Raphael Eisenhofer, Jeremiah J. Minich, Clarisse Marotz, Alan Cooper, Rob Knight, and Laura S. Weyrich

Contamination in low microbial biomass microbiome studies: issues and recommendations

Trends in Microbiology, 2019; 27(2):105-117

Final publication at: http://dx.doi.org/10.1016/j.tim.2018.11.003

Contamination in low-microbial biomass microbiome studies: issues and recommendations

Authors: Raphael Eisenhofer[1,2], Jeremiah J. Minich[3], Clarisse Marotz[4], Alan Cooper[1,2], Rob Knight[4,5,6], and Laura S. Weyrich[1,2]

1: Australian Centre for Ancient DNA, University of Adelaide, Australia

2: ARC Centre of Excellence for Australian Biodiversity and Heritage

3: Marine Biology Research Division, Scripps Institution of Oceanography, La Jolla, California, USA

4: Department of Pediatrics, University of California San Diego, La Jolla, California, USA

5: Center for Microbiome Innovation, Jacobs School of Engineering, University of California San Diego, La Jolla, California, USA

6: Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, USA

Keywords: Microbiome, Microbiota, DNA contamination, Laboratory contamination, Low-biomass, Methodology

Abstract

Next-generation sequencing approaches in microbiome research have allowed surveys of microbial communities, their genomes, and their functions with higher sensitivity than ever before. However, this sensitivity is a double-edged sword, as these tools also efficiently detect contaminant DNA and cross-contamination, which can confound the interpretations of microbiome data. Therefore, there is urgent need to integrate key controls into microbiome research to improve the integrity of microbiome studies. Here, we review how contaminant DNA and cross-contamination arise within microbiome studies and discuss their negative impacts, especially during the analysis of low-microbial biomass samples. We then identify several key measures that researchers can implement to reduce the impacts of contaminant DNA and cross-contamination during microbiome research. We put forth a set of minimal experimental criteria to improve the validity of future low-microbial biomass

research called the '**RIDE**' checklist.

Prospects and pitfalls of microbiome research

The completion of the Human Microbiome Project in 2017[1] was a major landmark in **microbiome** research. This research field has the potential to create novel therapies for human disease, aid in environmental conservation, improve agricultural outputs, understand our ancestor's lifestyles, and identify criminals in forensic casework, amongst many other areas[2–6].

Amplification-based methods that target hypervariable regions (*e.g.* PCR amplification of the 16S ribosomal RNA (rRNA) gene) account for the majority of studies exploring the **microbiota** because of their speed and inexpensive cost[7]. Shotgun sequencing has also become more popular in recent years due to decreasing DNA sequencing costs and the ability to obtain both species-level taxonomic resolution and functional genomic information. Both of these approaches rapidly illuminate uncultured microorganisms and allow researchers to compare and contrast microbial communities in diverse environments, including the human body, subglacial Antarctic lakes, NASA's space equipment, deep-sea hydrothermal vents, extinct hominids, and coral reefs[5,8–12].

Despite their benefits, the molecular methods used to investigate microbial communities have key limitations, including non-proportional target amplification and the inclusion of **contamination**. While tools to address non-proportional target amplification have been developed[13–15], strategies to limit contamination are less appreciated. Several studies have documented the routine amplification of contamination and its impacts on biological interpretations[16–24], but there is still no systematic requirement to examine or report contamination within microbiota or microbiome (hereby referred to as microbiome) studies. Here, we highlight how contamination has negatively impacted microbiome research, especially when assessing **low-microbial biomass samples**, and provide several recommendations to minimize the effects of contamination in future research.

Contamination in Microbiome Studies

Two key types of contamination can arise in microbiome studies: **contaminant DNA** and **cross-contamination**. Contaminant DNA can originate from many sources despite the utmost care in sample collection and preparation, including the sampling and laboratory environments[25–27], researchers, plastic consumables[28], nucleic acid extraction kits[5,19,23,24,29–32], laboratory reagents including PCR mastermixes[16–18,33–36], and cross-contamination from other samples and sequencing runs[37,38]. To date, over 30 common contaminant taxa have been identified in **DNA extraction blank controls** and **no-template controls** across multiple studies (Table 1). For example, Salter *et al.* found that several contaminant taxa were shared in blank controls across multiple studies, laboratories, and DNA extraction methods[19]. These widespread contaminant taxa appear to originate from common sources (*e.g.* kit and reagent manufacturing, human commensals on lab personnel, or thrive within laboratory environments). Despite the identification of some common contaminants, the types and abundance of contaminant taxa vary between extraction kits and laboratories[5,19,23,24] and even through time within the same laboratory[39].

Cross-contamination is another challenge during microbiome sample processing and includes the transfer of primary sample DNA, barcodes, or amplicons from neighboring wells or tubes to create “batch effects”[40]. Cross-contamination can occur at multiple steps throughout sample processing: sample DNA can be accidentally transferred during initial sample processing and placement into tubes or plates[41], from aerosolization during pipetting, or during plate cover removal[42]. Barcode cross-contamination may also occur when incorrect neighboring barcodes ‘jump’ into sample wells or tubes — a phenomenon known as ‘tag switching’[43]. Finally, cross-contamination can also occur on the sequencing instrument from barcode sequencing errors, residual amplicons from past sequencing runs, or “index hopping,” where some sequencing platforms mismatch indexing reads to sequencing reads [44,45]. Overall, both contaminant DNA and cross-contamination are dynamic and need to be consistently and routinely monitored.

Sample Types Most Affected by Contamination

The impacts of contaminant DNA and cross-contamination can vary between samples according to their levels of microbial biomass. The microbial biomass in a sample can be estimated by comparing the quantity of microbial DNA in samples (*e.g.* quantitative PCR of 16S rRNA amplicons) to that in DNA extraction blank controls[23]. Samples that typically contain high-microbial biomass include feces or soil, and will usually contain substantially more DNA than DNA extraction blank controls, while low-microbial biomass samples will contain DNA levels similar to DNA extraction blank controls and include glacial ice, air, rocks, the built environment, placenta, and blood. Lower levels of microbial DNA within low-microbial biomass samples allow contaminant DNA and cross-contamination (*e.g.* from high-biomass samples processed simultaneously) to easily outcompete and dominate the biological signal within samples[19,23,24,46].

