An interesting case study of metabarcoding revealing unexpected diet in an endangered marine bird

Dear Marina Querejeta and colleagues,

Your article has now been reviewed by two referees and myself.

Both referee have found your approach valuable and results of interest and they both have serious and complementary recommendations to ascertain or improve data and interpretations. Based on their reviews and my lecture, I suggest this preprint deserves a revision. I hope comments below and referee's detailed suggestions will help. I suggest you follow those very detailed suggestions.

Particularly, referee 1 suggest detailed paths to re-analyze data and make sure a large set of sequences could not have been lost in the first bioinformatic steps, introducing a bias in the prey recognized with the sequences left. Reading the MS, I found no information on the way primers were treated and understand this wonder particularly considering the 300bp limit and the 2M merged reads for nearly 10M raw ones. This raises the point of access to data and bioinformatic scripts, that will eventually be required and would ease the review process. Would referee1 be right on his guess, results may come strongly different in terms or RAA and dominant preys, thus I suggest to follow his recommendation and make data and script available for the next round of submission. Referee 1 also made a number of suggestions to improve the MS. Among other the use of eDNA is confusing when it comes to diet, for stomach content or poo are not exactly environmental. The use of dDNA was recently coined by Sousa et al (2019, see below) for dietary DNA, and may be a nice option?

Referee 2 is more worried by the dominance of Talitridae in the inferred diet. First of all, i) looking at Table 1, these assignments seem to suffer from a small % of identity (around 80% , which means it may be a different family?) and ii) they were performed on NCBI database that is not curated and contains sometimes sequences associated to highly misleading identifications. Considering the dominance of the sequence of these OTUs in the dataset, despite the acknowledgement for possible secondary predation in the discussion (that would, I agree with referee 2, still sound even more awkward from cephalopods and fishes) , it may be advisable to also blast the results on curated or at least reduced databases such as Silva or Midori, in order to check these assignments, or to blast them individually and check the status/level of confidence of the blasted sequences (for example are there holotypes among those?). Would such taxa remain as dominant in the revised results, discussion should emphasize this level of uncertainty to clarify the assignment does not necessarily imply the taxa at the origin of the sequence is a sandhopper or a grandhopper. Referee 2 also made suggestion to improve the indirect evidence offered by metabarcoding by producing predictions for prey size, that may be interesting to consider. Finally, it seems this family of amphipod has exhibited range shifts associated to climate change and/or human activities in the Northern hemisphere (the baltic sea). I wondered if such report would exist in the Southern hemisphere Sousa LL, Silva SM, Xavier R. 2019. DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems. Environmental DNA 1:199–214.

Thank you for your useful suggestions, together with the ones proposed by the two referees. We have followed your guidelines in detail in order to improve our manuscript and we have uploaded the bioinformatic filtering script.

As suggested by referee 1, we have re-filtered and re-analyzed the data and all the bioinformatic steps are detailed in "Material and Methods". As shown in Fig. 1, with this new analysis, talitrids remain the most important prey item although their Frequency of Occurrence (FOO) is now similar to Chordata (fish) and the Relative Read Abundance (RRA) is comparable in the different phyla detected as prey items (Arthropoda, Chordata and Mollusca). We also followed the rest of the suggestions by Referee 1, including the change of the term eDNA by dDNA (dietary DNA), which we agree is more accurate for our approach.

Regarding Referee 2, we agree that the percentage of taxonomic assignment of talitrids is low. Thus, we try classifying the sequences using MIDORI. However, this did not add any new assignment to our OTUs. In relation to the trait-based approach suggested, we agree that it could be informative but we find that it would introduce several major bias as we explain in detail in our response below. In addition to addressing the minor comments in detail, we also followed the suggestion to include the rarefraction plots and the ordination (in supplementary material) to help with the understanding of the beta-diversity and community composition. Moreover, we have included the potential explanation of the geographical range shifts in Amphipods.

Reviews

Reviewer 2

Reviewed by Francis John Burdon, 2020-12-06 09:54

The authors have used metabarcoding of environmental DNA from fecal samples of tāiko (Procellaria westlandica), an endangered New Zealand seabird, to better understand its foraging behavior and trophic ecology. Two seasons and two sub-colonies were compared to assess temporal and spatial variation in the potential diets of tāiko. The authors found that surprisingly, talitrid amphipods dominated both the frequency of occurrence and relative number of sequence reads in the fecal samples from tāiko. However, more consistent with expectations, fish and cephalopods were also significant components of tāikos diet – with an important commercial deep-sea species (hoki) featuring prominently, thus suggesting a potential conflict with the fishing industry due to bycatch.

