**Confidential: October 2023 Reviews of manuscript entitled "DNA sequences from two non-avian dinosaurs"**

Referee expertise:  
  
 Referee #1: aDNA  
 Referee #2: aDNA  
  
Reviewers' Comments:  
  
**Reviewer #1:  
Remarks to the Author:**  
Kim and colleagues report on biomolecular analyses done on two dinosaur species, and among other things claim to have recovered endogenous DNA. However, there is in my view nothing in the manuscript that supports this claim.  
  
The field of ancient DNA has progressed a lot over the last decades, and a number of recommedations for authentication have been established. While these do not necessarily comprise a list of rules that must be followed, there is an expectation that authors (especially those making extraordinary claims) provide a nuanced and objective assessment of whether their results are valid. The manuscript by Kim and colleagues does not contain any such assessment.  
  
First of all, it seems that the authors did not include any negative controls during DNA extraction and library build. The use of negative controls is important to monitor for background contamination from e.g. reagents and the environment. How were reagents and equipment sterilized to avoid contamination? The authors also do not present any DNA damage plots for the data mapped to the chicken genome. DNA damage at the end of reads would be expected given that the DNA extracts were not UDG treated.  
  
In addition, there is hardly any description of what bioinformatics methods were used when the reads were mapped to the chicken reference. How were the reads mapped? Were duplicates discovered/removed? What filtering was done, for example in terms of minimum fragment length? Finally, why isn’t there a more in-depth analysis of the reads that did map? Were these mapping to particularly conserved parts of the vertebrate genome? Why is it relevant that DNAs and peptides overlap in the genome? The authors state that the distribution of mapped reads “agreed with the expected background genome”. What does this even mean? Why is it relevant?  
  
Moreover, the results from the metagenomic analysis make little sense. Why would a majority of reads be assigned to reptiles rather than birds, even though the latter should be much more closely genetically related to dinosaurs? And, critically, how many reads were assigned during the metagenomic analysis?  
  
The manuscript is also contradicting in several cases, for example where it says that SRSLY DNA libraries were prepared in the main txt (i.e. single-strand libraries), whereas it says Omni-C in the SI and one of the figure legends. Omni-C library reads are used for genome scaffolding, and it is very unclear what the point would be of using this method on samples of this antiquity. Also, it is unclear what reference database was used for the metagenomic assignment. The SI refers to a supplementary table (a “full taxonomic table”) that doesn’t exist.  
  
Overall, both laboratory and computational methods used (where described at all) are completely unsuitable for ancient DNA analysis, and the results are not presented in a way that allows for an assessment of their validity. For anyone in the field of ancient DNA, the lack of clarity and lack of sanity checks of the data in the manuscript is highly disappointing. There thus is nothing in the presented results that even hints at the successful recovery of authentic dinosaur DNA. Because of of this, I unfortunately have little other choice other than to recommend that this manuscript is rejected.  
  
  
  
**Reviewer #2:  
Remarks to the Author:**  
I appreciate the opportunity to review this paper, which is an exciting new chapter to the long history of work done on these remarkably well preserved dinosaur bones. The preservation of these bones is compelling, and this is certainly the moment to return to them with new approaches to assess whether DNA is indeed preserved. Unfortunately, however, I was not convinced by the results as they are presented in this version of the manuscript. I am not dismissive – It is possible that this new approach has indeed recovered authentic ancient DNA. However, the absence of controls and missing methodological details made it impossible to fully evaluate the data. If authentic, this would be an important advance for ancient DNA and one that will receive considerable attention from both the media and, correctly, from the scientific community. The authors therefore should be strongly compelled to be transparent and thorough in what they present. This version of the manuscript is, however, not at all transparent, due to missing details, missing rationale, and insufficient validation.  
  
  
Below, I highlight several of the major problems encoutered while reviewing this manuscript. I keep this review at a high level as, given these major concerns, I felt that it did not make sense to go into more of the details at this point.  
  
  
Appropriate negative controls: There is a lack of proper negative controls in the cell isolation/DNA extraction, library prep, and sequence analyses components. Negative controls processed throughout the entire molecular and bioinformatic process are crucial for discerning between contamination and authentic signals. An additional negative control of a ‘blank’ sample, processed in the same way as the dinosaur samples, would further serve to show how a known non-dinosaur sample would map to this set of reference genomes. The listed negative controls of ancient human tooth and modern ostrich osteocytes have several issues. The human tooth library contains only human-aligned reads to begin with - a more suitable control would be the raw sequence data obtained from an ancient human tooth processed in the same way as the dinosaur samples. Additionally, the negative control of the modern ostrich osteocytes is not appropriate. Nearly all mapped reads from this sample map to the ostrich reference genome, as expected, and it is difficult to discern mapping statistics of the remaining reads which might represent contamination or mappability to different reference genome. And while framed as a ‘negative’ control, this sample serves as something closer to a positive control (see below).  
  
