Final Report

Brilliant Marine Research Idea 2017

1. General information

Title of the idea	Metabarcoding of museum samples stored in formalin: analysing the prey and microbiome composition of Antarctic fish (*Trematomus sp.*)
Name PhD student	Franz Maximilian Heindler
Name supervisor	Prof. dr. Filip Volckaert
Flemish University or Flemish University College	KU Leuven

2. Brilliant Marine Research Idea – Report about the activities

Abstract

We present the first study to assess the prey item and microbiome composition of Antarctic fish that have been stored in formalin for prolonged amounts of time. We applied a technique of treatment of the stomachs/hind guts with an alkaline solution and heat in order to extract DNA. We then amplified a region of the COI mitochondrial gene for stomachs and 16S rRNA nuclear gene for the hindgut samples to assess the prey item and microbiome composition, respectively. We employed extensive quality and contamination controls in order to ensure reliable results. We found that stomach samples were characterized strongly by contamination. In contrast, microbiome composition clustered differently from all our quality control samples. Therefore, we have high confidence in the quality of the microbiome composition data. Furthermore, we found a weak correlation between microbiome composition and standard length (size) of the fish. Due to high dropout rates among samples few were left and therefore the results are lacking statistical confidence. Results will be compared with contemporary samples once these have been further analyzed.

Intro

This PhD is embedded in the BELSPO project "RECTO" ("Refugia and Ecosystem Tolerance in the Southern Ocean"), in which causes and consequences of range expansions and reductions, respectively, are compared across various cold-adapted Antarctic taxa. My project in particular focuses on Antarctic icefish of the genus *Trematomus*, that have been driven into refugia in the past.

I am trying to understand how spatial distribution changed due to a changing climate and other factors, such as pollution or exploitation. For this I am combining biological data with state of the art genetic techniques enabling us to look back in time and make use of the vast amount of biological information that is archived in natural history museums worldwide. Specifically, I am comparing genetic variation of contemporary (2013 - present) samples from current research expeditions with historic samples (1899 – 1960) provided through collaborations with the Natural History Museums in Paris and London. We are furthermore collecting occurrence data from contemporary and historic databases (e.g. SCAR Biogegraphic Atlas of the Southern Ocean, biodiversity.aq) and match derived distribution patterns with catch records of museums all over the world.

This will yield insight in the past and current distribution patterns of the icefishes (notothenoids) of the genus *Trematomus,* which can be combined with the evaluation of the connectivity between different populations in space and time obtained from our genetic analysis. This study focuses on species of the genus *Trematomus*, which are amongst the most abundant members of the Antarctic icefish community. The genus is not exploited commercially and has a circum-continental habitat range.

My '*Brilliant Marine Research Idea'* was to use metabarcoding on the stomachs and guts of museum samples to make a snapshot of the prey item and microbiome composition at that time. This genetic multi-layer approach, understanding the spatial distribution of the predator and then correlate this to the diversity of prey items and microbial communities within the stomach and gut, has not yet been performed on museum samples. The composition of the microbiome within or on an organism can be strongly affected by environmental parameters, the host's diet as well as its behaviour, social interactions and stress level. This can be then put into context with the prey item composition.

By assessing these two aspects we aimed to understand the importance of climate change and pollution on Antarctic *Trematomus*.

Material & Methods

We sampled 200 fish from the Natural History Museum London. We extracted stomachs as well as hind gut samples to assess prey item and microbiome composition, respectively. We have tested a total of 6 protocols for DNA extraction and have found a combination of heating and submersion in an alkaline solution to yield the best results for extracting DNA from tissue that has been stored in formalin for a prolonged amount of time. After that, the Nucleospin (Macherey-Nagel) *DNA from Tissue Purification* kit was used following the manufacturer's protocol, with the sole exception that we used between 2 to 4 times of the chemicals, that are normally required. After DNA extraction we amplified a short region of the mitochondrial COI or nuclear 16S rRNA gene for the stomach or hind gut samples, respectively. This first Polymerase Chain Reaction (PCR) is also commonly referred to as Amplicon PCR, as the aim of it is to amplify the region (amplicon) of interest. To remove unspecific binding products as well as free nucleotides an AMPure (Agencourt) bead cleanup was performed. An indexing PCR was performed to attach indexing barcodes to the amplicons. Each individual sample receives unique barcodes, such that afterwards each sequenced read is assigned to a

specific sample. After the indexing PCR another AMPure Bead cleanup was performed. Quality was checked via gel electrophoresis and the concentration of the final PCR product was measured using the Quant-iT Picogreen (Thermo Fisher) kit. Samples of sufficient quality and quantify were then pooled and sent away for sequencing on an Illumina MiSeq Platform (Genomics Core, KU Leuven). Data was filtered, aligned and clustered using qiime and mothur through the online platform Nephele (www. nephele.niaid.nih.gov). Data was visualized in calypso (http://cgenome.net/).

