

REVIEW

Open Access

DNA extraction for human microbiome studies: the issue of standardization



K. Leigh Greathouse^{1*}, Rashmi Sinha² and Emily Vogtmann²

Abstract

Among the laboratory and bioinformatic processing steps for human microbiome studies, a lack of consistency in DNA extraction methodologies is hindering the ability to compare results between studies and sometimes leading to errant conclusions. The purpose of this article is to highlight the issues related to DNA extraction methods and to suggest minimum standard requirements that should be followed to ensure consistency and reproducibility.

Introduction

Standardization in microbiome analyses is fundamental to reliably measure the human microbiome. Variability in results can stem from all the steps in the human microbiome study process including sample collection, DNA extraction, sequencing, and bioinformatics. Of these, DNA extraction was identified by the MicroBiome Quality Control project (MBQC) [1], the International Human Microbiome Standards (IHMS) group [2], and others as contributing a majority of experimental variability. Thus, while each aforementioned aspect is important, in this commentary, we will only cover issues related to DNA extraction from different human sample types and the inclusion of appropriate reference material for human microbiome research.

DNA extraction from fecal samples

A variety of protocols are available to extract microbial DNA from human samples, but the literature has predominantly focused on protocols for extraction from fecal samples. The Human Microbiome Project (HMP) [3], MetaHIT [4], and the Earth Microbiome Project [5] and many other groups have published DNA extraction protocols for feces, as well as, other sample types.

Both the MBQC and IHMS evaluated the impact of different DNA extraction protocols for fecal samples, and they found that DNA extraction protocols had the largest impact on experimental variability [1, 2]. The source of variability between DNA extraction methods could be

related to multiple factors including reagent contamination, type of lysis (e.g., mechanical or enzymatic), differences between laboratory personnel or automation of DNA extraction, among many other factors. The IHMS study also proposed a DNA extraction protocol that maximized ease of use and reproducibility [2], although currently this method has not been automated. In a study directly comparing the HMP and MetaHIT extraction methods for fecal samples, the MetaHIT protocol yielded a higher number of read mapping to eukaryotic genomes while the HMP protocol had a greater number of reads mapping to bacterial genomes. Even within the analysis of bacteria, the two methods detected differing abundances of specific genera [6]. These studies indicate that a standardized protocol for DNA extraction from fecal samples may be ideal to decrease variability due to DNA extraction, although using standardized protocols will not prevent other inter-laboratory differences. Even though multiple potential standard protocols exist for fecal samples, the selection of a standardized protocol for all studies is still extremely difficult due to many factors, including a need for automation of DNA extraction methods for large sample sizes and a lack of understanding of the true composition of a fecal sample to serve as the ground truth for comparisons between methods. Inclusion of quality control samples as discussed below would help address some of these issues, but more studies to evaluate DNA extraction protocols for fecal samples are needed.

While some standard DNA extraction methods have been proposed for fecal samples, these methods are not always appropriate for other human sample types. For example, oral samples contain higher concentrations of human DNA compared to fecal samples [7, 8], which

* Correspondence: Leigh_Greathouse@baylor.edu

¹Nutrition Sciences, Robbins College of Health and Human Sciences, Baylor University, Waco, TX, USA

Full list of author information is available at the end of the article



can affect DNA extraction and downstream procedures. Methods are available to deplete host DNA during extraction [9]; however, these methods are quite costly and may also bias DNA extraction and affect the resulting microbial community. Overall, while protocols for DNA extraction have been studied for fecal samples, care should be taken when adapting these methods for other sample types, especially non-fecal and low-biomass samples.

DNA extraction from samples with low microbial biomass

Several human microbiome studies have also focused on samples with lower microbial biomass, such as human tissue samples and spinal and joint fluids. While samples with high microbial DNA biomass, such as fecal samples, are less subject to biases or false positives due to contamination during processing, low biomass samples are highly subject to contamination from exogenous sources [10, 11]. Two complementary methods are required to avoid most but not all significant errant results: (1) reduction of possible contaminants and (2) demonstration of the existence of microbial life beyond sequencing.

As it relates to contamination in low biomass samples, several previous studies have identified common contaminants emanating from DNA extraction kits and other laboratory reagents [11, 12], but beyond this list, researchers must carefully evaluate all possible sources of contamination at each stage of the study, ranging from sample collection to laboratory processing. As an example, cancer-associated microbiome studies using low biomass tissue such as lung tissue should start by including numerous samples of possible contamination and minimizing contamination at the source. In this scenario, one would begin with the origin of the sample tissue collection location, identify all possible environmental sources of contamination, and collect swabs of air, surgical tools, gloves, collection tubes, patient skin, and other samples, until all points of contacts have been identified for this tissue being collected immediately prior to storage or freezing. It is also ideal to collect low biomass samples in an environment designed to minimize as much contamination as possible. This can be established, for example, by creating a sterile pathology suite next to the operating room that is optimized for microbiome studies, including staff trained in sterile techniques. Specific protocols for DNA extraction from low biomass samples have been proposed to decrease contamination during extraction [13, 14], and a checklist has been recently proposed to guide microbiome studies of samples with low biomass to help minimize and identify contamination [10].

