Response to the reviewers

Dear Reviewers,

First, we would like to thank you for the time and effort you dedicated to reviewing our manuscript. We acknowledge the increase in the review requests and the strain that this causes admist our other obligations as scientists.

Please find our point-by-point response to your comments below. They have signficantly helped us to improve our manuscript.

[…]
R1> Overall, I found the manuscript thoroughly informative and easy to follow. The text and contents appear very well edited and I struggled to find any passages in need of changes. Due to the crisp nature of the present document, I can recommend this manuscript for immediate publication without revisions. I have nevertheless provided some minor suggestions for improvement.

Response:
Many thanks for your thoughful reading of our manuscript and kind words. We are glad to learn that you liked reading it.

General comments:

R1> In the abstract and later on, please consider how the species and subspecies types are described. You have, Linnaeus, Bell, and Nilsson included as uncited sources and I suggest you include the full references. Linnaeus 1758, Bell 1837, and Nilsson 1831. https://doi.org/10.2307/3504302

Response: Citations for the species descriptions are typically given only for taxonomic work. However, we are happy to include them as they add nice historical dimension to our genome note. The old original publications are easy to find online these days and reference accordingly. We were not sure if extra parenthesis would have been added to Nilsson, 1831, as the *sylvaticus* subspecies (‘varietet’) seems to be originally described under the junior synonym, *Lepus borealis*. We left them now out to avoid confusion.

R1> On line 35 in the abstract, this should be nominate* subspecies

Response: Thanks for pointing this out, corrected now throughout the text.

R1> The last sentence of the introduction takes up five lines and should be broken up for better clarity.

Response: Done.

R1> In your methods section, some additional clarity on the generation and vouchering of your cell line could be good. I am not personally very familiar with acceptable thresholds for genomic stability in cell lines after multiple passages, and I required some extra reading to understand if this was actually an acceptable passage number for an immortalized line. Examples of other recently accepted genomes using comparable cell-line sourcing could be useful for readers although I see you have addressed the concerns surrounding immortalization effects thoroughly in the discussion.

Response: This is a valid point, thank you. To be accurate, genomic stability of any cell line can be only assessed when the cells are karyotyped. We have done this here using a sequencing approach, which is an order of magnitude more informative than any chromosome-spreads done in the past. Not only the sequencing yields the expected karyotypes, the chromosomes have a good synteny with our brown hare reference genome (Figure 5). Although this genome too has been obtained from cultured cells, it would be highly unlikely that the two cell lines would have obtained similar chromosomal rearrangments. In this sense, our findings can be used as a benchmark for such studies in future. We have now added a couple of clarifying sentences.

R1> At several places in the manuscript, you mention that a mitochondrial reference has already been made available by a different project, but you reference a preprint. Please update this reference to the new peer-reviewed publication in Gene https://doi.org/10.1016/j.gene.2024.148644

Response: Yes, this manuscript was submitted already in July, so the referenced preprint has been published in the meanwhile. We have now updated this reference, thank you!

R1> Figure 1. No clear photo credit for image C, are these cellular microscopy photos from the line that was used for sequencing? Simply including "using the LT1 cell line *shown here*" would clarify this.

Response: The figure legend has been modified and credits included.

R1> It was a pleasure reading your work and I wish you luck in moving forward with it.

Response: Thank you again for your constructive comments.

Review by anonymous reviewer 1, 15 Nov 2024 02:13

[…]

R2> This preprint is extremely clear, and the authors were thorough in their approach. I enjoyed reading this article, and I think this new assembly is a valuable contribution.
My comments below are mainly regarding points where more clarification would be useful.

Response: Thank you for the kind words and detailed comments. Please see our point-by-point response below.

R2> Major comments

R2> Please add in details summarizing the proportion of k-mers in the reads that are in the final assembly (similar to Figure 2, but it would be useful to have the actual numerical estimate).

Response: We can do this, but we also feel that the numbers would really add anything. The k-mer analysis is a quality control step, not an end result. With a quick try, they add the number-heavyness of results section and are perhaps confusing. We could place them in a supplement, but these are a bit orphan data since no other supplementary data is included with our manuscript. We have left them now out, but are happy to include these if the reviewer insists.

R2> Please justify that the level of sequencing coverage of PacBio HiFi reads (21X) is sufficient to ensure an accurate assembly. Referring the reader to other assemblies with similar coverage, or to the author's own assessment of confidence in base calls (in general) would be useful here.

Response: PacBio HiFi reads have similar accuracy levels to Illumina short reads and to Sanger sequencing, with an accuracy of 99.9% 21X coverage will therefore meet the 1/10,000 nucleotide error rate standard for reference genome assemblies, as set by the Earth BioGenome Project (EBP). We have now added this into to the discussion, pointing out that also the other metrics of our genome assembly meet the EBP criteria.

R2> L100-109 disrupts the flow of the introduction, and I think would be better placed in the discussion.