Results/Conclusions

There was a large dropout rate at every stage of the laboratory protocol (Fig. 1), which illustrates the challenges to work with museum samples. Often, pre-selected fish from the catalogues are not suitable, due to size or they cannot be found at all. Then, many fish have missing stomachs or hindgut, because they have been either sampled before or the intestines have been removed completely already when sampled to ensure better preservation. Then, the fact that they have been stored in formalin complicates the laboratory work, manifested in the inability to amplify fragments and failed samples during sequencing. Lastly, due to the age and preservation method, we are working with low-template samples, meaning there is very little usable DNA in the samples. Therefore, contamination from the environment, other samples or researchers can occur easily, despite tremendous efforts to prevent such.

For the prey item composition (COI), we have found that museum samples show to be different from recent Antarctic fish samples (Fig. 2). We have implemented extensive quality controls, such as deliberate *Human Contamination*, *Workbench contamination* and *Blank* samples, in which human saliva swabs, swabs form the workbench and molecular grade water was used for DNA extraction, respectively. This was to ensure, that we are able to identify contamination from different sources. We found that museum samples were mainly characterized through reads that occurred in our contamination controls. When removing these reads we were left with too few reads to analyze the data. We conclude that metabarcoding of stomach samples is not possible using the methods used in this study.

Similar to the prey item composition data, we found that there is a distinct difference between recent and museum samples in terms of microbiome composition (Fig. 3). However, when analyzing the museum samples on themselves, we find that while some are contaminated (indicated by clustering together with the quality controls in the upper right corner), many are distinctly different (Fig. 4). While only few samples are left (n=27) after all dropouts have been accounted for, this is a remarkable result, as it confidently shows that the microbiome of museum fish is distinctly different from quality control samples. Furthermore, we find a weak biological relationship between the standard length of these fish and their microbiome composition. The fish are getting larger along the arrow. While this is not statistically significant, it might be due to the small leftover sample size.

We are currently working on contemporary *Trematomus* stomach and hind gut samples, to compare recent samples with the results found in this museum study. Once we have obtained data on current prey item and microbiome composition we will analyze these results further and make a detailed comparison between samples. Museum samples have time and again proven to be hard to work with, and this study is no exception. Resources, and time are in strong contrast to the little remaining samples sizes. However, museum samples do offer a unique opportunity to have a window into the past. Therefore, we strongly believe into the value of such studies, and in this one in particular, as it demonstrates the changes in diet and microbiome composition over time.

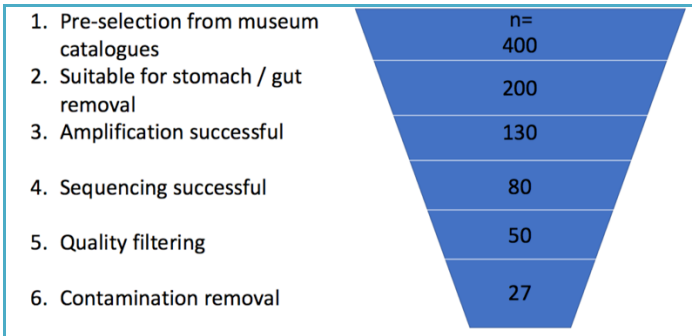

Figure 1. Sample size after each stage of the molecular laboratory protocol. With every step

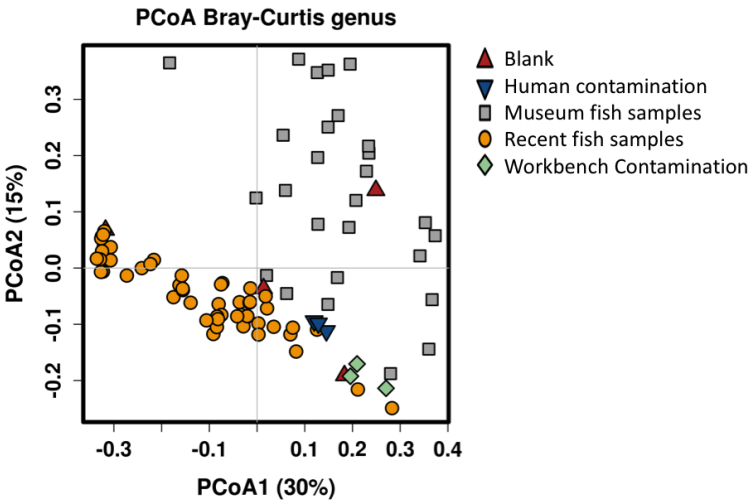

Figure 2. Principal Coordinate analysis of prey item composition from stomachs based on COI amplification of museum fish (*Trematomus sp.*) and other recent fish (various species). Blank, human contamination and workbench contamination samples are included in the plot.

