Article:
Taylor, M, Wood, HM orcid.org/0000-0003-3009-5904, Halloran, SP et al. (1 more author) (2017) Examining the potential use and long term stability of guaiac faecal occult blood test cards for microbial DNA 16srRNA sequencing. Journal of Clinical Pathology, 70 (7). pp. 600-606. ISSN 0021-9746

https://doi.org/10.1136/jclinpath-2016-204165

TITLE PAGE

(1) Title: Examining the potential use and long term stability of guaiac faecal occult blood test cards for microbial DNA 16srRNA sequencing

(2) Authors: Morag Taylor[1], Henry M Wood[1], Stephen P Halloran[2], Philip Quirke[1]

[1]Section of Pathology and Tumor Biology, Leeds Institute of Cancer and Pathology , University of Leeds, Leeds, United Kingdom; [2] NHS Bowel Cancer Screening Programme Hub, University of Surrey, Guildford, United Kingdom

Joint first authors; M Taylor and H M Wood contributed equally to the study

(3) Correspondence: Miss Morag Taylor, Section of Pathology and Tumour Biology, Level 4, Wellcome Trust Brenner Building, St James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK. Tel: +44 (0) 113 3438510; Fax: +44 (0) 113 3438431; email medmta@leeds.ac.uk

(4) Keywords: stool; bowel cancer screening; next generation sequencing; microbiome

(5) Abbreviations:

CRC: colorectal cancer;

gFOBt: guaiac faecal occult blood test;

NHSBCSP: national health service bowel cancer screening programme;

OTUs: operational taxonomy units;

rRNA: ribosomal RNA

PCR: polymerase chain reaction

ABSTRACT

Aims

With a growing interest in the influence the gut microbiome has on the development of colorectal cancer (CRC), we investigated the feasibility and stability of isolating and typing microbial DNA from guaiac faecal occult blood test (gFOBt) cards. This has the future potential to screen the microbial populations present in confirmed colorectal neoplasia cases with aims to predict the presence and development of CRC.

Methods

Fresh stool samples from three healthy volunteers were applied to gFOBt cards. DNA was extracted from both the cards and fresh stool samples. A series of additional cards were prepared from one volunteer, and extracted at time points between 2 weeks and 3 years. The V4 region of the 16srRNA gene was amplified and sequenced on an Illumina MiSeq at 2x250bp read lengths. Data was analysed using QIIME software.

Results

Samples were grouped both by volunteer and by type (fresh or gFOBt), and compared a variety of ways: visual inspection of taxa, alpha and beta diversity, intra-class correlation. In all comparisons, samples grouped by volunteer, and not by sample type. The different time points showed no appreciable differences with increased storage time.

Conclusions

This study has demonstrated that there is good concordance between microbial DNA isolated from fresh stool sample, and from the matched gFOBt card. Samples stored for up to 3 years showed no detrimental effect on

measureable microbial DNA. This study has important future implications for investigating microbial influence on CRC development and other pathologies.

INTRODUCTION

The human gut plays host to a large number of microbes, estimated to be between 1 and 10 times the number of human cells [1 2]. Known collectively as the gut microbiome, this ecosystem is diverse with only a small percentage of the species routinely cultured in the laboratory [3 4]. The emergence of next-generation sequencing has led to the deep sequencing of the 16s rRNA gene becoming the preferred method of studying microbial diversity [5 6]. This complex ecosystem has been widely studied and linked to many human diseases [7-9], including a growing interest in the influence the gut microbiome has on the development of colorectal cancer (CRC) [7 10 11].

CRC is currently the fourth most common cancer and second most common cause of cancer death in the UK, accounting for 16,187 deaths in 2012 [12] , with 94% of diagnoses in people over the age of 50 [13]. Since 2006, the National Health Service Bowel Cancer Screening Programme (NHSBCSP) in England has been issuing bowel cancer screening tests every two years to individuals between the ages of 60 and 74 (Scotland commenced in 2007 with an age range is between 50 and 74)[14]. The NHSBCSP utilises a guaiac faecal occult blood test (gFOBt) card in which participants complete the test, and mail the cards to an NHSBCSP screening hub[15].

