Choice of DNA extraction method affects stool microbiome recovery and subsequent phenotypic association analyses

Asier Fernández-Pato (asierfp1996@gmail.com)
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0001-6361-290X

Trishla Sinha
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0002-0992-7983

Ranko Gacesa
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0003-2119-0539

Milla F. Brandao Gois
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0001-7918-2481

Jody Gelderloos-Arends
University of Groningen, University Medical Center Groningen

Dianne B.H. Jansen
University of Groningen, University Medical Center Groningen

Martin Jaeger
Radboud University Medical Center https://orcid.org/0000-0001-6994-493X

Leo A.B. Joosten
Radboud University Medical Center and Iuliu Hatieganu University of Medicine and Pharmacy

Mihai G. Netea
Radboud University Medical Center and University of Bonn https://orcid.org/0000-0003-2421-6052

Rinse K. Weersma
University of Groningen, University Medical Center Groningen

Cisca Wijmenga
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0002-5635-1614

Hermie J.M. Harmsen
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0003-2725-6148

Jingyuan Fu
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0001-5578-1236

Alexandra Zhernakova
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0002-4574-0841

Alexander Kurilshikov
University of Groningen, University Medical Center Groningen https://orcid.org/0000-0003-2541-5627

Research Article

Keywords: gut microbiome, fecal sample, DNA extraction, shotgun metagenomics

Posted Date: August 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1967940/v1

Abstract

Background: Interest in microbiome research has increased exponentially in recent years. However, the growth in the number of studies has outpaced the standardization of the processing and analysis of microbiome samples. This lack of standardization represents a major limitation that hampers the replication of results across studies and the clinical translation of research findings. The major source of variation in microbiome results on the experimental side are differences in the methods of DNA extraction from fecal samples. In this study, we aimed to compare the metagenomic profiles obtained by using two commercially available DNA extraction kits, and their effects on microbiome diversity, composition and associations to phenotypes.

Methods and Results: We compared two commonly used DNA extraction kits, the AllPrep DNA/RNA Mini Kit (APK) and the QIAamp Fast DNA Stool Mini Kit (FSK), in 745 paired samples from two independent population cohorts: Lifelines-DEEP (LLD, n = 292), and 500 Functional Genomics project (500FG, n = 453). We evaluated the performance of both methods for DNA yield and quality and explored whether the DNA extraction protocol introduces heterogeneity in microbiota composition and diversity or in phenotype–microbiome associations. In both cohorts, APK protocol yields a higher DNA concentration and alpha diversity, with 25% and 10% more bacterial species being detected in comparison to the FSK method in LLD and 500FG cohorts, respectively. Both extraction kits result in markedly different community composition and microbial abundances, with >80% of species being differentially abundant in both cohorts. Species belonging to Firmicutes and Actinobacteria show increased abundances in the APK protocol, whereas Bacteroidetes and Proteobacteria are more prevalent in FSK samples. These differences lead to significant variations in the phenotypic association profile with gut microbes.

Conclusions: The results of this study further reinforce that choice of DNA extraction method impacts metagenomic profile of human gut microbiota. We demonstrate that accounting for differences in fecal sample processing is essential for improving the reproducibility of microbiome research findings.

Introduction

The human gut harbors diverse microbes that play a fundamental role in the well-being of their host. Numerous studies have identified gut microbiome activities that range from immune functions and protection against pathogens to roles in human metabolism, nutrition and brain function [1–4]. Recently, there has been a focus on gut microbiome research because of its role in influencing the development of numerous diseases [5–6]. The composition of the microbes in the gut has largely been evaluated using fecal samples, which are easy and non-invasive to obtain. However, many parameters of fecal sampling and processing affect the composition of the gut microbiota, including sample collection method, storage (solution composition, storage temperature), homogenization method for DNA extraction (e.g.

mechanical or enzymatic lysis), choice of PCR primers (in the case of 16S rRNA analysis) and sequencing method [7, 8].

Isolation and purification of DNA from fecal samples is a crucial initial step to ensure high yield and quality of isolated nucleic acids. Over the past decades, several DNA extraction kits have been developed that are now commercially available, with the goal of enabling rapid extraction of large numbers of fecal samples. Available kits differ in several steps of the extraction process, including the cell lysis procedure and DNA isolation and purification methods. Earlier research showed that different DNA extraction kits yield different results in terms of the amount and quality of DNA extracted and the bacterial community composition [9, 10], with the inclusion of a mechanical lysis step demonstrated to have a positive effect on the recovery of DNA from gram-positive bacteria [10–12]. This heterogeneity between methods causes substantial technical variation among cohorts, an issue that represents a major challenge in multicenter microbiome studies, and it may be partially responsible for the low replication rate in microbiome research [9, 13, 14]. In addition, the effects of DNA extraction methods on the microbial composition determined by shotgun metagenomic sequencing have not been extensively explored in large cohorts. In this study, we investigated the impact of two fecal DNA extraction methods on the yield and quality of isolated DNA and microbiome composition and diversity. We also assessed the implications of these effects for microbiome–phenotype associations. For this purpose, we applied two commercially available DNA extraction kits to fecal samples from two independent cohorts of 292 and 453 participants (745 pairs of samples in total) in which other sources of heterogeneity were reduced as much as possible, followed by shotgun metagenomic sequencing.

