis OE, Bendall M, Manimaran S, Hong C. 2013. Pathoscope: species identification and strain attribution with unassembled sequencing data. *Genome Res* 23:1721-1729. 10.1101/gr.150151.112

Gao W, Howden BP, and Stinear TP. 2018. Evolution of virulence in Enterococcus faecium, a hospital-adapted opportunistic pathogen. *Curr Opin Microbiol* 41:76-82. 10.1016/j.mib.2017.11.030

Godoy-Vitorino F, Rodriguez-Hilario A, Alves AL. 2017. The microbiome of a striped dolphin (Stenella coeruleoalba) stranded in Portugal. *Res Microbiol* 168:85-93. 10.1016/j.resmic.2016.08.004

Groussin M, Mazel F, Sanders JG, Smillie CS. 2017. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat Commun* 8:14319. 10.1038/ncomms14319

Gu S, Fang L, and Xu X. 2013. Using SOAPaligner for Short Reads Alignment. *Curr Protoc Bioinformatics* 44:11 11 11-17. 10.1002/0471250953.bi1111s44

Guo M, Chen J, Li Q, Fu Y, Fan G. 2018. Dynamics of Gut Microbiome in Giant Panda Cubs Reveal Transitional Microbes and Pathways in Early Life. *Front Microbiol* 9:3138. 10.3389/fmicb.2018.03138

Han YW. 2015. Fusobacterium nucleatum: a commensal-turned pathogen. *Curr Opin Microbiol* 23:141-147. 10.1016/j.mib.2014.11.013

Hernandez Rodriguez S, Martinez Gomez F. 1986. Elaeophora elaphi n. sp. (Filarioidea: Onchocercidae) parasite of the red deer (Cervus elaphus). With a key of species of the genus Elaeophora. *Ann Parasitol Hum Comp* 61:457-463. 10.1051/parasite/1986614457

Hooper R, Brealey JC. 2019. Host-derived population genomics data provides insights into bacterial and diatom composition of the killer whale skin. *Mol Ecol* 28:484-502. 10.1111/mec.14860

Hullar MA, and Fu BC. 2014. Diet, the gut microbiome, and epigenetics. *Cancer J* 20:170-175. 10.1097/PPO.0000000000000053

Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207-214. 10.1038/nature11234

K. C. Bierlich CM. 2018. Temporal and regional variability in the skin microbiome of humpback whales along the Western Antarctic peninsula. *ENVIRONMENTAL MICROBIOLOGY* 84:e02574-02517.

Kahn C. 2005. Infectious Necrotic Hepatitis (Black disease). *The Merck Veterinary Manual (9th ed) Whitehouse*

Station, New Jersey: Merck & Co ISBN 978-0-911910-50-6. OCLC 57355058.

Karen Evans MM, Mark Hinde. 2002. Three mass strandings of sperm whales (physeter macrocephalus) in southern Australian waters. *MARINE MAMMAL SCIENCE* 18(3):622-643.

Kuiken T HM. 1991. Proceedings of the First ECS Workshop on Cetacean Pathology : dissection techniques and tissue sampling. *Leiden* Netherlands: [European Cetacean Society]; .

LANGER P. 2001. Evidence from the digestive tract on phylogenetic relationships in ungulates and whales. *Institut fuÈr Anatomie & Zellbiologie* 39 (2001) 77-90.

Langmead B, and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357-359. 10.1038/nmeth.1923

Lusk RW. 2014. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS One* 9:e110808. 10.1371/journal.pone.0110808

Marino L. 2004. Cetacean brain evolution: multiplication generates complexity. *International Journal of Comparative Psychology* Vol 17(1), 2004, 1-16.