Whilst these are useful insights, the dominance of the Talitridae in the diet of tāiko is problematic. These amphipods, commonly known as sandhoppers or landhoppers, could be secondary prey (e.g., prey of fish) caught by tāiko, raising the specter of the troublesome "Russian dolls" problem in food-web ecology. However, that explanation is unsatisfactory, because it is difficult to imagine their fish and cephalopod prey relying on subsidies of terrestrial amphipods. I wondered if the problem could be more prosaic, reflecting a technical issue with the extraction of DNA. For instance, the chitonous

exoskeleton of the amphipods might not be fully digested by tāiko, meaning that this could lead to their DNA being overrepresented in the fecal samples. Such a problem could be exacerbated if the fecal samples were not fully homogenized before DNA extraction – it is unclear presently how the authors addressed this methodological issue (see L165-174). I also wondered how well described the Amphipoda are in the reference sequence databases used - could there be challenges on the bioinformatics side? If there was greater taxonomic resolution regarding the Talitridae, it might be possible to better understand where these putative prey come from – are they leaf-litter dwelling "landhoppers" or coastal "sandhoppers"?

Even if such technical issues do not explain the frequency and relative abundance of the Talitridae in environmental DNA from fecal samples of tāiko, it does highlight the potential limitations of this approach for characterizing food-web linkages. Other approaches, relying on biomarkers such as stable isotopes and fatty acids would be very useful for understanding the trophic ecology of this species by providing a time-integrative perspective. This approach might also be useful to understand the energy flow between foraging adults and recipient young - in the present study it is unclear if fecal samples were collected from adults or juveniles.

The use of prey traits could be another potential line of evidence that would be useful for better understanding the trophic ecology of tāiko. For instance, using optimal foraging models, it might be possible to predict the body size of prey most profitable for tāiko - which presumably is in a size range larger than the amphipods. A trait-based approach (e.g., using prey body size) would also help better understand how ontogeny might influence the diet of tāiko, since environmental DNA says nothing about the life stage of prey (i.e., are there size-refugia?). I am unware if such approaches have been applied to seabirds, but it would be a powerful way to leverage the novelty of metabarcoding environmental DNA data from fecal samples. For further information I can recommend Petchey et al. (2008). Size, foraging, and food web structure. Proceedings of the National Academy of Sciences 105:4191-4196. Alternatively, for a more simple, descriptive approach see Table 1 in van Donk et al. (2017) The most common diet results in low reproduction in a generalist seabird. Ecology and Evolution 7:4620-4629

The use of environmental DNA for understanding trophic ecology is an exciting new development in conservation. However, the present study possibly raises more questions than it answers, indicating that further analysis and follow-up studies are required. If tāiko are indeed strongly reliant on talitrid amphipods in their diets, then there could be concerns about whether this is sufficient to meet their nutritional needs (e.g., as opposed to a diet rich in polyunsaturated fatty acids from marine sources). As actionable suggestions, I recommend the authors consider if prey traits could add to their story (e.g., is it possible to model an optimal prey size for tāiko and then compare that information with the data generated by environmental DNA), and also explore if supplementary data (e.g., tracking data or direct observations) are suggestive of possible foraging by tāiko in the coastal environment or leaf-litter near their burrows, where they would most likely encounter talitrid amphipods. I found at least one publication where nearshore foraging activity by petrels was recorded, although it is by no means a silver bullet to the problem at hand. - see Thomas et al. (2006) Evidence for nocturnal inter-tidal foraging by European Storm-Petrels Hydrobates pelagicus during migration. Atlantic Seabirds 8:87-94.

We have appreciated the valuable suggestions and comments and we have followed them in as much detail as possible.

Following the referee's suggestions, we have re-filtered and re-analyzed the data. These results show that talitrids are still an important prey or secondary prey of Westland petrel. However, as shown in Fig.1, the Frequency of Occurrence (FOO) of Arthropoda (which are 99% talitrids) is similar to the other two phyla described, Chordata (fish) and Mollusca (cephalopods) and, in terms of abundance, talitrids and Chordata are similar, but much more abundant than Mollusca.

