What goes in, must come out: Evaluation of the DNA metabarcoding approach to analyse diet of threatened seahorses

*Tamara Lazic[1,2], Giuseppe Corriero[1], Bachir Balech[3], Frine Cardone[4], Michele Deflorio[5], Bruno Fosso[3], Carmela Gissi[3,6], Marinella Marzano[3], Graziano Pesole[3,6], Monica Santamaria[3], Michele Gristina[7], Cataldo Pierri[1]

[1] *Department of Biology, University of Bari, Bari, Italy*
[2] *CoNISMa, Rome, Italy*
[3] *National Council of Research (CNR), Institute of Biomembrane, Bioenergetics and Molecular Biotechnology (IBIOM), Bari, Italy*
[4] *Department of Integrated Marine Ecology, Zoological Station Anton Dohrn, Naples, Italy*
[5] *Department of Veterinary Medicine, University of Bari, Bari, Italy*
[6] *Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari, Bari, Italy*
[7] *National Council of Research (CNR), Institute of Anthropic Impacts and Sustainability in marine environment (IAS), Palermo, Italy*

***Abstract* – Seahorses are considered flagship species of the conservation efforts. Indeed, due to the worldwide decline of local populations during the last decades, all seahorse species were listed on the IUCN Red List and Appendix II of CITES. Because of the threatening status, improving knowledge of their dietary composition, while using a non-invasive approach, might be of great importance. Starting from faecal samples of the European seahorse *Hippocampus guttulatus* collected during feeding trials, we used, for the first time on these fish, metagenomic amplicon-based HTS (High Throughput Sequencing) approach. The findings indicated the reliability of the present molecular approach, allowing the characterization of the effectively ingested prey. Unlike traditional methods on faecal samples, this technique can identify items that might not leave solid remains. As only a small amount of starting faecal material is needed and the sampling procedure is neither invasive nor lethal, DNA metabarcoding appears to be useful in the investigation on threatened seahorse diet and, in future, could help to define management and conservation actions.**

I. INTRODUCTION

The knowledge of the species diet is a keystone to understand the way it exploits the environment and to develop adequate conservation actions both for species and total biodiversity [1]. The dietary composition is traditionally identified through visual morphological analyses of stomach, guts or faecal contents using light microscopy [2]. Although these techniques can be useful, they generally result in a poor resolution of determining taxa, possibly precluding the identification of food items that do not leave behind hard remains and sometimes require the sacrifice of animals [3]. Furthermore, some potential prey species, such as crustaceans, could be morphologically similar to one another [4], making the process taxonomically challenging [5]. In recent years, remarkable progress has been made towards developing accurate and non-invasive DNA metabarcoding strategies using highly degraded samples such as faeces, making it relevant for dietary studies even in threatened species. This strategy involves the amplification of a standard DNA barcode using universal or group-specific primers pairs targeting multiple taxa [6]. This approach, where complex mixtures of DNA are extracted and sequenced by employing Next Generation Sequencing (NGS) technologies, has been successfully applied to faecal dietary studies in many species, including fish, and had promising results [7] [8].

Seahorses (*Hippocampus* spp.) are small predatory fish with an almost worldwide distribution [9]. In the past decades, their populations have gone under severe declines in many areas, which led to the inclusion of all seahorse species, including the European seahorse *Hippocampus guttulatus*, on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species and in Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora [10]. Seahorses usually rely on their sight to capture prey [11]. They practice “sit-and-wait” predation strategy, which involves an examination of the environment from a hidden place and rapid execution of a surprise attack [12]. By

employing morphological examination of gut or stomach contents, stomach flushing or by biochemical means, studies on their diet have shown that seahorses feed mainly on small-sized crustaceans [13]. Despite the sensitive conservation status of seahorses, there have been no published studies demonstrating the use of non-invasive techniques to explore their diet. Indeed, it has been shown that the prey's DNA is recoverable from seahorse faeces [14], suggesting that the dietary DNA metabarcoding could be applied to foraging ecology of these fish.

Therefore, using faecal samples of the European seahorse *H. guttulatus*, the purpose of this study was to validate the effectiveness of DNA metabarcoding to characterize the seahorse diet, while the developed protocol would serve as a reference for the interpretation of diet and feeding behaviour of wild populations. To achieve this, the present study included feeding trials, for which it has been shown to be crucial in validating molecular techniques for prey detection [15]. Importantly, as *H. guttulatus* is a threatened species in the Mediterranean Sea and along the Italian coast [16], it is of value to develop a non-lethal and non-invasive method to study its diet.