McFall-Ngai M, Hadfield MG, Bosch TC. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* 110:3229-3236. 10.1073/pnas.1218525110

Meehan CJ, Langille MG, and Beiko RG. 2015. Frailty and the Microbiome. *Interdiscip Top Gerontol Geriatr* 41:54-65. 10.1159/000381162

Merson SD, Ouwerkerk D, Gulino LM. 2014. Variation in the hindgut microbial communities of the Florida manatee, Trichechus manatus latirostris over winter in Crystal River, Florida. *FEMS Microbiol Ecol* 87:601-615. 10.1111/1574-6941.12248

Moore SE. 2008. Marine mammals as ecosystem sentinels. *Journal of Mammalogy* 89:534-540. 10.1644/07-mamm-s-312r1.1

Muegge BD, Kuczynski J, Knights D. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 332:970-974. 10.1126/science.1198719

Mulcahy LR, Isabella VM, and Lewis K. 2014. Pseudomonas aeruginosa biofilms in disease. *Microb Ecol* 68:1-12. 10.1007/s00248-013-0297-x

Nelson TM, Rogers TL, Carlini AR, and Brown MV. 2013. Diet and phylogeny shape the gut microbiota of Antarctic seals: a comparison of wild and captive animals. *Environ Microbiol* 15:1132-1145. 10.1111/1462-2920.12022

Nurk S, Bankevich A, Antipov D. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* 20:714-737. 10.1089/cmb.2013.0084

Okwumabua O, Peterson H, Hsu HM, Bochsler P, and Behr M. 2017. Isolation and partial characterization of Streptococcus suis from clinical cases in cattle. *J Vet Diagn Invest* 29:160-168. 10.1177/1040638717690014

Pan H, Guo R, Zhu J, Wang Q, Ju Y, Xie Y. 2018. A gene catalogue of the Sprague-Dawley rat gut metagenome. *Gigascience* 7. 10.1093/gigascience/giy055

Pompa S, Ehrlich PR, and Ceballos G. 2011. Global distribution and conservation of marine mammals. *Proc Natl Acad Sci U S A* 108:13600-13605. 10.1073/pnas.1101525108

Price MN, Dehal PS, and Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641-1650. 10.1093/molbev/msp077

Prithiviraj B, Bais HP, Weir T. 2005. Down regulation of virulence factors of Pseudomonas aeruginosa by salicylic acid attenuates its virulence on Arabidopsis thaliana and Caenorhabditis elegans. *Infect Immun* 73:5319-5328.

10.1128/IAI.73.9.5319-5328.2005
Rognes T, Flouri T, Nichols B, Quince C, and Mahe F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. 10.7717/peerj.2584
Sanders JG, Beichman AC, Roman J, Scott JJ. 2015. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. *Nat Commun* 6:8285. 10.1038/ncomms9285
Signat B, Roques, C., Poulet, P., Duffaut, D. 2011. Role of Fusobacterium nucleatum in periodontal health and disease. *Curr Issues Mol Biol* 13, 25-36.
Skarin H HT, Westerberg J, Segerman B. 2011. Clostridium botulinum group III a group with dual identity shaped by plasmids phages and mobile elements. *BMC Genomics* 12:185.
Skarin H, and Segerman B. 2014. Plasmidome interchange between Clostridium botulinum, Clostridium novyi and Clostridium haemolyticum converts strains of independent lineages into distinctly different pathogens. *PLoS One* 9:e107777. 10.1371/journal.pone.0107777
Smith SC, and Whitehead H. 2000. The diet of galapagos sperm whales physeter macrocephalus as indicated by fecal sample analysis *MARINE MAMMAL SCIENCE* 16(2):3 15-325.
Susannah J Salter MJC, Elena M Turek, Szymon T Calus. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* 12:87.
Van Bressem MF, Duignan PJ, Banyard A, Barbieri M. 2014. Cetacean morbillivirus: current knowledge and future directions. *Viruses* 6:5145-5181. 10.3390/v6125145
Wesley C. Warren LK, Alana Alexander. 2018. The Novel Evolution of the Sperm Whale Genome. *Genome Biol Evol* 10:1063. 10.1093/gbe/evy069
Whittaker RH. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol Monogr* 30, 279–338.
Zhao L, Huang Y, Lu L, Yang W, Huang T. 2018. Saturated long-chain fatty acid-producing bacteria contribute to enhanced colonic motility in rats. *Microbiome* 6:107. 10.1186/s40168-018-0492-6
Zhu L, Wu Q, Dai J, Zhang S, and Wei F. 2011. Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc Natl Acad Sci U S A* 108:17714-17719. 10.1073/pnas.1017956108
Zhu W, Lomsadze A, and Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* 38:e132. 10.1093/nar/gkq275

Figure 1. **Host-associated microbial communities of sperm whale.** (A) A PCA of species with sperm whale blood, fecal and muscle samples. (B)The taxonomic distribution of microbes detected at least in two tissues by metagenomics analysis of sperm whale blood, fecal and muscle samples. The relative abundance genera from the top 20 abundant genera are shown.