How Contaminant DNA Influences Microbiome Studies

The amount and composition of contaminant DNA and cross-contamination can vary through time and location, generating signals within low-microbial biomass samples that can be easily perceived as biological; this concept is illustrated in Figure 1. Numerous studies have described contaminant DNA and demonstrated how it can skew results, including those in published low-microbial biomass studies[19,23,24]. For example, >95% of the taxonomic composition in a *Salmonella bongori* culture diluted to ~1,000 cells was revealed to be contamination using both amplicon and shotgun DNA sequencing[19]. The same authors also found that infant nasopharyngeal swabs clustered according to the DNA extraction kit lot number, demonstrating that contaminant taxa introduced during DNA extraction were driving the observed signal[19]. A comparison of low-microbial biomass placental samples with blank controls, saliva, and vaginal swabs revealed that 16S rRNA gene sequences in placental samples could not be distinguished from those in blank controls[23]. Lastly, an analysis of peripheral blood and submucosal tissue samples demonstrated that 99% and 95% of the respective identified sequences corresponded to contaminant taxa[24]. The impacts of contaminant DNA and cross-contamination are not limited to these ‘whistle-blower’ studies and have likely impacted each and every low-microbial biomass study published to date. Even if controls and low-microbial biomass samples can be distinguished using beta-diversity analyses (*e.g.* a PCoA plot of unweighted UniFrac distances), measures of alpha (within-sample) diversity and differential abundance can be confounded in microbiome studies due to contaminant DNA and cross-contamination. Together, these findings demonstrate that contaminant DNA and cross-contamination can have a severe impact on low-microbial biomass microbiota studies and will continue to pose a demonstrable threat to the integrity of the field if left unaddressed.

How Has DNA Contamination Already Impacted the Microbiome Research Field?

The failure to include controls to assess DNA contaminants and cross-contamination has resulted in several controversial studies. For example, a recent study identified a distinct microbial community within human placenta without publishing appropriate controls[47]. Bacterial DNA contribution from maternal blood was raised as an issue[48], and no evidence for a distinct placental microbiota was found when placental samples were compared with blank controls in a follow-up study[23]. A recent, comprehensive review concluded that current evidence does not support the notion that the human placenta harbors a distinct microbiota[49]. Nevertheless, the initial publication[47] spurred several subsequent studies[50–53] on the ‘placental microbiota’; all lacked appropriate controls and further perpetuated the notion that the placenta harbors a distinct microbiota. In addition to the placenta, there has been a recent surge of other low-microbial biomass microbiota studies, especially in clinical medicine, and include investigations of the microbial components of brain tissue[54], breast tissue[55,56], nipple aspirate fluid[57], intrauterine samples[58], and seminal fluid[59]. None of these studies included appropriate controls or an assessment of contaminant taxa and cross-contamination in their findings. Unsurprisingly, each of these studies identified common contaminant taxa from commercial extraction kits and molecular reagents as the taxa driving the observed biological signals. In addition, the studies failed to examine the limit of detection using their methodology – the critical first step when exploring low-microbial biomass communities. While it is possible that these are true biological signals, it is also possible that they arise from contaminant DNA, and additional experiments should be included to determine if such microbial DNA originates from living cells as opposed to contaminant DNA[60]. Together, these studies highlight the desperate need for the field to recognize and adhere to a minimum set of experimental criteria to ensure valid and reproducible findings.

Mitigating the Impacts of Contaminant DNA

To control for contaminant DNA and cross-contamination in low-microbial biomass microbiome studies, there are several measures that need to be taken to 1.) reduce all types of contamination and experimental bias, 2.) monitor and identify contaminant sources, and 3.) recognize and mitigate the effects of contaminant DNA and cross-contamination during analysis. In chronological order of how a study would be performed, we provide suggestions for each approach, and put forth minimum guidelines ('**RIDE**' checklist; Box 1.) to help researchers, editors, and reviewers manage the effects of contamination in future microbiome research (**Box 1**).

1.) *Reduce experimental bias and contamination during sampling and processing.*

Simple measures during sample collection and processing can be used to limit the introduction of contaminant DNA and cross-contamination and minimize their downstream effects (Figure 2). First, randomizing samples and treatments (*i.e.* collecting or processing samples from different treatments together) is an important experimental design consideration to prevent erroneous conclusions arising from batch effects or day-to-day variation of contaminant DNA (Figure 1). In addition, the same researcher, reagents, robots, and equipment should be used to process all of the samples in a specific study, if possible. To specifically avoid contaminant DNA, there are several key considerations. Samples should be collected in the cleanest available environment (*e.g.* inside a ship rather than on deck; in a wind protected area; etc), and personnel should wear protective clothing and equipment to cover all exposed human surfaces if possible (*i.e.* lab coats or cleanroom suits, face masks, hair nets, sleeves, and clean disposable gloves). Ideally, researchers should also process the samples in an isolated, low-contaminant, controlled environment (*e.g.* still-air cabinet or laminar-flow hood) where surfaces and equipment are treated with a ≥3% sodium hypochlorite solution and ultraviolet radiation to minimize and fragment environmental contaminant DNA[61]. Samples should be processed using reagents, lab ware, and sampling equipment that have the lowest levels of contamination possible. As consumables labeled 'DNA free' typically contain degraded microbial DNA[36], consumables with hard surfaces, such as plastic tubes and pipettes, can be decontaminated using ethylene oxide treatment[28], and reagents can be decontaminated by UV treatment that is optimized for each reagent (*i.e.* UV irradiation can destroy enzyme function)[62]. Ideally, a physically isolated workstation should also be used to aliquot stock reagents to limit contamination[63]. To minimize cross-contamination, there are additional steps to consider. Library preparation should be performed in a separate room from DNA extraction to minimize contamination from highly-amplified products (*i.e.* pre-PCR work should be physically isolated from post-PCR work). Filter tips and low-aerosol pipettes can also help in reducing cross-contamination[64]. The use of non-redundant dual indexing is strongly recommended to prevent index swapping during sequencing [65,66]. It is also important to perform the recommended bleach and maintenance washes in the DNA sequencer between sequencing runs, as this can reduce run-to-run cross-contamination in Illumina MiSeq studies by 100-fold (from 0.01% to 0.0001%) [67].

