Marine Ecology Enhancement Fund (MEEF)
Declaration

To: The Secretariat of the MEEF

Reference No.: MEEF2017004

Project Title: Revealing cryptic faunal biodiversity in Hong Kong western waters using environmental DNA approach

Name of Project Leader: Dr CHEANG Chi Chiu

I hereby irrevocably declare to the MEEF Management Committee and the Steering Committee of the relevant Funds including the Top-up Fund, that all the dataset and information included in the completion report has been properly referenced, and necessary authorisation has been obtained in respect of information owned by third parties.

Signature: _____
Project Leader, Dr CHEANG Chi Chiu

Date: 10/4/2019

香港教育大學
The Education University of Hong Kong

Reference No. MEEF2017004

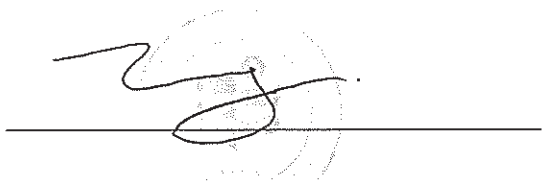
Marine Ecology Enhancement Fund (MEEF)

Revealing cryptic faunal biodiversity in Hong Kong western waters using environmental DNA approach

---

Final Report (2017-2018)

Signature:



Name of Applicant: Dr. CHEANG Chi Chiu

Name of Organization: The Education University of Hong Kong

Declaration

I hereby irrevocably declare, warrant and undertake to the MEEF Management Committee and the Steering Committee of the relevant Funds including the Top-up Fund, that I myself, and the Organisation:-

1. do not deal with, and are not in any way associated with, any country or organisation or activity which is or may potentially be relevant to, or targeted by, sanctions administered by the United Nations Security Council, the European Union, Her Majesty's Treasury-United Kingdom, the United States Department of the Treasury's Office of Foreign Assets Control, or the Hong Kong Monetary Authority, or any sanctions law applicable;

2. have not used any money obtained from the Marine Ecology Enhancement Fund or the related Top-up Fund (and any derived surplus), in any unlawful manner, whether involving bribery, money-laundering, terrorism or infringement of any international or local law; and

3. have used the funds received (and any derived surplus) solely for the studies or projects which further the MEEF Objectives and have not distributed any portion of such funds (including any 13 derived surplus) to members of the recipient organisation or the public.

Signature: _____

Name of Person-in-charge: Prof. Lui Tai Lok
Vice President (Research and Development)

THE EDUCATION UNIVERSITY OF HONG KONG C0099 香港教育大學

Date: 28 February 2019

Official chop: _____

Executive summary

Attributing to the influence of Pearl River and the underwater topology, the seawater in the western Hong Kong is usually low in salinity, turbid and fast flowing. The poor visibility and relatively dangerous condition in the western waters would introduce bias in the collection of underwater inventory data using visual census method in this region. In the present study, the biodiversity of Hong Kong western waters were studied using molecular technique - Environmental DNA (eDNA), a non-invasive and less labour intensive method is an encouraging way in aiding biodiversity monitoring. Seawater and sediment samples were collected from five water bodies in western Hong Kong: Tsing Lung Tau, Tai O, Lung Kwu Tan, the Brothers Island and Airport. eDNA were extracted from both seawater and sediment samples and undergo Polymerase Chain Reaction (PCR) amplification. By applying NGS technology (Miseq) on the analysis of environmentally derived samples, fish and crustacean biodiversity were investigated. 22 fish species from 17 families and 34 crustacean species from 27 families were identified in this study, including a common cultivated fish species *Epinephelus lanceolatus*. The results showed segregation among samples from different sites, which echoed the results aforementioned for individual datasets. ANOSIM showed that there is a statistically significant difference among different sites ($R=0.71$; Permutation: 9999; $p<0.05$). Based on ANOSIM of the combined fish and crustacean eDNA datasets, Airport and the Brothers Island are highly similar in terms of fish and crustacean species comp+osition. They shared *Coryphopterus tortugae* and *Paracalanus parvus* as most abundant fish and crustacean species, respectively. eDNA dataset revealed in the current study did not demonstrate a high recovery of species on records. This could be attributed to various factors including the incomplete and poorly curated barcoding dataset of local marine organisms in the DNA online database, mismatching of the sequences revealed and those on the online database, and misidentification of species represented by DNA barcode in the database. At the moment, eDNA technology could not replace other surveying methods completely. Nevertheless, present study also revealed some species that were not reported in previous biodiversity survey in Hong Kong, suggesting present field survey method adopted in Hong Kong might not be sufficient in detecting rare and cryptic species. The present study has demonstrated that this technology could supplement other conventional surveying methods such as trawling and underwater visual census.

Brief description of the project

This project aims at detecting any undiscovered (cryptic) biodiversity of fish and crustacean species using environmental DNA approach. Those cryptic species could not be accessed by the conventional ecological surveying method, particularly in the turbid and fast-flowing waters in the Hong Kong western waters.

Completed activities against the proposed Work Schedule

Original work schedule	Actual work schedule	Proposed work
July – Aug 17	July – Aug 17	Recruitment of full-time research assistant
July – Aug 17	July – Sep 17	Preliminary checking of the methodology e.g. determining the exact sites and amount of seawater needed to be filtered
Aug – Sep 17	Sep – Dec 17	Collection of wet season water samples
Sep – Dec 17	Sep 17 – Jan 18	Process of wet season water samples for NGS sequencing
/	Nov 17 – Mar 18	Negotiation with vendor for NGS service
Nov – Jun 18	Feb – Mar 18	Metabarcoding analysis of the NGS dataset
Jan – Feb 18	Feb – Mar 18	Submission of interim report
Feb – Mar 18	Mar – Mar 18	Collection of dry season water samples
/	Mar – Apr 18	Recruitment of full-time research assistant (due to the resignation of the original research assistant)
Mar – Jun 18	Apr – Aug 18	Process of dry season water samples for NGS sequencing
May – Jun 18	Jul – Oct 18	Contingency run of NGS datasets and comparison between the seasonal NGS datasets and inventory based on literature
Jun – Jun 18	Sep – Oct 18	Prepare and submit final report by 31 Oct 18

Introduction

1. **Biodiversity in the western waters**

   The hydrological parameters in eastern Hong Kong waters have long been known for their difference with those in the western waters, in which the seawater in eastern Hong Kong generally is more oceanic while that in the west is heavily influenced by the discharge of Pearl River (Yeung et al. 2014). Attributing to the influence of Pearl River and the underwater topology, the seawater in the western Hong Kong is usually low in salinity, turbid and fast flowing (Morton & Morton 1983). This difference between the eastern and western waters, in fact, resulted in the different habitats of the two regions, harboring significantly different kinds of marine species (Morton & Morton 1983). Western water is well-known as the habitats for Chinese White Dolphin, and yet quite a high marine biodiversity was recorded in the area. A large portion of the biodiversity records in the western waters comes from consultancy reports in recent years, for examples the Sham Tzeng Development EIA report, Hong Kong-Zhuhai-Macao Bridge EIA report. According to the most updated report, the third runway EIA report, there were 155 and 112 benthic macro-infaunal species collected in wet and dry seasons respectively. They were mainly the species of gastropod, polychaete and bivalve (Airport Authority Hong Kong 2014). In the vicinity of the airport, four estuarine fish species of conservation value, such as Spotted Seahorse *Hippocampus kuda* and seaweed pipefish *Syngnathus schlegeli* were found. For the fishery resources, 134 species of fish, thirty species of crabs and shrimps and 18 species of other taxa were recorded. Ten of those species were regarded as of conservation importance, for instance banded tuna *Scomberomorus commerson*, horseshoe crab *Carcinoscorpius rotundicauda* and yellow croaker *Larimichthys crocea* etc.

2. **Limitation of the conventional SCUBA diving survey in the western waters**

   Despite the relatively large amount of surveys done in the western waters, the poor visibility and relatively dangerous condition for SCUBA divers would hinder comprehensive ecological survey using visual inspection through SCUBA diving (Airport Authority Hong Kong 2014). This, perhaps, would introduce bias in the collection of underwater inventory data using visual census method in this region. For example, the surveyed sites for the Scleractinian corals in the western waters are much fewer than those in the east, which was associated with the finding that higher diversity of corals was observed in the east (Chan et al. 2005). Although this potential bias could be eliminated by other means of ecological survey in certain types of organisms such as the trawling survey on fish species, a more general sampling method with a relatively unbiased sampling, or at least a method that would generate a new dimension of

ecological datasets, is sought to supplement what we have known about the biodiversity in the western waters up-to-date, particularly for those groups, such as highly mobile crustacean, fish and infaunal mollusk, that could not be readily accessed by conventional surveying method in the circumstance of the western waters.

3. **Environmental DNA (eDNA)**

   Environmental DNA (eDNA), which is a mixture of different DNA molecules released from the organisms to the environment, was newly used in species detection and monitoring in the natural environment without actually having them observed (Herder et al. 2014). This approach can be applied to a wide range of habitats for detection of species and monitoring species composition (Metabarcoding). Targeted barcoding regions are amplified from the eDNA sample for species detection by checking against available databases like GenBank (e.g. Goldberg et al. 2011). This non-invasive and less labour intensive method is an encouraging way in aiding biodiversity monitoring (Bohmann et al. 2014, Thomsen & Willerslev 2015).

   Many studies have already proven the applicability of eDNA in detecting either rare species or invasive species in the freshwater environment (E.g. Ficetola et al. 2008, Wilcox et al. 2013, Tréguier et al. 2014). This approach, however, is not feasible to understand the biodiversity composition of an area. Instead, a universal primer for a conserved genetic marker would facilitate the amplification of DNA fragment from multiple species, and yet this mixture of DNA amplicons from diverse species could not be readily sequenced by typical PCR and Sanger sequencing.

