Ways in which contamination might be distinguished from authentic human eDNA

David S. Thaler

Biozentrum, University of Basel, Basel, Switzerland; Program for the Human Environment, Rockefeller University, New York, USA; Correspondence：davidsthaler@gmail.com

Draft 4 October 2022

Here we briefly classify several potential sources of contamination and how they might be identified and minimized, informed by developments in the relatively mature fields of ancient DNA, forensic DNA, and microbiome analysis.

Possible contamination comes from at least three distinct sources: (a) field and laboratory personnel, (b) reagents, and (c) DNA sources within the laboratory, e.g., from one sample to another or from amplicons generated in prior experiments. Each of these potential sources of spurious results is routinely considered in the fields of ancient DNA, forensic DNA analysis, and microbiome studies.

Consider potential approaches to identifying human DNA introduced during collection and processing. Human eDNA is likely to have been a relatively longer time away from its source than “fresh DNA” introduced by operators of collection and laboratory processing. Furthermore, human eDNA is in many cases likely to be longer away from its source and to have degraded differentially when compared with eDNA from other animals, e.g., fish in marine environments. A longer time from its source may lead to authentic human eDNA exhibiting characterizable degradation. At least four types of differential degradation could distinguish authentic human eDNA from operator contamination: (1) C > T transitions, (2) G > T transversions, (3) the ratio of short to long amplicons, and (4) ratio of mitochondrial to nuclear sequences.

It is currently unknown if the degradation suffered by human eDNA might be enriched for C > T transitions (1-3). The field of ancient human DNA makes important use of the chemical properties of 5-methylcytosine. 5-mC is susceptible to chemical deamination; deaminated cytosine becomes uracil which Watson-Crick pairs with adenine during the first cycle of PCR replication (4). During the next cycle, the adenine pairs with thymine. The net result is that ancient sequences display C > T transitions compared to fresh DNA.

Similarly, so far as we are aware, 8-oxoguanine and consequent G > T have not been assayed in the context of eDNA. The formation of 8-oxoguanine occurs in DNA by several chemical processes. Efficient enzymatic mechanisms repair these lesions inside living cells (5). However, circulating extracellular DNA in human serum accumulates these lesions, and they lead to G >T transversions during PCR (6).

Amplicon length ratios may also prove valuable in comparing relative residence times of eDNAs from multiple different species. Those more recently in the environment are predicted to have a higher ratio of long to short amplicons. Human eDNA, like all eDNA, is likely to undergo degradation that shortens the effective size of recoverable amplicons (Saito and Doi, 2021). Several physical and enzymatic processes approximate random hits on DNA, leading to a predictable ladder or ratio of amplicon sizes for a given amount of damage (7). Human DNA entered into an environmental sample by the person who collected it or by those who purified DNA in the laboratory is likely to have a ratio of short to long amplicons of close to 1:1. On the contrary, authentic eDNA is likely to show a preponderance of short amplicons.

With regard to the ratio of mitochondrial to nuclear DNA, this ratio contributes to estimates of contamination in ancient DNA studies (8). By hypothesis, recoverable nuclear sequences will be much less likely from authentic human eDNA than from operator contamination. To the best of our knowledge, none of the four differentiators has been applied to human eDNA in the context of environmental vertebrate surveys, and each is potentially an area for discovery lifting the power of DNA forensics.

With respect to our second and third classes of contamination, good laboratory practices include attention to the possibility that commercial reagents may contain

contaminating signals, as have been found with bacterial DNA (9). Molecular biology reagents seem less likely to be contaminated with human than with microbial DNA, but reagents must be tested and controls run. A systematic review entitled "The elephant in the lab (and field): Contamination in aquatic environmental DNA studies" warns that negative controls are not always included or rightly interpreted (10). Contamination in a PCR lab can follow from "loose amplicons" of previous experiments. To minimize this risk, standard precautions include handling post-amplification samples in a physically separate location from where new samples are prepared for amplification. More generally, careful attention to laboratory practices potentially can take care of most contamination problems even in resource-poor environments. The same laboratory care and experimental designs that include appropriate negative controls help generate trustworthy results for both human and other eDNA targets.