Minimum guidelines: Different sample groups or treatments should be randomized and not processed separately. Researchers should wear disposable lab gloves, face masks, and avoid exposed skin to reduce the introduction of contaminant DNA into the samples. As many procedures as possible (*e.g.* sample transfer, DNA extraction, library preparation, and sequencing) should be performed in a cleaned, isolated working environment with appropriately treated equipment and consumables.

2.) *Include controls from sampling to sequencing.*

Several types of controls should be included in every analysis to monitor contaminant DNA and assess the levels of cross-contamination between samples. These controls include both negative controls to monitor background levels of contaminant DNA: (1) sampling blank controls, (2) DNA extraction blank controls, and (3) no-template amplification controls. In addition, two types of positive controls across a titration (variable cellular or gDNA input) can be used to determine the limit of detection and ensure cross-contamination does not drive the results of the study: (4) DNA extraction positive controls and (5) amplification positive controls.

Negative controls

Three types of negative controls are minimally required to allow adequate monitoring of contaminants throughout sample handling and processing and provide the ability to detect when and how contaminants are introduced into biological samples. At least one of each type of negative control must be included per sampling, extraction, and amplification batch. Although we would recommend that two negative controls should be used and placed strategically to monitor contaminants from the start to the end of the process (*e.g.* the first tube should be negative control #1, the last tube should be negative control #2). For larger studies using robotic systems with plates, 8 of each negative control type should be minimally required per study [68].

(1) **Sampling blank controls** allow for detection of contaminant DNA introduced during the sampling procedure, including items used to collect the sample, such as swabs, gauze, or drills, and any reagents or preservatives used to store or transport the samples (*e.g.* media, alcohol, or RNA stabilizer). Material analyzed in sampling blanks should be collected in the same room and at the same time as biological samples and should undergo the same laboratory treatment as the biological samples, from collection to sequencing. While sampling controls will contain DNA from the extraction process, it will allow the researcher to discern which contaminants are specific to the sampling location and equipment versus the laboratory.

(2) **DNA extraction blank controls** monitor the contaminant DNA content in extraction kits, molecular reagents, and the laboratory environment through the DNA extraction process

and, as above, should be processed alongside the biological samples from extraction to sequencing.

(3) **No-template amplification controls** can monitor contaminant DNA present in reagents and the laboratory environment during library preparation and sequencing. All negative controls provide a semi-quantitative estimate of background contaminants and allow researchers to identify contaminants that can be used in downstream subtractive analyses. Finally, it should be noted that negative controls can contain too little DNA to be effectively processed. In these cases, the use of known carrier DNA in blank controls can help to efficiently amplify contaminants[69].

Positive Controls

Two types of positive controls can be included to determine the limit of detection and provide insight into the effects of cross-contamination during extraction, library preparation, and sequencing.

(4) **DNA extraction positive controls** monitor DNA extraction efficiency, determine the limit of detection, and examine levels of cross-contamination during DNA extraction. To include a DNA extraction positive control, a serial dilution of a known cell type(s) (*e.g.* 1, 10, 100, 1000, 10,000, 100,000 cells) should be extracted alongside samples and span the expected limit of detection of the assay (see Katharoseq below)[68]. Ideally, researchers should use a commercially available mixed community, such as the Zymo mock community (Zymo, D6300), as this enables standardization across different laboratories. Researchers can also consider including a range of positive titration spike-ins into liquid samples, such as blood, urine, or mucus, to evaluate the efficiency of extraction and the limit of detection, which is important as many sample types have inhibitors or chemicals that can increase the limit of

detection. The bottom line is to use a positive control of known concentration that is relevant to your study and experimental questions.

(5) The last recommended positive control is the **positive amplification control**, which is again a titration of DNA from known organism type(s) to be processed solely during the library preparation stage. This control enables a detection limit to be established for library preparation. Critically, both positive control types can be used to calculate the limit of detection within the laboratory techniques used and the levels of cross-contamination using novel bioinformatic approaches [68]. For example, Katharoseq utilizes differences in amplification efficiencies of true positives compared to negatives to mathematically determine a limit of detection by calculating cutoff scores to guide sample exclusion. In doing so, cross-contamination can also be evaluated, as positive controls from DNA extractions should be different from those used in library preparation.

Control samples often produce libraries of lower quantity and quality, but this should not prevent the control samples from being sequenced. Libraries should be quantified (*i.e.* using a PicoGreen or Qubit assay for amplicon studies or a TapeStation or BioAnalyzer for shotgun sequencing) and pooled at equal molarity (*e.g.* X ng per observed fragment lengths per sample). If amplified control samples contain significantly lower amounts of DNA compared to biological samples, they should be included in sequencing pools by pooling the controls at a certain maximum volume (*e.g.* 20 µl of each control). In addition, amplified biological samples with low amounts of DNA can be pooled at this same maximum volume as controls (*e.g.* 20 µl)[68]. Alternatively, all samples and controls can be pooled at equal volumes; however, this approach requires deeper sequencing because the higher-biomass samples will dominate the DNA sequencing effort. While not ideal, another option is to increase the number of PCR cycles for negative controls to gain more DNA for sequencing. For highly contentious sample types and claims (*e.g.* placenta), independent replication in another laboratory and the use of non-DNA sequencing approaches (*e.g.* FISH) for verification are highly recommended.

Minimum guidelines: One of each negative control type (sampling blank control, DNA extraction blank control, and no-template amplification control) must be included for each batch of samples, or a minimum of 8 negative controls per type per 96-well plate for studies using robotic systems. Controls must be processed alongside samples to account for contamination and should not be processed separately.

3.) *Critically* a*ssess and report contributions of contamination during analysis.*

The impacts of contaminant taxa must be assessed in the final analysis and interpretation of the data. Three different strategies currently exist to assess the impacts of contamination in microbiome datasets: (1) compare controls to biological samples; (2) subtract contaminants from biological samples; and (3) use predictive modeling to identify putative contaminants. Each method varies in its stringency and application.

