Ancient microbial DNA and dogmatic approaches: A Response to critical comments to the paper "Retroviral DNA..."
Ancient microbial DNA and dogmatic approaches: A Response to critical comments to the paper “Retroviral DNA Sequences as a Means for Determining Ancient Diets”

Jessica I. Rivera-Perez1*, Raul Cano2 and Gary Toranzos1

1Department of Biology, University of Puerto Rico, Rio Piedras Campus, San Juan, Puerto Rico. Address: PO Box 23360, San Juan PR 00931

2Center for Applications in Biotechnology, Biological Sciences Department, California Polytechnic State University, San Luis Obispo. Address: Center for Applications in Biotechnology, Biological Sciences Department, California Polytechnic State University, San Luis Obispo, CA 93407

*Corresponding author
E-mail: jessica.i.rivera88@gmail.com

Abstract
A recent publication by our team has raised concerns towards the validity of using ancient DNA, specifically endogenous retrovirus sequences, as a means of determining the diets of ancient ethnic groups. Our colleagues rely on proposed ideas to conclude that the isolation of DNA from high temperature areas is highly unlikely and fundamentally impossible because of degradation. Such ideas were published in several opinion pieces over a decade ago, but remain largely un-validated to this day. We clearly disagree with these statements, as would a rather large group of paleoscientists currently working with ancient DNA isolated from this geographical area. Additionally, our colleagues conclude that not having an in-house sequencing facility will result in contamination by extant DNA. This is an overly enthusiastic and generalized opinion that does not bid well for commercial DNA sequencing facilities being used for any type of work. We also take this opportunity to explain the procedures used in our study in further detail here, as we acknowledge that we mainly relied on referencing previous publications from our lab rather than including a lengthy explanation of the study controls. Our data clearly show, by what is present and what is not detectable, that we are not making conclusions based on contamination with extant DNA. Our methods include guidelines previously suggested for ancient DNA studies, combined with modifications addressing sample uniqueness. The detection of DNA from animals that are not part of our own diet, as well as previous cytochemical and paleogenetic analyses conducted on these samples clearly indicate we are working with ancient DNA. This response elucidates why we trust the analyses reported in this study, and why we are confident with the conclusions drawn from the data.
Retroviral DNA in ancient feces from Caribbean cultures

We recently published a paper in which we isolated DNA from Caribbean coprolites and analyzed these by next-generation sequencing. Most sequences belonged to feces-associated microbes, however others were surprisingly similar to certain plant and animal endogenous retroviruses. We then used these retroviral sequences to infer the diet of ancient cultures (1). The logic behind this analysis is that most animals and plants have been abortively infected by retroviruses at some point over millennia (2, 3). Thus, although the viral DNA sequences were retained as part of the viral hosts’ germ-line, the virus is no longer expressed in many cases (4). If food is consumed, the DNA of those plants/animals/insects, as well as the DNA from their associated endogenous retroviruses, will end up in the feces and could be detectable by sequencing. The sequences detected in these coprolites have allowed us to determine in a unique manner what was included in the diet of pre-Columbian Caribbean ethnic groups. In many cases, animal bones and other items found during archaeological searches are used as the basis for elucidating diet. So, to complement these analyses, we are proposing the use of endogenous retroviral DNA sequences present in coprolites or fecal masses to determine diet.

Fellows-Yates, et al. have recently reacted to our data raising concerns in regards to our study. Such concerns are actually expected in an area of research that is establishing its importance amongst other fields, and where there are currently no universally tested standard methods for the isolation of the DNA, etc. However, as several have argued, we should be wary of believing the extreme measures used to prevent contamination and validate ancient human DNA studies would work in the same way for ancient microbial DNA (5, 6). In the absence of such standard methods, we rely on combining previously published suggestions, our microbiological experience with aseptic techniques and the use of tools and areas dedicated specifically for ancient DNA work.