Figure 2. **Phylogeny tree of all strain sequences virulence genes found in *C. botulinum and C. novyi*.** (A) Phylogeny tree of four virulence genes found in *C. botulinum*. All *C. botulinum* strains genome from NCBI were shown. (B) Phylogeny tree of nine virulence genes found in *C. novyi*.

421 All *C. novyi* strains genome from NCBI were shown.

422

423 **Figure 3**. **Gut microbiomes of sperm whale fecal sample.** (A) Phylum composition distribution
424 of sperm whale and six marine mammals' gut microbiomes. Beluga whale (*Delphinapterus leucas*)
425 and dolphin (*Tursiops truncatus*) are tooth whale. Humpback whale (*Megaptera novaeangliae*),
426 sei whale (*Balaenoptera borealis*), and right whale (*Eubalaena glacialis*) are baleen whale.
427 Manatee (*Florida manatee*) is belongs to Sirenia. (B) Phylogenetic tree basic on Bray-Curtis
428 distance in mammalian gut microbiomes.

429

430

Figure 1(on next page)

Host-associated microbial communities of sperm whale.

(A) A PCA of species with sperm whale blood, fecal and muscle samples. (B)The taxonomic distribution of microbes detected at least in two tissues by metagenomics analysis of sperm whale blood, fecal and muscle samples. The relative abundance genera from the top 20 abundant genera are shown.