II. MATERIALS AND METHODS

A. Sampling and feeding trials

Four adult non-reproductive *H. guttulatus* specimens were collected by diving at Taranto Mar Piccolo (Southern Italy). Specimens were hand-caught and transported to the facilities of agricultural society ''Ittica Caldoli S.r.L.''where they were placed in individual 30 L aquaria. Seawater inside the aquaria was filtered through 0.2 µm pore-size polycarbonate filters. The seawater temperature was maintained at 18 ± 0.5°C, salinity at 36 ± 1‰ and the photoperiod was adapted to the natural day cycle. Except for commercially bought *Artemia* sp., other three prey items (*Gammarus* sp., *Palaemon* sp. and *Nereis* sp.) were collected at Taranto Mar Piccolo and taxonomically classified under the microscope. Before the beginning of the experiment and between successive feeds with different prey, seahorses were starved for 24 h to ensure an empty gut. Seahorses were fed simultaneously on a single prey species added daily at a single dose, according to the following sequence: *Gammarus* sp. at day 1, *Artemia* sp. at day 3, *Palaemon* sp. at day 5, and *Nereis* sp. at day 7. The faeces (n=10; three from the diet with *Artemia* sp., three from *Palaemon* sp., two from Gammarus sp. and two from *Nereis* sp.) produced in aquaria were immediately collected by syphoning, and together with four prey samples (P_ANF - *Gammarus* sp., P_AR - *Artemia* sp., P_PA - *Palaemon* sp., P_PL - *Nereis* sp.) were preserved in 96% ethanol and stored at -20°C. Faecal samples were named according to the individual from which they were collected (FT1-FT4) and diet applied (_ANF, _AR, _PA or _PL). At the end of the trials, all animals were released to the original capture site.

B. DNA extraction

Total genomic DNA was extracted from faecal and prey samples using FastDNA SPIN kit for soil (BIO 101, Carlsbad, Canada). Cell lysis was achieved by bead beating in a FastPrep Instrument (BIO 101) at speed 6 for 40 s. Negative extraction controls were added, and identical molecular analyses were performed upon these to monitor for possible contamination. The quantity and quality of the extracted DNA were assessed spectrophotometrically and by agarose gel (1%) electrophoresis, respectively. A skin filament tissue sample from *H. guttulatus* was processed under the same conditions to minimize host contamination.

C. Cox1 Library preparation and sequencing

An amplicon-based approach was applied to the extracted DNA. The sub-region of Cox1 gene was amplified using primers mlCOIintF and dgHCO2198 [17]. Amplicon libraries were prepared starting from 2 ng of DNA extracted from each sample. RNase/DNase-free Molecular Biology Grade water (Ambion) was used as a negative control of PCR amplification. The adopted strategy is described in details in [18]. Equimolar quantities of the purified amplicons were pooled and subjected to 2 × 250 bp paired-end sequencing on the Illumina MiSeq platform. To increase genetic diversity of the sequenced samples, as required by the MiSeq platform, a phage PhiX genomic DNA library was added to the mix and co-sequenced.

PCR products of two prey samples (*Artemia* sp. and *Palaemon* sp.) and seahorse's skin filament, obtained using the primer pair mlCOIint and dgHCO [17], were subjected to Sanger DNA sequencing by Eurofins Genomics (www.eurofinsgenomics.com). Due to uncertainty that more than one taxa were present in the samples of *Nereis* sp. and *Gammarus* sp., these were processed and sequenced by the Illumina MiSeq platform together with the faecal samples under the conditions described above.

D. Taxonomic analyses

Two Illumina MiSeq runs were performed to achieve an adequate number of Paired-Ends (PE) reads per sample. Given that the expected amplicons length (~400bp) was shorter than the total sequenced PE length (2 x 250bp), most of the generated readings were overlapping and consequently merged into contiguous consensus sequences using PEAR [19]. ASVs (Amplicon Sequence Variants) were defined by applying DADA2 [20]. Chimera removal was performed using the reference-based VSEARCH [21] procedure. The obtained ASV sequences were taxonomically annotated using BioMaS pipeline [22] and mapped on the BOLD-based reference collection using Bowtie2 [23]. Sequences matching, with an identity