In order to demonstrate “proof-of-life” in low biomass samples, especially those that have historically been considered “sterile,” the use of both microbial culture and fluorescent in situ hybridization (FISH) is typically recommended.

While it is likely possible that a microbe of interest, especially anaerobes, may not survive outside of the sample environment, utilizing RNA probes for FISH in a stabilized sample or culturing of specific microbes, such as *Fusobacterium*, should yield supportive evidence as to the existence of metabolically active microbes or polymicrobial communities [15, 16].

DNA extraction of non-bacterial microbes

The majority of DNA extraction protocols have focused on applications for 16S rRNA gene sequencing or whole genome shotgun metagenomic sequencing in fecal samples. Since the microbes in fecal samples are primarily comprised of bacteria [4], most of the protocols have been optimized for bacterial DNA extraction and may yield biased results for other microbes such as fungi, protists, and viruses. Some studies have evaluated the impact of DNA extraction methods on fungal communities [17, 18], but larger-scale, multi-laboratory studies are needed to evaluate optimal DNA extraction methods for studies of human fungal communities. For studies of viruses, specific DNA extraction methods have been proposed to isolate viral particles and remove contamination by human and microbial cells [19]. However, these methods are focused on solely investigating viruses and therefore remove the other microbes, which indicates a need for a viral DNA extraction protocol to complement DNA extraction for other microbes or a single optimized DNA extraction protocol for all microbes.

Use of quality control samples during DNA extraction

Due to a lack of standard reference materials for human microbiome studies, quality control samples are not regularly included in microbiome datasets, which can lead to the inability to compare or reproduce results [20, 21]. This can be particularly important when comparing studies utilizing differing DNA extraction methods. Microbiome analyses, at present, can typically use three types of positive reference material: (1) complex environmental samples (e.g., feces from a generous donor) [1, 22], (2) chemostat communities from in vitro microbial model systems [1, 23], and (3) artificial or mock communities [1]. Each type of reference material comes with pros and cons. An advantage of using complex environmental samples is that its composition will closely match the sample of interest which is particularly important for the microbial representation and abundances, in addition to the same environmental matrix. However, the ecosystem is complex with many low-abundance taxa and may be difficult to reproduce if more material is needed later. The advantage of a chemostat is that they approach the complexity of environmental samples and can produce large quantity of reference material, but the supply is still finite as chemostat batches cannot be

reproduced at a later time, and some specific taxa may be lost during sample creation. Pure microbial strains and mock communities can be created with known quantity of different bacteria, but they lack the complexity of the environmental samples, are not in the same environmental matrix as the human sample, and only contain culturable microbes. Complex environmental samples or chemostat communities can be used to evaluate reproducibility or consistency of DNA extraction, while artificial or mock community samples can serve as a known ground truth for accuracy. While no reference material is ideal, currently using these reference materials is a good option for standardization within- and between-studies.

Apart from positive controls, negative controls, as described previously in reference to low-biomass samples, are equally critical [1, 20]. Including blanks for each kit or laboratory reagent used during sample handling is key in identification of potential contamination prior to sequencing. At a minimum, both blanks and background environmental contamination controls should be included and analyzed throughout the pipeline. While it is likely not all contamination can be eliminated, identifying and quantifying the contaminants is key to accurate interpretation of results and comparison between studies. Including both positive and negative controls will improve our ability to normalize across users, DNA extraction batches, between experiments, and across laboratories.

Conclusions and future directions

Ideally in human microbiome studies, we want to extract all or specific microbial DNA found in a sample in order to accurately characterize the microbiota in a sample. Given the lack of a known ground truth for the composition of complex samples, it is still not possible to guarantee validity of a DNA extraction method, but we should at least aim for both consistency and reliability to the greatest extent possible. A great deal of research has been conducted to identify optimal DNA extraction methods for some sample types, particularly fecal samples, but larger-scale studies for optimization of DNA extraction methods for other sample types and non-bacterial microbes are needed. In the meantime, we propose that there are at least three minimal standards that should be followed for human microbiome studies: (1) detailed reporting of DNA extraction methods employed in a study such that another laboratory can easily reproduce all DNA extraction procedures; (2) inclusion and reporting of positive and negative controls in all DNA extraction batches with reporting of coefficients of variation or intraclass correlation coefficients for the positive controls, as well as, information regarding contamination in any blank samples; and (3) utilization of the same DNA extraction protocol across studies for institutions or multi-site studies that plan to