4. **Metabarcoding approach and next generation sequencing (NGS)**

   Next-generation sequencing (NGS) is a high-throughput sequencing technology, which could satisfy the need of handling complex and massive samples (Shokralla et al. 2012). For NGS detection of eDNA, universal primers for targeted taxonomic groups can be designed with ease by making reference to readily available sequence data on the databases. With its high throughput compatibility, NGS can then generate sequences from a mixture of different species. The application of NGS technology on the analysis of environmentally derived samples is extensive, and these studies focus on exploring what is actually present in the environment (Shokralla et al. 2012). The promising results from NGS technology open the gateway to enhance environmental and ecological research, as in the case of current proposed study.

5. **Taxonomic resolution of metabarcoding**

   A dilemma in obtaining data, however, occurs between maximizing the coverage of taxa and increasing the resolution of the taxa to be determined in a given biodiversity

monitoring (Leray & Knowlton 2016). Small Subunit (SSU) of ribosomal RNA is a common marker in revealing metazoan biodiversity to phylum level due to its generality in amplifying the target region against a diverse spectrum of animals by universal primers (Fonseca et al. 2010). Nevertheless, this board range of amplification hinders the identification of the eDNA into species level (Leray & Knowlton 2016). More specific markers in resolving the lower taxonomic level in specific animal groups are sought for a more meaningful biodiversity monitoring (Leray & Knowlton 2016). Instead of surveying the overall phylum presented in the western waters, the proposed study will investigate two selected faunal groups, which are fish and crustacean, in the western waters. The reason why the two faunal groups were selected is because of the relatively rich online database in Genbank for the identification of eDNA sequence, minimizing the occurrences of the unknown species. The visual census of these two faunal groups would be easily affected by the poor visibility in the western waters, and the tested PCR primer sets were available to detect a wide range of fish species (Miya et al. 2015) and the crustacean (Cheang unpub. data).

6. **Objectives of the proposed study**

   This project, thus, would adopt the eDNA approach to reveal the metazoan biodiversity of two selected faunal taxa, fish and crustacean, in the western waters, so as to detect any cryptic biodiversity in the region.

Materials and Methods

1. **Study sites**

   Water and sediment sampling were conducted at five sites in Hong Kong western waters, namely Tsing Lung Tau, Tai O, Lung Kwu Tan, the Brothers Island and Airport (**Fig. 1, Table 1**).

Fig. 1 Map of Hong Kong showing the location of the study sites

Table 1 DMS of sample location in wet and dry season		
Sites	Locality No.	DMS of the localities
Tsing Lung Tau (TLT)	TLT1	22° 21' 33.6" N 114° 2' 35.94" E
	TLT2	22° 21' 32.58" N 114° 2' 30.54" E
	TLT3	22° 21' 30.54" N 114° 2' 26.1" E
Tai O (TO)	TO1	22° 15' 8.64" N 113° 50' 47.22" E
	TO2	22° 15' 12.84" N 113° 50' 51.96" E
	TO3	22° 15' 15.24" N 113° 50' 55.32" E
Lung Kwu Tan (TM)	TM1	22° 22' 50.88" N 113° 54' 44.88" E
	TM2	22° 23' 1.08" N 113° 54' 40.56" E
	TM3	22° 23' 12.54" N 113° 54' 31.98" E
The Brothers Island (MT)	MT1	22°20'11.1"N 113°57'39.2"E
	MT2	22° 20' 17.04" N 113° 57' 49.2" E
	MT3	22° 20' 19.5" N 113° 57' 59.46" E
Airport (IA)	IA1	22° 16' 59.34" N 113° 53' 19.92" E
	IA2	22° 16' 56.52" N 113° 53' 14.1" E
	IA3	22° 16' 53.16" N 113° 53' 16.92" E

2. Sample collection

All of the equipment and tools were bleached in 10% bleach solution for at least 60 minutes before use to minimize the cross-site and exogenous DNA contamination. Seawater and sediment sampling (**Fig. 2**) were performed once a season from September 2017 to March 2018 (**Table 2**). In each season, 3 × 10L of seawater samples and 3 × 1.25g of sediment samples were collected from each study site (**Fig. 3**). Each seawater sample was a pool of 10 sub-samples of 1L seawater sampled from three water depths (surface, middle and bottom), while each sediment sample was a pool of 5 sub-samples of 0.25g benthic sediment. The seawater samples were filtered in the field using a sterile

47mm Glass Microfiber Filter (GE Healthcare, Life Science, Whatman™, UK) using pump (**Fig. 4**), each filter was then wrapped in commercial aluminium foil. Sediment samples were collected using sterile 15ml centrifuge tubes and spatula (**Fig. 4**). All filters and sediment samples were then kept in ice during the transportation from the study sites to the laboratory. A total of 29 seawater samples and 26 sediment samples were collected.

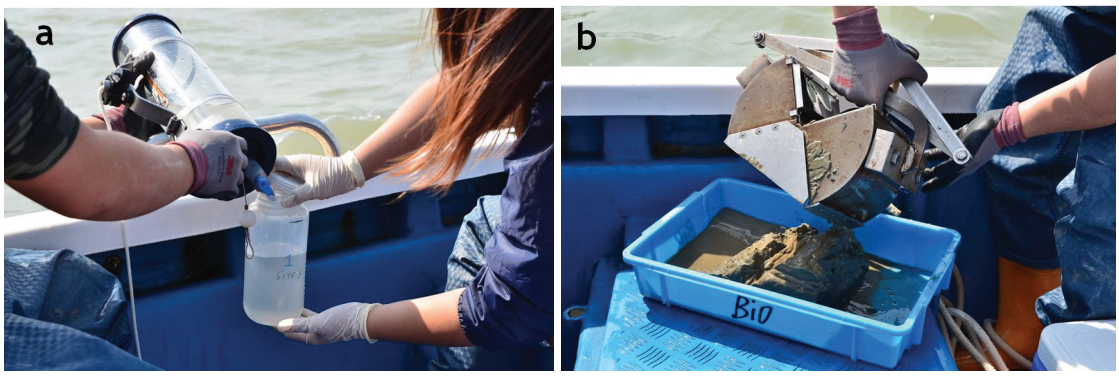


Fig. 2 a Photo showing seawater sampling. **b** Photo showing sediment sampling.

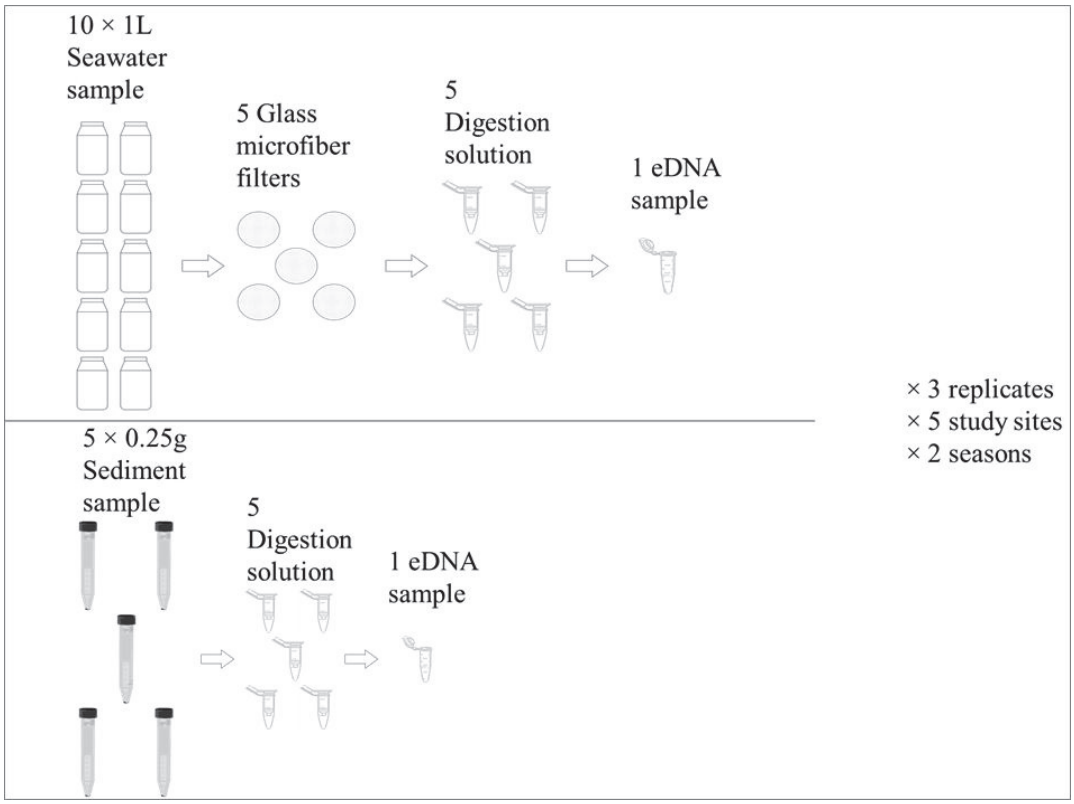


Fig. 3 Overview of the experimental design of water and sediment sampling.

Fig. 4 a Photo showing filtration of seawater samples. **b** Photo showing collection of the filter after the filtration using commercial aluminium foil and forcep. **c** Photo showing collection of the sediment samples using sterile 15ml centrifuge tube and spatula.

Table 2 Details of sample collection from individual study sites. *W: seawater sample; S: sediment sample.

Study sites	TLT		TO		TM		MT		IA	
Season	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry
Sampling date	20 Sep 2017	5 Mar 2018	13 Oct 2017	6 Mar 2018	31 Oct 2017	12 Mar 2018	15 Nov 2017	19 Mar 2018	8 Dec 2017	26 Mar 2018
No. of samples*	3W	3W and 3S	2W and 2S	3W and 3S	3W and 3S	3W and 3S	3W and 3S	3W and 3S	3W and 3S	3W and 3S

3. eDNA extraction from seawater samples

DNA extraction was performed immediately after the sample collection (**Fig.** 5) in order to minimize the degradation of eDNA. The protocol described by Miya et al. (2015) using QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany) was adopted with the following modifications: (i) each filter was folded twice into fan-shape and placed in a spin column, (ii) 900µl absolute ethanol was added after incubation, (iii) five separate microcentrifuge tubes with digested solution in each replicate were pooled into one QIAamp Mini Spin Column, and (iv) all samples were eluted with 100µl sterile DNA-free PCR Grade Water. All samples were kept in -20℃ before subsequent analyses.

