a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF BIOLOGY
Okanagan Campus

Irving K. Barber School of Arts and Sciences
3333 University Way
Kelowna, BC V1V 1V7

May 28, 2015

Dear Dr. Fonseca,

Attached please find our revision of the manuscript entitled "From promise to practice: pairing non-invasive sampling with genomics in conservation" for consideration for publication in *PeerJ*. We thank the editor and anonymous reviewers for your constructive feedback on the submitted version, which has improved the manuscript. We have thoroughly addressed all comments as detailed below, with reviewer points indicated in italics and line numbers according to the track change version of the revised manuscript.

Please feel free to contact me for any additional information. Thank you in advance for your time and consideration.

Sincerely,

Michael Russello, PhD
Associate Professor, Biology
michael.russello@ubc.ca

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF BIOLOGY
Okanagan Campus

Irving K. Barber School of Arts and Sciences
3333 University Way
Kelowna, BC V1V 1V7

Reviewer 1 (Anonymous)

Basic reporting
In this study the authors address a problem in the field of conservation genomics: the need to collect samples non-invasively does not always yield enough DNA for high-powered NGS methodologies. Both the problem and their planned approach are presented nicely and the paper is overall well-written. I also thought the figures and tables were chosen well to support their message.

RESPONSE: Thank you.

Experimental design
I believe (and the authors may correct me if I am wrong) that this paper is meant to be more of a 'proof of concept' type paper (re: non invasive sampling and NGS) than a study on pika genetics. Therefore, I believe it would be strengthened by some additional explanation of the methodology chosen and their evaluation of its success: for example, how were the 9bp selective sequences chosen, and were any attempts made to examine how well the loci they obtained from this method represent the entire genome?

RESPONSE: We have added more details as suggested. See specific responses below.

Validity of the findings
The quality control methods used to filter samples and loci appear sound and well considered. Sample sizes are small although for a concept study this is not a big deal. I thought that they actually sell themselves a bit short in their conclusions and could be a little more overt in stating the advantages of their approach relative to other methods like exon capture.

RESPONSE: We have added more details as suggested. See specific responses below.

Comments for the author
Specific comments

95 'RAD-seq' was mentioned earlier in the introduction, and 'nextRAD' here. These are two different approaches (and in fact the RAD stands for different things). Some clarification would be helpful for those who are new to these methodologies.

RESPONSE: Thank you for pointing out this oversight. We have now defined the nextRAD acronym in the text (line 105).

106-107 Does 'this area' refer to the transects, North Cascades Park, the Pacific northwest, etc?

RESPONSE: We edited the sentence for clarity, which now begins with "Sites within this national park…" (line 114).

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF BIOLOGY
Okanagan Campus

Irving K. Barber School of Arts and Sciences
3333 University Way
Kelowna, BC V1V 1V7

107 typo – should be 'affected'

RESPONSE: Done (line 117).

113 I am assuming 15m is outside the dispersal range of pika?

RESPONSE: It is not outside the dispersal range, but in our experience, it helps maximize the acquisition of samples from different individuals. This is subsequently assessed using genotypic data. No additional text added here.

114-115 How was this tested? What was the outcome? (i.e. conclusive determination that no individual was sampled more than once, or some kind of likelihood)

RESPONSE: We just used straight genotype matching to assess numbers of individuals. We conducted this analysis for a random subset of 100 loci. These methods and results are now reported, as is the corresponding probability of identity ($1.1x10^{-23}$ average PID within sampling site) (line numbers 189-193, 266-268).

130 How were the 9 nucleotides chosen? Why 9 (not 8, not 10)? Was there a mathematical procedure or simply trial and error to determine this sequence would select a good number of potential loci? Additional explanation would improve the ms especially since 'nextRAD' seems to be a newer method and potentially of interest to a lot of readers.

RESPONSE: We added details regarding the selection of the primer sequence (lines 156-160).

143-147 Does 'population' here refer to all of the reads within a single individual, or to the combined set of reads across all individuals examined? I am inclined to draw the latter meaning from the wording used, but hesitate at the phrase 'more than the expected maximum of two alleles.' Within an individual, two would be expected, but across an entire population there could easily be loci with more than two. A little clarification would aid understanding.

RESPONSE: Your understanding is correct. We have revised for clarity (lines 170, 178).