In recent years, gFOBt cards have been used to study the feasibility of detection of both molecular biomarkers of colorectal cancer and Escherichia

coli. Rennert et al.[16] demonstrated the extraction of human DNA and detection KRAS mutations from gFOBt cards. Grimes et al.[17] showed that gFOBt cards were a suitable alternative to collecting and transporting fresh stool samples. They also demonstrated that samples could be stored on these cards for up to 14 months with no detrimental effect. Dominianni et al.[18], investigated gFOBt cards as one of several different storage methods. They stored samples for a three day period before extraction, and did not find any ascertainable differences in the collection methods. Sinha et al [19] recently compared various sample processing protocols, and concluded that gFOBt cards produced microbiome data that was comparable to fresh stool samples.

Our aim was to confirm the feasibility of using the gFOBt cards for microbial analysis by profiling matched gFOBt and fresh stool samples from laboratory volunteers. We expanded on the work by Sinha et al [19] to demonstrate that these cards can be stored long term before use, by storing and extracting samples from one volunteer at time points of up to 3 years. Volunteers were included if they had no known gastrointestinal disorders, and considered themselves to have a healthy gut. This was a laboratory-based preliminary study which only included departmental technical and medical staff. Full consent was given for the samples to be used in this investigation, which also included a Quality Assurance of the hypothesis we set out.

This work has the potential long term goal of using these cards to investigate the value of microbial population data on the prediction of the presence of

colorectal neoplasia within the large bowel. This method might also be used for studying the epidemiology of the microbiome.

METHODS

gFOBt card vs fresh stool sample

Three healthy volunteers (A, B and C) each provided three fresh stool samples, which were collected over a five to ten day period. Volunteers A and B collected their stool samples at home, and stored the samples at 4˚C in biohazard bags overnight. Volunteer C collected samples the morning they were returned to the laboratory. All samples were transported at ambient temperature. Stool samples were processed as soon as they were received, where they were divided into two aliquots, from which duplicate DNA extractions were performed. Each stool aliquot was also applied in duplicate to a gFOBt card (Immunostics inc, USA). gFOBt cards were left at room temperature until three samples had been applied, and were then developed using the supplied Haema-screen hydrogen peroxide developer (Immunostics inc, USA) according to manufacturers' instructions before DNA extraction.

Long term storage on gFOBt cards

The gFOBt cards were treated as they would be if distributed to screening programme participants. One volunteer (A) applied fresh stool to each of the three card tabs on separate days to mimic typical stool collection. Cards were developed on receipt by the lab and stored at room temperature until extraction. Time point extractions DNA initially covered 2, 4, 6, 8, 10, and 12 weeks post development of the guaiac test. A second set of cards collected

over an additional three days were then stored for 7, 12, 24 and 36 months post developing, before DNA extraction

Microbial DNA Isolation and Illumina Sequencing

Microbial DNA Extractions were performed based on a method originally reported by Yu et al [20], with slight modifications to the protocol for fresh and gFOBt samples.

gFOBt cards

Each of the two windows of a tab were initially treated as separate samples, being excised from the gFOBt card with a scalpel, and then dissected into thin strips. All strips were placed into a 2mL microcentrifuge tube with 600µL of Buffer ASL from the QIAamp DNA Stool Mini Kit (Qiagen, Germany). Samples were placed on a shaker at 850rpm and 23℃ for 1 hour to allow faeces to detach from the paper. Samples were centrifuged briefly at between 850 and 2800g for 10 to 30 seconds. At this point, both windows of the same tab were transferred into one pathogen lysis tube (Qiagen, UK). Lysis tubes were placed onto a shaker for 10 minutes at between 1800 and 2000rpm, immediately transferred to a heated shaker at 95℃ for 15 minutes and 850rpm, centrifuged at 21910g for 1 minute, and supernatant removed to a clean 2mL microcentrifuge tube.

Fresh Stool Samples

Around 250mg of stool was taken from each aliquot, and weighed in a pathogen lysis tube. 1mL of the Buffer ASL was added to the tube and placed

onto a shaker for 10 minutes at between 1800 and 2000rpm, immediately transferred to 95℃ for 15 minutes and 850rpm, centrifuged at 22000 g for 1 minute, supernatant removed to a clean 2mL microcentrifuge tube and stored at room temperature. A further 300µL of Buffer ASL added to the pathogen lysis tube, and the above steps repeated with the supernatant being transferred to the 2mL microcentrifuge tube previously stored at room temperature.