Regarding the importance of talitrids in our results, we discuss the possibility that talitrids are in fact secondary prey and propose an alternative explanation for their over-representation in our results. Our results suggest that petrels probably feed heavily on bycatch, which often comprise large quantities of fish guts and thus increases the representation of secondary prey in the diet of petrels.

Prey DNA would have been homogenised during digestion and therefore, no further homogenisation step was performed prior to DNA extraction.

There is indeed challenge in identifying talitrids as all corresponding OTUs showed less than 85% identity with the best match we could find. This limitation is due to a lack of sequences in existing databases and restricts us from being any more conclusive with regards to the taxonomy. However, we also discuss the possibility that talitrids could be, in fact, a potential prey of taiko as it is not the first seabird that feeds regularly on amphipods. It is known that penguins and, also other Procellariiformes, such as, the Providence petrel and the Wilson's storm petrel feed on amphipods when krill is not available. Also, as suggested by the reviewer, European Storm-Petrels (and other Procellariiforme species) feed in the coastal environment where amphipods are present. So, it may be possible that Westland petrel do the same. However, the lack of complete databases on talitrids have not enabled us to go further on the classification of these amphipods, despite the fact that we tried to complement the taxonomic assignment against NCBI with MIDORI, although the results were not included as they did not add any new assignments to our data.

We agree with the reviewer about the limitations of our approach on this aspect. In our study, we did not collect fecal samples from inside the nest as this would have disturbed the juveniles. Since all fecal samples were very fresh and mostly line shaped, there were most likely produced by flying birds i.e. adults as they land/crash through the vegetation. Also, we do mention fatty acid and isotope -based approaches in the introduction of the manuscript, however, our aim here was to reach a high taxonomic resolution, which is why we chose a DNA metabarcoding approach.

A trait-based approach of the prey could be informative, however, it would inevitably face a major bias caused by age-related size variations in fish and cephalopods. Using the data at hand (diet DNA analysis), it would not be possible to propose such analysis. In Hoki for example (the main predated fish species), adult size varies from 60 to 120 cm, but juveniles can be much smaller, and may be more represented in by catch. In addition, pieces of bigger fish (e.g. guts) could also be found in large quantities in bycatch. So there is no way to estimate accurately the size of a predated item based on its molecular identification in the feces of the birds.

Despite the limitations of the taxonomic assignment in amphipods, we believe that our approach, together with the reviewers suggestions, add valuable information for understanding the trophic interactions of the ecology of Westland petrel. Although we cannot assure whether talitrids are primary, secondary prey, or both, we can confirm that they are key taxa in the flow energy through the food web. Finally, consistent with previous studies, we were able to detect that Actinopteriigy fish and cephalopods were also main prey items of Westland petrel.

With these major concerns raised, I have outlined some more minor comments and suggestions for the authors below.

Minor comments

L25 I think it might be not be totally accurate to call these impacts direct effects, since indirect effects are definitely plausible (e.g., changes in the diets of competitors could indirectly affect tāiko).

This suggestion has been included in the manuscript.

L27-29 I think this point could be made more effective by explicitly mentioning the spatial component (e.g. their mobility and foraging behavior enable tāiko to exploit prey distributed patchily in time and space)

This suggestion has been included in the manuscript.

L38 “geographical” = spatial

This suggestion has been included in the manuscript.

L48-51 I would consider breaking up this sentence e.g., “This work demonstrates the potential for environmental DNA to inform the conservation of endangered species with elusive foraging behaviors. In our example, environmental DNA provided valuable information regarding the diet preferences of an iconic species contributing to New Zealand’s unique biodiversity.”

This suggestion has been included in the manuscript.

L56 Could drop “in particular” since it is redundant

This suggestion has been included in the manuscript.

L61 Consider making a new paragraph here (also at L77, L98, and L136).

This suggestion has been included in the manuscript.

L144 The sub-colonies are only 1.5km apart yet there seems to be some differences in their diets. Has anyone compared the population genetics of these two populations – presumably there is quite a lot of mixing? Was any attempt made to identify the fecal samples to the individual birds – since this non-independence could skew differences

between sub-colonies (i.e., once individual differences were accounted for, there might be no difference between locations).

We included in the manuscript, that each faecal sample likely belonged to a different bird individual because, the collected material was very fresh and usually line-shaped, which could only correspond to feces produced by birds as they they landed on the previous day. Hence, each bird could only produce one of these faeces, and samples were considered independent. Very few no older faecal samples were observed on the sites, as these were probably rapidly washed away in this extremely rainy location.

