The genomic foundations of adaptation: evaluating the mountain hare

Jitendra Narayan (iD) *based on peer reviews by* **Theodore Squires** (iD) *and 1 anonymous reviewer*

Cite this recommendation as:
Narayan, J. (2024) The genomic foundations of adaptation: evaluating the mountain hare. *Peer Community in Genomics*, 100377. 10.24072/pci.genomics.100377

Fekete et al. (2024) generated a chromosome-level reference genome assembly for the mountain hare (*Lepus timidus*). This represents a significant advancement in genomic research for non-model organisms, achieving high quality through advanced sequencing and curation techniques. This achievement serves as a foundational blueprint for future efforts in other species, particularly those with ecological or evolutionary importance. The assembly has high continuity and completeness, with an N50 scaffold length of 125.8 Mb and a contig N50 of 4.9 Mb, meeting the Earth BioGenome Project's stringent criteria for reference-grade genomes (Mc Cartney et al., 2024). The combination of PacBio HiFi sequencing and Hi-C scaffolding techniques enabled robust assembly and chromosomal scaffolding of all 23 autosomes and the X and Y sex chromosomes. Additionally, manual curation enhanced the assembly quality, accurately representing genomic sequences. Although the genome provides valuable structural insights, the limited functional annotations highlight a need for further investigation into the genetic underpinnings of the ecological and adaptive traits of the mountain hare.

The ecological and evolutionary implications of resolving this genome are considerable, particularly given the mountain hare's adaptations to cold, snowy environments and its role in boreal ecosystems. The assembly facilitates the study of adaptations, such as camouflage and snowshoe-like feet, which are critical for survival in its rapidly changing habitat. Comparative genomic analyses reveal the evolutionary relationship between *Lepus timidus* and closely related species, such as the brown hare (*L. europaeus*) and Irish hare (*L. t. hibernicus*), providing insights into gene flow, hybridization, and speciation. These findings have practical implications for conservation genetics, particularly for subspecies threatened by habitat loss and climate change. However, the study does not identify specific adaptive loci or functional variants, limiting its immediate applicability

to understanding the molecular basis of traits crucial for survival in extreme environments. Expanding the functional annotation of this genome would significantly enhance its utility in conservation and ecological genomics. Moreover, the high repetitive element content (42.35%) underscores the need for detailed annotation to facilitate downstream studies. These issues suggest that additional refinement and validation are warranted. Despite these limitations, the assembly is invaluable for studying genetic adaptations, hybridization, and hare conservation. Future research should focus on functional annotation, population-level comparisons, and targeted studies of ecological traits to fully realize the potential of this high-quality reference genome.

References:

Fekete Z, Absolon DE, Michell C, Wood JMD, Goffart S, Pohjoismäki JLO (2024) Chromosome-level reference genome assembly for the mountain hare (*Lepus timidus*). bioRxiv, ver. 2 peer-reviewed and recommended by PCI Genomics. https://doi.org/10.1101/2024.06.10.598177

Mc Cartney AM, Formenti G, Mouton A, De Panis D, Marins LS, Leitão HG, Diedericks G, Kirangwa J, Morselli M, Salces-Ortiz J, Escudero N, Iannucci A, Natali C, Svardal H, Fernández R, De Pooter T, Joris G, Strazisar M, Wood JMD, Herron KE, …, Mazzoni CJ (2024) The European Reference Genome Atlas: piloting a decentralised approach to equitable biodiversity genomics. npj Biodiversity, 3, 28. https://doi.org/10.1038/s44185-024-00054-6

Reviews

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.1101/2024.06.10.598177
Version of the preprint: 1

Authors' reply, 12 December 2024

Dear Dr. Narayan,

Thank you for soliciting reviews for our manuscript. We acknowledge the effort required to secure reviewers, and we greatly appreciate their constructive feedback, which has significantly helped us improve the manuscript.

Please find our point-by-point response to the reviewer comments attached. I am also uploading a manuscript with tracked changes to show how these comments have been used to adjust it.

We hope that our revisions meet the expectations of both the reviewers. Please do not hesitate to reach out if further clarification is needed.

Thank you for your time and consideration, and we look forward to hearing your decision.

Sincerely yours,

Jaakko Pohjoismäki

Download author's reply

Download tracked changes file

Decision by **Jitendra Narayan** ⓘ**, posted 25 November 2024, validated 27 November 2024**

Fekete and colleagues present a high-quality chromosome-level genome assembly for the mountain hare (Lepus timidus), developed using PacBio HiFi sequencing at 21x depth and Hi-C sequencing for contact mapping.

The assembly adheres to European Reference Genome Atlas (ERGA) standards and qualifies as a gold-standard reference genome.

The manuscript is clear, well-organized, and highly informative, reflecting the authors' thorough approach to presenting their research. The reviewers have offered a number of thoughtful suggestions to enhance the clarity and impact of the manuscript, focusing on areas that could benefit from additional detail. I am currently awaiting the submission of the revised version, incorporating these changes. Once the updated manuscript is received, I will perform a final review to confirm that all suggested improvements have been addressed satisfactorily, after which a final recommendation will be made.

Reviewed by Theodore Squires ⓘ, 09 November 2024

In the present manuscript the authors have outlined a newly available reference genome for the mountain hare (Lepus timidus). Their new genome conforms to standards put forward by the European Reference Genome Atlas (ERGA) initiative and appears to be of acceptably high quality for consideration as a gold-standard reference genome. The reference genome was developed using fibroblast cell-line samples from a male hare collected in Finland. This reference is representative of the nominate subspecies and unlikely to show genetic introgression with other hares. Notable characteristics include a length of just under 2.7 Gb across 25 well supported chromosomes (including both sex chromosomes). The new reference genome was created using PacBio HiFi at 21x depth and is a marked improvement upon previous assemblies.

The mountain hare is useful as a species with multiple cold adapted lineages that may have important insights into evolutionary responses to climatic change. In addition to helping with understanding the mechanisms of local adaptation, the new reference genome provides myriad avenues for new biological discovery. All of the standard aspects of a reference genome (karyotyping, BUSCO, GC content, etc.) were well reported and showed appropriate values for cutting-edge sequencing efforts.

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.

General comments:

1.   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

2.   On line 35 in the abstract, this should be nominate* subspecies

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

4.   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.

5.   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

6.   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.

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

Reviewed by anonymous reviewer 1, 15 November 2024

Fekete and colleagues present their new chromosome-level assembly of the mountain hare, which was produced through HiFi PacBio sequencing in conjunction with Hi-C sequencing to get contact maps. The product is a high quality genome with all expected chromosomes assembled, and high syntenic agreement with the recently sequenced brown hare genome assembly the same group also recently sequenced.

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.

Major comments

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).

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.

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

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).

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).

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.

L219 – Define N50 (for unacquainted readers)

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

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.

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.

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

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.

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

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.

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.

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).

Minor comments

L29 – "chromosome" should be plural

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.

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.

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).

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).

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

L75 – Space missing after "assembly"

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

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.

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

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.

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

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

L154 – Rather than "access#" should be "accession".

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

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.

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

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

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.

L235 – Missing period after "Table 2".

L300-301 – Capitalization of "Chr" intended for rabbit genome only?

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

L304 – Space missing after "aligning"?

L305 – "doesn't" should be "does not"

L306 – "a higher amounts of" should be re-worded (currently grammatically incorrect).