Materials And Methods

Study population

Our study included fecal samples from two Dutch population cohorts: Lifelines-DEEP follow-up (LLD, a subcohort of the Lifelines biobank, Groningen, the Netherlands) and 500 Functional Genomics (500FG, Nijmegen, the Netherlands) [15, 16]. LLD comprises 338 individuals (55.6% female, 44.4% male, mean age: 51.7 years) sampled in 2017, 4–5 years after the original Lifelines-DEEP study [17]. The 500FG cohort consists of 534 healthy individuals (44.4% male, 55.6% female, mean age: 28.5 years), with stool samples taken between July 2013 and December 2014 available for 471 participants. Both cohorts combine deep phenotypic characterization with extensive collection of biological samples, including stool samples for metagenomic analysis. For the present study, we selected 292 fecal samples from LLD and 453 samples from 500FG that had all been isolated using two different DNA extraction protocols (Fig. 1).

Sample collection and DNA isolation

All stool samples included in this study were collected by participants at home and frozen within 15 minutes after production. Frozen samples were then collected by qualified personnel, transported on dry ice and stored at -80°C until DNA extraction.

We compared two commercially available DNA extraction methods used in the human microbiome research field: the AllPrep DNA/RNA Mini Kit (QIAGEN, Germany) and the QIAamp Fast DNA Stool Mini Kit (QIAGEN, Germany), hereafter referred to as APK and FSK, respectively. Samples from both cohorts were extracted shortly after sample collection using the exact same procedure, except for the 500FG samples processed with FSK, which were stored frozen for 4 years and the isolation was performed in a different laboratory. For both extraction kits, DNA extraction was performed according to the manufacturer's instructions with the following deviations: 1) we performed an enzymatic lysis (with TE buffer containing lysozyme and proteinase K) and a bead-beating step in the APK-based protocol and 2) the FSK protocol was automated using the QIAcube, increasing the temperature for the cell lysis step from 70°C to 95°C (See Additional File 1 for detailed methods).

Assessment of DNA yield and purity

The concentration of DNA extracted from LLD samples was measured spectrophotometrically using a NanoDrop ND-1000 (Thermo Fisher Scientific, United States). The DNA concentrations of samples from the 500FG cohort were assessed differently according to the DNA extraction method used. Samples extracted with the FSK protocol were evaluated with the NanoDrop ND-1000. Samples processed with the APK method were analyzed via Quant-iT PicoGreen dsDNA Assay (Thermo Fisher Scientific). In both cohorts, DNA purity was assessed with the NanoDrop ND-1000 and calculated as the 260/280 absorbance ratio.

Library preparation, sequencing and microbiome profiling

Following DNA extraction, differentially processed samples from both cohorts were sent to two alternative sequencing facilities (APK: Broad Institute of Harvard and MIT, United States, FSK: Novogene, China) to perform library preparation (FSK: NEBNext® DNA Library Prep Kit (New England BioLabs, United States), APK: Nextera® XT DNA Library Preparation Kit (Illumina, United States)) and whole-genome shotgun sequencing on the Illumina HiSeq 2000 platform. Following the filtration of low-quality reads at the sequencing facilities, we removed raw metagenomic reads that mapped against the reference human genome (GRCh37) or aligned to Illumina adapters using KneadData (v 0.7.4). The remaining reads were used as input to determine the taxonomic profiles of the samples using the Biobakery3 tool MetaPhlAn3 (v 3.0.7) [18].

Statistical analysis

Statistical differences in read depth, DNA concentration and purity measures were assessed using the Wilcoxon signed-rank test. The same statistical test was applied to determine microbial richness and diversity differences estimated with R package mia (v. 1.2.7) [19] (based on the vegan package) [20]. We then estimated the Pearson correlation of 1) alpha diversity and read depth and 2) alpha diversity between paired samples extracted with the different protocols. Differences in microbial community composition were analyzed and visualized via principal coordinate analysis (PCoA) based on Aitchison distances. The contribution of DNA extraction protocol and anthropometric variables to inter-individual microbiome community variation on species-level was determined by a PERMANOVA analysis using the

adonis function. A Mantel test was applied to evaluate the correlation of the distance matrices of samples extracted with each DNA extraction method. Differences in abundance of microbial taxa between paired samples were analyzed at phylum- and species-level using a paired t-test on centered log ratio (CLR)-transformed counts. We also assessed Pearson's correlations of log-transformed microbial abundances between APK and FSK. Finally, we determined the associations of the available metadata with microbial diversity measures (Pearson's correlation and PERMANOVA analysis for alpha and beta diversity, respectively) and abundances at species-level (linear regression). Only species present at a prevalence > 10% were considered in the association analysis. Unless otherwise specified, results were considered significant at a false discovery rate (FDR)-corrected p-value < 0.05. All analyses were done using the R programming language (v.4.1.2) (https://www.R-project.org/).

Results

Assessment of differences in read depth, DNA concentration and purity

We compared the metagenomic sequencing data from paired fecal samples in LLD and 500FG cohorts, where DNA from the same sample was isolated using two different methods – APK and FSK. Since sequencing depth can critically affect the microbiome profiling, we first evaluated the read depth values in differentially extracted samples from each cohort. Metagenomic sequencing yielded a significantly different mean read count in the 500FG cohort. After removal of adapter sequences and human reads, 500FG samples processed with the FSK protocol had a significantly higher read depth compared to samples extracted with APK (mean ± SD (million): APK: 11.59 ± 3.64, FSK: 25.19 ± 5.51, paired Wilcoxon test, $p < 0.001$, Fig. S1A). By contrast, we saw no significant differences between protocols in the LLD cohort (mean ± SD (million): APK: 13.33 ± 4.75, FSK: 12.91 ± 2.40, paired Wilcoxon test, $p = 0.332$, Fig. S1A).