L167 Here the authors state that a QIAamp DNA Stool Mini Kit was used. Are there any pros and cons of using this product? I ask because in a recent study I read, the authors instead used a DNeasy PowerSoil Kit for extracting DNA from scat samples – see Harper, L. R., et al. (2020). Using DNA metabarcoding to investigate diet and niche partitioning in the native European otter (Lutra lutra) and invasive American mink (Neovison vison). Metabarcoding and Metagenomics 4: e56087.

Although DNA from faeces samples can be extracted with a wide variety of kits, the QIAamp DNA Stool Mini Kit was used in this case, as its efficiency has been proved especially in bird diet analyses (Travelline et al., 2016, Travelline et al., 2018). This is potentially due to the InhibiteX tablets, which neutralize the possible inhibitors present in the scats.

Trevelline, Brian K., et al. "Molecular analysis of nestling diet in a long-distance Neotropical migrant, the Louisiana Waterthrush (Parkesia motacilla)." The Auk: Ornithological Advances 133.3 (2016): 415-428.

Trevelline, Brian K., et al. "DNA metabarcoding of nestling feces reveals provisioning of aquatic prey and resource partitioning among Neotropical migratory songbirds in a riparian habitat." Oecologia 187.1 (2018): 85-98.

L165-174 What was used to homogenise the samples – a tissue lyser?

A portion of the sample was taken with a cotton swab in order to extract DNA, and subsequent library preparation. No homogenization was conducted as the digested material was thoroughly mixed during digestion and the faeces had a homogeneous aspect.

L166 Did the authors consider using technical replicates from the samples before pooling?

No, we did not use replicates in this approach.

L177 It would be useful to state the full name of the primers used.

Original names of the primers are given in methods, together with its original reference.

L185 Did the authors consider using positive and negative controls with the PCR runs?

Negative controls were included in each PCR run. This is now mentioned in the manuscript. In addition, a negative control sample (DNA extraction from leaf litter collected on site) was included in the library and was also sequenced.

L248 What was the number of sequences that the samples were rarefied to? It could be useful to provide the rarefaction curves in the Supporting Information.

We did not rarefied the sequences other than removing singletons. The rarefaction curve originally cited on line 248 has a different purpose. It aims at estimating the proportion of the diet we were able to describe from the samples analysed. This figure is now available in the supplementary material (Figure S2). In addition, based on the reviewer's comment, we now provide the number of sequence reads per OTUs detected (Fig.S3) as well as the cumulative frequency of the OTUs detected (Fig.S4). These figures show that relatively few OTUs make up for the vast majority of reads.

L286 Did the removal of singletons have any effect on the results?

No, it did not have any effect on the results. However, we consider singletons (abundances equal to 1) as potential artefacts.

L288 Probably better to say "did not match"

This suggestion was included in the manuscript.

L293 So this is another way of saying that 12 samples only had unassigned OTUs, and thus were not used in the analyses?

Yes, this is now changed in the manuscript.

L295 I felt that the results could have done with some multivariate analyses to assess changes in overall prey communities – for instance Non-Metric Multidimensional Scaling plots with a PERMANOVA model to test differences between sampling dates and sub-colony location.

We did perform multivariate analysis, but we chose to use multiple GLM approach, which is model-based and improves power across a range of OTUs with different variances (as is the case here), compared to distance-based approaches such as NMDS for example. (See Fig. S5)

L483 Might be better to use concordance here instead of "coherence"

This suggestion was included in the manuscript.

Reviewer 1

Reviewing the: Investigating spatiotemporal variation in the diet of Westland Petrel through Metabarcoding, a non-invasive technique

Dear Editor and Authors

This study is a very good application of genetics to answer a specific ecological question of endangered seabirds in a non-invasive way. The feces of 99 Westland Pertles were analyzed with 16S barcodes using two primer set-ups. Results are presenting several unexpected fish species, cephalopods, and underlining the importance of crustacean prey, even if only assumed as secondary predation. Finally, indicated are the implications of human impact on the diet of these nearly extinct seabirds and answering ecological questions about their foraging and behavior between time and space.

