

Review comments

Dear Reviewers, Louise,

Thank you for taking the time to review this preprint; it has greatly improved the scientific component and clarity. We are very pleased with the overall positive feedback, and we recognise that this takes time from your busy schedules. We greatly appreciate it!

We have addressed all of the comments. Most of the recommendations have been incorporated into the revised manuscript, barring a few minor details which were either not feasible or conflicted with our personal preferences (more details in the response to the specific comments). We restricted the length of the added components due to the fact that the discussion is already a bit on the long side.

Best,
Bjorn (on behalf of co-authors)

Revisions requested for preprint #389

Recommender: Louise Le Meillour

The manuscript and its findings are of great interest, new and exciting for potential future applications. Even though I found the text well written, I noticed double spacing, extra point and typos here and there throughout the manuscript. The overall external comments were very good and the reviewers agree on re-working on specific points to ensure clarity. I would also argue in favor of some rephrasing (especially in the introduction and discussion) to maximise the impact of your study: to a non-specialist audience, I found some sentences a little difficult to follow.

- **[RESPONSE:** *I think the double spacing is because the text alignment is justified (please correct me if that's not what you meant). The typos and extra points that I could identify have been fixed. I have also rephrased the sections recommended by the reviewers, and some additional sections in the intro and discussion.*]

Reviews

Reviewed by Mario Zimmerman, 11 Sep 2023 22:03

The manuscript brought forth by Bartholdy et al. represents an important step forward in the study of secondary metabolite residues in archaeological human remains. Building on their own team's recent advances as well as research published by other scholarly groups, the authors address new questions – the correlation between the detection of different compounds, or the correlation between pathological conditions and residue signals – while

at the same time increasing the time depth for the application of a protocol previously only tested on contemporary samples. While contamination remains an issue to be aware of, Bartholdy et al. also provide suggestions as to how to distinguish diagenetic impacts and lab contaminants from metabolomic signals related to human substance consumption.

Specific suggestions for improvements:

It might be worthwhile to include King et al. (2017) in the discussion on the proportions of caffeine, theophylline, and theobromine.

King, Adam, Terry G. Powis, Kong F. Cheong, and Nilesh W. Gaikwad 2017 Cautionary tales on the identification of caffeinated beverages in North America. *Journal of Archaeological Science* 85:30–40. DOI:10.1016/j.jas.2017.06.006.

- **[RESPONSE:** *a short sentence has been added to this section.*]

As the authors mention the need to account for soil-calculus contamination, I would appreciate an indication as to the plans for control in upcoming analysis. Personal experience has shown this to be of utmost importance not only with regard to potential false positives for salicylic acid but also methylxanthines.

- **[RESPONSE:** *Sentence added to the end of the future directions paragraph, stressing the importance of sampling from the soil associated with the skeleton.*]

While quantitation of detected compounds is expressly not a focus of this study – and I agree with the authors regarding the taphonomic complexity – I believe a comparison of nicotine/cotinine peak area values between positive individuals with pipe notches and those without this dental pathology would further augment the value of the authors' protocol accuracy statement.

- **[RESPONSE:** *The absolute quantity (in ng) of compounds is equivalent to the peak area normalised against the internal standard, and these are reported in the raw data, which are publicly available. We indeed wanted to do a comparison of tobacco-positive individuals with and without pipe notches, but there is only one tobacco-positive individual without a pipe notch, and only five individuals in the sample without pipe notches. In hindsight we should have included more individuals without pipe notches, but we were limited by the material that had previously been sampled combined with our sampling criteria of middle adult males (and adult male individuals without pipe notches are hard to come by in this particular population).*]

Reviewed by anonymous reviewer, 26 Aug 2023 20:39

This is an interesting article building on previous studies on alkaloid preservation in dental calculus. It will make an important contribution to the growing field of dental calculus studies, and I recommend it be published in a journal. The study is well-designed and reaches some

interesting conclusions regarding this 19th century rural population from the Netherlands. I don't have any major critiques or criticisms.