Prevention and assessment of contamination

We welcome the comments provided by Fellows-Yates, et al. as we recognize that we have not described in full the extent of the controls used throughout this new study, and we have relied on detailed explanations published in previous articles by our team. We agree that it is vital to clarify the validity of the DNA obtained, and now realize that all details of previous controls conducted on these coprolites should be detailed in every one of our papers, since we cannot assume that all colleagues read all our publications. First, to prevent cross-contamination, we processed samples independent of one another in assigned sterile biosafety hoods. Rigorous procedures required for ancient DNA studies were used, including but not limited to: use of protective clothing, autoclave and baking of non-disposable equipment, UV-light and ethanol or chlorine sterilization of working areas, separation of operators, stations and equipment for pre- and post-PCR work, aerosol resistant pipette tips, a previously unused DNA extraction kit dispensed into single-use aliquots and a thorough effort in using aseptic techniques during all procedures. This is part of any protocol prior to sending the isolated DNA for sequencing. No heat-producing
protocols were used during sample processing (e.g., drilling) as they could lead to heat-induced DNA fragmentation and less detection efficiency (7). DNA extractions were repeated independently three times for each sample. Extraction blanks were included as negative controls in each DNA extraction session meant for shotgun metagenome sequencing (methods described in Cano et al. 2014 were used for 16S and 18S rDNA microbial profiling). Prior to amplification, no DNA was detectable in the negative controls using NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE) nor BioPhotometer (Eppendorf, Biotech, AG). Controls were then subjected to PCR amplification for bacterial, archael (16S rDNA) and eukaryote (18S rDNA) DNA using standard primers and a high fidelity polymerase (Q5® High-Fidelity DNA Polymerase; NE BioLabs, MA, USA) (8, 9). Reactions were prepared according to polymerase specifications using previously unused, molecular biology-grade solutions. To account for possible PCR inhibitors, serial amplifications were conducted three times and bovine serum albumin was included. Reactions were done in triplicate. As expected, no visible PCR product was obtained from any of the controls. In addition, PCR control reactions showed no presence of DNA when re-evaluated by spectrophotometry. In addition, extracted DNA was subjected to speed-vac drying (without heat) to reduce the possibility of contamination, molecule mobility and enzymatic activity during storage.

As part of their concerns, Fellows-Yates, et al., reference Warinner et al., (2015) promoting the use of in-house sequencing and warning against the use of commercial facilities for sequencing ancient DNA because of the possibility of contamination. Although this might make some sense, our research group is not fortunate enough to have its own sequencing facilities. In any case, this approach would limit this type work to a few laboratories around the world. Additionally, this is a highly subjective opinion, and the suggestion goes against the use of commercial facilities for any type of sequencing work. We assume that any sequencing facility that cannot currently keep samples from being contaminated would quickly be out of business in such a competitive field. We have complete confidence in the facility we rely on for sequencing. We work one on one with this facility, and not only have a commercial understanding with its operatives, but they are also fully aware of the uniqueness of the samples we send for sequencing analyses.

Fellows-Yates, et al. also condemn the use of DNA pre-amplification in our methodology, ignoring its verified effects in substantially increasing ancient DNA detection rates in previous studies with extremely low DNA yields (similar to ours) (10). In addition, they highlight the use of DNA shearing in our library preparation protocols, however we should note that recent findings of well-preserved DNA from ancient bison samples also implemented this step in their procedures (11). Of course, whenever we do ancient DNA work one of our main concerns is avoiding human DNA contamination at each point. Although this includes the process of obtaining the samples from archaeological excavation sites, contamination associated with the removal of the sample from its origin and later handling is always a risk, and the same has been criticized for laboratories using museum
pieces for ancient DNA work (12). For this study, the outside layers of the coprolites were carefully removed and only the cores used for DNA extraction. We should also note that our previous publication included procrustes analyses comparing the microbiomes associated to the outside layers and cores of the coprolites. Results determined the DNA sequences associated to the two regions of the samples were significantly different to each other; this indicates the coprolite cores remained largely untouched by environmental contamination (Cano et al. 2014, Fig.3). Although we do not present the data in the questioned publication, human sequences are also present, as would be expected in fecal samples. However, it is becoming apparent that the extreme measures suggested by main ancient DNA labs for human DNA are not necessary when handling microbial DNA (5, 14).

