Editor's assessment:
I do agree with the reviewers this is a nice piece of work that should be published as soon as the issues raised by the reviewers are addressed. I saw the data have been deposited in the NCBI databases, and are under embargo until February 2021 currently. I hope this manuscript will be accepted for publication sooner than that, so I would like to ask you to remember changing the embargo dates. In addition, I have a few minor issues I would also like you to address:
We thank the editor for taking the time to assess our manuscript and provide feedback/suggestions.

We have released the embargo such that the data is publicly available now (BioProject ID PRJNA663974).

To improve the reproducibility of our study and make it more open we have also created a github repository for this manuscript that contains all of the supplementary figures, files, GPS coordinates and detailed local maps of samples collected, and scripts used in the analyses (Jupyter notebook for QIIME2 BASH code, Rmarkdown for the R code):
https://github.com/EisenRa/2020_SHNW_Faecal_16S

Lines 64, 68. XXXX [12] demonstrated… XXX [13] showed…
Fixed

Line 78: there is an extra space between goronoruym) and [15]
Fixed

Line 146: this statement requires a reference.
Added the Song. et al. reference:
Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, et al. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. mSystems [Internet]. 2016

Line 406: move reference to the beginning of the sentence: ...other mammals [56], the...
We have moved the reference to the start of the sentence as suggested.

Figure 2: specify the type of habitat of each sampling area, so that the figure is as descriptive as possible.
The legend for figure 2 now reads:
“Microbial diversity and composition of SHNW faecal samples from different populations: Kooloola and Brookfield (degraded habitat), Wonga (intact habitat), and captive.”

We have also added this information to the other appropriate figure legends for clarity (tracked changes in revised manuscript).

Figure 3: polish the representation of taxonomy. Ideally, colours with similar tones should also represent major bacterial clades. In the Venn diagram, it would also be preferable the size of each circle to represent the diversity (grey circle smaller than the green circle). Also, and ideally, if the colour that will represent wild populations is green, it should also be green (different tones) in Figure 2 and 4.
We have remade the Venn diagrams in R for reproducibility, and provided the code in the .rmd file on the github repository. The Venn diagram for figure 3 (captive vs. wild) is now scaled by the number of ASVs. However, we could not do this for the 3-way wild population comparison (Figure 5) as some circle configurations for 3-way Venns are mathematically impossible! We have changed the colour of the wild circle in figure 3 from green to white to keep things simple. For the taxonomic bar plot in figure 3, we have reduced the number of families represented to 20 (from 25) for clarity, and changed the colour palette. We would prefer, if possible, to keep the colours as different as possible (not different shades of a single colour for each phylum), as we feel this allows for better discrimination.

Figure 2 and 4 are redundant. Remove one, or move it to supplementary materials.
We would prefer to keep both figures if possible, the reason being that the inclusion of the captive samples in figure 2 has a massive impact on the ordination that 'masks' the differences observable between the wild populations (whereas we observe distinct clustering of Kooloola and Brookfield populations in figure 4).

Reviewer reports:

Reviewer #1:
The authors of this paper investigate microbiota differences between captive and wild Southern Hairy-nosed wombats (SHNW) as well as differences between three populations of wild SHNW inhabiting habitats differing in their degree of native flora. The findings of this very relevant study could potentially lead to improved management of endangered species being bred in captivity with the aim of stocking wild populations. Overall the manuscript is very well written and I have only a few comments, mainly to the method section of the manuscript.
We thank the reviewer for their kind assessment of our paper.

In the methods section under Sample collection I would like the authors to first of all specify in the text (and not only in the figure legend) how many samples they collected and to specify whether each sample originates from an individual wombat, a burrow or the entire area with the chance of different samples origination from the same individual. If the latter is the case, then it should be addressed that the intra-habitat

similarity between samples could be caused by the samples originating from the same individual.
We have added the sample numbers and warrens numbers for each site in the methods section as suggested, and the number of different warren complexes sampled from each site:
Kooloola = 11 warrens sampled (21 samples)
Brookfield = 8 warrens sampled (17 samples)
Wonga = 11 warrens sampled (23 samples)
SHNW warrens can be large and complex with multiple entrances and SHNWs living in each warren. We are unable to determine whether some wild samples originated from the same individual, though given the very small home range of SHNWs 2-4 Ha (Finlayson et al. 2005) it is unlikely that the same individual was sampled from different warrens. However, it is possible that the same wombat could be sampled more than once from the same warren complex, though we attempted to mitigate this by sampling from entrances that were as far away from each other as possible. In the extreme hypothetical scenario where all samples from the same warren came from the same animal (unlikely given compositional differences observed between most samples from the same warrens -- SI file 1 viewed with 'Warren' metadata labelling), this would not change the overall results and clustering of samples between populations.

We have also created detailed maps for each site showing GPS coordinates for each warren and sample that was collected:
(https://github.com/EisenRa/2020_SHNW_Faecal_16S/tree/main/Site_maps).