Besides minor comments, there are some arguments absent in the introduction, and discussion could improve the manuscript. Likewise, the use of the term eDNA is questionable as the analyses taking place with an eDNA pipeline, but analyzed is DNA of feces. Finally, I have some major technical considerations regarding the bioinformatics pipe and the following data analyses. Based on the given methods description and the surprisingly low merged sequence outcome, I suggest the primer sequence were not cut off, therefore a lot of fragment was accidentally filtered by length. Suggesting the analyzed data are not complete. Besides, data presented as one experiment, as they are one gene region. Still, used are two primer set-ups, with different fragment lengths, which affect the sequencing outcome. This is left out in the given analyses, results, and discussion. Both points could explain why the suggested secondary predation presented as the most consumed prey. I would suggest to reanalyze the data with the recommendations given in detail below and to include this information in the results and discussion. Redoing the analyses is a lot of work, but it would exclude doubts of possible bioinformatics flaws and improve the outcome, to be the outstanding study it can be.

Detailed revisions

General and major concerns

1. You using an eDNA approach but you do not analyses eDNA itself. This is a question of how to define eDNA. Over the last years, a lot of development was going into analyzing eDNA by Metabarcoding and it is great to see the broad usage for more ecological studies and specific questions like this. It is worth mentioning the use of an eDNA pipeline, but it is not eDNA analyses, this are not mixed samples from the enviroment, you analyzed specifically faces. You used the litter, which is eDNA but as a negative control to exclude potential eDNA contamination. It was a thoughtful way to exclude potential contaminations. However, it is missing in your bioinformatics descriptions of how you used this negative sample, and your positive bulk sample, which is not essential but would improve to understand the workflow. But there is more than eDNA, there is bDNA for bulk samples, aDNA for ancient. Why not starting something new, DNA for feces, or more general dDNA for dietary. There will be coming more studies about Metabarcoding of feces and gut content, so it would make sense

2. Another point is a little reshaping regarding soft tissue prey; only some sentences confusing the reader through the whole manuscript. You found cephalopods indicating soft tissue prey,

which is amazing, but I believe you did not found them on purpose. It is a little roller coaster for a reader. In the introduction you explain this advantage, your primer set up explaining specifically to exclude everything else than fish and crustacean, then you describe a special identification set up for Mollusca, and in the discussion finally only mentioned that is known prey. I give line-specific lines below where I trembled.

3. I am wondering, you did not mention how you trimmed the primer sequences. In line 176 you talking about demultiplexing and adapter trimming, which are the Illumina adapters binding on the flow cell and their barcodes for multiplexing. I don't know any company removing primer sequences. And you using a two-step primer system, so you have also additional adapters included, that concerns me a lot. Normally I would suggest you forgot to mention it, but after reading your results, it's the best explanation about the things that did not make up.

I have a suggestion, it is written online that the 16S fragment for this primer is ~180-270bp long. I suggest you have forgotten to remove your primers sequences. Therefore, your fragments for chordate were actually over 300 bp long and you cut them off in the merging process. It would also affect the taxonomic assessment.
Most commonly used is the program cutadapt. The best would be to merge the sequences, cut of primer, and then cut off sequences by length
https://cutadapt.readthedocs.io/en/stable/guide.html#trimming-paired-end-reads.
Because you using 2 step PCR for library prep you have to use cutadapt only once, however, if your library is prepared with ligation you have to use it twice. And don't forget to use reverse your reverse primer. It sounds stupid but is probably the most common mistake in analyzing Metabarcodes. Check your outcome, using Miseq allowing up to 2-3 mismatches I will normally be able to recover 96-99% of the sequences.
And if you redo your analyses please include a Q score cut off. The standard for Illumina based reads is 30, to my knowledge, there are a lot of journals seeing this as the minimum standard and do not accept anything below.
4. Another advantage of cutatapt is you can analyze your barcodes separately by primer. I have a major problem with the fact that you show no results separately by primer because they have major fragment size differences. You assume that those identifications based on the primers, but which primer sequenced the Mollusca? This is an assumption and not a result. By cutting primers, which you need to do anyway, you should separate bioinformatics pipe for both. You still can analyze them together, but you need to show it also separately. The fact crustacean is most dominant, and their RRA higher than fish can be explained by two simple methodical facts.
1; Primer affiliation, one set up can simply be worked better, you did not equalize the amount of PCR before pooling. Because of the magnetic bead cleaning, you have automatically a separation of length. Depending on your ratio smaller or larger fragments a preferred.
2 Illumina prefers short sequences. Therefore, to calculate the read abundances or RRA based on actually two experiments is questionable. Which not says you can't do it, but you have to acknowledge that.