  
Appropriate positive controls: Dinosaur microbial reads are listed as a positive control, but it is not clear for what they are a positive control. It is not possible to have a real positive, as there is no dinosaur DNA. For the bioinformatic steps, however, there are some better controls than are currently used that might simplify interpretation of the results. Here, for example, you map ostrich to itself and show that all reads map to ostrich. Instead, why not map a divergent bird to the ostrich genome? Better yet, a version of the chicken genome with sequence read lengths that match what is recovered, perhaps with damage and a significant portion of human contamination, and map that? That will show what the expectation might be when mapping divergent, damaged, and contaminated reads from a given library to modern genomes.  
  
  
Justification for methodological choices: Some unorthodox choices are made in both the molecular and bioinformatic methods which might warrant an explanation or justification. These include: Why was a version of Omni-C chosen as the library preparation method? Why was a dual-unique indexing scheme (presumably) not chosen to prevent index-hopping when multiplexing libraries in sequencing? Why were PCR duplicate reads not removed bioinformatically? What is the explanation for the high level of G repeats that are found in the SRSLY libraries (possibly some Illumina failure leading to reads consisting only of G bases)? Why was the entire library exhausted in sequencing, instead of preserving this as a future resource for additional study? What filtering steps were used and why? Was there DNA damage on “authentic” reads that map to chicken? Which of the various data sets were used with which analyses? What was the purpose of using reads from cancer cells to increase sequence diversity rather than standard phiX?  
  
  
Data availability and usage: Some analyses are unclear with respect to the raw data used. Were the non-dinosaur fastq data processed in the same way or were alignment/vcf files from previous studies used? Were the filters applied to the dinosaur sample the same as those used to generate ostrich/human data? There is no reference to mapping quality filters or other types of filters prior to alignment such as PCR duplicate removal, and to which samples these filters were applied. It is mentioned on line 96 that “insert-containing fragments averaged 360–390 bp per sample”. This is extraordinary for ancient DNA and it is unclear whether this is referring to the mappable reads or if the majority of this is adapter-only reads, or likely contaminants. Modern DNA would be expected to have larger insert sizes, so the insert sizes of the mappable reads should be reported here.  
  
  
Interpretation of key results: As stated above, evaluation of these results is difficult given the lack of details in the methods. However several of the results claiming to support the presence of endogenous dinosaur DNA are puzzling. For example, a key piece of evidence for the presence of authentic dinosaur DNA is the proportionate mapping of reads to different functional parts of the genome (Extended data Fig. 4A). However, it is unclear whether this is a null expectation given the distribution of these functional categories in these assemblies. How would a negative control (‘blank’) map to these categories? Or the non-human reads from the human tooth sample? One can imagine many possible controls here to determine whether these results are truly indicative of endogenous DNA, or whether they are the result of simple sample contamination. Similarly, for the correlations shown in Extended Data Figure 4B, there is an obvious relationship between bases in the two tissue types from the same species, but what would be expected from authentic DNA, rather than contamination or spurious alignments? If anything, should these correlations not be higher, given that the DNA in the two tissues are the data from the same species? Also, for this analysis, it is stated that “loci were mapped by at least two samples and with a depth > 10 counts”- given the extremely low coverage recovered from the dinosaurs, wouldn’t the subset of bases covered by more than 10 reads be extremely low, and likely to be highly repetitive and spuriously aligned? It is also mentioned a few times that reads are “high quality”. What does this mean? Is it in reference to mappability?  
  
  
While Fig. 2 does a helpful job of depicting the sampling study design, it misses a key opportunity for comparison of library characteristics for modern DNA, ancient DNA, and fossil DNA. In panel C, it would be useful to see the distribution of read lengths coming from a raw ancient DNA sample, for example an ancient human tooth, or even a blank sample (see ‘Negative controls’). Instead, we see a modern vs. fossil comparison, giving us no context in which to place the purported fossil DNA. Furthermore, the info in panel A indicates that such samples are available (‘Extinct bird of prey, bone), yet they are not plotted for comparison in panel C where it would be useful. If this is taken to be evidence of endogenous DNA recovery, then we need to see comparisons of the read length distribution to proper controls.  
  
  
Figure 3 seems to be the key figure, but it is tough to interpret. The pie slices are not comparable (see above wrt controls) so why present them as if they are? This seems misleading – perhaps not deliberately so, but it does seem to convey a fundamental misunderstanding of the data. Additionally, it’s unclear if the proportion of reads mapping to each category might be related to the quality of the single genome assembly representing each clade. One can imagine more/spurious mapping to a lower quality (e.g. alligator) genome.  
  
  
What point are you trying to convey in Fig. 4? And if the SRSLY prep yields higher quality data, why are the main analyses not conducted with this dataset, or using all the data from the two library prep methods? And why is there no description of the SRSLY preparation in the methods section? Oddly, it seems from extended data fig 4C, there is a huge overlap in segments mapping between the two species from SRSLY libraries. Given what is presumably extremely low coverage, why would you see this much overlap? Could this be some sort of bioinformatic artifact?  
  
  
Again, the comments above are not exhaustive of the issues that emerged when attempting to make sense of this potentially remarkable data set, but simply highlight some of the major concerns mainly in the lack of crucial details provided. We note, however, that the absence of a negative control is the most problematic aspect of this work and one that, if no negative was processed alongside these experiments, may be impossible to correct. Unfortunately, use of a negative control is among the few unbreakable rules in ancient DNA research and precisely why so many false reports, several of dinosaur DNA, were published during the early phase of this research field.