Response: We find it relevant to mention the existing pseudoreference genome assembly for the mountain hare also in the introduction. We now made adjustements to improve the flow.

R2> In the methods it is unclear to me in places whether the authors did the hunting and original cell line isolation, or whether that was done previously. Using active voice (i.e., "We hunted… We isolated, etc.") at the beginning of each methods paragraph would be good to make sure this is not ambiguous (although passive voice is fine for method sections that are the not ambiguous).

Response: Thank you for these suggestions. We have now added active voice.

R2> L176 – Please specify what you mean by "and assembly parameters adjusted based on the expected genome size and coverage." (also missing "were" before adjusted there).

Response: The parameters we used are specified L177-179. "were" has been added. Thank you!

R2> L213 – The results section starts very abruptly with the genome accessions. I suggest these be moved to the end of the results (after the assembly steps are described). However, if the authors strongly disagree then they can keep the accessions listed there.

Response: The accession numbers have been now added to the end of the genome assembly result section.

R2> L219 – Define N50 (for unacquainted readers)

Response: -We have now added a clarifying sentence "the length of the shortest read at 50 % of the total sequence length" here. We hope that this is clear enough?

R2> Please comment on why the expected genome size from the literature and observed here differs (in context of "Genome assembly" section of results).

Response: While the first estimated haplotype length by Vinogradov (1998) is longer than our assembly, this estimation comes from flow cytometry data, without DNA isolation or sequencing. His genome size calculation is also based on an unrelated species (Rana temporaria) as reference. However, our assembly size is very similar to the previous mountain hare reference genome, assembled by Marques et al. (2019), at 2.7 Gbp. We also think that genome assemblies based on long molecule sequencing give much more accurate genome sizes than the old guestimates based flow cytometry or similar analyses. We have now added a few lines into the discussion.

R2> L238 – Please describe in words what the BUSCO categories refer to (e.g., "Fragmented" is not obvious) and remind the reader that this refers to expected single-copy genes. Similarly, clarify what "groups" you are referring to on L241.

Response: Done.

R2> L242 – Briefly expand on the T-antigen vector insertions (as many readers may not follow why these were expected). You should at minimum make it clear that these were expected due to the fibroblast cell line they DNA was derived from (which readers may have missed at this point in the manuscript). Also, re-word "As of note", which is not grammatically correct.

Response: Reworded the section and added a few lines about the T-antigen vector and its purpose.

R2> Figure 2 legend – Explain what the different categories are in the legend in panel A. Also define "read-only" and "shared" in panel B.

Response: We have now explained these.

R2> Figure 3 – I do not find these plots intuitive and I think many readers will not understand this, even with the description. I suggest you give some examples in the legend for what particular parts of the graph correspond to. “For example, the N50 line covers 50% of the sequenced assembly, and covers X GB, as this represents…” and “the record lengths increase in a jagged pattern because…”. I think comments like that could help readers new to these plots.

Response: We have to agree that the snail plots are really not intuitive, they have just become a quite standard way to display the genome assembly metrics in a single figure (see e.g. any genome note from the Darwin Tree of Life project). However, once the reader is familiar, they do make the comparison of the different genome assemblies easy just by eyeballing. We have now added some more explanation to the figure legend.

R2> Please adjust Table 2 so that it is entirely on a single page.

Response: Our apologies, but we find this unnecessary as the tables will be copyedited for the published version. In the manuscript, these will only confuse the line numbering.

R2> I do not follow the authors’ argument for why it makes sense that the sequence identity is lower for the brown hare vs mountain hare despite higher synteny. Could this observation not also be due to errors in the assemblies, or do the authors believe they can reject that possibility? It would be good to have explicit clarification on this point.

Response: The synteny differences were explained at the beginning of the section: the earlier mountain hare genome was assembled with the rabbit genome as a reference for scaffolding, which results in “chromosomal rearrangements”. So yes, technically it is due to assembly errors of the earlier assembly, although we didn't mention it specifically as an error.
Similary, we feel that the sequence similarities and differences are also explained in this section. Higher sequence similarity between subspecies compared to between species is expected, and we have also found this. We now reference this in the figure 5 legend.

R2> In Figure 5, the colour key needs units. However, I think the colours is too difficult to read if they are only on the line anyway. A different visualization should be used to more clearly display the differences in percent identity. For instance, boxplots showing the distribution of the mean percent identity per query would be much clearer. Also, the text at the top is too small to read, and should be removed.

Response: Units have been added – they had for some reason dropped out from the previous version. Small top text has been removed.

R2> Also, in Figure 5, it would be useful to have the common names listed for each species as well (as that is what is referred to in the text).

Response: The scientific names have been now replaced with the common names in the figure.

R2> **Minor comments**
R2> L29 – "chromosome" should be plural

Response: Corrected, thank you.

R2> Generally the percent symbol ("%") should follow directly after the number, so "95.1%" for instance, rather than "95.1 %". I suggest this be changed, but if it is an issue to do with the authors' word processor (e.g., with LaTeX), or if they strongly disagree, then it is not necessary.