Not all conditions favor preservation of eDNA; it will not always be detectable in soil cores or sediments. Nevertheless, when eDNA is preserved, the ratio of human-specific sequences to those of other vertebrates could provide a means to document and possibly quantify the relative human contribution to an ecosystem. Depending on the state of samples and concomitant degradation of eDNA, different specifics of assay protocol such as primer choices and amplicon size may be used (11). Studies are likely to be commensurable as long as their methods report the ratio of human: total vertebrate eDNA. The use of blocking primers against human sequences (12) decreases a study's value in this regard and can also distort nontarget detections.

Commensurability of studies for large scale mapping as well as the characterization of chemistries used to distinguish contamination from bona fide environmental DNA will both be aided by absolute rather than relative quantification. Absolute quantification has been achieved with microbial samples (13) and should be looked upon as a goal for human eDNA, and may be possible by spiking metabarcoding PCRs with known concentrations of DNA standards.

Incorporating standards for normalization (14, 15) and reporting human / total vertebrate ratios seem the most technically robust means to allow combining independent measurements into a worldwide citizen science eDNA biological diary of the Anthropocene.

1 Smith, R. W., Monroe, C. & Bolnick, D. A. Detection of Cytosine 277 methylation in ancient DNA from five native american populations using bisulfite sequencing. PLoS One 10, e0125344, doi:10.1371/journal.pone.0125344 (2015).

2 Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences of the United States of America 111, 2229-2234, doi:10.1073/pnas.1318934111 (2014).

3 Peyregne, S. & Peter, B. M. AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. Genome biology 21, 246, doi:10.1186/s13059-020-02123-y (2020).

4 Frederico, L. A., Kunkel, T. A. & Shaw, B. R. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry 29, 2532-2537, doi:10.1021/bi00462a015 (1990).

5 Markkanen, E. Not breathing is not an option: How to deal with oxidative DNA damage. DNA repair 59, 82-105, doi:10.1016/j.dnarep.2017.09.007 (2017).

6 Arbeithuber, B., Makova, K. D. & Tiemann-Boege, I. Artifactual mutations resulting from DNA lesions limit detection levels in ultrasensitive sequencing applications. DNA Res 23, 547-559, doi:10.1093/dnares/dsw038 (2016).

7 Spangler, R., Goddard, N. L., Spangler, D. N. & Thaler, D. S. Tests of the single-hit DNA damage model. Journal of molecular biology 392, 283-300 (2009).

8 Furtwangler, A. et al. Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. Sci Rep 8, 14075, doi:10.1038/s41598-018-32083-0 (2018).

9 Spangler, R., Goddard, N. L. & Thaler, D. S. Optimizing Taq polymerase concentration for improved signal-to-noise in the broad range detection of low abundance bacteria. PLoS One 4, e7010 (2009).

10 Sepulveda, A., Hutchins, P., Forstchen, M., Mckeefry, M. & Swigris, A. The Elephant in the Lab (and Field): Contamination in Aquatic Environmental DNA Studies. Front. Ecol. Evol. 8, 609973, doi: https://doi.org/10.3389/fevo.2020.609973 (2020).

11 Jo, T., Takao, K. & Minamoto, T. Linking the state of environmental DNA to its application for

biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. Environmental DNA (Wiley) 00, 1-13, doi: https://doi.org/10.1002/edn3.253 (2021).

12 Ruppert, K., Kline, R. & Rahman, M. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation 17, e00547 (2019).

13 Cruz, G. N. F., Christoff, A. P. & de Oliveira, L. F. V. Equivolumetric Protocol Generates Library Sizes Proportional to Total Microbial Load in 16S Amplicon Sequencing. Frontiers in microbiology 12, 638231, doi:10.3389/fmicb.2021.638231 (2021).

14 Ushio, M., Murakami H., Masuda, R., Sado, T., Miya Mi., Sakurai, S., Yamanaka, H., Minamoto, T., Kondoh, Mo. . Quantitative monitoring of multispecies fish environmental DNA using high throughput sequencing. Metabarcoding and Metagenomics, 2 (2018).

15 Sato, M. et al. Quantitative assessment of multiple fish species around artificial reefs combining environmental DNA metabarcoding and acoustic survey. Sci Rep 11, 19477, doi:10.1038/s41598-021-98926-5 (2021).