The authors also mention that some captive individuals are sampled up to three times. Are these samples then treated as all the other samples collected or are they pooled before downstream analysis? In general there needs to be more information about the likely origin of the samples.
For the captive individuals that multiple samples collected, we processed each sample separately, then randomly selected one sample per animal for subsequent analyses. We have clarified this further in the methods.

Another point related to sample collection that I would like the authors to address in the discussion section is that the samples from the degraded habitats are collected three months earlier in the year than the samples from the intact habitat. I am not an expert in the seasonal climatic variation of Southern Australia, but for many animals seasonal variation (e.g. temperature and food availability) are likely to play a major role in shaping the gut microbiota. It would be great if the others could include this in the discussion.
This is a very good point, and we have added a paragraph in the discussion about it. While we were unable to collect the Wonga samples contemporaneously with the

degraded habitat samples, we would not expect this three month difference to have a substantial impact due to the paucity of rain received at these sites in the first half of 2019. See the new discussion paragraph for more information.

A final comment to the methods section is regarding the removal of the seven outlier samples, which is also linked to the missing information on individuality of the samples. If these seven samples are collected outside the burrow where other non-outlier samples were also collected, then the author's explanation about taphonomy seems very likely. The readers need to know if this is the case to access if the author's decision to remove these samples are valid. There is also a sample from the captive animals that seems to be an outlier. Why is this sample not excluded like the others?
These outlier samples came from the four Eastern-most warrens labelled M, P, O and Q in this figure: https://github.com/EisenRa/2020_SHNW_Faecal_16S/blob/main/Site_maps/BrookfieldMap.png
We have added this information to the appropriate method section. We are still puzzled as to why these samples are so distinct to other samples. Their diversity is far greater (mean 1256 ASVs vs. 828 ASVs for other Brookfield samples). However, the near complete lack of *Spirochaetaceae* in these samples compared to its high prevalence and abundance in all other samples (mean relative abundance in other wild samples of 11.6% and captive samples at 5.1%) suggests something unusual. However, there could be biological differences in these samples that we cannot determine (host health/disease, etc.). We have added this information and alternate explanations into the method section.

We were on the fence about removing the captive outlier sample. While this sample had lower observed ASVs than other captive samples (463 vs mean of 750), compositionally it was intermediate between captive and wild samples, and we didn't want to throw it out just because it didn't fit with the other samples in the captive vs. wild trend.

In the discussion section the authors mention in line 389 that a previous functional profiling of a captive SHNW showed substantial capacity to degrade complex plant polysaccharides. Wouldn't this suggest that important microbial functionalities endures captivity even though the taxonomic composition changes? It seems like this is in contrast to the point that the authors are trying to make.
Our intention was to illustrate that there are indeed functions within the SHNW gut microbiome for plant fibre degradation. All SHNWs in captivity would likely be receiving some amount of complex fibre, with the types and quantities varying from place to place (the diet of the SHNW individual referred to is not reported). While Shiffman et al. reported such functions in a captive SHNW, we don't currently have a wild comparison to determine whether captivity results in a loss of carbohydrate

active enzymes (CAZymes). We believe this would be an interesting thing to investigate in a future study, and we would predict that captivity results in a loss of CAZyme diversity. We have added a line to clarify this point in the discussion (see tracked changes):
"While such functions were found in a captive SHNW, we would predict that captivity reduces the diversity of such genes, which could impact the ability of SHNWs to derive energy from wild diets."

Besides the comments above, the manuscript is a pleasure to read and the authors manage to present the methodology and results in easy-to-read and precise manner. In addition the discussion includes some very relevant topics.
Thank you!

Minor changes:
L145-146: This might need a reference.
We have added the Song et al. reference here:
Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, et al. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. mSystems [Internet]. 2016

L428: Change Near-threatened to near-threatened or "Near-threatened".
We have changed this to "Near-Threatened".

We would like to thank the reviewer for their constructive feedback and suggestions.

Reviewer #2:
To the Authors: I would like to thank you for this very thorough and well-written manuscript. I read it with great interest, being in a very closely related field, and I recommend it for publication in Animal Microbiome, as it directly fits the scope of this journal. Your manuscript would be a great addition to the scarce literature on the microbiota dynamics between mammal hosts in natural to degraded habitats, and shows more evidence of the effect of captivity on endangered mammals. The accentuation of limiting cross-contamination between samples during bioinformatic treatment is of great importance this manuscript offers more attention for the reader. I only add a few comments for the authors that I think should be addressed and listed them below:
We would like to thank the reviewer for their positive assessment of our paper.

-It did not appear clearly if all samples were included in the analyse for captive individuals: did the authors used the 3 samples/animal or only one? And if yes, was that component included in the ANCOM and PERMANOVA analysis?

We apologise for the confusion here. We only used 1 sample per animal for the captive individuals, such that we did not violate the assumption of sample independence. Our Jupyter notebook explains this, but we should have clarified this in the methods. We have now added this information to the methods section:
"For captive animals that had multiple collections we randomly selected one sample for further downstream processing (see the Jupyter notebook for details https://github.com/EisenRa/2020_SHNW_Faecal_16S/blob/main/SHNW_Gut_16S_2019.ipynb)"

-I was wondering if the authors could specify the sequencing depth for the sequencing, and if the addition of a rarefaction curve plot could be done as a supplementary file.
The sequencing depth is reported in the first line of the results:
"DNA sequencing of the 97 samples resulted in 13,545,820 reads (mean of 139,647)"

The rarefaction plot is available in the Jupyter notebook:
https://github.com/EisenRa/2020_SHNW_Faecal_16S/blob/main/SHNW_Gut_16S_2019.ipynb
We have added reference to the rarefaction plot in the methods.