(1) Comparisons of biological samples to the controls can be used to assess the level of contamination and the types of contaminant taxa. The level of contamination (*i.e.* background levels of contaminant DNA) must be determined per batch of samples, as level of contaminant DNA can vary based on different methodologies and through time[5,19,23,24,39]. Quantitative PCR (qPCR) can be used to determine the level of

contamination by comparing abundances in negative controls to biological samples [23]. Alternatively, we recommend that positive controls coupled with the limit of detection approach can be used to calculate a sample exclusion value (*e.g.* $K_{1/2}$ value)[68], and samples with fewer reads than the exclusion value should be discarded [68]. Taxa detected in negative controls must be reported. This is especially important to ensure that the significant differences in taxa abundances or composition between sample types or treatments are not driven by contaminant taxa. We provide a table containing taxa that have been detected in the negative controls from two or more studies (Table 1). While we do not recommend that researchers throw away any significant result driven by the taxa in this table, researchers and reviewers should be extra cautious of such findings.

(2) Contaminant taxa detected in negative controls can also be subtracted (filtered) from biological samples during analysis. One approach is to remove all taxa found within negative controls from the biological samples. This is an extremely conservative approach that can result in the loss of biological signal due to cross-contamination of DNA from biological samples into negative controls. In addition, taxa closely related to common contaminant taxa can be truly present in a biological specimen (*e.g. Pseudomonas*), and would be removed by this approach. We would instead recommend the use of more nuanced filtering approaches that have been developed to help in situations where cross-contamination is high or when taxa closely related to common DNA contaminants are thought to be present in biological samples [70–73]. Finally, should contaminant taxa still be driving biological signal after filtering, they should be verified using a different approach such as an effectively used and validated Fluorescent In-Situ Hybridization (FISH) assay [74,75]

(3) Bioinformatic modeling has been developed to estimate the source and proportions of contaminant taxa within biological samples. For example, SourceTracker analysis uses Bayesian modeling to estimate the proportion of potential contaminant taxa from a data set[76]. To do this, the blank controls can serve as contaminant 'sources' and the biological samples as 'sinks' to estimate the origin and abundance of contaminant taxa within biological samples. Subsequently, the relative contributions of contaminant DNA within the samples can be factored into downstream analysis and data interpretation. However, it should be stressed that sufficient cross-contamination can confound SourceTracker analysis.

Minimum guidelines: The level of contamination must be determined for each batch of samples. Biological samples must be compared to negative controls and taxa identified in negative controls must be reported. The approach taken to identify and minimize the effects of contaminant DNA during analysis should be clearly reported to enhance reproducibility and allow such approaches to be critically evaluated by others.

Concluding Remarks

Microbiome research holds great promise for multiple fields, but methodological pitfalls can easily undermine the progress and reputation of this developing research area. Therefore, these pitfalls must be recognized and explicitly addressed at each phase of the scientific process by researchers, reviewers, and editors alike. Here, we present the '**RIDE**' checklist for contaminant assessment to be applied across a wide-range of disciplines interested in exploring the microbial communities in low-microbial biomass samples (see **Box 1** for our

'**RIDE**' minimum standards checklist). Failure to take these caveats into account is likely to waste valuable time and money and erode the credibility of microbiome research. The current situation is similar in many ways to the methodological issues in ancient DNA research recognized over 20 years ago. A series of high-profile publications based on PCR amplification of short sequences were used to support remarkable findings, including the reported recovery of DNA more than 40 million years old[77–79] – well beyond the theoretical limit of DNA survival of around one million years[80]. Although these findings were heavily criticized by other ancient DNA researchers[81–85] and are now recognized as erroneous, these publications nevertheless damaged the credibility of the ancient DNA field. As a direct result, a set of ancient DNA authentication criteria was formulated and widely adopted[63]. These standards, improved techniques, and greater attention to the issue of contaminant DNA dramatically improved the credibility of ancient DNA research. In microbiome research, similar standards need to be established to improve scientific integrity and secure the credibility of such research. It is important to note that the minimum set of guidelines and the 'RIDE' checklist that we propose (**Box 1**) will not guarantee that all contamination can be accounted for or removed, nor will it provide a solution for every contaminant problem. Complementary approaches for verifying results such as replication in independent laboratories and using non-DNA sequencing techniques such as FISH should also be considered. As new methods and analyses for microbiome analysis are also developed, novel solutions to account for contaminant DNA and cross-contamination will need also to be established (see Outstanding Questions). In the meantime, it is imperative that low-microbial biomass research generates sufficient control data and that researchers develop and maintain a critical mindset when dealing with low-microbial biomass microbiome samples. In this regard, we hope that the guidelines introduced in this article will help authors, reviewers, and editors monitor and protect the future of the microbiome field.

Declarations

Acknowledgments

We would like to thank Alan W. Walker, Bastien Llamas, Jessica L. Metcalf, Kieren J. Mitchell, and Matilda Handsley-Davis for their feedback and suggestions.

Funding

LSW and RE were funded from the Australian Research Council grants: DECRA (DE150101574) and ARC Centre of Excellence CABAH (CE170100015).

Competing interests

The authors declare that they have no competing interests.

Glossary:

Contamination: An umbrella term encompassing both contaminant DNA and cross-contamination (see below).

Contaminant DNA: DNA from sources other than the sample(s) under study (*e.g.* DNA from reagents or researchers performing laboratory work).

Cross-contamination: DNA exchange between samples within a study (*e.g.* accidental movement of DNA between different sample tubes during DNA extraction).

DNA extraction blank control: A negative control consisting of an empty tube/well that is processed alongside biological samples during DNA extraction and allows for the detection of contaminant DNA introduced during DNA extraction.

DNA extraction positive control: A positive control consisting of serially diluted cells of known type(s) that is processed alongside biological samples during DNA extraction and allows for determination of the limit of detection, monitoring of extraction efficiency, and quantification of cross-contamination during DNA extraction.

Low-microbial biomass samples: A biological sample that contains similar quantities of target microbial DNA in the sample compared to negative controls (e.g. ≤10,000 microbial cells — [19]).

Microbiome: The microorganisms of a specific habitat, their genomes, and the surrounding environmental conditions[86].