4. eDNA extraction from sediment samples

All sediment samples were kept in -20℃ before eDNA extraction. eDNA was extracted using DNeasy® PowerSoil® Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol except (i) five separate microcentrifuge tubes with digested

solution in each replicate were pooled into one MB Spin Column and (ii) all samples were eluted using 50µl sterile DNA-free PCR Grade Water. All samples were kept in -20℃ before subsequent analyses.

Fig. 5 Photo showing DNA extraction

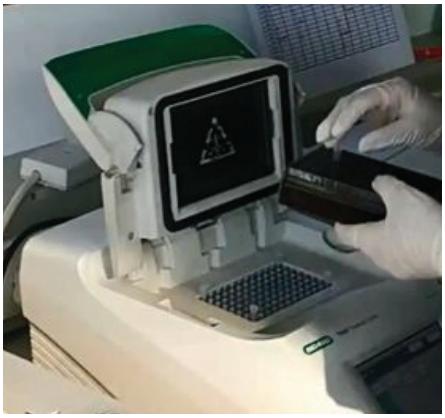


Fig. 6 Photo showing PCR

5. Purification of eDNA samples before PCR

To further reduce the contaminant in eDNA samples, QIAamp® DNA Stool Mini Kit (Qiagen, Hilden, Germany) was used to clean the previously extracted DNA samples. 50 µl of eDNA from seawater samples and 25 µl of eDNA from sediment samples were purified using the kit following the manufacturer's procedure except (i) 1350 or 1375 µl of ASL buffer was added, (ii) no Proteinase K digestion was performed, and (iii) all samples were eluted using a final volume 50 or 25 µl sterile DNA-free PCR Grade Water. All samples were kept in -20 ℃ before PCR amplification.

6. PCR amplification

The 18S rRNA and cytochrome *c* oxidase subunit I (COI) region were analyzed in the present study. 18S were amplified with universal primers 18S_3ndf: 5′-GGCAAGTCTGGTGCCAG-3′ (Scoble & Cavalier 2014) and 18S_ER1: 5′-GACTACGACGGTATCTRATCRTCTTCG-3′ (Ho et al. 2015) with a amplicon size of approximately 500 bp fragment for18S; while COI fragment was amplified using metazoan universal primers mICOIintF: 5′-GGWACWGGWTGAACWGTWTAY-CCYCC-3′ (Leray et al. 2013) and jgHCO2198: 5′-TAIACYTCIGGRTGICCRAARAA-YCA-3′ (Geller et al. 2013) with a amplicon size of 313 bp fragment of COI. The total PCR mixture was 100 µL in volume, which included 53.6 µL nuclease free water, 10 µL 10X buffer (Takara, USA), 2 µL $MgCl_2$, 10 µL dNTP (Invitrogen, USA), 4 µM of each forward and reverse primer, 1 U Taq DNA polymerase (Takara, USA), 0.4 µL BSA, 15 µL of extracted eDNA. Each reaction mixture was then separated into five individual reactions to serve as technical replicates (**Fig. 6**). The PCR cycle profile was set with an

initial 2 min denaturation at 95 °C; followed by 35 cycles of 20 sec denatuFigration at 95 °C, 50 sec annealing at 46 °C, and 45 sec extension at 72 °C; and a final 3 min extension at 72 °C. The five replicate tubes from each sample were subsequently pooled together for gel electrophoresis. The DNA mixture extracted from several fish species was included as positive control. A negative control without addition of DNA was included as well.

7. **Agarose Gel Electrophoresis**

The success of PCR amplifications were checked using 2% agarose gel (**Fig. 7**) prepared with 1X TAE (Tris-acetate-EDTA) buffer (included of Tris-base, acetic acid and EDTA). 5 µL PCR product was mixed with 1 µL 6X DNA loading dye (GeneDireX, Taiwan) and then loaded into the well of 2% TAE gel. A 2000 bp DNA ladder (GeneDireX, Taiwan) was also loaded together as size standard. The gel was ran for about 40 min at 80 volts and then checked by UV light box (Topbio, Taiwan).

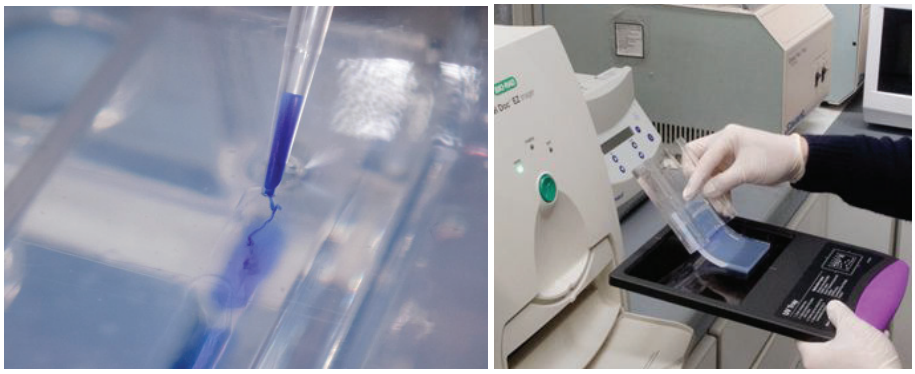


Fig. 7 Photo showing agarose gel electrophoresis

8. **Next Generation Sequencing (NGS) - Miseq**

Library preparations and Illumina MiSeq sequencing were conducted at GENEWIZ, Inc. (Suzhou, China). DNA samples were quantified using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). 50-100 ng DNA was used to generate amplicons using a panel of primers designed by GENEWIZ (GENEWIZ, Inc., South Plainfield, NJ, USA). The amplified DNA samples were ligated with 6-bp sample-specific barcode on both ends. They were further ligated with Illumina sequencing adaptor primers for Miseq, allowing uniform amplification of the library with high complexity ready for downstream NGS sequencing on Illumina Miseq platform. DNA libraries were validated by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). DNA libraries were multiplexed and loaded on an Illumina MiSeq instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was performed using a 2x300 paired-end (PE)

configuration; image analysis and base calling were conducted by the MiSeq Control Software (MCS) embedded in the MiSeq instrument.

9. **Data Analysis**

The QIIME data analysis package was used for 18S rRNA and COI data analysis. The forward and reverse reads were joined and assigned to samples based on barcode and truncated by cutting off the barcode and primer sequence. Quality filtering on joined sequences was performed and sequence which did not fulfill the following criteria were discarded: sequence length < 600-bp, no ambiguous bases, mean quality score ≥ 20. Then the sequences were compared with the reference database (RDP Gold database) using UCHIME algorithm to detect chimeric sequence, and then the chimeric sequences were removed. The effective sequences were used in the final analysis. Sequences were grouped into Molecular Operational Taxonomic Units (MOTUs) using the clustering program VSEARCH（1.9.6）against the Silva 119 database pre-clustered at 97% sequence identity. The Ribosomal Database Program (RDP) classifier was used to assign taxonomic category to all OTUs at confidence threshold of 0.8. The RDP classifier uses the Silva 132 database which has taxonomic categories predicted to the species level.
Sequences were rarefied prior to calculation of alpha and beta diversity statistics. Alpha diversity indexes were calculated in QIIME from rarefied samples using for diversity the Shannon index. This is to determine if the number of copies sequences has a good estimate on the observed MOTUs.

The fish and crustacean datasets were combined and analyzed to form a pair-wise distance (Bray-Curtis dissimilarity) matrix among samples, so as to reveal the compositional difference of fish and crustacean communities among sites. This analysis was conducted using PRIMER-E Ver. 6. Non-metric multidimensional scaling (nMDS) was used to provide visualisation of the difference among samples. ANOSIM was carried out to verify if there is any statistically significant difference among sites in terms of the species composition.

Results

The recovery of 12S marker, as suggested originally in proposal, was not succeeded for both fish and crustacean eDNA samples. Research team have further tested 18S and COI markers to reveal the species of the two taxa. PCR products from these two markers were in good quality and quantity, which render us proceed to the NGS sequencing.

1. **Determining how good the estimate of the sequences per samples to reveal MOTUs**
   Rarefaction curve (**Fig. 8**) between observe sequences per samples and observed MOTUs reach the plateau in COI, indicating that the number of sequences generated per samples is good enough for the estimate of MOTUs.

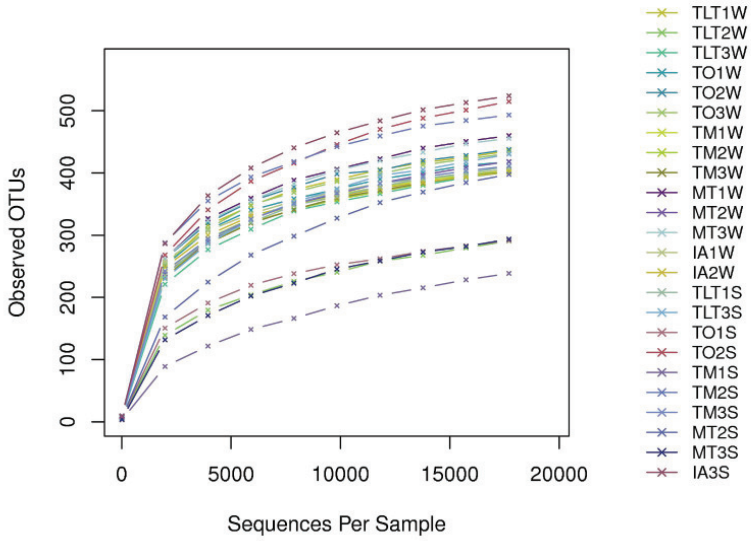


Fig. 8 Observed MOTUs rarefaction curve.