Table 1. Shannon alpha diversity index and Chao1 species richness estimator among analysed samples. Samples: FT1, FT2, FT3 and FT4 refer to one of four individuals used in feeding trials, while ANF, AR, PA and PL stand for the prey given to seahorses (ANF – Gammarus *sp., AR –* Artemia *sp., PA –* Palaemon *sp., PL –* Nereis *sp.)*

SAMPLE	SHANNON INDEX	CHAO1 INDEX
FT2_ANF	0.354	7.000
FT3_ANF	2.018	19.000
FT1_AR	2.245	16.333
FT2_AR	2.019	35.000
FT3_AR	2.155	24.000
FT2_PA	3.109	78.000
FT3_PA	3.527	34.000
FT4_PA	4.848	113.333
FT2_PL	3.803	164.100
FT4_PL	5.000	104.500

percentage of at least 97%, were directed to the genera classification [24], while others were classified at higher taxonomic levels. Unassigned sequences were taxonomically investigated using the BLAST tool [25] [26] against the nucleotide collection at NCBI.

Alpha diversity index (Shannon Index, H Index) and the species richness estimator (Chao1) were calculated using R phyloseq package [27] at the level of ASVs. The diversity among samples' composition (beta diversity) was compared through Principal Coordinate Analysis (PCoA), based on the Bray-Curtis dissimilarity matrix, using Vegan R package [28].

III. RESULTS

Amplicon libraries (from 10 faecal and 2 prey samples) were successfully sequenced on the MiSeq platform using a 2 x 250 bp paired-end sequencing strategy. A total of 4,669,953 paired-end (PE) reads, ranging from 296,279 to 519,847 per sample, were obtained. Approximately 99 % of the PE reads were merged into a consensus sequence, maintaining good quality to pass the quality filtering step. A total of 492 ASVs were identified, ranging from 8 (FT2_ANF) to 224 (FT2_PL) per sample. The values of the Shannon index ranged from 0.354 in sample FT2_ANF to 5.0 in sample FT4_PL (Table 1). Chao1 index varied from 7 in FT2_ANF to 164,1 in FT2_PL (Table 1). Beta

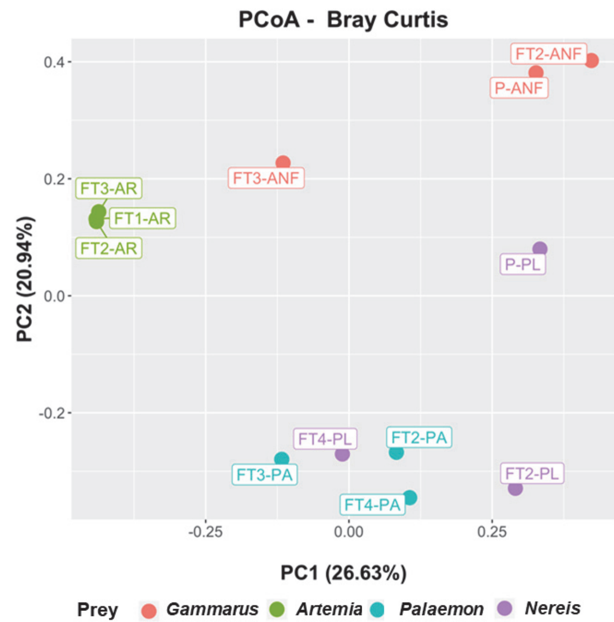

Fig. 1. PCoA obtained using Bray Curtis matrix on ASV. Samples: FT1, FT2, FT3 and FT4 refer to one of four individuals used in feeding trials, P refers to the prey sample, while ANF, AR, PA and PL stand for the prey given to seahorses (ANF – Gammarus *sp., AR –* Artemia *sp., PA –* Palaemon *sp., PL –* Nereis *sp.)*

diversity assessed from Bray-Curtis dissimilarity matrices indicated that the samples were grouped mostly according to the diet applied: samples from the diet with *Gammarus* sp. and *Artemia* sp. were well separated from other samples (*Nereis* sp. and *Palaemon* sp.), while these latter were clustered together (Fig.,1).

