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.

Detailed

Abstract

L 48 as seen above its fecal DNA analyzed by Metabarcoding

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.

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

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.

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

L170 “only” one centrifugation, rather than “two”

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.

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.

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

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.

L187, which two-step protocol is used?

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.

L191 delate (600 cycles)

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.

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

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

L 207 what is pre clustering?

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.

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.

226-227 useless sentence, delete

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

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.

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

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.

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.

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.

L290 define contaminants

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

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?

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?

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

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

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

562 I like this paragraph, for me a highlight

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

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

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.