2. **Successful amplification of 18S**
   PCR were successfully amplified in three samples (**Fig. 9** two seawater samples and one sediment samples) and sequenced. A total of 400,710 reads were obtained from the three samples (ranged from 113,290 to 147,896 reads per sample) of which delineating into 868 MOTUs. However, the sequencing results showed majority of the reads belong to non-metazoan taxa that is not the major target group of the present study (**Fig. 10**) possibly attributed to the universal applicability of 18S primers across eukaryote. Hence, we decided to focus on applying COI gene sequences in the subsequent eDNA analyses.

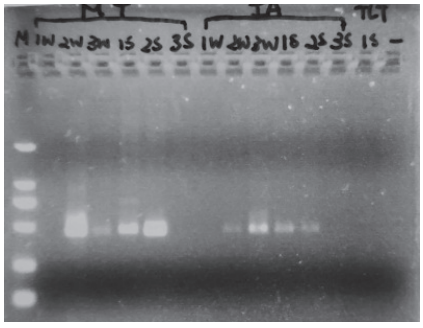


Fig. 9 Agarose gel electrophoresis results of PCR amplicons with 18S primer

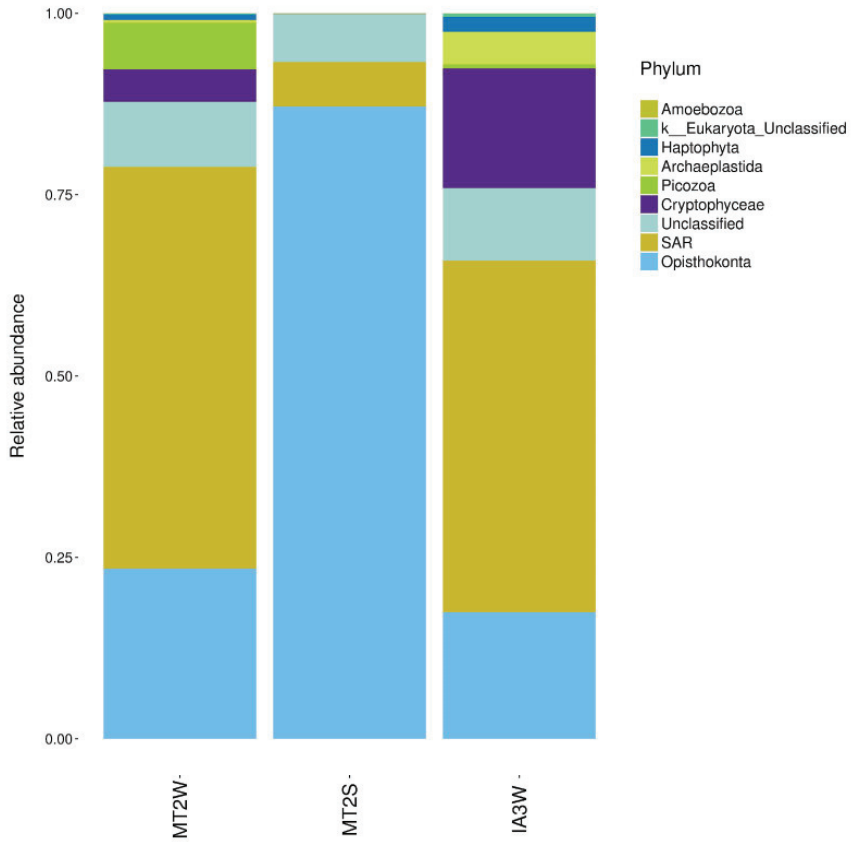


Fig. 10 Relative read abundance of each sample at phylum level of 18S NGS results.

3. Successful amplification of COI

PCR were successfully amplified in 24 samples (**Fig. 11**; 14 seawater samples and 10 sediment samples) and sequenced. A total of 3,418,724 reads were obtained from 24 samples (ranging from 80,178 to 269,398 reads per sample) of which delineating into 1,142 MOTUs (**Figs. 12 & 13**).

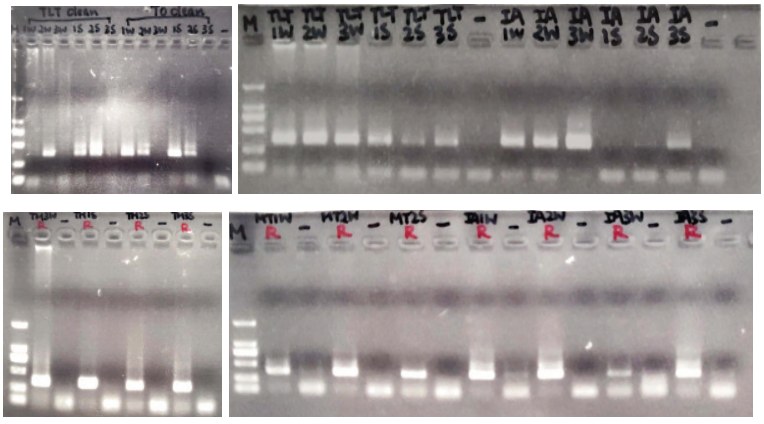


Fig. 11 Agarose gel electrophoresis results of PCR amplicons with COI primer

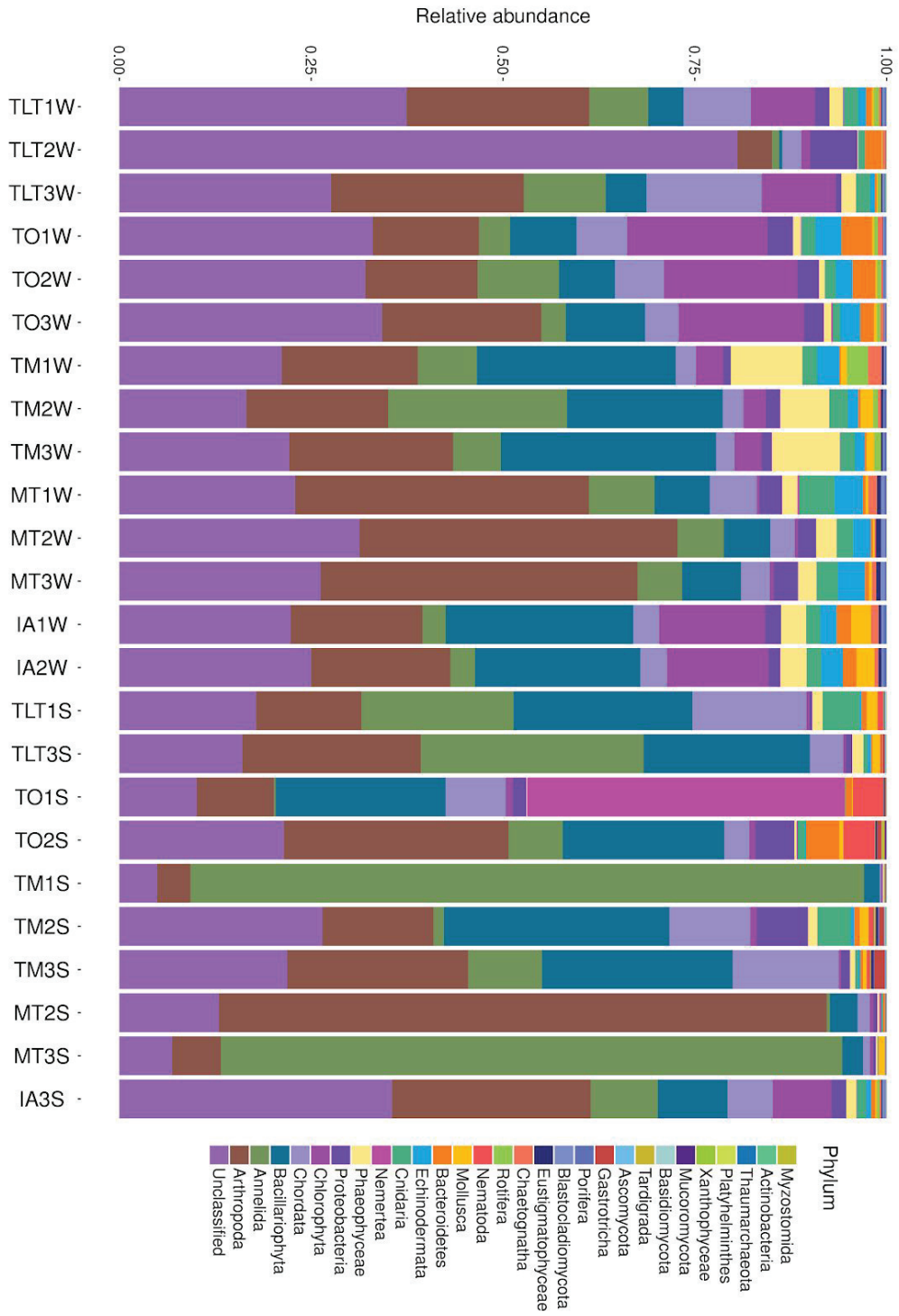


Fig. 12 Relative read abundance of each sample at phylum level of COI NGS results.