Microbiota: The assemblage of microorganisms present in a defined environment[86].

No-template amplification control: A negative control made by preparing an amplification or library preparation reaction without input template (*i.e.* sample DNA) that is processed alongside biological samples and allows for the detection contaminant DNA during library preparation/PCR amplification.

Positive amplification control: A positive control consisting of serially diluted DNA from known organism type(s) that are processed alongside biological samples during amplification or library preparation and allows for determination of the limit of detection, monitoring of library preparation efficiency, and quantification of cross-contamination during library preparation.

RIDE: **R**eport methodology, **I**nclude controls, **D**etermine the level of contamination, and **E**xplore the impacts of contamination in downstream analysis. Minimum standards checklist for low-microbial biomass microbiome studies.

Sampling blank control: A negative control consisting of an empty tube that is processed alongside the collection of biological samples. Allows for the detection of contaminant DNA introduced during the sampling procedure (*e.g.* airborne, swabs, preservatives).

Genus	Reference
Actinomyces	[23][24][39]
Corynebacterium	[19][24][68]
Arthrobacter	[19][24]
Rothia	[23][24]
Propionibacterium	[19][23][24][68]
Atopobium	[23][24]
Sediminibacterium	[23][39]
Porphyromonas	[23][24]
Prevotella	[23][24][68]
Chryseobacterium	[19][39]
Capnocytophaga	[23][24]
Chryseobacterium	[19][24]
Flavobacterium	[19][21][23][39]
Pedobacter	[19][39]
unclassifiedTM7	[23][24]
Bacillus	[19][24][39]
Geobacillus	[24][39]
Brevibacillus	[19][24]
Paenibacillus	[19][24][39]
Staphylococcus	[24][39][68]
Abiotrophia	[19][24]
Granulicatella	[23][24]
Enterococcus	[23][24][39]
Lactobacillus	[23][24][39]
Streptococcus	[19][23][24][39][68]
Clostridium	[24][39]
Coprococcus	[23][24]
Anaerococcus	[23][24]
Dialister	[23][24]
Megasphaera	[23][24]
Veillonella	[23][24]
Fusobacterium	[23][24]
Leptotrichia	[23][24]
Brevundimonas	[18][19]
Afipia	[19][24]
Bradyrhizobium	[19][21][24][39]
Devosia	[19][39]
Methylobacterium	[18][19][23][39][68]
Mesorhizobium	[19][39]
Phyllobacterium	[19][24]
Rhizobium	[18][19][21]
Methylobacterium	[19][24]

Phyllobacterium	[19][24]
Roseomonas	[19][24]
Novosphingobium	[19][39]
Sphingobium	[19][39]
Sphingomonas	[18][19][21][39]
Achromobacter	[21][39]
Burkholderia	[19][21][24][39]
Acidovorax	[18][19]
Comamonas	[18][19][24][39]
Curvibacter	[19][24]
Pelomonas	[19][24][68]
Cupriavidus	[18][19][39]
Duganella	[16][19]
Herbaspirillum	[16][18][19][24]
Janthinobacterium	[19][24]
Massilia	[18][19][24]
Oxalobacter	[19][24]
Ralstonia	[17][18][19][21][39]
Leptothrix	[16][19]
kingella	[19][24]
Neisseria	[23][24]
Escherichia	[16][18][19][21][24][68]
Haemophilus	[23][24][68]
Acinetobacter	[16][18][19][23][39][68]
Enhydrobacter	[19][24][39]
Pseudomonas	[17][19][21][24][39][68]
Stenotrophomonas	[16][17][18][19][21][24][39]
Xanthomonas	[17][19]

Boxes and Figures:

Table 1: Taxa previously identified in negative controls from multiple studies
Taxa identified in the negative controls of more than one study are listed. Taxa listed in this table that are found to be driving significant results in a study should be treated with extra skepticism and evidence should be provided by researchers to prove that such findings are not due to contamination.

Box 1: For authors, reviewers, and editors, the 'RIDE' minimum standards checklist for performing/reviewing low-microbial biomass microbiome studies.

- ✓ **R**eport experimental design and approaches used to reduce and assess the contributions of contamination.
- ✓ **I**nclude controls to assess contaminant DNA. One of each type of negative control (*s*ampling blanks, DNA extraction blanks, and no-template amplification) must be included per sampling, extraction, or amplification batch.

- ✓ **D**etermine the level of contamination by comparing biological samples to controls.
- ✓ **E**xplore contaminant taxa within each study and report their impacts on the interpretation of biological samples.