However, why do you have more prey than a predator? That is the point, which makes it clear for me that you probably lost most of your fish sequences and/ or cannot equalize data into one RAA calculation from varying fragment sizes of different primer setups, which can lead to this misinterpretation about what is the dominant prey.

We have appreciated the valuable suggestions and comments and we have followed them in as much detail as possible.
Regarding the fact that we are exploring the diet of the Westland Petrel using a DNA metabarcoding approach, we fully agree that using the term dietary DNA (dDNA) is more correct and we have included this term in the manuscript instead of using environmental DNA (eDNA). Moreover, we have included the corresponding reference: Sousa LL, Silva SM, Xavier R. (2019). DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems. Environmental DNA 1:199-214.

Concerning the soft tissue prey detection, we agree that reshaping the manuscript was necessary to avoid misunderstanding of the reader. Although the primer from Deagle et al (2009) is called Malacostraca, it has been proved to amplify efficiently Cephalopod DNA (Olmos-Pérez et al., 2017). We have now included that information in the section Material and Methods within the manuscript to make clear from the beginning that we were looking for cephalopods and we found them.

Olmos-Pérez, Lorena, et al. "Diet composition and variability of wild Octopus vulgaris and Alloteuthis media (Cephalopoda) paralarvae: A metagenomic approach." Frontiers in Physiology 8 (2017): 321.

Regarding the filtering and analysis of the data, we have re-analyzed everything from the beginning and we have detailed all the bioinformatic steps in Material and Methods. In short, we analyzed the data of each pair of primers separately, however, only the data from the Malacostraca primers were used in the subsequent analyses as the Chordata primers did not perform well enough to include in the results. Nevertheless, the Malacostraca primers performed very well and we were able to detect an estimated 90% of the prey items within the diet of the Westland petrels. Thus, all the results in this manuscript belong to amplicons obtained with the Malacostraca pair of primers.

Concerning the steps followed, we included the trimming of the primers with cutadapt software and we raised the Phred score threshold to 30 while merging forward and reverse reads, which was performed using PEAR software. The rest of the filtering performed by a toolbox are detailed in “Material and Methods” with some modifications in comparison from the first manuscript submitted. Filtering scripts have also been made available.

As a result of this new analysis, we were able to merged almost 85% of the data. Although the global composition of the diet of the Westland petrel did not change and talitrids remain one of the most important prey, there are less differences in terms of Frequency of Occurrence (FOO) with fish prey items (Chordata) and in terms of Relative Read Abundance (RRA) with the other

two phyla (Chordata and Mollusca). Moreover, these re-analyses made the differences between seasons even more obvious as, for instance, fish is the most common prey before hatching (BH) compared to chick rearing season (CR) when talitrids are the most common prey. So, at this point, and with only one set of primers where the differences in fragment sizes come from the variation of 16S, we could potentially say that our OTUs are representative of the biological community of prey within the diet of Westland petrel. The suggestions followed and included in the manuscript were incredibly valuable to improve the quality of the manuscript, and again, we would like to thank the reviewers for their insight.

Detailed

Abstract

L 48 as seen above its fecal DNA analyzed by Metabarcoding

This suggestion has been included in the manuscript.

Introduction

L98. Remark that this is an advantage but you found accidentally soft tissue prey, your method was planned on even this biased characterization; There coming comments later but I would start here reshaping a little to not confuse readers and make it a little more smooth to follow. It would better fit in the discussion.

DNA of cephalopods was not found by chance in the faecal samples. The so-called Malacostraca primers by Deagle et al. (2009) are not specific for Malacostraca as they are also able to amplify Cephalopod DNA as shown in Olmos-Pérez et al. (2017). Also, previous studies regarding the diet of the Westland petrel found cephalopod in its diet and that is the main reason why we specifically looked for this group (Imber, 1976 & Freeman, 1998).

Freeman, A.N.D., 1998. Diet of Westland Petrels *Procellaria westlandica*: The importance of fisheries waste during chick-rearing. Emu 98, 36–43.

Imber, M.J., 1976. Comparison of prey of the black *Procellaria* petrels of New Zealand. New Zeal. J. Mar. Freshw. Res. 10, 119–130.