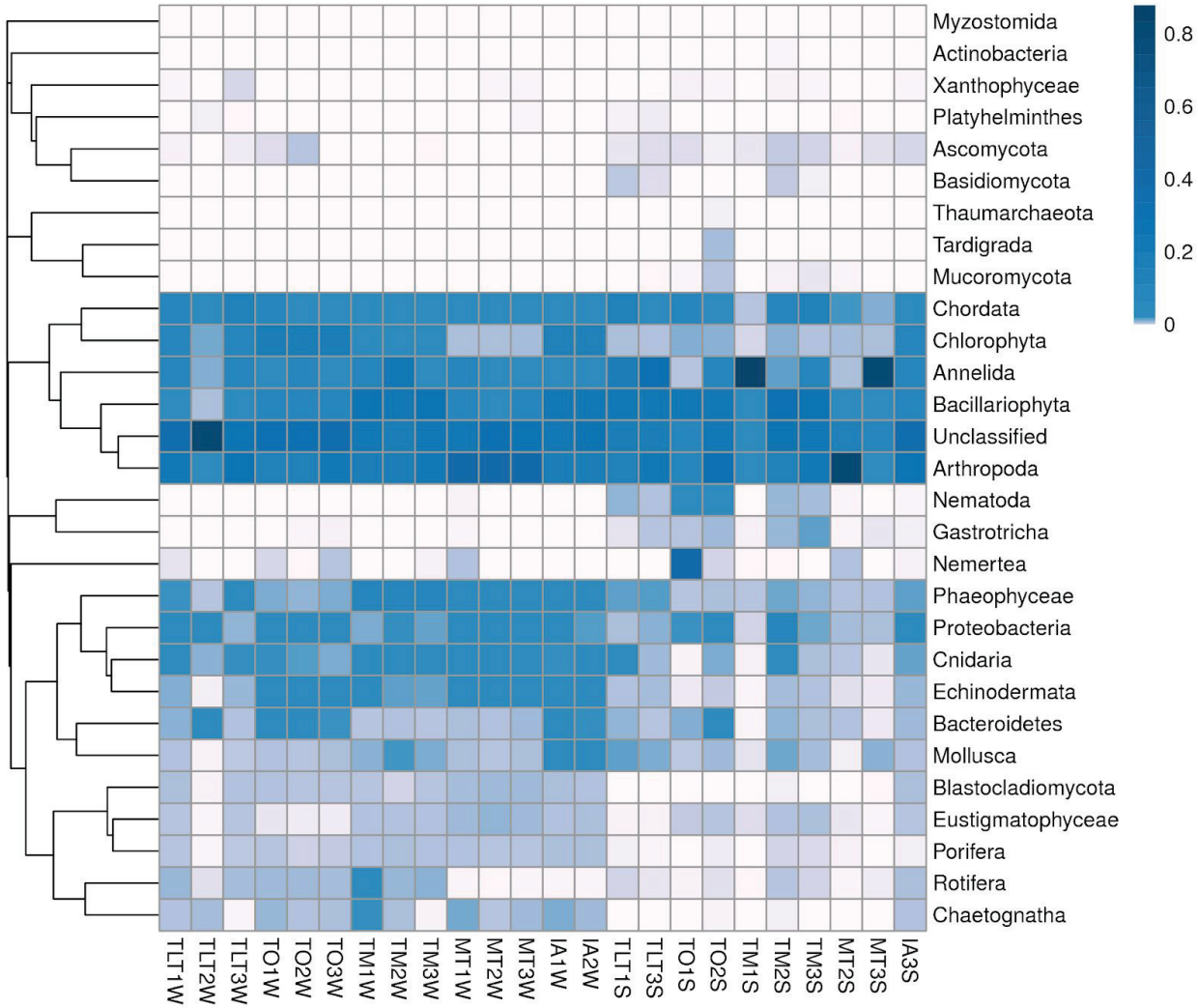


Fig. 13 Distribution heatmap of the top 30 phylum in the COI NGS results

Diversity and abundance of fish revealed by eDNA

The sequencing count delineating 26 MOTUs produced 18,060 reads of fishes, ranging from 8 to 7,765 reads per sample. Among all fish species, 7 MOTUs with 7,765 reads (43%) from Gobiidae, 3 MOTUs with 3,381 reads (18.7%) from Cichlidae, 1 MOTU with 2,187 reads (12.1%) from Cottidae, 1 MOTU with 1,839 reads (10.2%) from Rhinochimaeridae, 1 MOTU with 906 reads (5.0%) from Amblycipitidae, 2 MOTU with 776 reads (4.3%) from Cyprinidae; and 11 MOTUs with 706 reads from the other 11 fish families (**Fig. 14**).

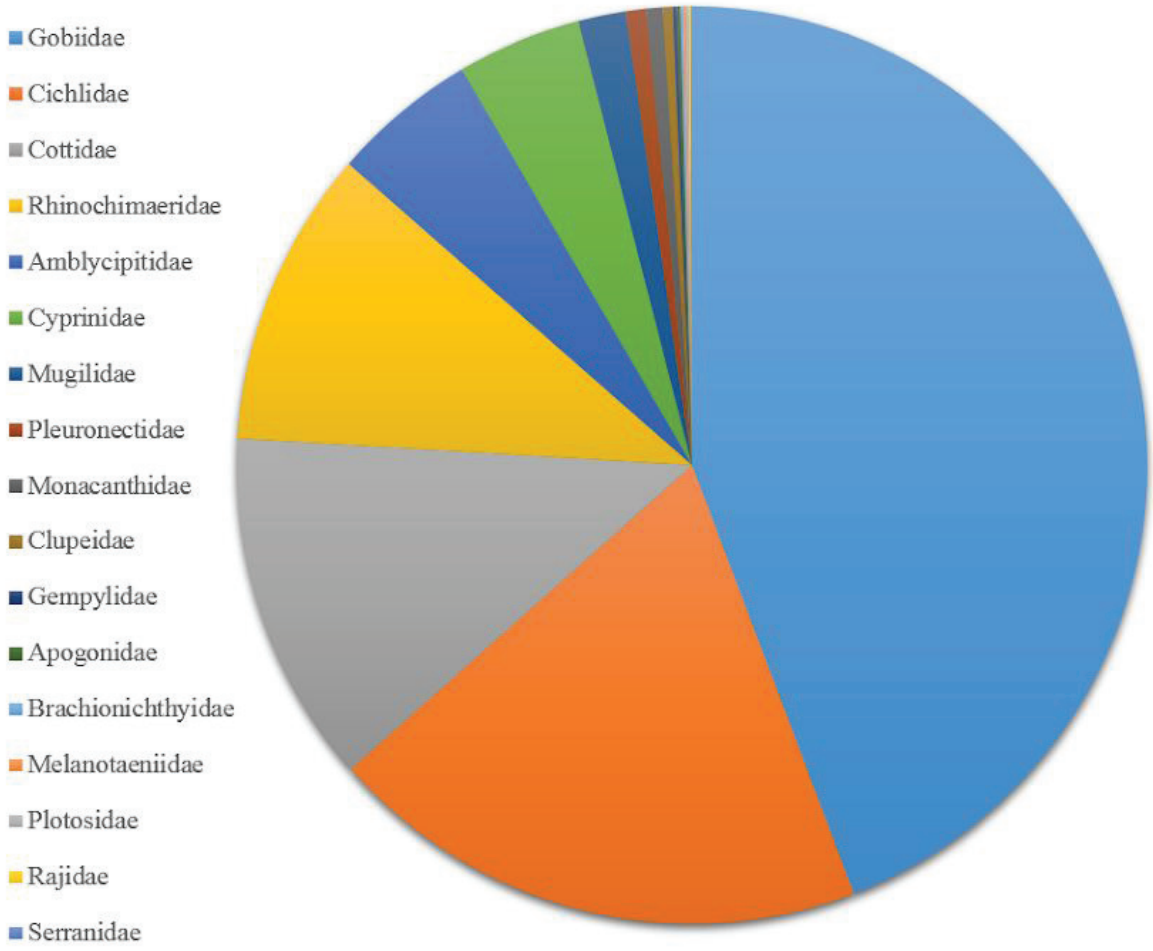


Fig. 14 relative abundance of fishes in family level

We then identified them to species (or genus) level using a sequence similarity threshold of 90%. This value was adopted instead of a more stringent value of 97 or 98% that employed to delineate species in other barcoding study because the COI database for local aquatic species is far from comprehensive and hence, there were quite a number of sequences could not identify to a match with similarity > 95%. Therefore, we adopted a lower threshold here for a pilot eDNA metabarcoding biodiversity study in Hong Kong. We believed this represents an approach that is robust enough to reveal diversity and effective survey method.

We identified 22 fish species from 17 families in the our data based on this approach. Among twenty species identified (**Table 3**), only six (occurred at 574 reads; 3.18%) could be found in HK Fish Net (AFCD) and FishBase. In terms of the higher taxonomic level, four fish families could be found in local fish databases, while seven fish families were not recorded in Hong Kong before. Among the six local species found, two of them were reef-associated (*Pervagor janthinosoma* and *Cheilodipterus artus*) and the other four species could be found in Hong Kong

western waters. Fifteen fish species (16 MOTUs; 12,262 reads; 67.90%) were marine species, while the other four species (severn MOTUs; 5079 reads; 28.12%) are freshwater species. The remaining three species (three MOTUs; 219 reads; 1.21%) living in both marine and freshwater environment.

Table 3 overview of fish species in COI NGS results

Family	Species	Local/ non-local	Freshwater / Marine	No. of MOTUs	Reads	Remarks	%
Gobiidae	*Barbulifer ceuthoecus*	NL	MW	1	258	Ornamental	43.00
	Coryphopterus tortugae	NL	MW	2	7399		
	Ctenogobiops tongaensis	NL	MW	1	10	Congener in HK	
	Eviota shimadai	NL	MW	1	6	Congener in HK	
	Sicyopus zosterophorum	NL	FW & MW	1	8	Found in Guangdong	
	Tridentiger bifasciatus	L	MW	1	84		
Cichlidae	*Laetacara thayeri*	NL	FW	3	3381		18.72
Cottidae	*Clinocottus analis*	NL	MW	1	2187		12.11
Rhinochimaeridae	*Harriotta raleighana*	NL	MW	1	1839	Occur in continental Shelf	10.18
Amblycipitidae	*Liobagrus mediadiposalis*	NL	FW	1	906	Aqua-culture	5.02
Cyprinidae	*Labeo lineatus*	NL	FW	2	776	Aqua-culture	4.30
Mugilidae	*Liza affinis*	L	MW	1	296		1.64

Pleuronectidae	*Platichthys stellatus*	NL	MW	1	127		0.70
Monacanthidae	*Pervagor janthinosoma*	L	MW	1	101	Common species in HK	0.56
Clupeidae	*Konosirus punctatus*	L	MW	1	62		0.34
Gempylidae	*Thyrsites atun*	NL	MW	1	26		0.14
Apogonidae	*Cheilodipterus artus*	L	MW	1	23	Common species in HK	0.13
Brachionichthyidae	*Brachionichthys hirsutus*	NL	MW	1	18		0.10
Melanotaeniidae	*Melanotaenia praecox*	NL	FW	1	16	Ornamental	0.09
Plotosidae	*Cnidoglanis macrocephalus*	NL	FW	1	15		0.08
Rajidae	*Hongeo koreana*	NL	MW	1	14		0.08
Serranidae	*Epinephelus lanceolatus*	L	MW	1	8		0.04

The species composition (**Fig. 15**) among sites varied a lot. *Coryphopterus tortugae* was the most dominant species in the eastern sites, which are TLT, MT and IA, whereas *Laetacara thayeri*, a non-local marine cichlid, dominated the western sites, TO and TM, with three MOTUs. *Epinephelus lanceolatus* is a common species that is cultivated in fish farms in Hong Kong. It occurs in low eDNA copies in all sites except airport where no eDNA copies was detected (**Fig. 16**). Only the compositions at The Brothers Island (MT) and Airport (IA) were relatively similar with majority of species being *C. tortugae.* Other three sites are distinct in terms of fish compositions. For examples, 60.83 % of the fish species reads found in Lung Kwu Tan (TM) is *L. thayeri* that is not found abundantly in other sites. On the other hand, *Clinocottus analis* is only common in Tsing Lung Tau (TLT), represented by 39.57% of the fish species reads, while it is either very rare or not found in all other study sites.