218 It would be useful to refer to Table 2 here – so the reader can keep the distribution of sample sizes in mind going forward.

RESPONSE: Done (line 261).

230 typo – 'functions' not 'function'

RESPONSE: Done (line 287).

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF BIOLOGY
Okanagan Campus

Irving K. Barber School of Arts and Sciences
3333 University Way
Kelowna, BC V1V 1V7

Table 1 For those unfamiliar with outlier detection using Bayescan it would be helpful to indicate in the table legend or a footnote what these FST values mean (as opposed to traditional FSTs): basically, if I am understanding correctly, they correspond to how strongly the locus itself deviates from neutrality (so that neutral loci would have FST values of 0)?

RESPONSE: The FST reported is averaged over populations within transect. All reported outliers are those with a FST significantly higher than neutral expectations. This is now indicated as a footnote to Table 1.

Table 2 The higher elevation sites seem to have more individuals sampled on average. I am curious if the authors explored whether this could have contributed to the gradient in P, for example? (Perhaps by including n as an explanatory variable in the regression?)

RESPONSE: There was no correlation between sample size and elevation or sample size and gene diversity. We did find a correlation between proportion of polymorphic sites and sample size. The methods were revised accordingly (line 226) and results reported (lines 308-310).

280 Do the 3803 SNPs spread out evenly across chromosomes? Are some areas more vs. less represented, for example due to repetitive sequence?

RESPONSE: Unfortunately, we do not have a reference genome and so are unable to address this aspect.

282-287 Are there particular traits exhibited by pikas that could explain these findings? (e.g. social structure, dependence on heat-intolerant food sources). I know that this isn't the point of the paper, but perhaps the authors could refer to a key paper(s) containing this information for readers that are interested.

RESPONSE: Done. In response to this comment and those from reviewer # 2, we added a paragraph (and appropriate references) discussing how our results relate to those previously reported for the American pika (lines 350-380).

299-310 This discussion begs the question of why a conservation geneticist would choose to do nextRAD if exon-capture can also be us.ed with low amounts of starting material, eliminates the problem of non-target contamination, and doesn't require a reference genome. I suspect the answer is time and cost, but it would strengthen the authors' assertions about the utility of their approach to state this more overtly.

RESPONSE: Good point. We added the following sentence "Yet, these approaches are substantially more costly and, in the case of exon-capture, limited to expressed regions in the genome.". (lines 395-396)

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF BIOLOGY
Okanagan Campus

Irving K. Barber School of Arts and Sciences
3333 University Way
Kelowna, BC V1V 1V7

315 'such of parameters' – eliminate 'of'

RESPONSE: Done (line 415).

Reviewer 2 (Anonymous)

Basic reporting

No comments

Experimental design

Paragraph (Line 84-94)
The authors mention the use of AFLP's in some of their previous studies and list their undesirable properties but there was no discussion about how much better the nextRAD technology performs compared with the old AFLP or microsatellite data. Is there comparable data generated with microsatellites or AFLP's from the same localities?

RESPONSE: Unfortunately, we do not have comparable data at the same sites. The previous work was completed towards the northern range margin of the Cascade lineage of American pikas. The current study more likely approximates core sites within the distribution of the Cascade lineage, and does not lend itself to direct comparison.

Methods (Line 111-117)
Also, most validation studies for non-invasive samples use some kind of tissue or blood sample as a control to measure efficacy and reliability of the low quantity/quality DNA sample. Why was there not a control sample or samples with blood or tissue vs. hair comparison in this study? This would also give the reader a sense of how much allelic drop-out and genotyping error there is with hair samples.

RESPONSE: Agreed. Unfortunately, this type of control was not built into the current study given that we did not have the permission or opportunity to trap animals during that sampling period to collect higher quality sources of DNA.

DNA isolation section (Line 119-123)
There were no specifications given as to any precautions taken to avoid contamination during lab procedures such as use of negative controls and any other precautions taken during DNA extraction of samples to avoid contamination?

RESPONSE: We added additional details to this section, including the use of negative controls and a pre-PCR lab space (lines 138-140).

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF BIOLOGY
Okanagan Campus

Irving K. Barber School of Arts and Sciences
3333 University Way
Kelowna, BC V1V 1V7

Population genetics metrics (Lines 284-287)
How do the population genetics metrics compare to results of their previous Pika studies obtained usin AFLP's or microsatellites?