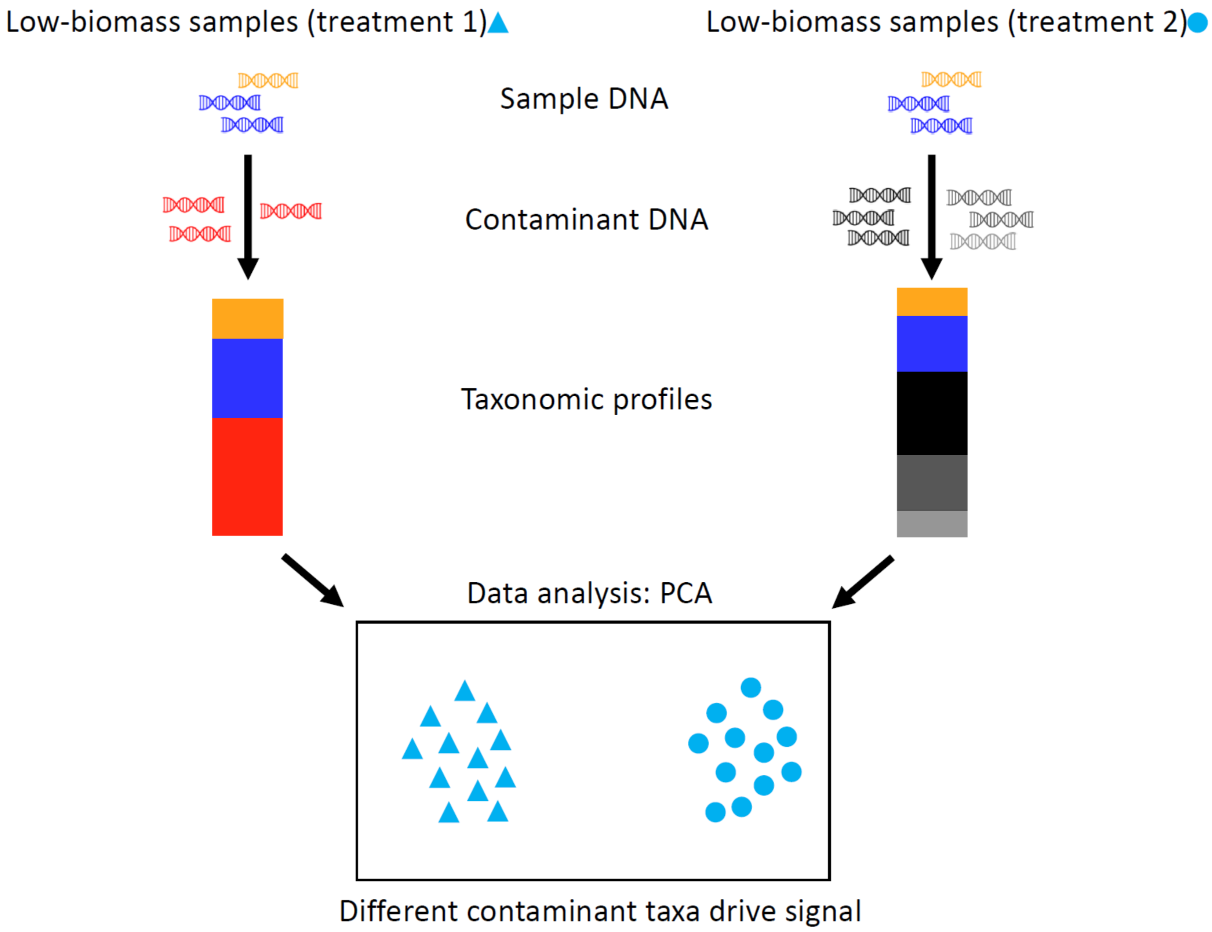

Figure 1: Illustration of how contaminant DNA can influence interpretations of low-microbial biomass microbiome data.
Both treatment groups (triangle vs. circle) of low-microbial biomass samples are not different in microbial composition (sample DNA colors are same, blue and orange). However, because treatment groups were processed on separate days, differences in the types and abundances of contaminant DNA (in this case, red vs. black) drive the signal, leading to the conclusion that the treatment groups have different microbial compositions. Proper randomization of sample collection/processing would eliminate this artifact. Abbreviation: PCA, principal component analysis.

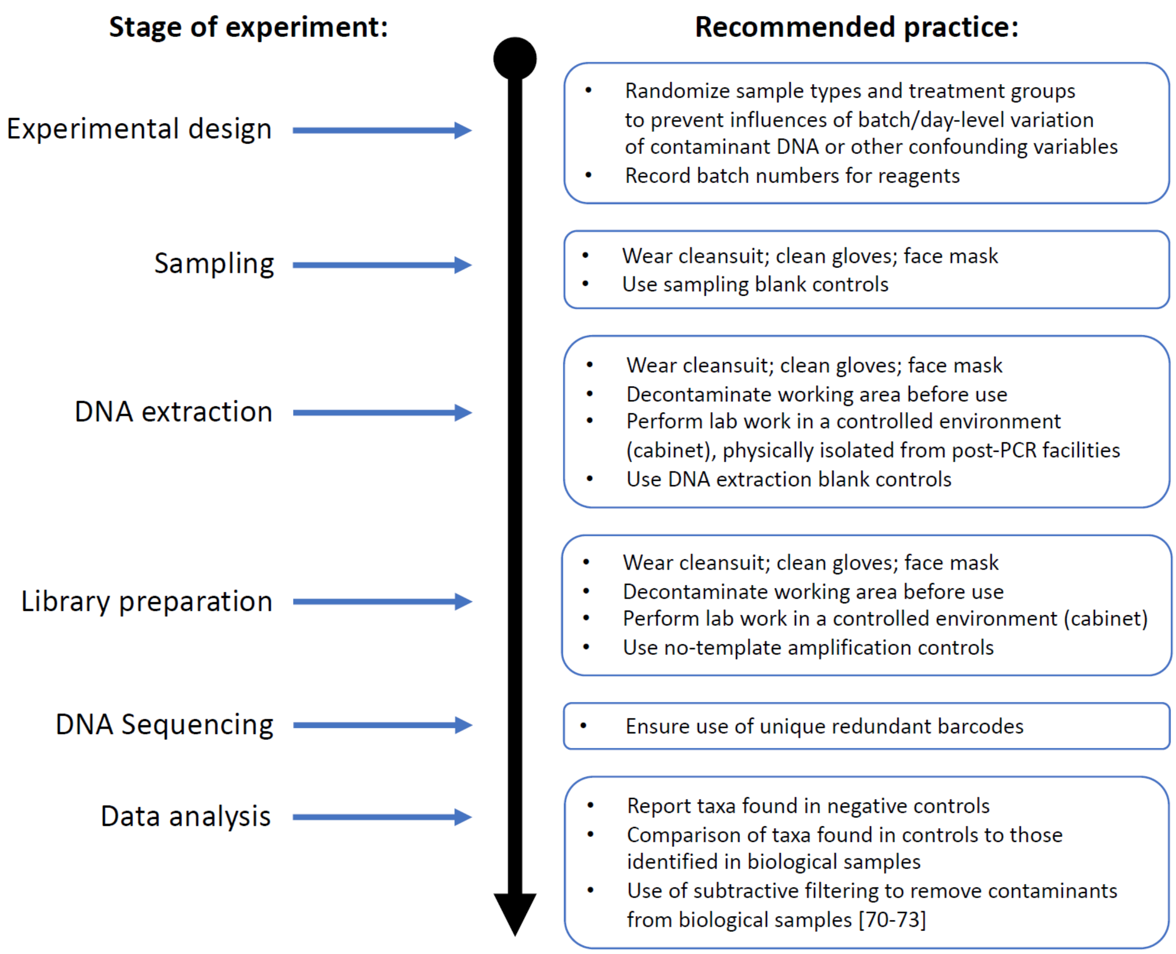

Figure 2: Flowchart of methods to minimize influence of contaminant DNA in low-microbial biomass samples. Measures to reduce experimental bias and the introduction of contaminant DNA in low-microbial biomass microbiome studies.