All extractions

260µL of 10M ammonium acetate was added to the supernatant which was vortexed briefly then placed on ice for 5 minutes. Samples were centrifuged for 5 minutes at 22000g, and the supernatant split between two 1.5mL microcentrifuge tubes with an equal volume of Isopropanol (Propan-2-ol). Tubes were vortexed, placed on ice for 30 minutes, centrifuged for 10 minutes at 22000g and the supernatant removed and discarded. 1mL of 70% ethanol was added to each tube and samples were centrifuged again for 5 minutes at 22000g. The supernatant was discarded and a further 500µL of 70% ethanol added to the tubes before being centrifuged again at 22000g for 5 minutes. The supernatant was discarded and the tubes left to air dry. The pellet was re-suspended in 100µL of 1x Tris EDTA Buffer and both tubes were pooled again. Following this step, samples were processed using the standard QIAamp DNA Stool Mini Kit protocol starting with the addition of 15µL proteinase K and 200µL of Buffer AL. Sample were eluted into 100µL of UV treated molecular grade water. Full lists of samples are given in supplementary table S1.

16s rRNA V4 amplification and Library Preparation

DNA was quantified by nanodrop (Labtech, USA) and diluted to 20ng/µL for polymerase chain reaction (PCR). PCR was performed in a 50µL reaction volume with the following final concentrations: 1µL of 20ng/µL template DNA, 10µL 5X QC Reaction Buffer, 0.5µL Q5 Hot Start High Fidelity DNA Polymerase, 10µL 5X Q5 High QC Enhancer, 1µL 10mM dNTPs (all New England Biolabs, USA), 2.5µL of each 10µM Primers. Primers reported by Claesson et al [21] to the V4 region V4F-5'-AYTGGGYDTAAAGNG, V4R- 5' – TACNVGGGTATCTAATCC were used. Thermal cycler conditions were 98℃ for 30s, 30cycles of 98℃ for 5s, 42℃ for 10s, 72℃ for 20s with a final extension of 72℃ for 2min. Successful amplification was confirmed by gel electrophoresis before samples were cleaned using the MinElute PCR Purification kit (Qiagen, Germany). PCR products were eluted into 10µL of Buffer EB and 2µL was used to quantify with the Quant-iT™ dsDNA Assay Kit, broad range (Life Technologies, USA). Up to 120ng of PCR products were carried forward to library preparation with the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® and NEBNext® Singleplex Oligos for Illumina® (New England Biolabs UK). Unique in house designed index primers (Integrated DNA Technologies Inc, UK) were used to allow multiplexing of samples. Twelve cycles of enrichment PCR were performed, and final libraries cleaned with AMPure Beads (Beckman Coulter, UK). Successful libraries were confirmed by DNA1000 bioanalyzer chips, or DNA Analysis screen tapes (Agilent, UK). Quantification was performed with Quant-iT™ dsDNA Assay Kit, broad range. 30ng of each library was pooled and sequenced on an Illumina MiSeq (2x250bp).

Sequence analysis

Demultiplexed FASTQ files were trimmed of adapter sequences using cutadapt [22]. Paired reads were merged using fastq-join (https://code.google.com/archive/p/ea-utils/wikis/FastqJoin.wiki) under default settings and then converted to FASTA format. Consensus sequences were removed if they contained any ambiguous base calls, two contiguous bases with a PHRED quality score lower than 33, or a length more than 2bp different from the expected 240bp. Further analysis was performed using QIIME [23]. Operational taxonomy units (OTUs) were picked using Usearch [24], and aligned to the Greengenes [25] reference database using PyNAST [26]. Taxonomy was assigned using the RDP 2.2 classifier [27].

As well as comparisons of observed taxa, rarefication was performed to various levels to compare alpha diversity for different sample groupings. All groups were rarefied to the lowest read number, and beta diversity calculated using weighted and unweighted UniFrac [28] as well as the non-phylogenetic Bray-Curtis dissimilarity measure [29]. Beta diversity was compared using principle component analysis on all samples, and two-sided Student's two-sample t-tests to compare different groupings.