RESPONSE: In responses to this comment and reviewer #1, we added a paragraph comparing our results based on SNPs to other published work at different sites and at different scales using other markers (lines 350-380).

(Line 214-215) I was confused about this statement "Three samples had high bacterial contamination (Spermophilus; TL1A13b, PP3A02a, 215 PP1A24a) as evidenced by the number of matching sequence reads (15-22%)". Spermophilus is a ground squirrel not a Pika or bacteria?

RESPONSE: This was an unfortunate error on our part. We have revised the text indicating the correct interpretation of the *Spermophilus* sequences (lines 255-258).

Validity of the findings

The authors evaluated the feasibility of using nextRAD genotyping-by-sequencing of noninvasively collected hair samples to simultaneously identify and genotype SNPs in American pikas from two transect along an elevational gradient in the North Cascades National Park. They highlight a range of issues that must be considered when pairing genomic data collection with non-invasive hair sampling, and analytical approaches for maximizing cost-effectiveness and information content of recovered genomic data. The only concerns I had were that this paper really just addressed the problems associated with noninvasive hair sampling and the results are not really applicable to the other more commonly used method of scat (feces) for noninvasive studies. Therefore, I suggest that the authors make this clear in their title, their introduction and discussion and perhaps refer the reader to other approaches used for capturing and sequencing of endogeneous DNA from fecal samples. They do mention that "exon-capture, in particular, has been effectively applied to historical DNA collected from museum specimens (Bi et al. 2013), which typically yield DNA of lower quantity and quality similar to non-invasively collected starting materials" but I think that they need to clarify that this capture approaches are necessary for fecal samples:
Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. Molecular Ecology, 19(24), 5332–5344. doi:10.1111/j.1365-294X.2010.04888.x

Also, I think this argument can be substantiated by referencing the below paper that actually does an exhaustive survey of different types of methods and more recent than the Waits & Paetkau 2005 that they cite:
Beja-Pereira, A., R. Oliveira, P.C. Alves, M.K. Schwartz, and G. Luikart. 2009. Advancing ecological understandings through technological transformations in noninvasive genetics. Molecular Ecology Resources 9:1279–1301.

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF BIOLOGY
Okanagan Campus

Irving K. Barber School of Arts and Sciences
3333 University Way
Kelowna, BC V1V 1V7

RESPONSE: We believe that many of our observations and recommendations apply to non-invasively collected materials other than hair, yet we see the reviewers point regarding lack of reference to specialized needs of some other materials, such as feces. We added a sentence suggesting that use of capture approaches may be obligatory for other starting materials, including feces, as well as the accompanying reference suggested (Perry et al 2010; lines 395-398). We also added a reference for an alternative approach for sequence capture (line 392) and referenced the newer contribution from Beja-Pereira et al 2009 in the manuscript (lines 64 & 72). Given the broader applicability beyond hair, our preference is to retain the title as is.

Comments for the author

Other topics that were not discussed:

In their outlier detection analysis they identified 37 loci along the TL transect and 18 loci along the PP transect and none of which were shared. Why don't they share outlier loci? This should be briefly clarified in the discussion.

RESPONSE: The possible explanations for this pattern really span the biological to the technical. Without a reference genome, it is very difficult to annotate and map the outliers. What we know is what we reported in the form of annotations where available and the observation that none were in linkage disequilibrium. As the primary purpose of the paper was more proof-of-concept rather than a definitive treatment of pika population history and behaviour, we have not added any additional text addressing this point.

The proportion of polymorphic loci varied across the sampling sites, with the lower elevation sites (PP1, PP2, TL1) exhibiting substantially lower numbers than found at the mid- and high-elevation sites for both transects. Any speculation as to why the lower elevation sites are more inbred? What do the results obtained in previous Pika studies with microsats or AFLP's reveal in terms of genetic diversity in lower vs higher elevation?

RESPONSE: We added text discussing our results to the only other comparable study (lines 352-354).

Also, what differences do they find in terms of fine-scale structure resolution with SNP's vs. AFLP's or microsats?

RESPONSE: We added additional text comparing our results based on SNPs to recently published work at different sites and at different scales using other markers (lines 350-380).