L102 I would not say that these are recent methods, they are standards, or do you mean they only started to use them recently in this field of research? Please give some references about the use and the outcome of switching feeding sites

We agree that they cannot be called recent methods and the references have been included in the manuscript.

L141 The use of 16S would not be my first choice, you selected your genetic approach based on several factors. Can you please include some context about that in the introduction? Then it is

easier to understand how you chose your approach/gene regions for hopefully a lot of following fellows.

This suggestion has been included in the manuscript.

Methods
You wrote especially the bioinformatics detailed, that anyone understands it. I appreciate this, but because of that, it is sometimes more confusing when you explain a logical consequence. For example, you remove chimeras, and then you say "the single file without chimeras". So I am confused why you have to say that, did you split files, did you need other commands to remove them? Hope to get my point

This was changed in the manuscript.

L170 "only" one centrifugation, rather than "two"

This suggestion has been included in the manuscript.

LL.170, it is focusing to follow without the protocol in place, and they can change over time. You can explain the steps in more detail or give exact detail about the version of the protocol you working on.

We used the QIAamp Fast DNA Stool Mini Handbook from 03/2014, (reference: 1081060_HB_LS). This is now included in the manuscript.

L176-182 confusing…, you write you have Chordata and specific Malacostraca Primer that is not a wide range. Secondly, the results you talk about arthropods in general and detections of mollusks. Thirdly, you talking about the detection of soft tissue prey in the introduction, so the choice of these primers and target groups is not understandable. If you chose, primers based on the described diet, but found unexpected prey, then structure your whole manuscript in this way to stay in a logical flow.

As stated before, we were specifically looking for cephalopods using the so-called Malacostraca primer. This information is now included in the manuscript.

L 186 gives the protocol for the bead cleaning. Which concentration and which beads did you use.

This suggestion has been included in the manuscript.
L 186 you mean both gene regions? So the primers were not tagged? If your Primers were added with any nucleotides etc. for following the second stage PCR you need to mention it.

Yes, both gene regions were pooled and tagged with the same adaptor during the second PCR (ligation).

L187, which two-step protocol is used?

Amplicon library preparation using Nextera barcode indices, followed the Illumina "16S Metagenomic Sequencing Library Preparation Manual" Rev B. This has now been included in the manuscript.

L190 good that you include a mock but describe mock communities and their role in methods in detail. Are they PCR products with the same primers? How, many species, etc.

The inclusion of a mock community was standard practice by the sequencing company, but this mock community is designed for microbial studies. Un surpisingly, this mock community did not produce any reads, which is why we did not give any further information about it.

L191 delate (600 cycles)

Deleted.

L 195 The adapter trimming and demultiplexing is normally is done automatically by the Miseq machine. I would delete that sentence because it is indicating this is done in a special and not following the normal Illumina protocol. If that is the case, please give more details.

Deleted.

L198, 25 is high, only out of curiosity, why, because your fragments are short? Because the untrimmed end otherwise inhibit merging? because this troubles me in line 284

The pipeline was changed and the merging was performed with the software PEAR, so this part has been removed from the manuscript.

L 203-204 That is not important and can be deleted.

The pipeline was changed and this part is now different.

L 207 what is pre clustering?

Pre clustering is grouping the sequences using a 98% of similarity before chimera detection following the *vsearch* pipeline. However, we have used another toolbox this time and this was deleted.

L 210-212 This sentence is confusing, of course, there are no chimeras or singletons, you removed them. And I hope it is only a single file after the second demultiplexing, otherwise, it

means you deleted replicates in different samples in the step l 206. So this is another pipeline or what you explained before? Than at the beginning not at the end. Besides, it would be better to give your Perl script, or simply “based on”, the word customized is troubling me a little.

Another bioinformatic toolbox was used in this version of the manuscript. Thus, this was deleted.

224-226 I am not sure what you mean with equal to 1? All singletons where removed, you mean OUT’s with a lower percentage per sample/ overall than 1% is deleted? Which is reasonable, but not written here.

We rewrote this (L305), but basically, we removed singletons, that is to say, read abundances equal to 1.

226-227 useless sentence, delete

Done.

227 it is not or secondary prey, until discussion everything is potential prey. To delete “or secondary prey”

Done, it is now only present in discussion.