Intra-class correlation coefficients (ICCs) were calculated for fresh and gFOBt samples, as well as all time points using methods similar to those described in Sinha et al [19]. For various metrics, the ICC was defined as:

$$p = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2}$$

where σ_b^2 was the inter-volunteer variability and σ_ϵ^2 was the intra-volunteer variability. For 500 iterations, two randomly chosen samples from one volunteer were compared to each other, as were two samples from different

volunteers. The metrics compared where the relative abundances of Actinobacteria, Bacteriodetes and Firmicutes, the beta diversity measures of weighted and unweighted unifrac, and Bray-Curtis, and the alpha diversity measures of observed species and Shannon index. For the time course comparisons, as only one volunteer provided samples, these were compared to randomly selected time=0 samples from other volunteers

RESULTS

114 samples were sequenced across 3 volunteers. Between 10192 and 85790 sequenced reads were obtained (median 44198). Sequencing statistics are in supplementary table S2. Raw sequence data is deposited in the European Nucleotide Archive, http://www.ebi.ac.uk/ena/data/view/PRJEB14174.

Alpha diversity rarefaction plots are shown in figure 1. Initially, all samples not part of the time course experiment were compared, grouped either by volunteer or by sample type (fresh of gFOBt). There were more differences seen between sample types than between volunteers. When all the samples from the time course experiment were compared, there was no appreciable difference between any of the time points.

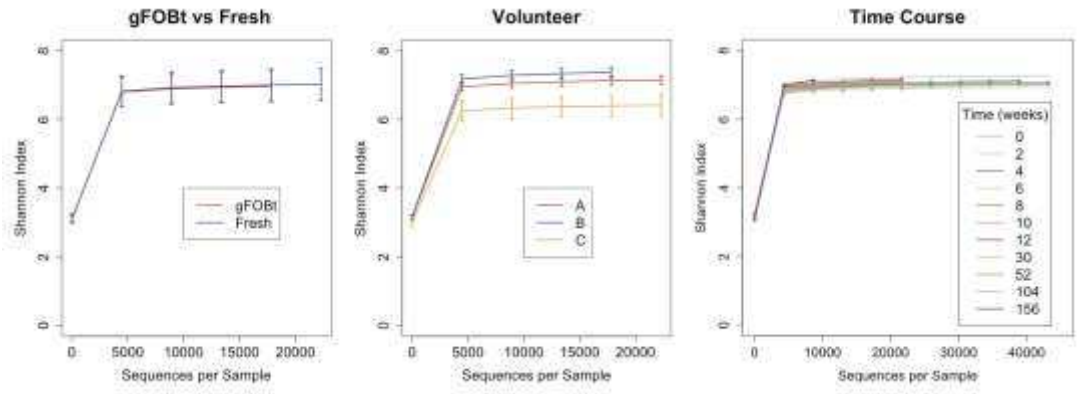


Figure 1: Alpha rarefaction plots comparing sample type (left), different volunteers (middle) and time points (right). For each grouping, the samples were rarefied to gradually decreasing sequencing depths, and the Shannon diversity index calculated for all samples within that group.

Taxonomic comparisons of the fresh and gFOBt samples from each volunteer, and from the time course experiment are shown in figure 2. Again, the differences between the fresh and gFOBt samples for each volunteer are smaller than the differences between volunteers. Although slight differences in observed taxa were seen at different times, there was no trend of any taxonomic groups increasing or decreasing with increased storage time, and was more likely to a result of small sample numbers for each time point.