Table 1 Recommendations for improved reliability and consistency in human microbiome studies

#1	Collect and process negative controls at each of the following points: during sample collection, DNA extraction, and sequencing
#2	Include at least one or more of the following three types of positive controls in each experiment, depending on study: a complex environmental sample, chemostat community, or mock community
#3	Report in detail the DNA extraction methods followed and sufficient information regarding the results from both positive and negative controls in a manner that allows for peer-review and reproducibility
#4	Utilize the same DNA extraction protocol across studies for multi-institute or multi-site studies
#5	Demonstrate “proof-of-life” beyond sequencing in low-biomass studies using microbial culture and/or FISH

pool data in the future (Table 1). Towards this effort, journals have the responsibility to ensure that these aforementioned minimal standards are met prior to publication, as has been implemented in other journals as standard reporting guidelines (e.g., STAR methods in the journal *Cell*). It may never be possible to have a standardized protocol for all human microbiome studies, but scientists, reviewers, and journal editors should strive to improve our current protocols and reporting criteria to increase reliability and consistency between laboratories and studies.

Acknowledgements

The authors appreciate the editor’s request of this important review article and the suggestions for key topics for discussion.

Authors’ contributions

All authors contributed equally to the writing and editing of this manuscript. All authors read and approved the final manuscript.

Funding

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Nutrition Sciences, Robbins College of Health and Human Sciences, Baylor University, Waco, TX, USA. ²Metabolic Epidemiology Branch, Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

Received: 26 June 2019 Accepted: 30 September 2019

Published online: 21 October 2019

References

- Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat Biotechnol*. 2017;35:1077–86.
- Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Herczeg R, Jung FE, et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol*. 2017;35:1069–76.
- The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207–14.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464:59–65.

5. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined metagenomics of diverse environmental samples. *Biotechniques*. 2017;62:290–3.
6. Wesolowska-Andersen A, Bahl MI, Carvalho V, Kristiansen K, Sicheritz-Ponten T, Gupta R, Licht TR. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome*. 2014;2:19.
7. Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, et al. Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A*. 2014;111:E2329–38.
8. Horz HP, Scheer S, Huenger F, Vianna ME, Conrads G. Selective isolation of bacterial DNA from human clinical specimens. *J Microbiol Methods*. 2008;72:98–102.
9. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*. 2018;6:42.
10. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol*. 2019;27:105–17.
11. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87.
12. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog*. 2016;8:24.
13. Minich JJ, Zhu Q, Janssen S, Hendrickson R, Amir A, Vetter R, Hyde J, Doty MM, Stillwell K, Benardini J, et al. KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. *mSystems*. 2018;3.
14. Morales E, Chen J, Greathouse KL. Compositional analysis of the human microbiome in cancer research. *Methods Mol Biol*. 1928;2019:299–335.
15. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, Peterson SN, Snesrud EC, Borisy GG, Lazarev M, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc Natl Acad Sci U S A*. 2014;111:18321–6.
16. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, Neuberger D, Huang K, Guevara F, Nelson T, et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science*. 2017;358:1443–8.
17. Huseyin CE, Rubio RC, O'Sullivan O, Cotter PD, Scanlan PD. The fungal frontier: a comparative analysis of methods used in the study of the human gut mycobiome. *Front Microbiol*. 2017;8:1432.
18. Rosenbaum J, Usyk M, Chen Z, Zolnik CP, Jones HE, Waldron L, Dowd JB, Thorpe LE, Burk RD. Evaluation of oral cavity DNA extraction methods on bacterial and fungal microbiota. *Sci Rep*. 2019;9:1531.
19. Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, McDonnell SA, Nolan JA, Sutton TDS, Dalmasso M, et al. Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome*. 2018;6:68.
20. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, et al. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*. 2017;5:52.
21. Hornung BVH, Zwitterink RD, Kuijper EJ. Issues and current standards of controls in microbiome research. *FEMS Microbiol Ecol*. 2019;95. <https://doi.org/10.1093/femsec/fiz045>
22. Sinha R, Ahsan H, Blaser M, Caporaso JG, Carmical JR, Chan AT, Fodor A, Gail MH, Harris CC, Helzlsouer K, et al: Next steps in studying the human microbiome and health in prospective studies, Bethesda, MD, May 16-17, 2017. *Microbiome* 2018, 6:210.
23. Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. *Genome Biol*. 2015;16:276.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.