Next, we compared DNA concentration and purity between paired samples. LLD samples showed significantly higher DNA concentrations when processed with the APK protocol in comparison to the FSK protocol (mean ± SD: APK: 205.20 ± 73.33 ng/µl, FSK: 64.91 ± 35.70 ng/µl, $p < 0.001$, Fig. S1B). Although we saw similar results for the 500FG cohort (mean ± SD: APK: 179.26 ± 87.80 ng/µl, FSK: 106.64 ± 54.92 ng/µl, $p < 0.001$, Fig. S1B), the use of different measurement methods for DNA concentration prevents direct comparison of these results. Regarding DNA quality, the 260/280 absorbance values were significantly different in LLD, where APK-extracted samples showed a ratio closer to 1.8, while FSK samples displayed a lower DNA purity and a higher standard deviation (mean ± SD: APK: 1.89 ± 0.05, FSK 1.99 ± 0.13, $p < 0.001$, Fig. S1C). Conversely, no significant differences in mean DNA quality were found for 500FG samples (mean ± SD: protocol APK: 1.90 ± 0.06, protocol FSK: 1.91 ± 0.11, $p = 0.113$, Fig. S1C). Table S1 provides detailed information on read number, DNA concentration and purity.

DNA extraction method significantly affects gut microbial diversity

Species richness analysis highlighted significant differences between the APK and FSK protocols in the number of species observed in both LLD (mean ± SD (species/sample): APK: 87.91 ± 13.95, FSK 68.25 ± 12.17, $p < 0.001$, Fig. 2A) and 500FG samples (APK: 78.93 ± 12.43, FSK 73.58 ± 14.16, $p < 0.001$, Fig. 2A), with APK yielding a higher number of species. Similarly, both the Shannon and Inverse Simpson diversity indices showed significantly higher diversities in APK-extracted samples in both LLD and 500FG ($p < 0.001$, Fig. 2B–C). All diversity values are summarized in Table S2. To disentangle the possible influence of differences in read depth on observed diversity, we analyzed the correlation between the alpha diversity indices and read depth, but this did not yield significant results (Fig. S2). Remarkably, despite these differences, we found a significant positive correlation of species richness and alpha diversity values between paired samples extracted with APK and FSK protocols in each cohort ($p < 0.001$, Fig. 2D–F), with the only exception being the Inverse Simpson diversity values, which did not reach statistical significance in the LLD samples ($p = 0.15$, Fig. 2F).

Microbial community composition reflects differences driven by DNA isolation protocol

We assessed the differences in observed microbial communities between samples extracted with each protocol. PCoA plots based on Aitchison distances showed clear clustering according to the DNA extraction method (Fig. 3A–B), with unpaired samples collected from different participants showing smaller distances that paired samples from the same individuals. PERMANOVA analysis confirmed the significant effect of DNA isolation protocol on microbial community composition (species-level Aitchison distance, LLD: $R^2 = 10.48\%$, 500FG: $R^2 = 7.86\%$, $p < 0.001$, permutations = 1000). The dissimilarities between both groups were mainly driven by differences in bacteria with alternative Gram-staining based classifications: gram-positive bacteria appeared to be enriched in samples extracted with the APK protocol, whereas gram-negative bacteria appeared to be enriched in FSK samples (Fig. 3C–D). Although we observed significant effects of other phenotypes including sex, age and body mass index (BMI) in shaping microbial community structure, their effect sizes were considerably smaller (Fig. 3E, Table S3).

Additionally, we evaluated the correlation of the distance matrices of samples extracted with the APK and FSK methods separately for each cohort. This analysis yielded a significant correlation between the distances of differentially extracted LLD and 500FG samples, indicative of a limited sensitivity of inter-individual distances to the heterogeneity introduced by the DNA extraction method (LLD: $r_{mantel} = 0.74$, $p < 0.001$, 500FG: $r_{mantel} = 0.84$, $p < 0.001$, permutations = 1000, Fig. 3F). These results were then further confirmed using the Bray-Curtis dissimilarity index (Fig S3, Table S3).

DNA extraction method has a large effect on observed microbial abundances

We evaluated the abundance of microbial taxa at several taxonomic levels. Phylum-level analysis pinpointed a significant difference between the APK and FSK protocols in the abundances of all microbial phyla identified, with the only exception being Verrucomicrobia in 500FG (paired t-test, $q < 0.05$). For

instance, we observed higher abundances of Firmicutes and Actinobacteria in samples extracted with APK, whereas samples extracted with FSK showed an increased abundance of Bacteroidetes and Proteobacteria phyla, among others. These observations also showed high consistency between LLD and 500FG cohorts (Fig. 4A, Table S4).

Analysis at higher taxonomic resolution revealed a higher number of both genera and species detected in > 1% of participants for the APK protocol in both LLD (APK: 130 genera, 371 species; FSK: 101 genera, 281 species) and 500FG (APK: 116 genera, 307 species; FSK: 106 genera, 279 species). Similar to what we observed at phylum-level, we found that species abundances were markedly different between the DNA extraction methods, with 283 and 260 species being differentially abundant in LLD and 500FG, respectively (paired t-test, $q < 0.05$, Table S5). Nonetheless, most species showed a positive and significant correlation between both extraction protocols (Table S6). When restricting the analysis to species with a relative abundance > 5% in each cohort, samples extracted with the FSK protocol displayed an increased presence of *Alistipes putredinis*, *Bacteroides uniformis* and *Prevotella copri*, whereas *Faecalibacterium prausnitzii* was more abundant in samples extracted with the APK protocol ($q < 0.05$, Fig. 4B–C). Again, however, the abundances of these species were positively correlated for both DNA extraction methods ($p < 0.001$, Fig. 4D).