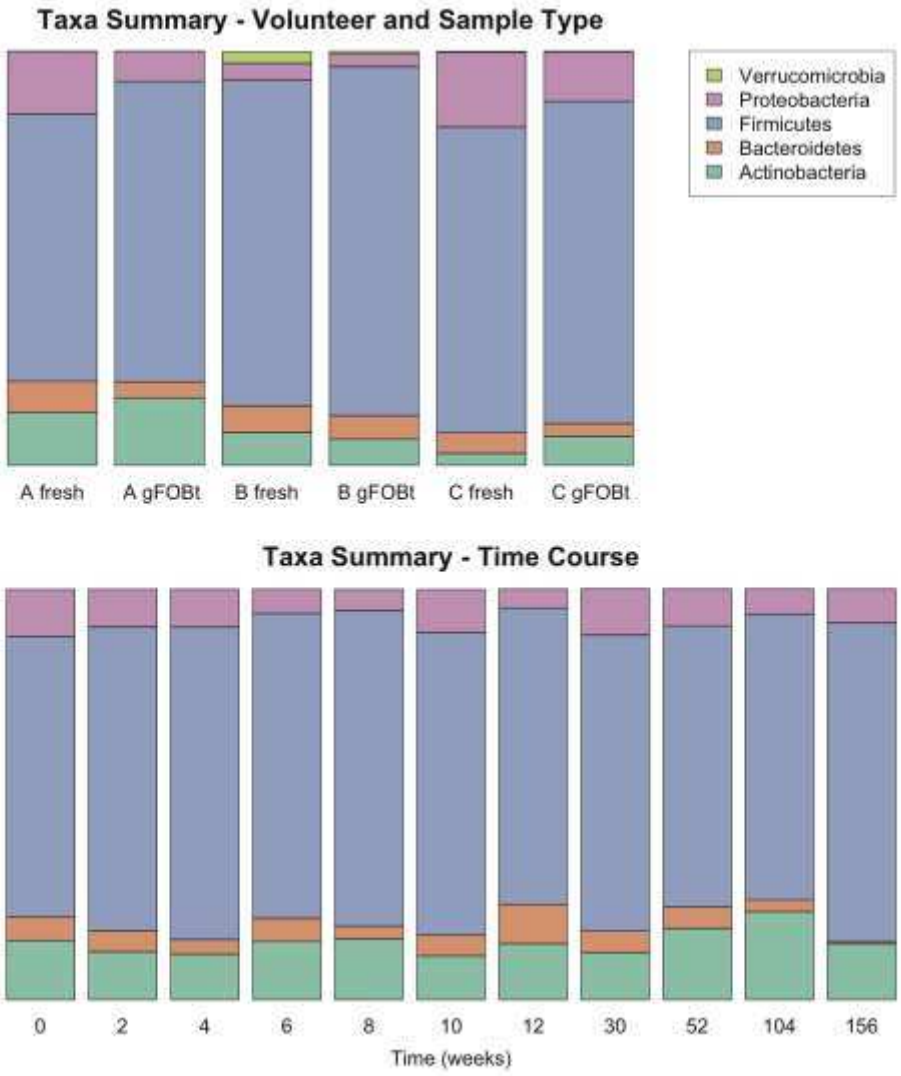


Figure 2: Taxa summary at phyla level. The top plot shows taxa for each volunteer (A, B or C) for each sample type (fresh or gFOBt). The bottom plot shows the time course experiment. For simplicity of plotting, only taxa present in over 1% of reads for at least one of the sample groupings are shown. Full lists of taxa are given in supplementary tables S3 and S4.

Principle component plots (figure 3) confirm this. With weighted and unweighted Unifrac, as well as Bray-Curtis comparisons, the samples grouped by volunteer rather than sample type. Comparing the time course

samples showed no particular effect of storage time. The differences between the fresh and gFOBt samples were greater than the differences across time points for gFOBt samples. However, when the samples from the other volunteers were also included in the analysis, these differences were dwarfed by the differences between volunteers.