**Processing and interpretation of data**

In this study we followed standard guidelines for the analyses of shotgun metagenomic DNA datasets; these guidelines have been used for both modern and ancient human microbiome studies (15–18). We used Diagrid (http://diagrid.org) and MetaVir (http://metavir-meb.univ-bpclermont.fr/) search software against the NCBI non-redundant and NCBI RefSeq Viral Genomes databases to identify the sequences obtained from coprolite cores. Contrary to what Fellows-Yates, et al., boldly state in their comments, results could not have been manipulated or biased by our group, as these programs automatically select the best hits for each submitted sequence. Furthermore, both programs identified DNA of the mentioned possible diet-associated viruses. Fellows-Yates, et al., provide results from a nucleotide blast analysis with default parameters conducted on the sequences in question (Supplementary Materials), however, human microbiome and other metagenome studies yield more reliable information when searching for protein-encoding sequences (15, 19–21). This is in part due to expected sequence polymorphisms, the ambiguity of amino acid translation as well as the bias of information observed in current databases, where a handful of organisms represent the total of information available and insufficient or incorrect annotations on the rest (22, 23). This is especially true for viruses, given their possible high mutation rates and the lack of information on viruses in public data repositories (24). As a result of these observations, metagenome software often implement a variety of methods for sequence comparisons, including a translated nucleotide query search and *k*-mer cluster analyses (25–27).

Fellows-Yates, et al., also raise concerns on the possibility of these retroviral sequences being inserted into the human genome, ignoring viral host-range specificity. Currently, the insertion of retroviral DNA in the host genome is not believed to be random. Once they are a part of the host genome, retroviral sequences remain in the host lineage for millennia. For these reasons, endogenous viruses (including retroviruses) are amongst the non-human sequences proposed by human microbiome specialists as phylogenetic markers for migrations of the host species (28). In the case of murine retroviruses, several have been induced to infect other mammalian cell lines in vitro, most specifically infect
rodent cells; moreover, strict strain bias has also been observed in murine retroviruses (29, 30). The possible rodent retroviral protein-coding sequences described in this study seem to be only found in Rodentia, as reported by the NCBI database. This suggests these retroviral-like sequences are exclusively from rodent genomes preserved in the coprolites. We include the XML file of identical protein report from NCBI in the supplementary materials of this rebuttal. We also detected rodent DNA in our data, further suggesting the origin of these endogenous murine retroviral sequences. Results from a nucleotide whole genome alignment to Mus musculus, the rodent with the most molecular data publicly available to date, are included in the supplementary materials of this rebuttal. We are confident that the possibility that these are contaminants is nil, since neither the sequencing facility workers, nor we include rodent meat as part of our diet or have them as pets. To further validate the antiquity of this DNA we include in this rebuttal an example of the results we have obtained from degradation pattern analyses conducted on coprolite DNA using MapDamage (31). Results obtained using various viral genomes as reference indicated higher Cytosine substitutions in the terminals of the DNA fragments, indicative of DNA degradation patterns. One such example can be seen in Figure 1A, generated using Enterobacteria phage PhiX 174 as reference. Data used for this analysis is found in the Supplementary Materials included in this publication. It’s important to point out that although a small segment of PhiX 174 DNA is used during Illumina sequencing, Figure 1B shows we detected the phage’s entire genome, and thus this DNA could not be a result of PhiX controls remnant from sequencing.
ancient DNA degradation behavior. However, degradation patterns are consistent with proposed DNA fragments. In addition, higher Cytosine base mismatches were observed. Low levels of higher frequencies of base pair misincorporations were observed in the terminals of the DNA fragments. In addition, higher Cytosine base mismatches were observed. Low levels of damage were apparent, however degradation patterns are consistent with proposed ancient DNA degradation behavior.

Figure 1: Fragmentation and base misincorporations observed in DNA isolated from Saladoid coprolite cores. Plots were generated using MapDamage (Ginolhac et al., 2011).
Defying dogmatic stringency on preservation

The authors of the commentary mention that “DNA has been shown empirically to be undetectable in high temperatures”, with one reference backing such an imperative claim (Smith 2001). Nevertheless, one study should not be deemed as representative of what can be found in the entire Caribbean. As a side note, the reference apparently has not gone through the peer-review system as of yet, and thus we cannot currently either accept or reject their arguments. We are confident, however, on the possibility of isolating DNA from "high temperature" climates. Although paleogenetic studies are scarce in tropical regions, Fellows-Yates, et al., notably fail to mention the various peer-reviewed studies where ancient DNA was successfully obtained from Caribbean samples (32–36). One recent study even managed to obtain mitochondrial DNA from Nala, a 12,000 year old skeleton that was completely submerged in the warm waters of Yucatan, Mexico, near the Caribbean Sea (37). Such findings defy current opinions on the possibility of finding well-preserved samples in tropical areas, and thus call for more studies on DNA degradation and preservation in these regions.