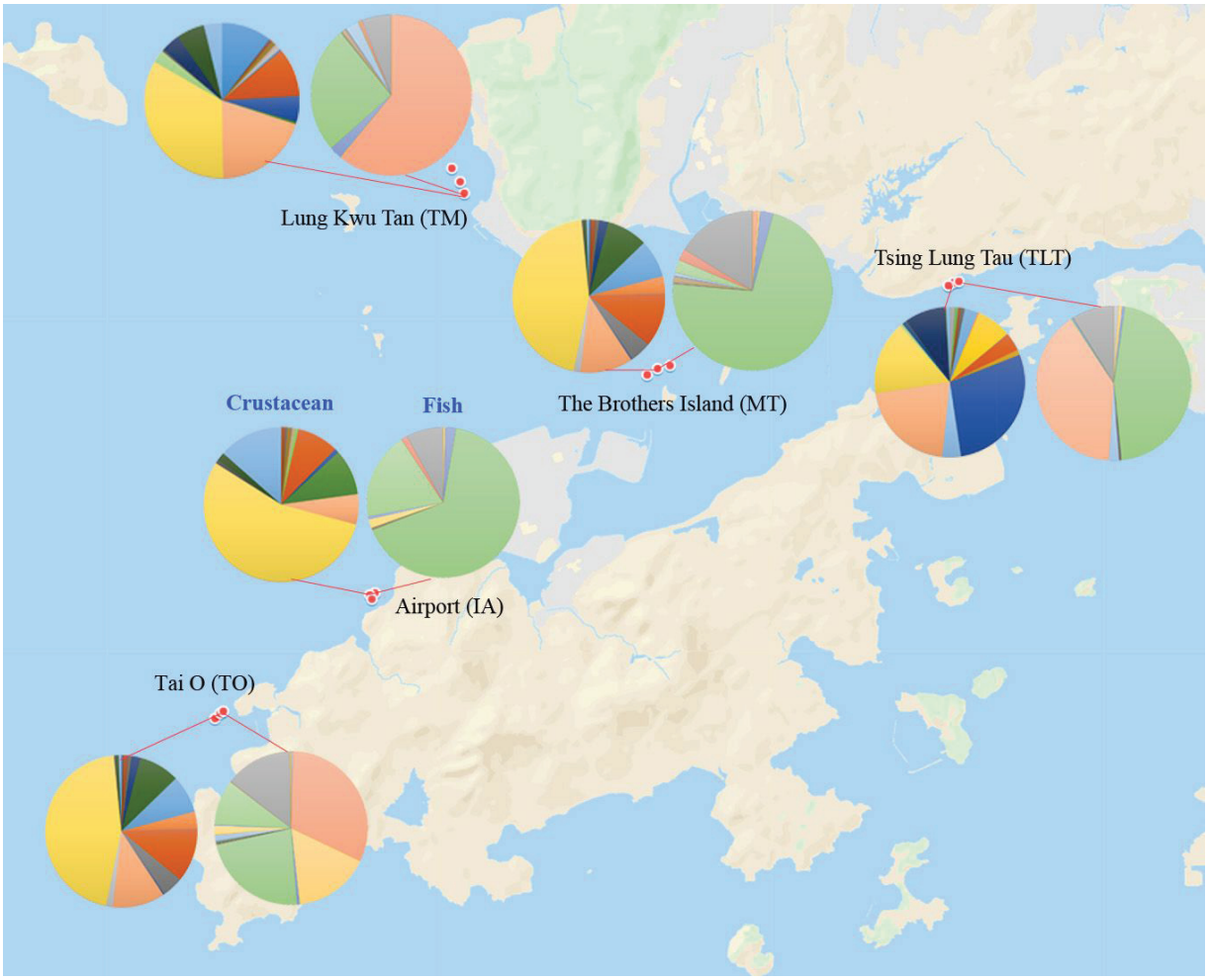


Fig. 15 Map showing the composition of crustacean (left) and fish (right) species inferred from the relative read abundance of COI data in the present study. Please refer to **Fig. 14** and **Fig. 17** for the color coding for different families.

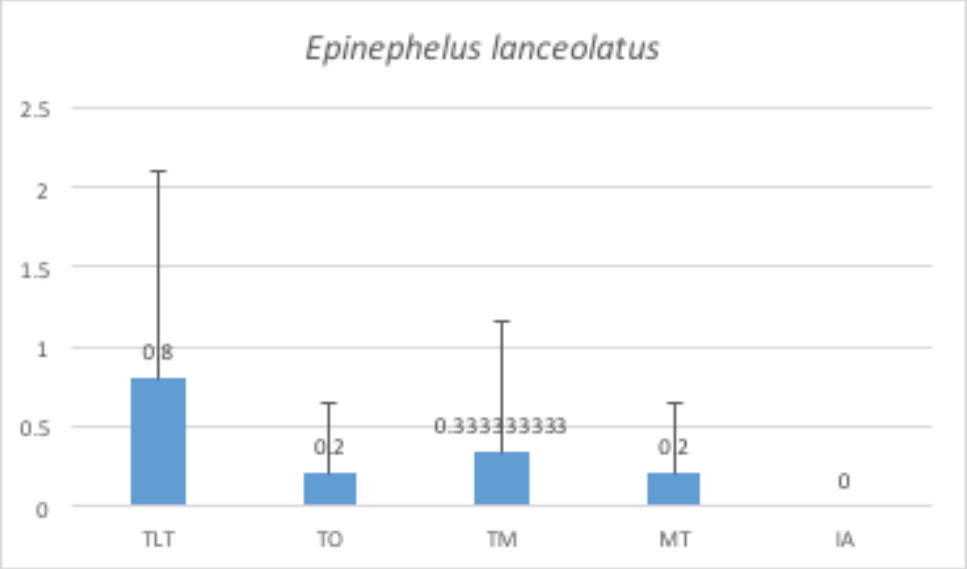


Fig. 16 The average eDNA reads of *Epinephelus lanceolatus* per samples

Diversity and abundance of crustaceans revealed by eDNA

The sequencing count delineating 56 MOTUs yielded 25,904 reads of crustacean, ranging from 1 to 3,371 reads in each sample, identifying 34 crustacean species from 27 families. Among all crustacean species, 6 MOTUs with 10,607 reads (40.95%) from Paracalanidae, 7 MOTUs with 3,371 reads (13.01%) from Centropagidae, 4 MOTU with 2,864 reads (11.06%) from Acartiidae, 2 MOTU with 2,234 reads (8.62%) from Ligiidae, 3 MOTU with 1,508 reads (5.82%) from Atyidae, 2 MOTU with 1,120 reads (4.32%) from Podonidae; and 32 MOTUs with 4,200 reads from another 20 crustacean families (**Fig. 17**). Zooplankton (e.g. species from Calalonida, Branchipoda) represented the most dominant crustacean found in the samples which constituted >70% of the relative read abundance. This was not surprising as they are dominant members in the open water.