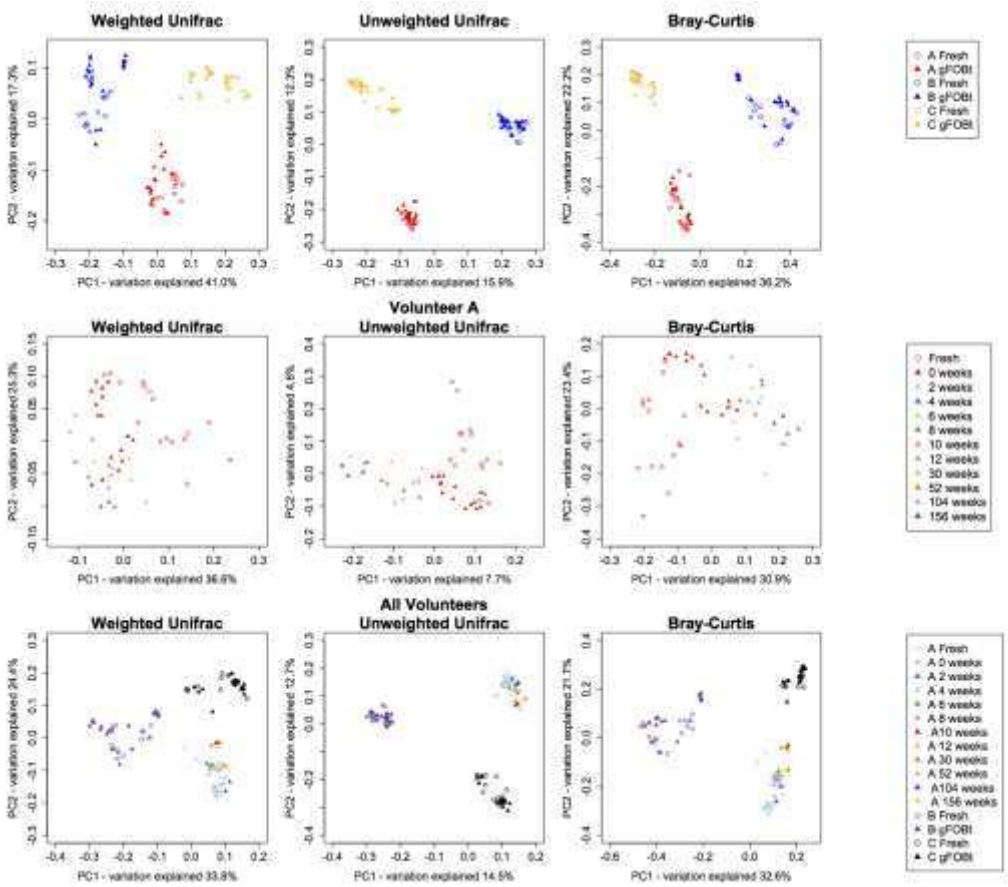
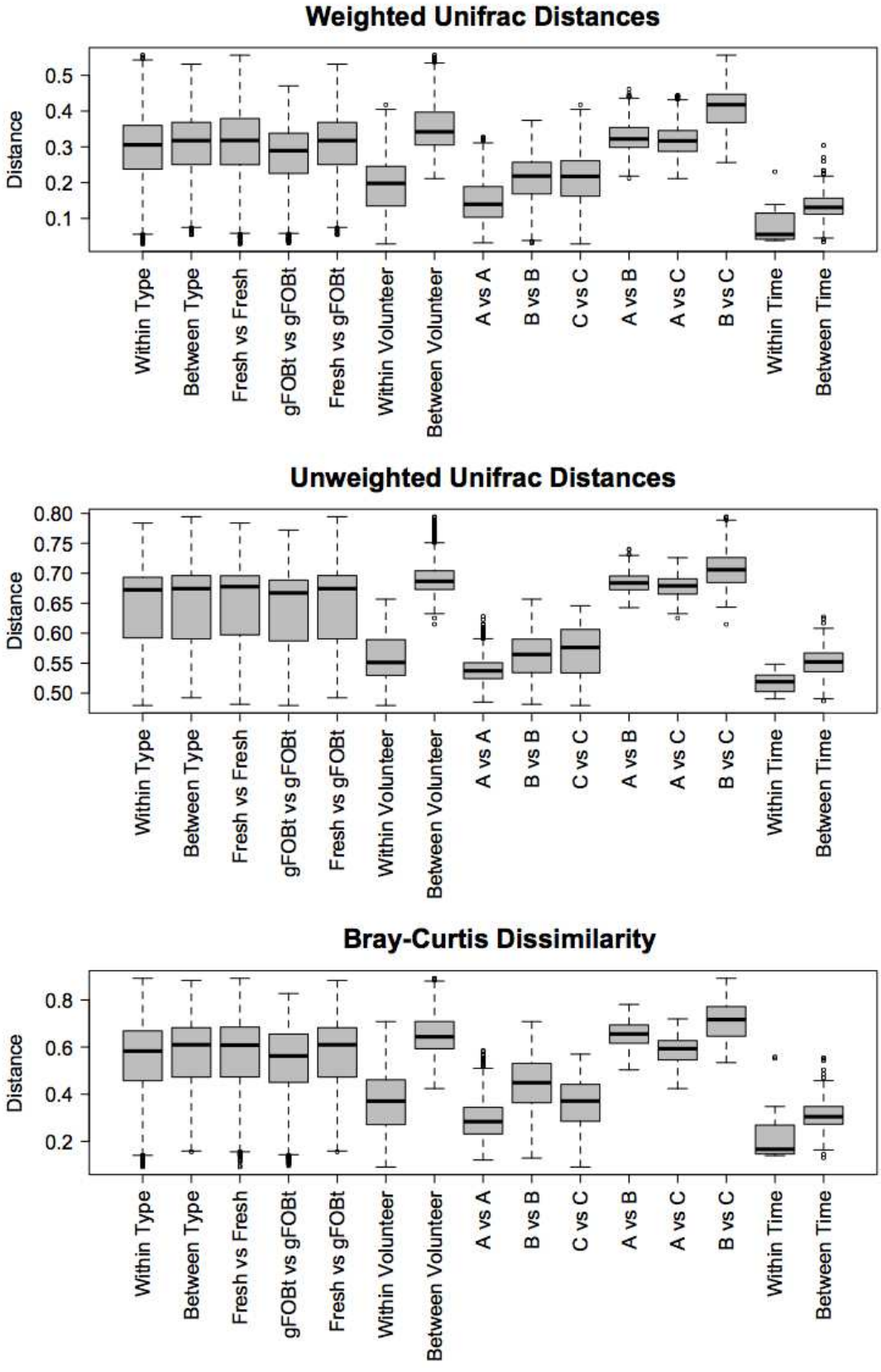