To further characterize the differences between protocols, we performed a core taxa analysis separately in each cohort. This revealed a different pattern of core taxa between APK and FSK, with a higher number of microbial species reaching 50% prevalence in samples extracted with APK (LLD APK: 70 species, FSK: 56 species; 500FG APK: 63 species, FSK: 60 species). The alteration profile of core taxa according to relative abundance thresholds was also more alike between samples processed with the same kit, with FSK samples consistently displaying a smaller number of core taxa (Fig. S4).

DNA extraction procedure also alters phenotypic associations with microbial features

Further, we analyzed whether different DNA extraction protocols affect the observed phenotypic associations with microbial diversity measures and taxa abundances. For this, we assessed anthropometric phenotypes (age, sex and BMI) known to be linked to characteristic gut microbiome features and environmental exposures including dietary information and lifestyle habits. Interestingly, we found at nominal significance ($p < 0.05$) that all associations with alpha (Shannon Index) and beta diversity were conserved between both protocols in the 500FG cohort, except for consumption of vegetables, which was only associated to microbial composition in FSK samples (Tables S7, S8). Conversely, we observed larger differences in the LLD cohort. Significant correlations between alpha diversity and smoking habits, alcohol, meat, vegetables, and fruit consumption found with use of the APK protocol were not replicated in FSK-extracted samples, whereas an association with BMI observed in FSK-extracted samples was not significant in samples processed with APK. Despite not reaching statistical significance in both protocols, some phenotypes showed opposite direction in the association (smoking habits, alcohol consumption and BMI). Similarly, sex and cholesterol level were only found to be

associated with microbial community composition in samples extracted using the FSK protocol, while glucose level, physical activity and fruit consumption only showed significant results in APK samples (Table S8). Remarkably, we observed that correlations with bacterial species abundances were also influenced by DNA extraction method. Associations of *Intestinimonas butyriciproducens* and *Anaerotruncus colihominis* with glucose level in APK samples from LLD were not replicated in FSK samples (Tables 1, S9). In 500FG, the associations of *Lawsonibacter asaccharolyticus* with age and of *Collinsella stercoris* with BMI were preserved in both DNA extraction protocols, but other significant associations were either exclusive to APK-processed samples (*Barnesiella intestinihominis*, *Alistipes putredinis* and *Ruminococcus torques* with sex, *Firmicutes bacterium CAG:94* with age and *Eubacterium eligens* with fruit consumption) or to FSK-processed samples (*Prevotella copri* with sex, *Butyricimonas virosa* and *Dorea longicatena* with age and *Pseudoflavonifractor sp An184* with BMI) (Tables 2, S10).

Table 1
Significant associations found between microbial taxa abundances at species-level and available metadata for LLD cohort.

		APK		FSK	
Species	**Phenotype**	**q-value**	**R^2**	**q-value**	**R^2**
I. butyriciproducens	Glucose level	2.49E − 02	1.10E − 04	1	1.99E − 03
A. colihominis	Glucose level	3.74E − 02	4.73E − 03	1	3.80E − 04

Linear model FDR-adjusted p-values (q-value) and R-squared values are shown separately for each association in APK and FSK protocols.

Table 2
Significant associations identified between microbial taxa abundances at species-level and available metadata for 500FG cohort.

		APK		FSK	
Species	**Phenotype**	**q-value**	**R^2**	**q-value**	**R^2**
L. asaccharolyticus	Age	3.11E − 09	1.60E − 01	1.60E − 03	5.34E − 02
C. stercoris	BMI	4.42E − 03	5.02E − 02	2E − 02	4.39E − 02
B. intestinihominis	Sex	1.65E − 02	4.38E − 02	1	2.54E − 02
A. putredinis	Sex	4.12E − 04	5.89E − 02	2.77E − 01	3.22E − 02
R. torques	Sex	3.57E − 02	4.06E − 02	1	5.16E − 03
Firmicutes bact CAG:94	Age	1.15E − 02	4.53E − 02	1	2.25E − 02
E. eligens	Fruit consumption	2.19E − 03	5.74E − 02	1	1.33E − 02
P. copri	Sex	1	2.55E − 02	1.91E − 04	6.21E − 02
B. virosa	Age	5.21E − 02	3.90E − 02	2.62E − 02	4.19E − 02
D. longicatena	Age	1	1.27E − 02	1.49E − 02	4.42E − 02
Pseudoflavonifractor sp An184	BMI	1	1.82E − 02	7.11E − 03	4.82E − 02

Linear model FDR-adjusted p-values (q-value) and R-squared values are shown separately for each association in APK and FSK protocols.

Discussion

In this study, we compared two commercially available DNA extraction methods used in microbiome research: the AllPrep DNA/RNA Mini Kit (APK) and the QIAamp Fast DNA Stool Mini Kit (FSK). Using shotgun metagenomic sequencing, we assessed differences in DNA yield and quality and taxonomic composition while trying to limit the effects of other sources of heterogeneity. We also evaluated whether associations with several phenotypes were affected by the DNA extraction protocol.

The APK extraction method yielded a higher DNA concentration than FSK. Although only a few studies have compared these two protocols, this difference can be explained by the inclusion of a bead-beating step in the APK procedure. This mechanical disruption has been shown to improve DNA extraction efficiency, independent of the commercial kit used [9]. Although previous studies have described that use of a heating step in combination with the enzymatic lysis used in FSK can also favor bacterial cell lysis by denaturizing the membrane proteins [21], our findings suggest inefficient bacterial DNA recovery with this procedure. While automatization of the FSK protocol could have also contributed to DNA yield dissimilarities, prior research has not found significant differences in nucleic acid concentration and quality between automated and manual methods [12]. Additionally, our results show that the increased DNA concentration in samples extracted with APK resulted in higher microbial diversity and species richness. These findings further support earlier evidence that suggested a higher DNA yield and species diversity in bead-beaten samples [9, 11, 12], pinpointing alpha diversity as an appropriate benchmark for the DNA extraction performance. These results, however, could only be validated in the LLD cohort due to technical variability in the DNA concentration measurements in 500FG. In addition, we did not find a significant correlation between alpha diversity indices and read depth, discarding a potential impact of read depth differences on observed diversity values.