-I add trouble finding additional files. Especially because the ones in the manuscript (in the text and at the end) do not match and numbers overlap with the .qzv files. It is only a problem of presentation, but readers might not be aware of the use of qiime2view. I suggest that those files should added "physically" to the manuscript but I understand that it is just a matter of point of view. I understood what which supplementary figures were based on file names, but captions should be included.
We have added a figure captions files that describes each supplementary figure in detail (SI-captions.docx). We have also uploaded the supplementary files to a GitHub repository:
https://github.com/EisenRa/2020_SHNW_Faecal_16S/Supplementary_files/
and have given readers instructions for how to use/view the interactive qiime2view files. We are happy to take guidance from the editor on how many of the supplementary files should be uploaded to the journal vs. publicly available in the paper's GitHub repository.

-The authors mention a trim of 150bp for their reads, but the targeted region from the EMP primers is around 250bp. Could the authors explain this choice of trimming? Why was it only conducted on forward reads if the sequencing was paired-end? I was missing some information on the demultiplexing and merging of read.
ASVs from a target region can be directly compared across studies assuming they are trimmed at the same length. Most studies to date using the V4 region on QIITA (https://qiita.ucsd.edu/) have used a cutoff of 150 bp, so we wanted our ASVs to be comparable to the bulk of V4 16S rRNA gene studies to date.

Some of the reasons for using only R1 and 150 bp include:

1) Cheaper by not sequencing R2
2) R2 typically has lower quality scores on illumina runs
3) The information gained from the extra 100 bp is not substantial for most taxa, see figure 1 from Wang et al. 2007 (DOI: 10.1128/AEM.00062-07)
4) Merging reads can result in a loss of data (magnitude of the loss depending on the strictness of the merging parameters)
5) Another computational step, requiring time, and with the potential to lead to artefacts

I think that this is partly why the Earth Microbiome Project changed the location of the barcodes from the reverse primer to the forward primer (see https://earthmicrobiome.org/protocols-and-standards/16s/). As we have the older V4 region primers (with the barcodes on the reverse primers), we had to generate paired end reads.

Previously, I ran the analysis using merged reads with a 250 bp cutoff, but found the results/interpretations to be nearly identical to when using 150 bp of R1.

-Figure 3: are the not recorded 10% below the prevalence threshold? Maybe it should be added in the caption to minimized the reader's surprise.
We chose to display the top 25 most abundant families to reduce the complexity of the plot (bar plots can be quite messy with lots of features!). The top 25 most abundant families accounted for ~90% of the total relative abundance of microbial families in the dataset. We have further clarified this in the figure legend.

-L342: I think there is a typo here, is it not supposed to be Kooloola versus Brookfield since Wonga and Brookfield results appear in the previous sentence?
We thank the review for finding this error! We have fixed it.

-Host sex was not possible to include based on the authors sampling methods. However, it is know to have an impact on the gut bacteria of mammals. I understand it is out of scope from the manuscript here but maybe it would be important to acknowledge this fact.
This is a good point, and we have added the following line in the discussion:
“Finally, host sex has been shown to influence the gut microbiota of vertebrates [63], and while our study could not determine the sex of wild animals, it appears that capitivity and population are more important drivers of the SHNW gut microbiota.”

-This last comment is more for my curiosity to be satisfied but it could be something to mention in the authors' discussion: Neither wild populations experienced captivity so some changes in microbiota are strongly impacted by the animal's environment,

but as the authors mention it, is the reduction in ASVs richness and changes in composition in degraded areas detrimental to their host, or the result of local adaptations?

At this point we can't answer this question, though we are very keen to test this exact question in a future study using shotgun metagenomics combined with host physiological measurements (condition score, etc.). We also believe that understanding population histories will be very important in answering this question (see last paragraph of discussion), as factors such as animal dispersal and 'microbiota drift' could also be responsible for richness/compositional differences between populations (for more information on this concept, we refer the reviewer to Groussin et al. 2020, reference 66).

is there any fitness related data or mortality causes for the host species in the wild areas sampled or in general to support either explanations?

Yes, there is fitness data (from before our sample collections took place), but it is in preparation for a separate publication at present. We have data that shows that the animals at Kooloola in degraded habitat are in poorer condition across both Spring and Autumn periods, and across all cohorts, compared to animals from Wonga in intact habitat, across the same period, which we mentioned in the discussion:

"These differences also align with significant differences in body condition score between wombats from Wonga and Kooloola across all cohorts, with animals from Wonga (intact habitat) in significantly better condition, at all times of the year, than those from Kooloola (degraded habitat) (Taggart et al. in preparation)".

We would like to thank the reviewer for their constructive comments and suggestions.