Figure 3: Principle component analysis comparing weighted (left) and unweighted (middle) UniFrac distances and Bray-Curtis dissimilarity (right) between all samples. The first two principle components are compared to each other. In the top row, samples are coloured by volunteer, with point shape showing sample type. In the middle row, all the time course samples are compared to the other samples from volunteer A. In the bottom row, all

samples are compared. Samples are coloured by time and volunteer, with fresh samples distinguished by shape.

When comparing all possible Unifrac distances or Bray-Curtis dissimilarity between samples, grouped either by sample type or volunteer (figure 4), all within and between sample type comparisons had overlapping distributions. The between sample type distributions were significantly different for the weighted and Bray-Curtis comparisons, but not unweighted. However, the various between and within volunteer combinations showed visibly different distributions to each other. The calculated p-values after Bonferroni correction were far lower when comparing within and between volunteers (all $< 1 \times 10^{-300}$) than between and within sample types (6.7×10^{-7}, 1 and 7.1×10^{-6}) for weighted and unweighted Unifrac and Bray-Curtis respectively. Full lists of p values for all comparisons are given in table S5. The between and with time point comparisons showed slightly different distributions, but had far lower distances and any other comparisons expect for the A vs A comparison. The small sample numbers at each time point made reasonable statistical comparisons of the distances between time points difficult.

Figure 4 (overleaf): Box plots showing distributions of weighted UniFrac (top), unweighted UniFrac (middle) distances and Bray-Curtis dissimilarity (bottom) between groups of samples. Distances within and between sample types, and within and between all and individual volunteers are shown, as are the distances within and between samples from different time points.



When ICCs for various different metrics were compared, four metrics (relative abundances of Actinobacteria, Bacteriodetes, and the two alpha diversity

meaures) were higher in gFOBt samples than fresh, while four metrics (abundance of Firmicutes and the three beta diversity measures) were lower in gFOBt sample. This suggests that neither sample type demonstrable inferior. All ICC metrics were very stable across the time course experiment, with the taxa abundance measures being lowest, the beta diversity measures highest, and alpha diversity between.