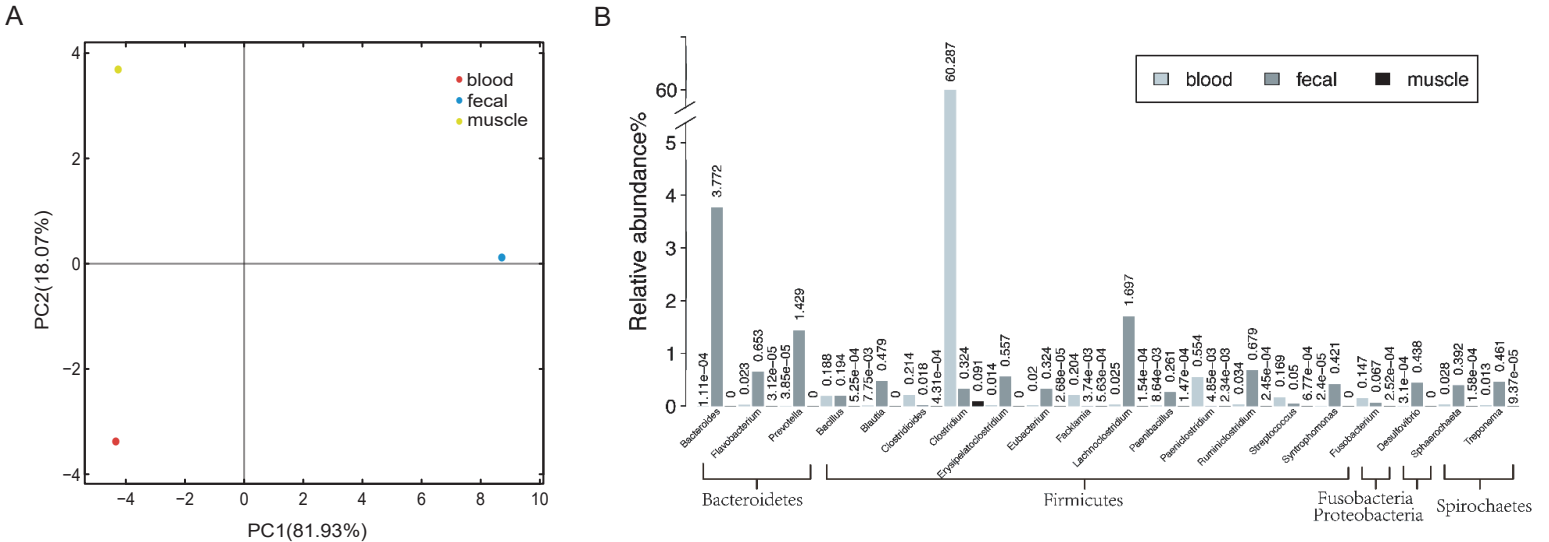

A
blood
fecal
muscle
PC2(18.07%)
PC1(81.93%)
B
blood
fecal
muscle
Relative abundance%
Bacteroides
Flavobacterium
Prevotella
Bacillus
Blautia
Clostridioides
Clostridium
Erysipelatoclostridium
Eubacterium
Facklamia
Lachnoclostridium
Paenibacillus
Paeniclostridium
Ruminiclostridium
Streptococcus
Syntrophomonas
Fusobacterium
Desulfovibrio
Sphaerochaeta
Treponema
Bacteroidetes
Firmicutes
Fusobacteria
Proteobacteria
Spirochaetes

Figure 2(on next page)

Phylogeny tree of all strain sequences virulence genes found in *C. botulinum and C. novyi*.

(A) Phylogeny tree of four virulence genes found in *C. botulinum*. All *C. botulinum* strains genome from NCBI were shown. (B) Phylogeny tree of nine virulence genes found in *C. novyi*. All *C. novyi* strains genome from NCBI were shown.