The applied approach permitted the identification of one genera in the samples of prey (e.g. *Nereis* sp. and *Gammarus* sp.) while in the faecal samples, all prey items (*Gammarus* sp., *Artemia* sp., *Nereis* sp. and *Palaemon* sp.) given to the seahorses during the trials (Fig.,2A, Fig.,2B). Interestingly, *Gammarus* sp. (applied on day 1) DNA was also detected in two samples after feeding with *Artemia* sp. (day 3) and *Palaemon* sp. (day 5). Unknown taxa represented the most abundant group with Phaeophyceae, Hexanauplia and Branchiopoda present at low abundances. Among them, the presence of contaminant taxa has also been observed considering the detection of classes such as Gastropoda and Insecta (Fig.,2A).

IV. DISCUSSION

Using faecal samples of *H. guttulatus* produced during the feeding trials with known ingested prey, the present study represents the first attempt to investigate the diet of seahorses by DNA metabarcoding. The results demonstrated the reproducibility and sensitivity of a developed molecular protocol, based on a metagenomic

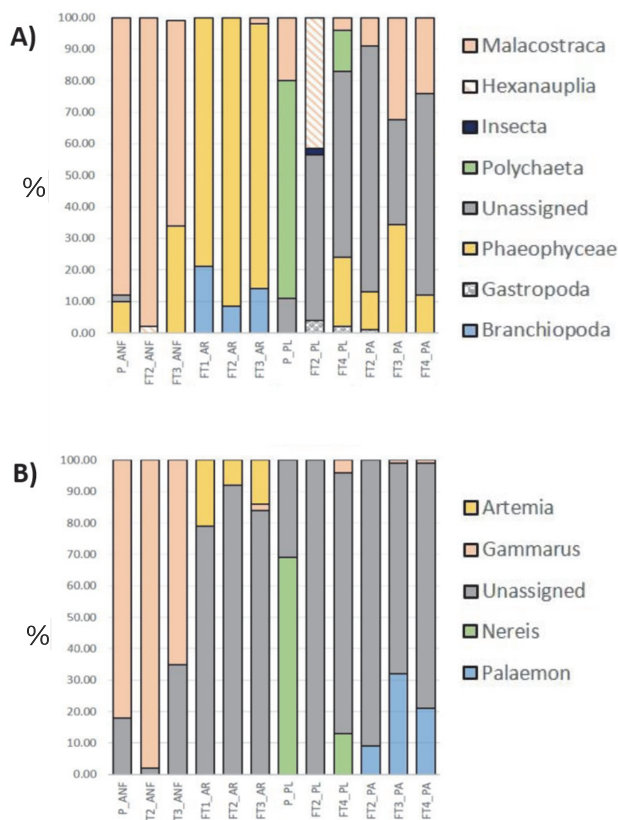

Fig. 2. Taxonomical assignment of ASVs at class (A) and genus (B) levels. Only groups with relative abundances ≥1% are presented. Samples: FT1, FT2, FT3 and FT4 refer to one of four individuals used in feeding trials, P refers to the prey sample, while ANF, AR, PA and PL stand for the prey given to seahorses (ANF – Gammarus *sp., AR –* Artemia *sp., PA –* Palaemon *sp., PL –* Nereis *sp.)*

amplicon-based HTS approach, to systematically detect the ingested prey. Importantly, DNA metabarcoding offered a non-invasive and non-lethal method allowing the identification of soft-bodied prey that might have been difficult or even impossible to detect through morphological analysis of faecal samples, such as *Artemia* sp. and *Nereis* sp.

DNA extraction, preparation and sequencing of amplicon libraries, and sequences analysis were carried out according to consolidated procedures [18]. The number of obtained sequences was high enough to cover the diversity of each sample. Chao1 and Shannon indices displayed a similar pattern among faecal samples containing both same and different prey ingested, indicating the efficiency of the applied approach to capture the samples' biodiversity. Beta diversity, based on a Bray–Curtis dissimilarity matrix and calculated via PCoA, revealed that the sample distribution corresponded to the diet applied. Taxonomical assignment of ASVs permitted the identification of all prey items effectively ingested during the trials. The results were in accordance with the morphological analysis on prey samples under the microscope, revealing the same taxa. The presence of remains of *Gammarus* sp. in faeces of seahorses fed with *Artemia* sp. and *Palaemon* sp. might indicate the longer retention of this prey in *H. guttulatus* guts probably due to the poorly digestible chitanaceous exoskeleton. Regarding unassigned sequences, it should be stressed that invertebrates, as main seahorse prey, are a diversified and relatively unstudied group, for which adequate information in reference databases are frequently absent, preventing their full taxonomical annotation. Among taxa detected in seahorse faeces, some of them could presumably result from contamination of different sources, such as aquarium water and prey items. Especially having in mind that most prey items were collected in the wild, detection of taxa such as Phaeophyceae is not surprising given that this alga is consumed by *Artemia* sp. and *Gammarus* sp. Moreover, the incidence of contaminants is even more relevant when analyses are carried out with high-sensitivity techniques, such as Next Generation Sequencing.