References

1 Lloyd-Price, J. *et al.* (2017) Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550, 61–66

2 Kassam, Z. *et al.* (2013) Fecal Microbiota Transplantation for Clostridium difficile Infection: Systematic Review and Meta-Analysis. *Am. J. Gastroenterol.* 108, 500–508

3 Bahrndorff, S. *et al.* (2016) , The Microbiome of Animals: Implications for Conservation Biology. , *International Journal of Genomics*. [Online]. Available: https://www.hindawi.com/journals/ijg/2016/5304028/. [Accessed: 27-Jul-2017]

4 Sessitsch, A. and Mitter, B. (2015) 21st century agriculture: integration of plant microbiomes for improved crop production and food security. *Microb. Biotechnol.* 8, 32–33

5 Weyrich, L.S. *et al.* (2017) Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* 544, 357–361

6 Fierer, N. *et al.* (2010) Forensic identification using skin bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* 107, 6477–6481

7 Caporaso, J.G. *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624

8 Consortium, T.H.M.P. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214

9 Christner, B.C. *et al.* (2014) A microbial ecosystem beneath the West Antarctic ice sheet. *Nature* 512, 310–313

10 Checinska, A. *et al.* (2015) Microbiomes of the dust particles collected from the International Space Station and Spacecraft Assembly Facilities. *Microbiome* 3, 50

11 Anantharaman, K. *et al.* (2016) Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. *ISME J.* 10, 225–239

12 Kelly, L.W. *et al.* (2014) Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc. Natl. Acad. Sci.* 111, 10227–10232

13 Lozupone, C. and Knight, R. (2005) UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* 71, 8228–8235

14 Aird, D. *et al.* (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12, R18

15 Gagic, D. *et al.* (2014) Improving the genetic representation of rare taxa within complex microbial communities using DNA normalization methods. *Mol. Ecol. Resour.* DOI: 10.1111/1755-0998.12321

16 Tanner, M.A. *et al.* (1998) Specific Ribosomal DNA Sequences from Diverse Environmental Settings Correlate with Experimental Contaminants. *Appl. Environ. Microbiol.* 64, 3110–3113

17 Grahn, N. *et al.* (2003) Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiol. Lett.* 219, 87–91

18 Barton, H.A. *et al.* (2006) DNA extraction from low-biomass carbonate rock: An improved method with reduced contamination and the low-biomass contaminant database. *J. Microbiol. Methods* 66, 21–31

19 Salter, S.J. *et al.* (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12, 87
20 Lusk, R.W. (2014) Diverse and Widespread Contamination Evident in the Unmapped Depths of High Throughput Sequencing Data. *PLOS ONE* 9, e110808
21 Laurence, M. *et al.* (2014) Common Contaminants in Next-Generation Sequencing That Hinder Discovery of Low-Abundance Microbes. *PLOS ONE* 9, e97876
22 Adams, R.I. *et al.* (2015) Microbiota of the indoor environment: a meta-analysis. *Microbiome* 3, 49
23 Lauder, A.P. *et al.* (2016) Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 4, 29
24 Glassing, A. *et al.* (2016) Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* 8, 24
25 Willerslev, E. *et al.* (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* 19, 141–147
26 Witt, N. *et al.* (2009) An Assessment of Air As a Source of DNA Contamination Encountered When Performing PCR. *J. Biomol. Tech. JBT* 20, 236–240
27 Llamas, B. *et al.* (2017) From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR Sci. Technol. Archaeol. Res.* 3, 1–14
28 Motley, S.T. *et al.* (2014) Improved multiple displacement amplification (iMDA) and ultraclean reagents. *BMC Genomics* 15, 443
29 Fierer, N. *et al.* (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc. Natl. Acad. Sci.* 105, 17994–17999
30 Dunn, R.R. *et al.* (2013) Home Life: Factors Structuring the Bacterial Diversity Found within and between Homes. *PLOS ONE* 8, e64133
31 Naccache, S.N. *et al.* (2013) The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns. *J. Virol.* 87, 11966–11977
32 Adams, R.I. *et al.* (2014) Airborne Bacterial Communities in Residences: Similarities and Differences with Fungi. *PLOS ONE* 9, e91283
33 McFeters, G.A. *et al.* (1993) Distribution of bacteria within operating laboratory water purification systems. *Appl. Environ. Microbiol.* 59, 1410–1415
34 Nogami, T. *et al.* (1998) Estimation of bacterial contamination in ultrapure water: application of the anti-DNA antibody. *Anal. Chem.* 70, 5296–5301
35 McAlister, M.B. *et al.* (2002) Survival and nutritional requirements of three bacteria isolated from ultrapure water. *J. Ind. Microbiol. Biotechnol.* 29, 75–82
36 Shen, H. *et al.* (2006) Sensitive, real-time PCR detects low-levels of contamination by Legionella pneumophila in commercial reagents. *Mol. Cell. Probes* 20, 147–153
37 Seitz, V. *et al.* (2015) A new method to prevent carry-over contaminations in two-step PCR NGS library preparations. *Nucleic Acids Res.* 43, e135
38 Ballenghien, M. *et al.* (2017) Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC Biol.* 15, 25
39 Weyrich, L.S. *et al.* (2018) Laboratory contamination over time during low-biomass sample analysis. *In Review*