- **[RESPONSE:** *We are happy to hear that there are no major critiques!]*

Reviewed by anonymous reviewer, 11 Sep 2023 13:57

Bartholdy et al. investigate the utility of a new method for detecting alkaloids, previously validated using recently deceased individuals, on archaeological dental calculus. The manuscript is well-written, the methods are sound, and the conclusions are relevant to the data presented. I only have minor comments for further clarification of the manuscript as detailed below.

Throughout the manuscript, the authors state that “well-preserved” individuals have higher yields of substances of interest. It would be helpful to briefly define what they mean by well preserved.

- **[RESPONSE:** *a short sentence on preservation has been added to the Methods section under Skeletal analysis.]*

L48: “The relation to plasma is why there is often a close correlation between the presence (not concentration) of drugs in oral fluid and blood” This sentence was a little confusing to me, can the authors clarify?

- **[RESPONSE:** *Completely agree. I have reworded this for clarity.]*

L82: “To reduce the number of potentially confounding factors to account for in the analysis, we preferentially selected males from the middle adult age category” Can the authors expand on this choice? Is there previous literature that details how age/biological sex are confounding factors?

- **[RESPONSE:** *I have added a study on the gendered division of labour at the site in the Materials section. More generally I don't think we know enough about the differences in calculus formation and drug metabolism between age/biological sex, so this made most sense to control for in our experimental design. I have expanded on this a little in the text (also in the Materials).]*

L98: Is this sentence missing “allow for statistical analysis”?

- **[RESPONSE:** *Yes, corrected.]*

L196: Here it states that the accuracy of the method is 59.3% but in the abstract, it is 60% if I'm interpreting these statements correctly?

- **[RESPONSE:** *actually the abstract is just stating the number of individuals with a pipe notch and detected tobacco, which is 56% (and which is still wrong in the abstract). I have corrected this number in the abstract.]*

L206: Please match these numbers to the same rounding point as found in Table 2 (e.g., 0.982 should be 0.98? 0.507 should be 0.51?)

- **[RESPONSE:** *Yes, definitely (my high school maths teacher would be disappointed in me). Corrected.*]

Table 2: In general, this table asks the reader to do much of the work to pull out the important points. Maybe highlight those correlations that are most interesting?

- **[RESPONSE:** *I have bolded the moderate and strong correlations to hopefully make it easier for the reader.*]

Figure 3: Again, this figure is a little hard to interpret. From my understanding of the text, the scatter plots are supposed to illustrate the correlation between weight and quantity of the substance – are there corresponding statistics that could be shown (e.g., R² values?). Similarly, are the amounts shown in the violin plots statistically different from one another?

- **[RESPONSE:** *I have added linear trendlines and r values to the plots to improve interpretation. Whether or not there are any ‘significant’ differences between the preservation and quantity is not really informative. We’re mostly just interested if there is an increased extraction quantity, which is clear (at least I think) from the plots. An additional note: I think there is often a little too much value placed on ‘statistical significance’, so make a point to avoid it (in some situations).]*

Figure 4: Similar issue here, distinguishing differences between the sizes and color gradient of the circles are a little difficult for me to see. Could again highlight important correlations with some statistics maybe?

- **[RESPONSE:** *The important correlations (moderate and strong) are mentioned in the text already. I changed the colours of the plot to improve readability (the yellow was admittedly a poor choice).]*

L269: “salicylic acid is a very mobile organic acid” Can you expand on this?

- **[RESPONSE:** *I have reworded this section to increase clarity.*]

L295: “It has been shown that, while abundant in opium, morphine degrades rapidly...” This sentence seems to be missing a statement. What is abundant in opium?

- **[RESPONSE:** *I have reworded this to make it clear that it is morphine that is abundant in opium.*]

L308: “We suspect that the original detection of cocaine was a result of lab contamination during analysis” Was this also detected in batch one (which was presumed contaminated)?

- **[RESPONSE:** *It was only detected in batch one, which is why we suspect contamination]*

L360: “be around 59.3%” Again, should this be 60%?

- **[RESPONSE:** *This 'accuracy' is different from the abstract; it includes the combined lack of pipe notch and absence of tobacco as a correct assignment, and is different from the statement in the abstract (now corrected to 56%) which only considers present pipe notch and successful detection of tobacco. This is also why it doesn't explicitly mention 'accuracy' in the abstract.]*