This study confirms that DNA metabarcoding on faeces is an effective tool for studying the seahorse diet, following past studies on other fish species [7] [8]. Hence, the technique could be used in further studies of *H. guttulatus* dietary composition in captivity but also in the wild. However, based on sampling experience, it is recommended that only thick faeces should be used, thus increasing the quantity of the extracted DNA and the possibility to detect prey. The results of DNA metabarcoding analysis, associated with the taxonomical classification of prey items under the microscope, showed that the developed molecular protocol ensures correct identification of prey fed to *H. guttulatus.* These findings, together with the advantage of being non-intrusive and especially, non-lethal, make DNA metabarcoding on faecal samples a good candidate to study diet of threatened seahorses in both captivity and wild.

REFERENCES

1. M.E.Soulé, J.A.Estes, J.Berger, C.M.Del Rio, ''Ecological effectiveness: conservation goals for interactive species'', Conserv. Biol., vol.17, No.5, October 2003, pp.1238–1250.
2. L.L.De Sousa, S.Marques Silva, R.Xavier, ''DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems'', Environ. DNA, vol.1, No.3, August 2019, pp.199 – 214.
3. M.A.Forin-Wiart, M.L.Poulle, S.Piry, J.F.Cosson, C.Larose, M.Galan, ''Evaluating metabarcoding to analyse diet composition of species foraging in anthropogenic landscapes using Ion Torrent and Illumina sequencing'', Sci.Rep., vol.8, November 2018, article No.17091.