Response: We discussed this among our authors from different language backgrounds and we have all learned that % represents a unit or a word and therefore there should be a space between the number and the symbol (same as for mm or kg). When researching the usage of %-symbol in English, there seems to be conflicting information in different sources. **However**, the brochure of the International System of Units (SI) in chapter 5 (page 44) states quite clearly: "In mathematical expressions, the internationally recognized symbol % (percent) may be used with the SI to represent the number 0.01. Thus, it can be used to express the values of dimensionless quantities. **When it is used, a space separates the number and the symbol %**."
https://web.archive.org/web/20171120061639/https://www.nist.gov/sites/default/files/documents/2016/12/07/sp330.pdf

R2> L31 – I would just say "based on mammals" or the equivalent, rather than "mammalia_odb10 database". This detail can be presented in the methods rather than the abstract, as many readers will not be familiar with what you're referring to.

Response: We have modified this accordingly in the results, along with the following BUSCO-related comment.

R2> L31 – Similarly, the reader would have to be familiar with the BUSCO categories to interpret "Complete", "Fragmented", etc. This should be re-written more clearly, keeping readers unfamiliar with BUSCO in mind (and mentioning this specific tool/database is not necessary in the abstract).

Response: Please see earlier comment to L238. We think it's enough to describe these in one place, and this probably shouldn't be the abstract.

R2> L36 – "The published genome assembly can" should be changed to "This published genome assembly could" (or "will", depending on the authors' confidence in this claim).

Response: Thank you for the suggestion. We changed this to "will", as we ourselves are already utilizing the genome for these purposes.

R2> L38-39 – I would split the long final sentence into two sentences.

Response: The sentence is restructured. We hope that it reads now better.

R2> L75 – Space missing after "assembly"

Response: Corrected, thank you.

R2> L82 – "was collected" is grammatically incorrect here. Needs to be re-written (or could be "…, which was collected…"

Response: Sentence modified, thank you.

R2> L117 – "Convention on" should be in front of "International" (for the acronym to make sense). I also do not think listing the acronyms "CITES" and "CBD" are necessary, since you are not using them again, unless you think readers will not know what you are referring to otherwise.

Response: The sentence here was meant to refer to the activities under CITES-regulations. This is now written more clearly. We would like to keep the acronyms, as they are recognizable for most readers.

R2> L139 – Add "done" before "previously" and "the" before "DNA Sequencing and Genomics Laboratory"

Response: Corrected, thank you.

R2> L139 – Also, the authors mention that the DNA was sequenced by this lab at the University of Helsinki on a PacBio Sequel II and then describe all the sequencing prep steps. If these were done by that lab then this should be specified and made clear. I would then mention how it was sequenced after describing the sequencing prep steps, for clarity.

Response: This is now clarified.

R2> L152 – "genomes" should be singular (or explained what the authors mean if not).

Response: The typo is now fixed.

R2> L153 – Should cite the published version of Tapanainen et al. 2024 now rather than preprint

Response: Published version cited.

R2> L154 – Rather than “access#” should be “accession”.

Response: Corrected, thank you.

R2> L157 – Should specify that the cut-out is Finland (just as a reminder to the reader). For example: “The geographic location in Finland and…”

Response: Corrected, thank you.

R2> L158 – Panel B description’s is interesting, but I would first mention that this is a picture of Lepus timidus (as many readers simply skim articles). Similarly, on L160, regarding panel C, the authors should specify that this is an actual image of the cell line that was used or not.

Response: Thank you for the suggestion. We have now added the relevant clarifications.

R2> L159 – Rather than “from e.g. ear clippings” I would re-write as “from, for example, ear clippings…” (or simply change to “…, e.g., …“)

Response: Sentence modified.

R2> L173: cutadapt should be cited (https://doi.org/10.14806/ej.17.1.200)

Response: Reference added, thank you.

R2> L219 – I would say “produced reads” rather than “produced data”, as many readers may not be used to N50 measures being used to describe reads, and mistakenly think these are assembled contigs.

Response: Great suggestion, thank you.

R2> L235 – Missing period after “Table 2”.

Response: Period added.

R2> L300-301 – Capitalization of “Chr” intended for rabbit genome only?

Response: Well spotted. These are now all capital.

R2> L302 – Should be Michell et al. 2024, not 2023

Response: Corrected, thank you.

R2> L304 – Space missing after “aligning”?

Response: There is a space, just quite small due to the text justification.

R2> L305 – “doesn’t” should be “does not”

Response: Corrected.

R2> L306 – “a higher amounts of” should be re-worded (currently grammatically incorrect).

Response: Sentence modified, thank you. There was a similar sentence in the discussion, which we also corrected.

Thank you again for the thoughtful reviews and excellent suggestions to improve our manuscript.

On the behalf of all coauthors,

Jaakko Pohjoismäki