Our analysis revealed that DNA extraction method is a major driver of community differences and that its explanatory power (LLD: 10.48%, 500FG: 7.86%) was considerably higher than that of other anthropometric variables including sex (LLD: 0.34%, 500FG: 0.68%), age (LLD: 0.50%, 500FG: 0.67%) and BMI (LLD: 0.49%, 500FG: 0.35%). In contrast with earlier findings [9, 12, 22–25], paired samples isolated with different protocols showed lower similarity than unpaired samples extracted with the same protocol, as seen in the clustering on our PCoA. This discrepancy may be partially explained by the limited numbers of human participants included in the earlier comparative analyses (which ranged from one to 18 participants) compared to the 745 paired stool samples included in our analysis. These low numbers of participants substantially limited the ability to compare inter-subject and technical variation, thus hampering extrapolation of conclusions to large-scale analyses. Nevertheless, some studies have already reported results where the technical variability could be at least comparable to inter-individual variation at taxonomic [26] and functional level [9]. In addition, we found that the inter-individual distances had a limited sensitivity to the heterogeneity introduced by the DNA extraction method, as distance matrices of differentially extracted samples were positively correlated.

Several studies have previously described differences in taxonomic composition associated with DNA extraction protocol. However, they have mainly been restricted to genus-level taxonomic resolution due to the limitations of 16S rRNA gene amplicon–based analysis, the predominant method used in existing literature [8, 10, 12, 27, 28]. Evidence for substantial species-level heterogeneity in the human microbiome, and even for the presence of strain-specific phenotypes and functional profiles, highlights the need to gain deeper insights into lower taxonomic levels [29]. Therefore, in the present study, we report a massive alteration of species-level abundances in stool samples processed with APK and FSK, with a large proportion of the identified species being differentially abundant according to the DNA extraction protocol used. Interestingly, due to the compositional nature of microbiome data, protocol-dependent efficiencies

in the disruption of the cell walls of gram-positive bacteria resulted in a considerable fraction of the differentially abundant species being gram-negative bacteria. Nonetheless, our findings broadly support previous work showing an increased abundance of gram-positive bacteria after the inclusion of a mechanical lysis step. Remarkably, we also found that methodological differences in DNA isolation impacted phenotypic associations with diversity measures. For instance, DNA extraction with the APK method led to a loss of significance in the associations between microbial composition and sex, cholesterol level (LLD cohort) and consumption of vegetables (500FG cohort). In contrast, LLD samples extracted with this protocol showed a significant association of several lifestyle and dietary habits with microbial diversity (smoking, alcohol, meat, vegetables and fruit consumption) and composition (glucose level, physical activity and fruit consumption) that were not observed in FSK samples. Additionally, the contrasting DNA extraction efficiency of both protocols resulted in notable alterations in significant phenotypic associations with microbial taxa abundances. When focusing on prevalent bacterial species, we found that the differential abundances of *Prevotella copri* and *Alistipes putredinis* in samples extracted with both protocols (both species were more abundant in FSK) led to changes in the ability to identify associations with human phenotypes. Indeed, we observed a significant association between *P. copri* and sex only in FSK samples, whereas *A. putredinis* was associated to the same phenotype only in APK.

This study has several limitations. Despite our efforts to limit technical variability that could complicate the assessment of differences directly caused by DNA extraction protocol, our samples are subject to technical heterogeneity at several levels. Firstly, while the isolation of LLD samples was done at the same time and place for both protocols, FSK samples' isolation with the APK and FSK methods was done in different laboratories and four years apart. Although only a minor effect size of storage time of extracted DNA has been previously reported [30, 31], inter-laboratory differences have been shown to impact microbial profiles, while having a limited effect on relative diversity levels [13]. Due to the absence of technical replicates, we cannot evaluate the variation introduced by the cross-lab effect. Secondly, effectively assessing the impact of the bead-beating step would require a comparative analysis of both DNA extraction protocols with and without this additional step, so the design of the current study did not allow us to disentangle the effect of this step from that of the rest of the extraction procedure. Lastly, samples extracted with each protocol were sequenced in two different sequencing centers. While both centers applied the same sequencing technology, this difference could still represent another layer of technical variation, although previous studies have described sample sequencing to have a smaller impact than DNA extraction method [14].

Notwithstanding these limitations, our study expands upon previous findings that the DNA extraction procedure used has a large impact on the gut microbial diversity and structure recovered. To our knowledge, our analysis is the largest study to assess the impact of DNA extraction method and fecal sample processing on recovered gut microbiome profiles, and our considerable sample size overcomes the statistical power limitations of earlier comparative analyses. This, in combination with the use of shotgun metagenomics and the inclusion of the latest Metaphlan database release in our analysis, provides evidence of disruption of the species-level microbial profile by alternative DNA extraction

methods. Although we only tested the differences in stool samples, it is likely that sufficient detection power will make it possible to unravel a similar effect in other microbiome samples, including those with lower biomass. In addition, we have demonstrated how the technical variability effect translates into the phenotype association analysis, pinpointing the influence of DNA extraction methodology on biological conclusions. This finding may help explain the considerable heterogeneity and low replication rate found in microbiome studies to date, an issue that has greatly hampered clinical translation of microbiome research findings.

Conclusions

Altogether, our results highlight the need to account for the DNA extraction protocol used as a confounding factor in microbiome and exposome analyses. This is especially critical for multicenter studies and meta-analyses of data from multiple cohorts, an increasing focus of the microbiome research field.