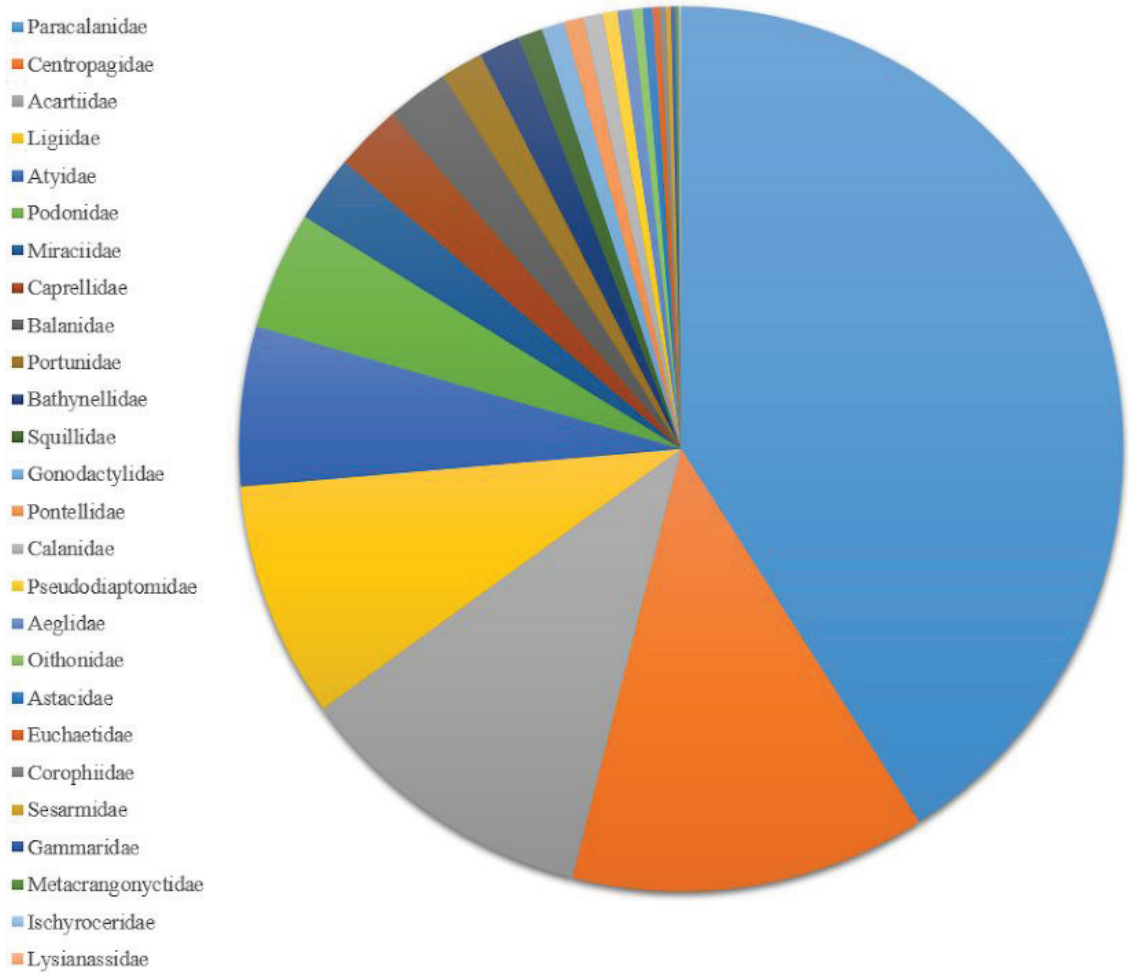


Fig. 17 relative abundance of crustacean in family level

We classified and identified the crustacean species using the approach adopted in fishes as described above. A total of 34 species from 27 families were identified in the 56 MOTUs (**Table 4**). Copepoda are the dominant taxa in term of both species and families recovered (13 species from 10 families) and abundance as revealed by relative read abundance (> 65%). Decapoda represents the second most common taxa with 9 species from four families were found (8.48% relative read abundance). Amphipoda ranked the third in the number of species found (six species) but it constituted a higher number of families (four families) and approximately 3% of read. Only one to two species were found in our samples for other commonly encountered marine Crustacea, like Stomatopoda, Isopoda and Maxillopoda.

Table 4 overview of crustacean species in COI NGS results						
Family	Species	Local/ non-local	Freshwater / Marine	No. of MOTUs	Reads	%
Paracalanidae	*Paracalanus parvus*	L	MW	5	10606	40.95
	Paracalanus sp.	NA	MW	1	1	
Centropagidae	*Centropages tenuiremis*	L	MW	7	3371	13.01
Acartiidae	*Acartia pacifica*	NL	MW	2	1748	11.06
	Acartia spinicauda	L	MW	2	1116	
Ligiidae	*Ligia baudiniana*	NL	MW	2	2234	8.62
Atyidae	*Atya scabra*	NL	FW	1	762	5.82
	Atyopsis spinipes	NL	FW	1	534	
	Halocaridina rubra	NL	FW	1	212	
Podonidae	*Pseudevadne tergestina*	L	MW	2	1120	4.32
Miraciidae	*Wellstenhelia qingdaoensis*	L	MW	2	631	2.44
Caprellidae	*Paracaprella tenuis*	NL	MW	1	628	2.42
Balanidae	*Balanus trigonus*	L	FW/MW	1	595	2.30
Portunidae	*Portunus hastatoides*	L	MW	1	58	1.56
	Portunus sanguinolentus	L	MW	1	345	
Bathynellidae	*Bathynellidae sp.*	NA	MW	2	392	1.51
Squillidae	*Squilloides leptosquilla*	L	MW	1	229	0.88
Gonodactylidae	*Gonodactylellus affinis*	NL	MW	1	212	0.82
Pontellidae	*Labidocera rotunda*	L	MW	1	188	0.73
Calanidae	*Neocalanus cristatus*	NL	MW	1	181	0.70

Pseudodiaptomidae	*Pseudodiaptomus annandalei*	L	MW	1	83	0.54
	Pseudodiaptomus tollingerae	NL	MW	1	58	
Aeglidae	*Aegla uruguayana*	NL	FW	1	140	0.54
Oithonidae	*Oithona similis*	NL	MW	1	97	0.37
Astacidae	*Austropotamobius italicus*	NL	FW	1	95	0.37
Euchaetidae	*Paraeuchaeta norvegica*	NL	MW	1	69	0.27
Corophiidae	*Corophium multisetosum*	NL	MW	1	53	0.20
Sesarmidae	*Perisesarma guttatum*	NL	MW	1	8	0.19
	Sesarma reticulatum	NL	MW	1	40	
Gammaridae	*Echinogammarus sp.*	NA	MW	1	44	0.17
Metacrangonyctidae	*Metacrangonyx samanensis*	NL	FW	1	32	0.12
Ischyroceridae	*Ericthonius punctatus*	NL	MW	1	16	0.06
Lysianassidae	*Scopelocheirus schellenbergi*	NL	MW	1	5	0.02
Cletodidae	*Enhydrosoma gariene*	NL	MW	1	1	0.00

Species composition differences among sites based on combined datasets

There is segregation among samples from different sites, which echoed the results aforementioned for individual datasets. ANOSIM showed that there is a statistically significant difference among different sites (R=0.71; Permutation: 9999; $p<0.05$). Tsing Lung Tau (TLT) is the most distantly related sites, while Airport (IA) and The Brothers Islands (MT) are closely related (**Fig. 18**).

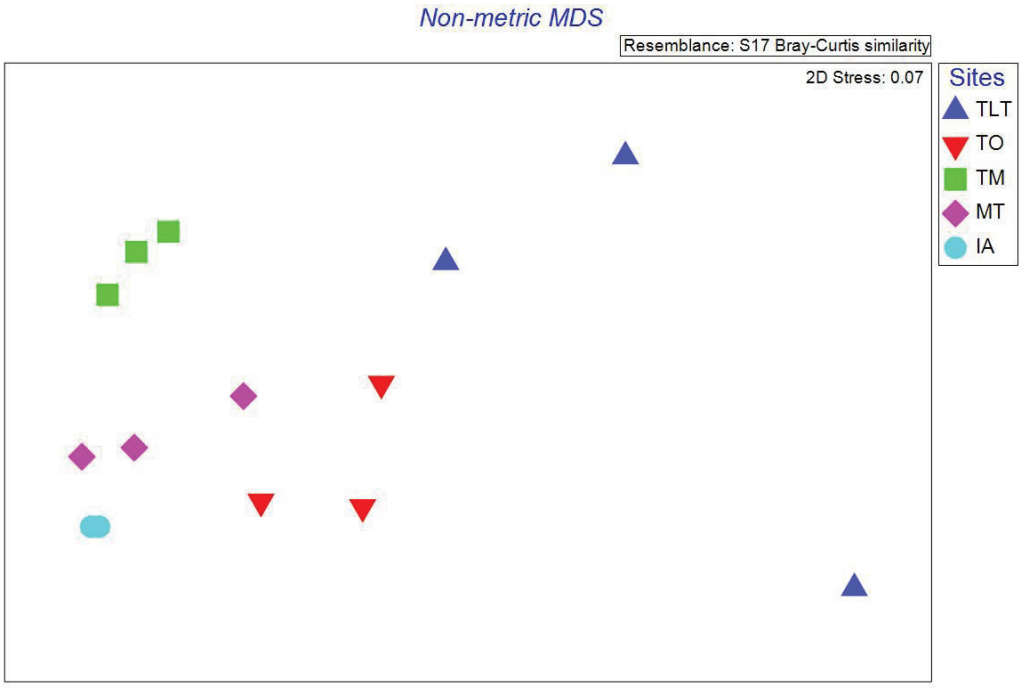


Fig. 18 The non-metric multidimensional scaling showing the difference among each sample from different sites in terms of the species composition of both fish and crustacean datasets. (R=0.71; p<0.05)

Discussion

1. **Ability of eDNA in detecting fish and crustacean species recorded in western waters**
   Current study has revealed comparable number of fish MOTU (28 MOTU) with other eDNA field studies around the world (**Table 5**), albeit with varying sampling effort across studies. However, there was no study focus chiefly on crustacean to the best of our knowledge. Comparison of the results between the current study and the eDNA studies in other countries on crustacean could not be conducted.

   In terms of species identification however, this study has revealed relatively few (22 species in present study compared to 33 to over 100 species in the others) but we recovered comparable number of fish families (17 families) with other studies (6-29 families). The differences in number of fish species and families recovered could be attributed to various factors, including the usage of different markers, the geographical regions studied (Tropical versus Temperate versus polar areas), and the species diversity.

Table 5. Comparison of the number of eDNA species revealed between the current study and those field studies in literature

Location	Studied taxa	Marker	No. of species revealed	Sampling effort/ range of sampling	Reference
Hong Kong, Tropical area	Fish, Crustacean	COI	22 species (Fish) from 17 families, 34 species (Crustacean) from 27 families	Twenty nine 10-L seawater and twenty six 1.25-g sediment each from 5 sites spanning 20 km area	This study
Denmark, Temperate area	Fish	cyt b	15 species from 11 families	One 1.5-L seawater each from 3 sites spanning 800 m area	Thomsen et al. 2012
US, Monterey Bay, Subtropical area	Fish	12S	72 species from 7 family	Six 1-L seawater each from 11 sites spanning 100 km	Andruszkiewicz et al. 2017
Japan, Maizuru Bay, Temperate area	Fish	12S	128 species	Two 1-L seawater each from 47 sites spanning 6 km	Yamamoto et al. 2017
Greenland, Subartic area	Fish	12S	37 taxa from 29 families	Twenty one 2-L seawater each from 21 sites spanning 400 km	Thomsen et al. 2016
US, Monterey Bay, Temperate to Subtropical area	Fish	12S	33 taxa from 6 families	One 3-L seawater each from 12 sites spanning 3 km	Port et al. 2016
US, Hudson River Estuary, Temperate area	Fish	12S	51 species	Two 0.5-L seawater each from 2 sites spanning 3 km	Stoeckle et al. 2017

When we compared the fish species found in this study with the AFCD Fish database, only six out of twenty two species (27%) could be explicitly found in the records. However, considering the facts that incomplete DNA barcoding dataset for local fish species in the online database GenBank; and some species identified now, in fact, have congeneric member in Hong Kong, we believe that the number of recorded fish species that can be revealed by eDNA technique could be more. This question could be better addressed only when a more comprehensive DNA barcoding database for local marine organisms have been completed. The baseline information would be very critical for any future works on eDNA application in the surveying marine biodiversity.