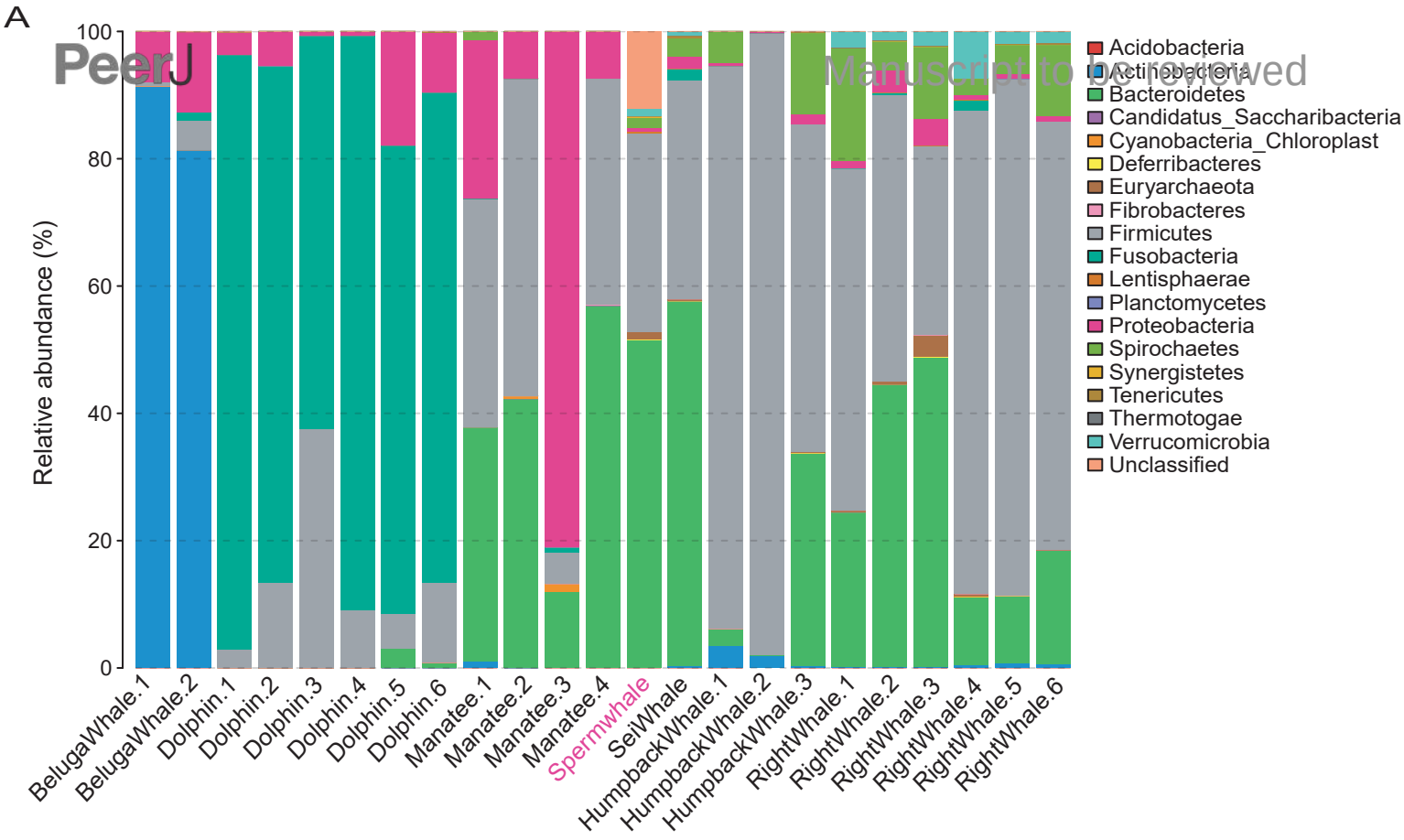
A
C.botulinum 58752-CD
C.botulinum 43243-CD
C.botulinum 12LNRI-CD
C.botulinum BKT015925
C.botulinum 58272-CD
C.botulinum 71840-CD
C.botulinum 69285-CD
C.botulinum 50867-CD
C.botulinum 55741-CD
C.botulinum 29401-CD
C.botulinum 12LNR10-CD
C.botulinum 48212-CD
0. 009
C.botulinum D str. 1873
C.botulinum C str. Stockholm
Sperm whale meta
C.botulinum C str. Eklund
C.botulinum C/D str. BKT12695
C.botulinum D str. 16868
C.botulinum C/D str. It1
B
C.novyi B str. NCTC 9691
C.novyi B str. ATCC 27606
C.novyi A str. GD211209
Sperm whale meta
C.novyi A str. 4540
C.novyi NT
C.novyi A str. NCTC 538
C.novyi A str. BKT29909
C.novyi A str. 4552
C.novyi A str. 4570
0. 02