Abbreviations

PCoA
Principal Coordinates Analysis
PERMANOVA
Permutational multivariate analysis of variance
CLR
Centered log-ratio
FDR
False discovery rate
BMI
Body mass index
q
FDR adjusted p-value
IQR
Interquartile range.

Declarations

Acknowledgements

The authors would like to thank all volunteers from the 500FG cohort and the Lifelines-DEEP participants, as well as the staff of the Lifelines study site and the contributing research centers delivering data to Lifelines for their collaboration. We thank Kate Mc Intyre for editing our manuscript. We also thank BioRender.com for providing the icons used in the Figure 1.

Authors' contributions

AF analyzed and interpreted the data and wrote the manuscript. TS, RG, MFBG and AK collaborated in the data analysis. JG and DBHJ performed the experimental data generation. TS, RG, AZ and AK contributed to review and edit the manuscript. AF, LABJ, MGN, CW, JF, AZ and AK were involved in the conception and design of the study. LABJ, MGN, RKW, CW, HJMH, JF and AZ contributed to the funding acquisition. JF, AZ and AK supervised the project and data analysis. All authors have read and approved the final manuscript.

Funding

HJMH, RKW and LABJ are supported by the collaborative TIMID project (LSHM18057-SGF) financed by the PPP allowance made available by Top Sector Life Sciences & Health to Samenwerkende Gezondheidsfondsen (SGF) to stimulate public–private partnerships and co-financed by health foundations that are part of the SGF. TS hold the scholarships from the Junior Scientific Masterclass, University of Groningen. AZ is supported by the European Research Council (ERC) Starting Grant 715772, the Netherlands Organization for Scientific Research (NWO) VIDI grant 016.178.056 and NWO Gravitation grant ExposomeNL 024.004.017. JF is supported by NWO Gravitation grant Netherlands Organ-on-Chip Initiative 024.003.001, ERC Consolidator grant 101001678 and NWO VICI grant VI.C.202.022. MGN, LABJ, JF and AZ are supported by the Netherlands Heart Foundation CVON grant 2018-27. RKW is supported by the Seerave Foundation and the Dutch Digestive Foundation (16-14). CW is supported by NWO Gravitation grant 024.003.001 and NWO Spinoza Prize SPI 92-266.

Availability of data and materials

The datasets supporting the conclusions of this article are available in A) LLD: the European Genome-Phenome Archive (EGA, https://ega-archive.org) via accession number EGAD00001006959, https://ega-archive.org/datasets/EGAD00001006959; B) 500FG: Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) via accession number PRJNA319574, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA319574. Due to informed consent regulations, phenotypic data of the Lifelines-DEEP cohort are available upon request to Lifelines (https://www.lifelines.nl/researcher). Datasets can be made available under a data transfer agreement, and the data usage access is subject to local rules and regulations.

Ethics approval and consent to participate

The Lifelines-DEEP study was approved by the ethics committee of the University Medical Center Groningen, document number METC UMCG LLDEEP: M12.113965. All participants signed an informed consent form prior to study enrollment. The 500FG study was approved by the Ethical Committee of Radboud University Nijmegen (NL42561.091.12, 2012/550). Inclusion of volunteers and experiments were conducted according to the principles expressed in the Declaration of Helsinki. All volunteers gave written informed consent before any material was taken.

Competing interests

The authors declare that they have no competing interests.

References

1. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun. 2019;10(1):4505.
2. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. BMJ. Jun 13 2018;361:k2179.
3. Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The Microbiota-Gut-Brain Axis. Physiol Rev. 2019;99(4):1877–2013.
4. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157(1):121–41.
5. Stiemsma LT, Michels KB. The role of the microbiome in the developmental origins of health and disease. Pediatrics. 2018;*141*(4):2017–2437.
6. Vijay A, Valdes AM. Role of the gut microbiome in chronic diseases: A narrative review. Eur. J. Clin. Nutr. 2022;*76*(4):489–501.
7. Wu WK, Chen CC, Panyod S, Chen RA, Wu MS, Sheen LY, et al. Optimization of fecal sample processing for microbiome study - The journey from bathroom to bench. J Formos Med Assoc. 2019;118(2):545–555.
8. Panek M, Cipcic Paljetak H, Baresic A, Peric M, Matijasic M, Lojkic I, et al. Methodology challenges in studying human gut microbiota - effects of collection, storage, DNA extraction and next generation sequencing technologies. Sci Rep. 2018;8(1):5143.
9. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol. 2017;35(11):1069–1076.
10. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PLoS One. 2012;7(3):e33865.
11. Zhang B, Brock M, Arana C, Dende C, van Oers NS, Hooper LV, et al. Impact of Bead-Beating Intensity on the Genus- and Species-Level Characterization of the Gut Microbiome Using Amplicon and Complete 16S rRNA Gene Sequencing. Front Cell Infect Microbiol. 2021;11:678522.
12. Lim MY, Song EJ, Kim SH, Lee J, Nam YD. Comparison of DNA extraction methods for human gut microbial community profiling. Syst Appl Microbiol. 2018;41(2):151–157.