Nevertheless, the sequences of eDNA collected in this study is very precious in the sense that we have collected this baseline data in the current timeframe. With more and more DNA barcoding sequences generated in the future on local marine species, the identity of the uncertain or unidentified species studied so far would be unraveled.

2. **Spatial variation of species composition as revealed by eDNA datasets of fish and crustacean**

Based on ANOSIM of the combined fish and crustacean eDNA datasets, Airport and the Brothers Island are highly similar in terms of fish and crustacean species composition. They shared *Coryphopterus tortugae* and *Paracalanus parvus* as most abundant fish and crustacean species, respectively. One of the possible reasons for this could be the similarity of both sites situated in the central region of the western waters. On the other hand, the composition of Lung Kwu Tan is more similar to Tai O, compared to Tsing Lung Tau. It could be attributed to the shared effect of the freshwater outflow from the Pearl River on TM and TO. Many species revealed in Lung Kwu Tan and Tai O are introduced species or freshwater species e.g. *Labeo lineatus* in Tai O and *Laetacara thayeri* in Lung Kwu Tan, while those found in Tsing Lung Tau in the east were mainly marine species.

3. **Unexpected taxa revealed by eDNA datasets**

a. Occurance of Freshwater fish

Referring to FishBase, 16 MOTUs with 12262 reads (67.90%) from 15 marine fish species were found, while 7 MOTUs with 5079 reads (28.12%) from 4 freshwater fish species and 3 MOTU with 219 reads (1.21%) from 3 fish species were found living in both marine and freshwater environment. Among the four freshwater species identified (**Table 6**), *Labeo lineatus* (carp) and *Liobagrus mediadiposalis* (catfish) are common freshwater cultivation species in southern Guangdong Province. The tissue or remains of those species may be transferred to western waters, mediated by the freshwater outflow from Pearl River. *Laetacara thayeri* and *Melanotaenia praecox*, in contrast, are two

species that commonly found in the trading of aquariums pet. They could be released by local aquarium hobbyists or originated from Pearl Rivers as well. Further study is needed to further verify this findings.

Table 6 freshwater fish species obtained from COI NGS data	
Family	Species
Cichlidae	*Laetacara thayeri*
Amblycipitidae	*Liobagrus mediadiposalis*
Cyprinidae	*Labeo lineatus*
Melanotaeniidae	*Melanotaenia praecox*

b. Detection of *Epinephelus lanceolatus* grouper and the potential usage of eDNA in environmental monitoring of invasive species

Epinephelus lanceolatus is a common species that will be cultivated in the fish cages in Hong Kong. The nearest fishing culture zone in the region is the one in Ma Wan, which is geographically near to the site, Tsing Lung Tau. It would be possible that the tissue of the neighbouring cultivated giant groupers has been transported in the western waters reaching, as far as, Tai O. However, further information on what species and when did the fish cultivating firm cultivated the fish should be obtained before this hypothesis could be tested.

Another possible scenario should not be overlooked. *E. lanceolatus* has been crossed with various other *Epinephelus* species such as *Epinephelus fuscoguttatus* to produce fast-growing hybrid, for instance Sabah Grouper, for culture purpose and widely sold in the HK local fish market (Luin et al. 2013). The species was also released by public because of religious practice and has caused ecological problem. In the present study, the eDNA approach adopted has detected the species in the water sample, suggesting the potential of the method in future ecological monitoring of the spread and demography of the species in HK water.

4. Limitations and potential application of eDNA approach on ecological study

eDNA dataset revealed in the current study did not demonstrate a high recovery of species on records. This could be attributed to various factors including the incomplete and poorly curated barcoding dataset of local marine organisms in the DNA online database, mismatching of the sequences revealed and those on the online database, and

misidentification of species represented by DNA barcode in the database. At the moment, eDNA technology could not replace other surveying methods completely.

Nevertheless, present study also revealed some species that were not reported in previous biodiversity survey in Hong Kong, suggesting present field survey method adopted in Hong Kong might not be sufficient in detecting rare and cryptic species. The present study has demonstrated that this technology could supplement other conventional surveying methods such as trawling and underwater visual census.

Summary and Way Forward

NGS-based eDNA approach represents a powerful tool to detect and trace the presence of introduced species. As demonstrated in the current study, a trace amount of DNA coming from introduced species could be readily amplified by PCR and sequenced by NGS method. It is worthy to explore the applicability of eDNA technique in the early detection of marine invasive species in the future. eDNA approach also appears to be promising in revealing and comparing species composition efficiently across sites, which might generate empirical data that could potentially inform conservation policies.

Acknowledgements

We are much grateful to Hong Kong Airport Authority for their support to the current study financially. The gratitude is also extended to Prof. TSANG, Ling Ming of the Chinese University of Hong Kong for his assistance in certain part of molecular analysis and comments on the final report; Mr IP, Ho Yeung and Ms LEE, Bo Yee for being research assistance in the whole project. Special thanks goes to Mr YIU, Wai Hong and Mr WONG, Tsz Chiu for their assistance in the field work.

References

Airport Authority Hong Kong (2014). Expansion of Hong Kong International Airport into a Three-Runway System Environmental Impact Assessment Report. (http://www.epd.gov.hk/eia/register/report/eiareport/eia_2232014/html/ES_Rev%20C.pdf)

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PloS one*, *12*(4), e0176343.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W. & de Bruyn, M. (2014) Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology and Evolution, 29(6): 358-367.

Chan, A. L. K., Chan, K. K., Choi, C. L. S., McCorry, D., Lee, M. W., & Ang, P. J. (2005). Field guide to hard corals of Hong Kong. Agriculture, Fisheries and Conservation Department, Hong Kong SAR Government and Cosmos Book Ltd., Hong Kong.

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. Biology Letters, 4(4): 423-425.

Fonseca, V. G., Carvalho, G. R., Sung, W., Johnson, H. F., Power, D. M., Neill, S. P., ... & Creer, S. (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nature Communications, 1: 98.

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Molecular ecology resources, 13(5), 851-861.

Goldberg, C. S., Pilliod, D. S., Arkle, R. S. & Waits, L. P. (2011). Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PLoS ONE, 6(7): e22746.

Herder, J. E., Valentini, A., Bellemain, E., Dejean, T., van Delft, J.J.C.W., Thomsen, P. F. & Taberlet P. (2014). Environmental DNA - a review of the possible applications for the detection of (invasive) species. Stichting RAVON, Nijmegen. Report: 2013-104.

Ho, T. W., Hwang, J. S., Cheung, M. K., Kwan, H. S., & Wong, C. K. (2015). Dietary analysis on the shallow-water hydrothermal vent crab Xenograpsus testudinatus using Illumina sequencing. Marine biology, 162(9), 1787-1798.

Leray M, Yang JY, Meyer CP et al. (2013) 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. Frontiers in zoology, 10, 34.

Leray, M., & Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first century. Philosophical Transaction of Royal Society B, 371(1702): 20150331.

Luin, M., Fui, C. F., & Senoo, S. (2013). Sexual maturation and gonad development in tiger grouper (Epinephelus fuscoguttatus) x giant grouper (E. lanceolatus) hybrid. *J Aquac Res Development*, *5*(213), 2.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... & Kondoh, M. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society open science, 2(7), 150088.

Morton, B., & Morton, J. (1983). The sea shore ecology of Hong Kong (Vol. 1). Hong Kong University Press.

Port, J. A., O'Donnell, J. L., Romero‐Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., ... & Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular ecology*, *25*(2), 527-541.

Scoble, J. M., & Cavalier-Smith, T. (2014). Scale evolution, sequence phylogeny, and taxonomy of thaumatomonad Cercozoa: 11 new species and new genera Scutellomonas, Cowlomonas, Thaumatospina and Ovaloplaca. European journal of protistology, 50(3), 270-313.

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next‐generation sequencing technologies for environmental DNA research. Molecular Ecology, 21(8): 1794-1805.

Stoeckle, M. Y., Soboleva, L., & Charlop-Powers, Z. (2017). Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PloS one*, *12*(4), e0175186.

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS one, 7(8), e41732.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–an emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation, 183: 4-18.

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PloS one*, *11*(11), e0165252.

Tréguier, A., Paillisson, J. M., Dejean, T., Valentini, A., Schlaepfer, M. A., & Roussel, J. M. (2014). Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish Procambarus clarkii in freshwater ponds. Journal of Applied Ecology, 51(4): 871-879.

Wilcox, T. M., McKelvey, K. S., Young, M. K., Jane, S. F., Lowe, W. H., Whiteley, A. R., & Schwartz, M. K. (2013). Robust detection of rare species using environmental DNA: the importance of primer specificity. PloS ONE, 8(3): e59520.

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., ... & Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific reports*, *7*, 40368.

Yeung, C. W., Cheang, C. C., Lee, M. W., Fung, H. L., Chow, W. K., Ang P. Jr. (2014). Environmental variabilities and the distribution of octocorals and black corals in Hong Kong. Marine Pollution Bulletin 85 (2): 774-782.

Appendix I List of project assets bought under the project

Project assets and expenditure details are not disclosed due to confidentiality reason.