Although DNA is arguably better preserved in cold, dry environments, this should not eliminate the possibility of DNA preservation in other climates as improbable does not mean impossible. We have observed from some of our previous work that in terms of the physical-chemical variables and their importance for DNA resilience, relative humidity plays a much more important role in DNA resilience. DNAses will be active over a very large temperature gradient, but their activity is more likely to be impaired under low, or very low humidity. This is also true in cold environments, where the water availability for enzymatic reactions is considerably reduced (38). In our case, the very formation and resilience of coprolites are demonstrative of a low humidity environment in Sorcé, since it would be impossible for the feces to undergo mummification under high humidity. Similarly, we conclude that even bone DNA would be much more resilient under low relative humidity. This observation is supported by several paleogenetic studies that have obtained quality ancient DNA of coprolites from regions with low humidity, but are exposed to extremely high temperatures and UV radiation (i.e., areas commonly deemed as incapable of preserving quality, amplifiable DNA) (39–46). Although coprolites have been collected in other Caribbean islands such as Cuba and Jamaica (47, 48), to our knowledge, the coprolites used in our study are the only Caribbean palaeofeces molecularly analyzed to date.

In a previous publication we presented results from cytochemical staining conducted on these coprolites before DNA extraction (see Santiago-Rodriguez et al. 2013, Fig.2) (49). In this study we confirmed the presence of nucleic acids, proteins and lipids in the inner cores of samples using this procedure, a criterion commonly used to suggest the DNA fragments detected were not modern contaminants (50). The preservation of these molecules in the coprolites also helps validate the preservation state of the DNA fragments obtained from the Huecoid and Saladoid samples in Vieques. Furthermore, coprolites used
in this study also contained intact parasite eggs, as discussed in previous publications (13). This also suggests a high level of preservation of the organic matter in these coprolites. Although paleogenetic studies on ancestral Caribbean populations have indeed noticed differences in DNA amplification success depending on the micro-environment responsible for sample preservation, amplifiable DNA was still obtained in most cases (51). Together these studies show that, although an arduous task, obtaining amplifiable ancient DNA from archaeological samples preserved in the Caribbean is indeed possible. In fact, there is very little empirical evidence supporting the widely believed claim that the preservation of organic molecules in the tropics is impossible. Based on these criteria and our own findings, we challenge the stringent views accompanying the apparent dogma that exists on the preservation and indeed the resilience of DNA for centuries.

Concluding remarks

We conclude by thanking our colleagues for their comments as they provided an opportunity to further detail our methods. We should note that the use of metagenomic sequencing in paleostudies is a completely new area of science and it will have its share of detractors. We agree that all ancient DNA studies should adhere to as many of the basic ‘checklist’ of suggestions and precautions listed by pioneers that have dealt with DNA degradation and modern contaminants in their paleosamples (50, 52, 53). However, as these same pioneers have said, several protocols allow us to effectively obtaining DNA from paleosamples, and there are currently no universal standard methods for the isolation of nucleic acids from ancient artifacts. Furthermore, reports comparing these methods in osteological samples have not been conducted on paleofaeces (54). We are convinced that there may never be one single Standard Operating Procedure for the extraction of ancient DNA, since the artifacts are so varied and unique that perhaps an overall procedure should always be accompanied with alterations pertinent to the uniqueness of the sample, archaeological site or DNA yield, together with the proper controls. We should point out that other colleagues in this field have also mentioned this (55).

As a group of scientists that has been very active in this area, we welcome all comments that are based on data; however, comments based on opinions, however authoritative they may be are harder to accept. The latter approach will do a disservice to our area of research. We are aware of the uniqueness of the samples, and as such, we need to be careful in the inclusion of controls. The most we can do is to carry out our research under the most stringent of conditions as possible. We cannot please all skeptics, and working in this area of research we are aware that there are always angles that can be criticized, but as long as this is constructive criticism we welcome this, and we are sure that these comments will help our area of science to elucidate much more about ancient ethnic groups. We refrain from including our own prejudices in our work, and our group is constantly in awe at the data we are obtaining, since we started out at the same point where many of our colleagues seem to be; that DNA in these samples would not be resilient
enough to be isolated and sequenced. Our past prejudices are by no means unique, and since our data are questioning the status quo, we should be ready for critiques. We have several manuscripts in preparation looking at different aspects of not only coprolites, but also fecal masses in mummies; we expect, and welcome all comments. Lastly, let us also not forget that all these manuscripts go through a thorough peer-review step prior to publication; we have answered many of the questions regarding possible contamination during the peer-review processes.

References
80(9):2648–2655.