13. 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.
14. Greathouse KL, Sinha R, Vogtmann E. DNA extraction for human microbiome studies: the issue of standardization. Genome Biol. 2019;20(1):212.
15. Tigchelaar EF, Zhernakova A, Dekens JA, Hermes G, Baranska A, Mujagic Z, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ Open. 2015;5(8):e006772.
16. Netea MG, Joosten LAB, Li Y, Kumar V, Oosting M, Smeekens S, et al. Understanding human immune function using the resources from the Human Functional Genomics Project. Nat Med. 2016;22(8):831–833.
17. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352(6285), 565–569.
18. Beghini F, McIver LJ, Blanco-Miguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife. 2021;10.
19. Ernst FGM, Shetty SA, Borman T, Lahti L. mia: Microbiome analysis. R package version 1.2.7. https://github.com/microbiome/mia. 2022.
20. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre R, Minchin PR, et al. vegan: Community Ecology Package. R package version 2.6.2. https://CRAN.R-project.org/package=vegan. 2022.
21. Islam MS, Aryasomayajula A, Selvaganapathy PR. A review on macroscale and microscale cell lysis methods. Micromachines. 2017;*8*(3):83.
22. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol. 2017;35(11):1077–1086.
23. Wagner Mackenzie B, Waite DW, Taylor MW. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. Front Microbiol. 2015;6:130.
24. Ducarmon QR, Hornung BVH, Geelen AR, Kuijper EJ, Zwittink RD. Toward Standards in Clinical Microbiota Studies: Comparison of Three DNA Extraction Methods and Two Bioinformatic Pipelines. mSystems. 2020;5(1).
25. Lim MY, Park YS, Kim JH, Nam YD. Evaluation of fecal DNA extraction protocols for human gut microbiome studies. BMC Microbiol. Jul 17 2020;20(1):212.
26. Bartolomaeus TUP, Birkner T, Bartolomaeus H, Löber U, Avery EG, Mähler A, et al. Quantifying technical confounders in microbiome studies. Cardiovasc Res. 2021;117(3):863–875.
27. Fiedorova K, Radvansky M, Nemcova E, Grombirikova H, Bosak J, Cernochova M, et al. The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery. Front Microbiol. 2019;10:821.

28. Gerasimidis K, Bertz M, Quince C, Brunner K, Bruce A, Combet E, et al. The effect of DNA extraction methodology on gut microbiota research applications. BMC Res Notes. Jul 26 2016;9:365.
29. Segata N. On the Road to Strain-Resolved Comparative Metagenomics. mSystems. 2018;3(2).
30. Sui HY, Weil AA, Nuwagira E, Qadri F, Ryan ET, Mezzari MP, et al. Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing. Front Microbiol. 2020;11:953.
31. Kia E, Wagner Mackenzie B, Middleton D, Lau A, Waite DW, Lewis G, et al. Integrity of the Human Faecal Microbiota following Long-Term Sample Storage. PLoS One. 2016;11(10):e0163666.

Supplementary Information

Additional file 1 (.docx): Text S1. Detailed step-by-step description of DNA extraction protocols.

Additional file 2 (.pdf): Figure S1. Read depth, DNA concentration and purity analysis. Comparison of (A) the number of reads after contaminant removal, (B) DNA concentration (ng/µl) and (C) DNA purity measured as 260/280 absorbance ratio between samples extracted with the APK (brown) and FSK (green) protocols from each cohort (LLD: lighter colors, 500FG: darker colors). Each dot represents one sample. Boxplots illustrate the median (middle horizontal line), IQR (boundaries of the boxes) and values within 1.5 times the IQR (whiskers) (*** p-value < 0.001, rank-based Wilcoxon test).

Additional file 3 (.pdf): Figure S2. Correlation analysis between alpha diversity and read depth. Correlation plots between species richness (A–B), Shannon (C–D) and Inverse Simpson (E–F) diversity indices and the number of reads after contaminants removal (log10 scale). Samples from each cohort are displayed (LLD: blue, 500FG: violet). Each circle is one sample. Coefficients and p-values from the Pearson correlation are shown.

Additional file 4 (.pdf): Figure S3. Beta diversity analysis based on Bray-Curtis dissimilarity. PCoA plot for (A) LLD and (B) 500FG samples extracted with each DNA isolation method (APK: brown, 500FG: green). Each datapoint represents the microbial community composition of one sample. Ellipses illustrate the standard deviation of samples belonging to each group. (C–D) Model coefficients estimated by PERMANOVA analysis for the species that exhibited the largest differences between samples extracted with APK and FSK protocol in each cohort. The top 15 species are shown ordered according to the coefficient value (continuous color scale: positive (increased in APK) in light blue and negative (increased in FSK) in dark blue). (E) Individual effect sizes, estimated by PERMANOVA, of DNA extraction protocol, sex, BMI and age on microbiome community variation (LLD: blue, 500FG: violet). (F) Correlation plot of Bray-Curtis distances between differentially extracted samples from each cohort (LLD: blue, 500FG: violet). Each data point represents one sample. Coefficients and p-values from the Mantel test are shown.

Additional file 5 (.pdf): Figure S4. Core taxa analysis. Each sub-cohort is analyzed separately: (A) APK-LLD, (B) FSK-LLD, (C) APK-500FG and (D) FSK-500FG. Heatmaps show the variation in taxa prevalence (continuous color scale, 1: red, 0: blue) according to relative abundance thresholds (y-axis). Only

microbial species with a prevalence > 50% are shown. Line plots show the alteration in the number of core species (y-axis) with changes in relative abundance (x-axis) for the 50% and 90% prevalence.

Additional file 6 (.xlsx): Tables S1-S9. Table S1. Read depth, DNA concentration and DNA purity values of samples from the 500FG and LLD cohorts. **Table S2.** Species richness and alpha diversity values of samples from the 500FG and LLD cohorts. **Table S3.** PERMANOVA analysis results of DNA extraction protocol and anthropometric variables in the 500FG and LLD cohorts. **Table S4.** Differential abundance analysis at phylum-level of paired samples from the 500FG and LLD cohorts. **Table S5.** Differential abundance analysis at species-level of paired samples from the 500FG and LLD cohorts. **Table S6.** Correlation analysis of species abundances between paired samples from the 500FG and LLD cohorts. **Table S7.** Association analysis between alpha diversity and available metadata for 500FG and LLD cohorts. **Table S8.** Association analysis between beta diversity and available metadata for 500FG and LLD cohorts. **Table S9.** Association analysis between microbial taxa abundances at species-level and available metadata for the LLD cohort. **Table S10.** Association analysis between microbial taxa abundances at species-level and available metadata for the 500FG cohort.