231 and here as a reader I gasping, you going for Chordata and crustacean, and know there are Mollusca, but you treat them differently. I think I know why, but please clarify this in the text. What is with the other stuff you did not expect? It is normal to find something different, but as a reader, I do not get your logic of processing here.

This was changed as we used another method for assigning taxonomically Mollusca.

232 “distance matrix of the alignment of the sequences”, don’t understand what this is.

This sentence has been removed from the manuscript.

L251 Deagle et al. counted only presence if minimum 1 %, did you do the same? It is important because Deagle et al. use this rule only for presence-absence, not for semi-quantitative analyses as RRA.

No, we simply discarded the singletons, as we considered them as potential contaminations. We only took out data that represented less than 1% to prepare the plots of abundance and presence absence, in order to have a better visualization of our results.

266 Be aware, people can confuse the abundance with the RRA and occurrence with FOO from the part before. So, I would advise you to slightly restructure the text. So that it is clear, you

have two tables, one with abundance and the same but as presence-absence. With these two tables and your factors, you do two analyses, one calculating the RRA, etc., and secondly the GLMs. Which is a nice strategy by the way.

We rewrote and we think it is clearer now.

Results
284 I am wondering, you have a short fragment, sequenced completely by one side, and even can merge partially over full length, also you allow a minimum of 10% error in merging (25 of theoretic 250bp), much more for the smaller region. But you have less than 50% merging success. Would be good for Nanopore sequencing but not miseq. Something is wrong.

This was redone and the read merged are different in this version of the manuscript (84.97% of the reads were merged).

L290 define contaminants

We considered contaminants all the taxa that was sequenced but cannot be a potential prey (such, for instance insects). Taxa that were doubtful were included but discussed with caution. This is, for instance, the case of talitrids.

L338 proxy of biomass is discussable, especial using your set up, I would suggest leaving it out here, and discuss this later

Done.

Discussion
Your discussion is detailed in the different prey species, and to explain the possibilities of how you could find them. But you argue is purely ecological based not on the methodology itself. Metabarcoding gives us many possibilities but it is not flawless. If you exclude this fact, it can look more like a forced explanation of your data, not a discussion that allows questioning your results. Especially as you underline you are the first doing this approach in this field, so I would expect feedback about that.
Points of suggestion
-The different length of your fragment, different primer see Who is eating what: diet assessment using next-generation sequencing from FRANCOIS POMPANON et al. 2011, or the various studies of Deagle
-why did you find soft tissue prey even if not expected, or how can you improve future studies to further analyses this issue. For example how about potential gelatinous prey, by including 18S setups.
L450, if you knew the eat cephalopods why did you not include them into your setup?

They were included in the setup, as the so-called Malacostraca primer has been proved to amplify Malacostraca in an efficient manner (Olmos et al, 2013). As stated before in the general

answer to your review, all the results shown in terms on abundance, occurrence and OTU diversity come from the same pair of primers (Malacostraca)

L460 This is likely, and we have similar results in a submitted study for a fish top predator. However, why do you have more prey than a predator?

As you suggest below, one potential explanation is that Petrel feed extensively on fish leftovers, in particular the stomach and guts, that are discarded in the water by fishermen. We have now included this potential explanation in our discussion.

L488 It is an important point, but what is your suggestion with that, please finish your thought. Is it likely they eat more amphipods because of insects because of that? Or connect it to the next paragraph

We finish this paragraph and we think it is now clearer.

492 Funny enough; scavenging the leftover fish stomach thought into the water by fisherman would explain the high amount of amphipods, or not?

It is, indeed, a possible explanation, together with the fact that amphipods are also prey from some fish and cephalopods detected as prey of the Westland petrel. We cannot assure what is the real explanation. It could also be a mix of both, and, it could be in part also primary prey.

523 please give the RRA; it is a much more reliable indicator here.

Done.

562 I like this paragraph, for me a highlight

Thank you.

579, Could these sub-colonies be genetic haplotypes? I think this is a nice fact that can indicating follow-up questions and already mentioned here. But this is only a suggestion

We agree, it is a possibility and it would be interesting as a continuation of this work.

599 Put this sentence somewhere else, it is important, but you building up for the next sentence line 602-604.

Done.

602-604. Rethink your structure for building up for this conclusion. I think it is an important point but is a little detached from your findings. I suggest putting it in connection with the

evidence your study gives or underlines like finding deep-sea species, scavenging, commercial fishing of known prey, and cephalopods.

We have re-structured the conclusion.