40 Nguyen, N.H. *et al.* (2015) Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytol.* 205, 1389–1393
41 Tamariz, J. *et al.* (2006) The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for the amplification of low copy number DNA. *J. Forensic Sci.* 51, 790–794
42 Joung, Y.S. *et al.* (2017) Bioaerosol generation by raindrops on soil. *Nat. Commun.* 8, 14668
43 Carlsen, T. *et al.* (2012) Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol.* 5, 747–749
44 Illumina Inc. (2017) Effects of Index Misassignment on Multiplexing and Downstream Analysis. Publication No. 770-2017-004-C QB # 5420, 4
45 Larsson, A.J.M. *et al.* (2018) Computational correction of index switching in multiplexed sequencing libraries. *Nat. Methods* 15, 305–307
46 Eisenhofer, R. *et al.* (2017) Reply to Santiago-Rodriguez et al.: proper authentication of ancient DNA is essential. *FEMS Microbiol. Ecol.* 93,
47 Aagaard, K. *et al.* (2014) The placenta harbors a unique microbiome. *Sci. Transl. Med.* 6, 237ra65
48 Kliman, H.J. (2014) Comment on "The placenta harbors a unique microbiome." *Sci. Transl. Med.* 6, 254le4-254le4
49 Perez-Muñoz, M.E. *et al.* (2017) A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome* 5, 48
50 Antony, K.M. *et al.* (2015) The preterm placental microbiome varies in association with excess maternal gestational weight gain. *Am. J. Obstet. Gynecol.* 212, 653.e1-653.e16
51 Zheng, J. *et al.* (2015) The Placental Microbiome Varies in Association with Low Birth Weight in Full-Term Neonates. *Nutrients* 7, 6924–6937
52 Amarasekara, R. *et al.* (2015) Microbiome of the placenta in pre-eclampsia supports the role of bacteria in the multifactorial cause of pre-eclampsia. *J. Obstet. Gynaecol. Res.* 41, 662–669
53 Bassols, J. *et al.* (2016) Gestational diabetes is associated with changes in placental microbiota and microbiome. *Pediatr. Res.* 80, 777–784
54 Branton, W.G. *et al.* (2013) Brain Microbial Populations in HIV/AIDS: α-Proteobacteria Predominate Independent of Host Immune Status. *PLOS ONE* 8, e54673
55 Xuan, C. *et al.* (2014) Microbial Dysbiosis Is Associated with Human Breast Cancer. *PLOS ONE* 9, e83744
56 Hieken, T.J. *et al.* (2016) The Microbiome of Aseptically Collected Human Breast Tissue in Benign and Malignant Disease. *Sci. Rep.* 6, 30751
57 Chan, A.A. *et al.* (2016) Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Sci. Rep.* 6, 28061
58 Fang, R.-L. *et al.* (2016) Barcoded sequencing reveals diverse intrauterine microbiomes in patients suffering with endometrial polyps. *Am. J. Transl. Res.* 8, 1581–1592
59 Javurek, A.B. *et al.* (2016) Discovery of a Novel Seminal Fluid Microbiome and Influence of Estrogen Receptor Alpha Genetic Status. *Sci. Rep.* 6,
60 Vaishampayan, P. *et al.* (2013) New perspectives on viable microbial communities in low-biomass cleanroom environments. *ISME J.* 7, 312–324
61 Champlot, S. *et al.* (2010) An Efficient Multistrategy DNA Decontamination Procedure of PCR Reagents for Hypersensitive PCR Applications. *PLoS ONE* 5, e13042

62 Woyke, T. *et al.* (2011) Decontamination of MDA reagents for single cell whole genome amplification. *PloS One* 6, e26161
63 Cooper, A. and Poinar, H.N. (2000) Ancient DNA: Do It Right or Not at All. *Science* 289, 1139–1139
64 Rouzic, E.L. (2006) Contamination-pipetting: relative efficiency of filter tips compared to Microman|[reg]| positive displacement pipette. *Nat. Methods Appl. Notes* DOI: 10.1038/nmeth887
65 Meyer, M. and Kircher, M. (2010) Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.* 2010, pdb.prot5448
66 Costello, M. *et al.* (2018) Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* 19,
67 Illumina Inc. (2013) Reducing Run-to-Run Carryover on the MiSeq Using Dilute Sodium Hypochlorite Solution. DOI: https://doi.org/10.25909/5b85eb0b95552
68 Minich, J.J. *et al.* (2018) KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples. *mSystems* 3, e00218-17
69 Xu, Z. *et al.* (2009) Improving the sensitivity of negative controls in ancient DNA extractions. *ELECTROPHORESIS* 30, 1282–1285
70 Jervis-Bardy, J. *et al.* (2015) Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome* 3, 19
71 Ozkan, J. *et al.* (2017) Temporal Stability and Composition of the Ocular Surface Microbiome. *Sci. Rep.* 7,
72 Davis, N.M. *et al.* (2018) Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. DOI: 10.1101/221499
73 Karstens, L. *et al.* (2018) Controlling for contaminants in low biomass 16S rRNA gene sequencing experiments. DOI: 10.1101/329854
74 Russell, J.H. and Keiler, K.C. (2012) RNA Visualization in Bacteria by Fluorescence In Situ Hybridization. In *Bacterial Regulatory RNA* pp. 87–95, Humana Press, Totowa, NJ
75 Kostic, A.D. *et al.* (2012) Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* 22, 292–298
76 Weiss, S. *et al.* (2014) Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol.* 15, 564
77 Cano, R.J. *et al.* (1993) Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature* 363, 536–538
78 Woodward *et al.* (1994) DNA sequence from Cretaceous period bone fragments. *Science* 266, 1229–1232
79 Cano, R.J. and Borucki, M.K. (1995) Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. *Science* 268, 1060–1064
80 Allentoft, M.E. *et al.* (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. Lond. B Biol. Sci.* 279, 4724–4733
81 Austin, J.J. *et al.* (1997) Problems of reproducibility – does geologically ancient DNA survive in amber–preserved insects? *Proc. R. Soc. Lond. B Biol. Sci.* 264, 467–474
82 Hedges, S.B. and Schweitzer, M.H. (1995) Detecting dinosaur DNA. *Science* 268, 1191–1192
83 Henikoff, S. (1995) Detecting dinosaur DNA. *Science* 268, 1192–1192

84 Zischler, H. *et al.* (1995) Detecting dinosaur DNA. *Science* 268, 1192–1193

85 Yousten, A.A. and Rippere, K.E. (1997) DNA similarity analysis of a putative ancient bacterial isolate obtained from amber. *FEMS Microbiol. Lett.* 152, 345–347

86 Marchesi, J.R. and Ravel, J. (2015) The vocabulary of microbiome research: a proposal. *Microbiome* 3, 31