Figures

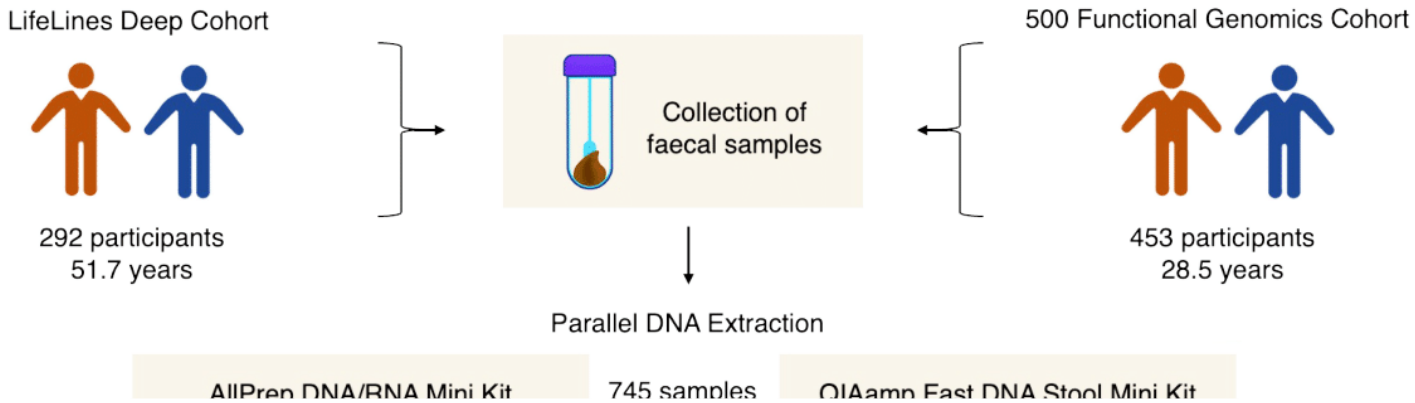

Figure 1

Experimental study design. Fecal samples collected from participants of two different cohorts were included in the present study. In total, 745 samples were isolated using two different DNA extraction kits: AllPrep DNA/RNA Mini Kit and QIAamp Fast DNA Stool Mini Kit (QIAGEN, Germany). After measuring DNA concentration and quality, extracted samples underwent library preparation and sequencing using the same technology (Illumina Hiseq 2000) at two different sequencing facilities. Finally, microbiome profiling and statistical analysis of sequencing data was performed.

Figure 2

Alpha diversity analysis. Comparison of species richness **(A)**, Shannon **(B)** and Inverse Simpson **(C)** diversity values between samples extracted with APK (brown) and FSK (green) from each cohort (LLD: lighter colors, 500FG: darker colors). Each dot represents one sample. Boxplots represent the median (middle horizontal line), the IQR (boundaries of the boxes) and values within 1.5 times the IQR (whiskers). Correlation plots of species richness **(D)**, Shannon **(E)** and Inverse Simpson **(F)** diversity indices between differentially extracted samples from each cohort (LLD: blue, 500FG: violet). Each data point represents one sample. The coefficients and p-values from the Pearson correlation are shown accordingly (*** $p < 0.001$, rank-based Wilcoxon test).

Figure 3

Beta diversity analysis. PCoA plot based on Aitchison distances for **(A)** LLD and **(B)** 500FG samples extracted with each DNA isolation method (APK: brown, 500FG: green). Each data point represents the microbial community composition of one sample. The ellipses illustrate the standard deviation of samples belonging to each group. **(C–D)** Model coefficients estimated by PERMANOVA analysis for species that exhibit the largest differences between samples extracted with the APK and FSK protocol in each cohort. The top 15 species are displayed in order according to the coefficient value (continuous color scale, positive (increased in APK): light blue, negative (increased in FSK): dark blue). **(E)** Individual effect sizes of DNA extraction protocol, sex, BMI and age on microbiome community variation (LLD: blue, 500FG: violet, PERMANOVA on Aitchison distance). **(F)** Correlation plot of Aitchison distances between differentially extracted samples from each cohort (LLD: blue, 500FG: violet). Each data point represents one sample. The coefficients and p-values from the Mantel test are shown accordingly.

Figure 4

Differential microbial abundance analysis. (A) Relative abundances for the microbial phyla visualized as a density plot over a log-scaled axis. Phyla are ordered from top to bottom by decreasing relative

abundance. **(B,C)** Strip chart showing the relative abundances for the microbial species with a mean relative abundance > 5% in the LLD **(B)** and 500FG **(C)**cohorts. Each dot represents a sample. Boxplots represent the median (middle horizontal line), the IQR (boundaries of the boxes) and values within 1.5 times the IQR (whiskers). Relative abundances are colored according to DNA extraction protocol (APK: brown, FSK: green) and cohort (LLD: lighter colors, 500FG: darker colors). **(D)** Correlation plot of log-transformed relative abundances (for species defined in (C)) between differentially extracted samples from each cohort (LLD: blue, 500FG: violet). Each data point represents one sample. The coefficients and p-values from the Pearson correlation are displayed.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- FigureS11.png
- FigureS21.png
- FigureS31.png
- FigureS41.png
- SupplementaryTables.xlsx