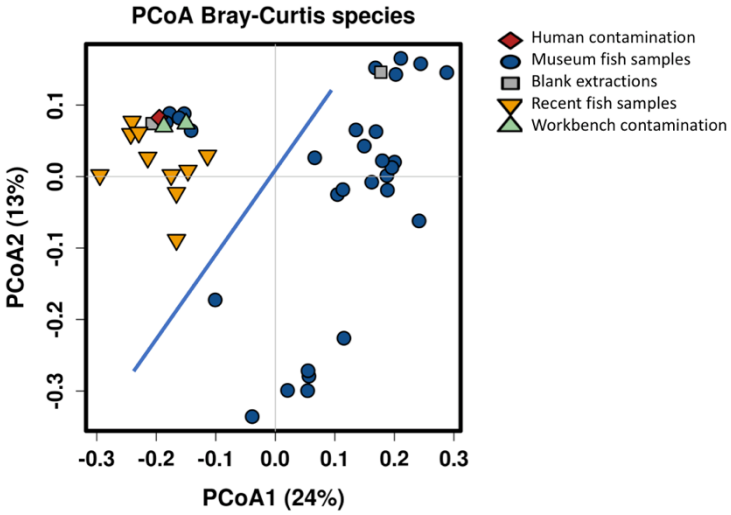

Figure 3. Principal Coordinate analysis of prey item composition from hind guts based on 16S rDNA amplification of museum fish (*Trematomus sp.*) and other recent fish (various species). Blank, human contamination and workbench contamination samples are included in the plot.

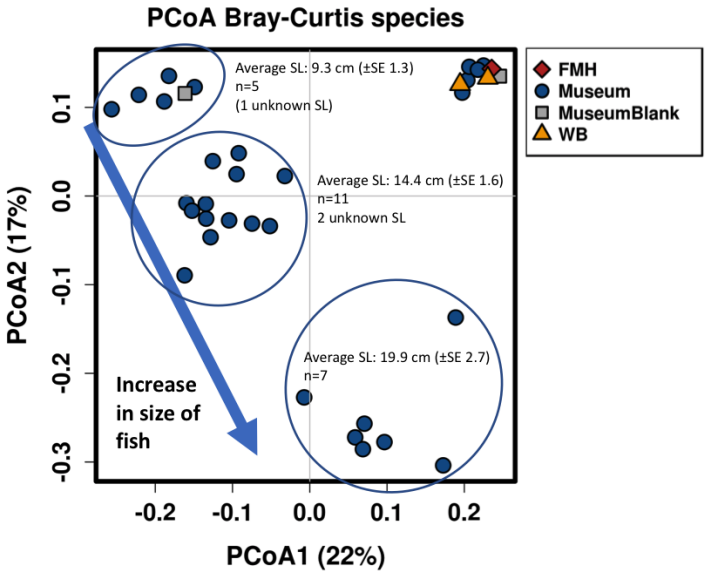

Figure 4. Principal Coordinate analysis of prey item composition from hind guts based on 16S rDNA amplification of museum fish (*Trematomus sp.*) and quality controls (Blank, human contamination and workbench contamination samples).

3. Overview of the expenditures

Describe in detail how the requested fund was spent within the implementation period (1 March 2017 and 28 February 2018). Be as specific as possible.

Even though we did not manage to process as many samples (n = 200) as initially anticipated (n = 400), the cost of the project remained the same, as we underestimated to cost per sample. Initial method testing was a lot more time and material intensive than expected. We had to repeatedly test six different extraction protocols. Conclusively, DNA extractions required between 2 and 4 times the amount of chemicals normally used.

Table. 1 Breakdown of costs in euro

Quantity	Description	Cost (euro)
1	Magnetic Beads Kit (PCR products, Size selection)	692.01
1	Mytaq (Ready reaction mix, 25ml)	809.50
1	MySeq Run (Sequencing)	2,130.00
3	Quality PCR	105.00

3	Bluepippin	225.00
	Total	3,961.51