Figure 3(on next page)

Gut microbiomes of sperm whale fecal sample.

(A) Phylum composition distribution of sperm whale and six marine mammals' gut microbiomes. Beluga whale (*Delphinapterus leucas*) and dolphin (*Tursiops truncatus*) are tooth whale. Humpback whale (*Megaptera novaeangliae*), sei whale (*Balaenoptera borealis*), and right whale (*Eubalaena glacialis*) are baleen whale. Manatee (*Florida manatee*) is belongs to Sirenia. (B) Phylogenetic tree basic on Bray-Curtis distance in mammalian gut microbiomes.

PeerJ Manuscript to be reviewed



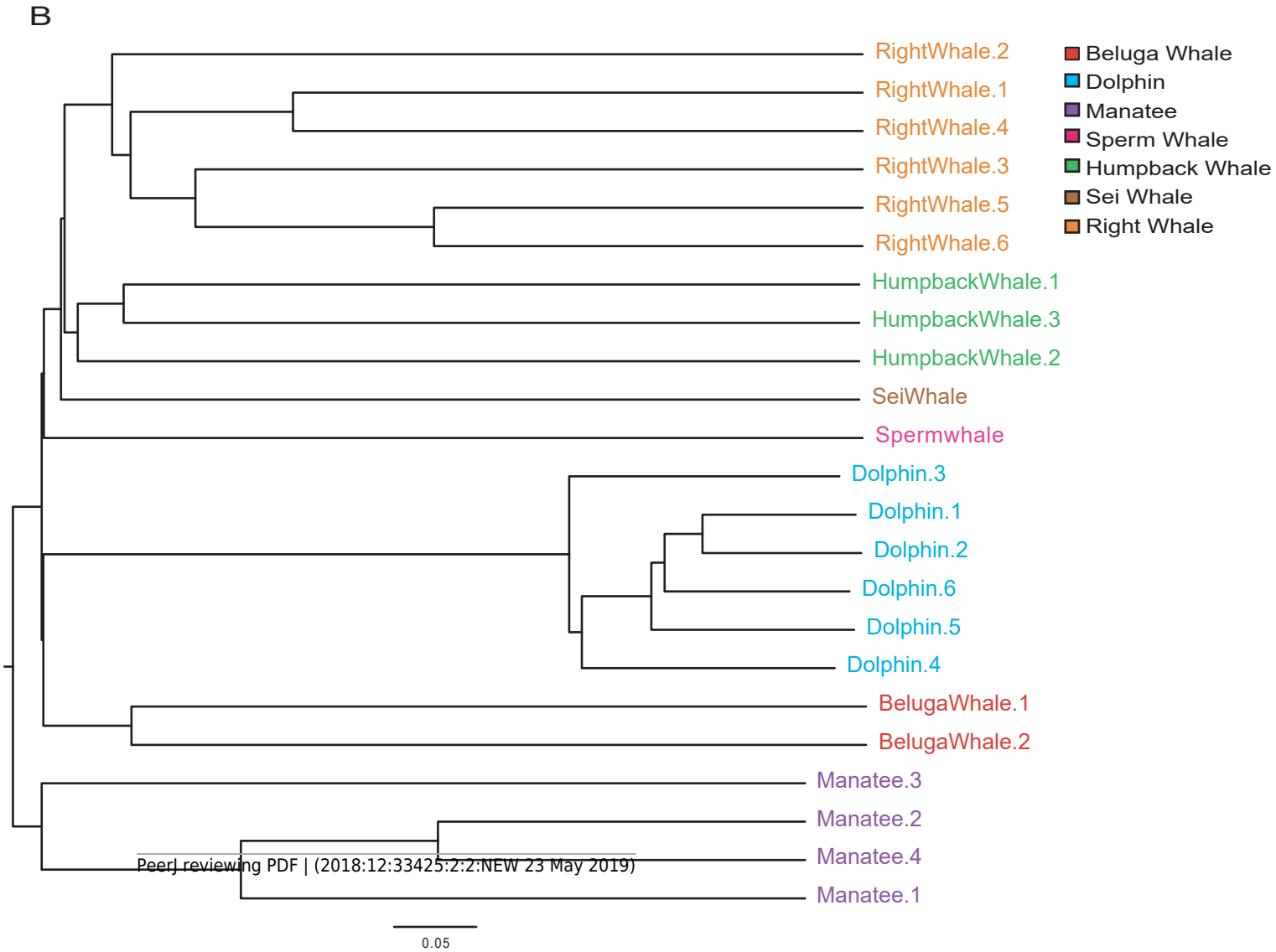


Table 1(on next page)

Statistics of metagenomics assembly for three tissues.

1 **Table 1**. **Statistics of metagenomics assembly for three tissues.**

Statistics	Blood		Fecal		Muscle	
	Contig	CDS	Contig	CDS	Contig	CDS
Total number	40,086	61,992	287,646	628,309	155,131	174,718
Total length(Mb)	37.16	27.83	470.47	375.63	151.50	43.71
N50 length (bp)	1,930	717	5,153	963	2,060	306
N90 length (bp)	309	213	498	294	314	123
Max length (bp)	488,134	10,443	546,844	29,826	112,716	147,900
Min length (bp)	200	57	300	57	200	57
GC content	32.73%	32.94%	48.30%	49.18%	49.97%	52.26%
Mapped reads	91.43%	66.33%	88.87%	57.86%	18.57%	7.11%

2

Table 2(on next page)

Detection of pathogenic bacteria and nematodes in fecal, blood and muscle samples.

1 **Table 2**. **Detection of pathogenic bacteria and nematodes in fecal, blood and muscle samples.**

Species	Blood	Fecal	Muscle
Clostridium baratii	+	+	+
Clostridium botulinum	+	+	+
Clostridium novyi	+	+	+
Clostridium perfringens	+	+	+
Clostridium tetani	+	+	+
Enterococcus faecium	+	+	-
Fusobacterium nucleatum	+	+	+
Pseudomonas aeruginosa	+	+	-
Streptococcus anginosus	+	+	-
Streptococcus pneumoniae	+	+	-
Streptococcus suis	+	+	-
Bacteroides fragilis	-	+	-
Enterococcus faecalis	-	+	-

2