4. A.E.Radulovici, B.Sainte-Marie, F.Dufresne, ‘’DNA barcoding of marine crustaceans from the Estuary and Gulf of St Lawrence: A regional-scale approach’’, Mol. Ecol. Resour., vol.9, No. s1, May 2009, pp.181–187.
5. T.A.Berry, S.K.Osterrieder, D.C.Murray, M.L.Coghlan, A.J.Richardson, A.K.Grealy, M.Stat, L.Bejder, M.Bunce, ‘’DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (*Neophoca cinerea*)’’, Ecol. Evol., vol.7, No.14, July 2017, pp.5435 – 5453.
6. G.K.Zhang, F.J.J.Chain, C.L.Abbott, M.E.Cristescu, ‘’Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities’’ Evol. Appl., vol.11, No.10, December 2018, pp.1901 – 1914.
7. E.Corse, C.Costedoat, R.Chappaz, N.Pech, J.Martin, A.Gilles, ‘’A PCR-based method for diet analysis in freshwater organisms using 18S rDNA barcoding on faeces’’, Mol. Ecol. Resour., vol.10, No.1, January 2010, pp.96–108.
8. T.Taguchi, Y.Miura, D.Krueger, S.Sugiura, ‘’Utilizing stomach content and faecal DNA analysis techniques to assess the feeding behaviour of largemouth bass *Micropterus salmoides* and bluegill *Lepomis macrochirus*’’, J. Fish Biol., vol.84, No.5, May 2014, pp.1271–1288.
9. S.A.Lourie, R.A.Pollom, S.J.Foster, ‘’A global revision of the seahorses *Hippocampus* Rafinesque 1810 (Actinopterygii: Syngnathiformes): taxonomy and biogeography with recommendations for further research’’, Zootaxa, vol.4146, No.1, August 2016, pp.1–66.
10. A.C.J.Vincent, S.J.Foster, H.J.Koldewey, ‘’Conservation and management of seahorses and other Syngnathidae’’, J. Fish Biol., vol.78, No.6, Jun 2011, pp.1681-1724.
11. R.H.Kuiter, ‘’Seahorses, pipefishes and their relatives: a comprehensive guide to Syngnathiformes’’, TMC Publishing, Chorleywood, UK, 2000.
12. K.Tipton, S.S.Bell, ‘’Foraging patterns of two syngnathid fishes: importance of harpacticoid copepods’’, Mar. Ecol. Prog. Ser., vol.47, No.1, July 1988, pp.31-43.
13. F.Ape, G.Corriero, S.Mirto, C.Pierri, T.Lazic, M.Gristina, ‘’Trophic flexibility and prey selection of the wild long-snouted seahorse *Hippocampus guttulatus* Cuvier, 1829 in three coastal habitats’’, Estuar. Coast. Shelf S., vol.224, August 2019, pp.1-10.
14. E.Corse, S.Valladares, M.Planas, A.Chammoro, J.Pintado, ‘’Analysis of the diet of the long-snouted seahorse *Hippocampus guttulatus* by 18SrDNA amplification of prey in faeces’’, Aquac. Nutr., vol.21, No.5, October 2015, pp.528-540.
15. S.K.Sheppard, K.D.Sunderland, J.Fenlon, D.Skervin, W.O.Symondson, ‘’Detection of secondary predation byPCR analyses of the gut contents of invertebrate generalist predators’’, Mol. Ecol., vol.14, No.14, December 2005, pp.4461–4468.
16. R.Pollom, ‘’*Hippocampus guttulatus*. The IUCN Red List of Threatened Species 2016: e.T41006A90859949’’, available at https://www.iucnredlist.org/species/41006/90859949
17. M.Leray, J.Y.Yang, C.P.Meyer, S.C.Mills, N.Agudelo, V.Ranwez, J.T.Boehm, R.J.Machida, ‘’ A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents’’, Front. Zool., vol.10, No.1, June 2013, pp.34.
18. C.Manzari, B.Fosso, M.Marzano, A.Annese, R.Caprioli, A.M.D’Erchia, C.Gissi, M.Intranuovo, E.Picardi, M.Santamaria, S.Scorrano, G.Sgaramella, L.Stabili, S.Piraino, G.Pesole, ‘’The influence of invasive jellyfish blooms on the aquatic microbiome in a coastal lagoon (Varano, SE Italy) detected by an Illumina-based deep sequencing strategy’’, Biol. Invasions, vol.17, March 2015, pp.923-940.
19. J.Zhang, K.Kobert, T.Flouri, A.Stamatakis, “PEAR: a fast and accurate Illumina Paired-End reAd mergeR”, Bioinformatics, vol.30, No.5, March 2014, pp.614–20.
20. S.Weiss, Z.Z.Xu, S.Peddada, A.Amir, K.Bittinger, A.Gonzalez, C.Lozupone, J.R.Zaneveld, Y.Vazquez-Baeza, A.Birmingham, E.R.Hyde, R.Knight, ‘’Normalization and microbial differential abundance strategies depend upon data characteristics’’, Microbiome, vol.5, March 2017, article No.27.
21. T.Rognes, T.Flouri, B.Nichols, C.Quince, F.Mahe, “VSEARCH: a versatile open source tool for metagenomics”, PeerJ, October 2018, article No.4:e2584.
22. B.Fosso, M.Santamaria, M.Marzano, D.Alonso-Alemany, G.Valiente, G.Donvito, A.Monaco, P.Notarangelo, G.Pesole, ‘’BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS’’, BMC Bioinformatics, vol.16, July 2015, article No.203.
23. B.Langmead, S.L.Salzberg, “Fast gapped-read alignment with Bowtie 2”, Nat. Methods, vol.9, No.4, March 2012, pp.357–359.
24. D.R.Mende, S.Sunagawa, G.Zeller, P.Bork, “Accurate and universal delineation of prokaryotic species”. Nat. Methods, vol.1, July 2013, pp. 881–884.
25. D.W.Mount, “Using the Basic Local Alignment Search Tool (BLAST)”, CSH Protoc, July 2007, article No.pdb top17.
26. S.F.Altschul, G.Gish, W.Miller, E.W.Myers, D.J.Lipman, “Basic local alignment search tool”. J.

Mol. Biol., vol.215, No.3, October 1990, pp.403–410.
27. P.J.McMurdie, S.Holmes, ‘’phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data”, PloS ONE, vol.8, No.4, April 2013, article No.e61217.
28. J.Oksanen, F.Guilaume Blanchet, R.Kindt, P.Legendre, R.B.O’hara, G.L.Simpson, P.Solymos, M.H.H.Stevens, H.Wagner, ‘’vegan: Community Ecology Package: R package version 2.5–3”, available at https://CRAN.R-project.org/package=vegan.