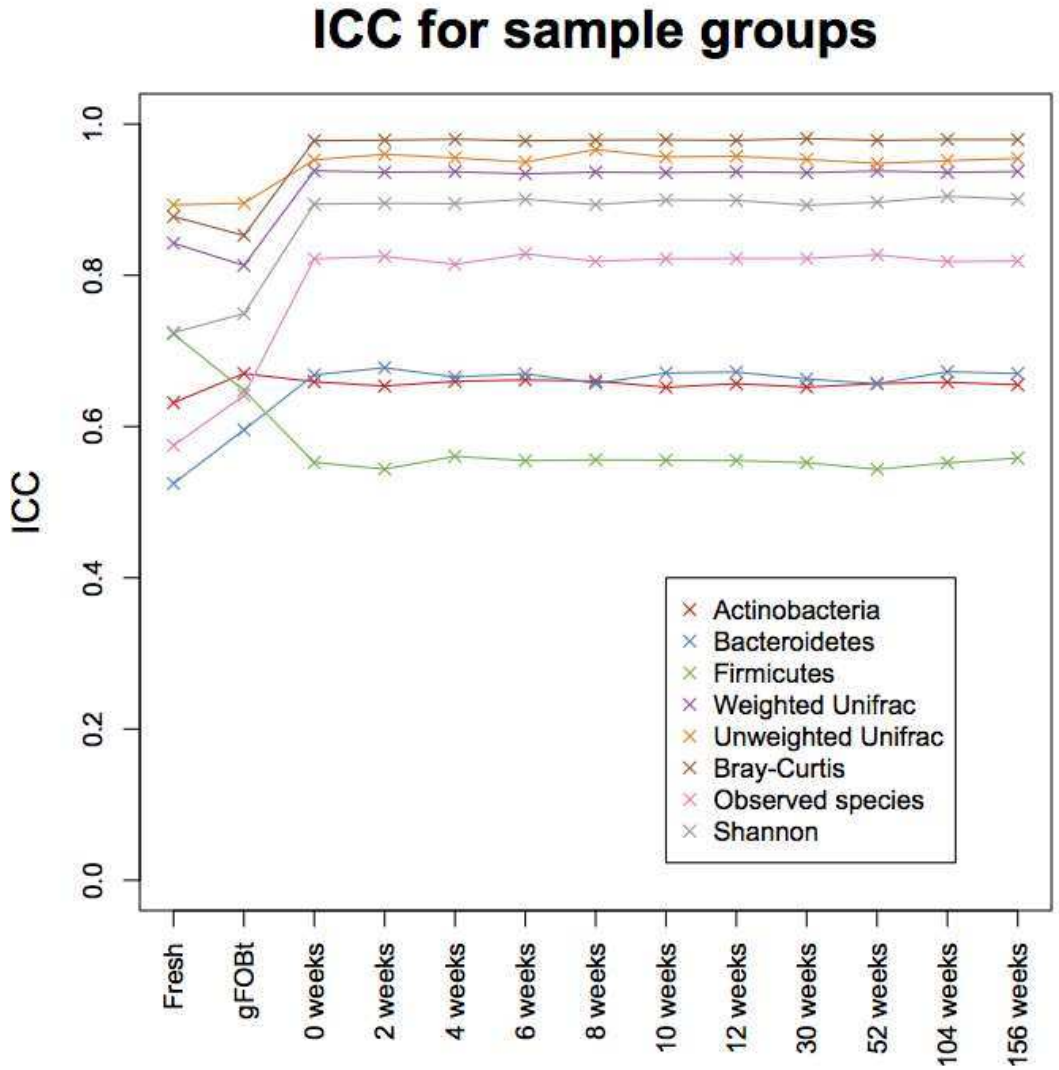


Figure 5: ICC calculations comparing different sample groupings with different measures, namely relative abundances of Actinobacteria, Bacteriodetes and Firmicutes, the beta diversity measures of weighted and unweighted unifrac,

and Bray-Curtis, and the alpha diversity measures of observed species and Shannon index. Samples were grouped by sample type or time stored.

DISCUSSION

The initial goal of this project was to assess the utility of gFOBt cards as an acceptable alternative to fresh stool samples when studying the faecal microbiome. The ability to use these cards to store stool samples would have many advantages. For defined prospective studies, it is usually much more convenient for subjects to use gFOBt cards to collect samples than to send or bring fresh or frozen stools to the testing laboratory. Secondly, if appropriate ethical approval were in place, it would enable population studies using the many millions of gFOBt cards returned to national screening programmes. Thirdly, if studies using either gFOBt cards or fresh stools were able to show a link between a particular microbial signature and any clinical disease characteristics, whether it were the presence of colorectal neoplasia, other gastrointestinal and liver diseases, obesity, drug response, or even mental health [30], then the existing screening programmes could be expanded to screen for such diseases or the microbial signature could be used as an additional biomarker in the programme.

The fresh faecal and gFOBt samples were examined using 16S sequencing from three healthy volunteers. The microbial profiles were studied in a variety of ways. The alpha (within sample) diversity was examined for each sample. Lists of the main taxa present in each sample grouping were compared. Unweighted (the variety of OTUs present) and weighted (the relative numbers

of those OTUs within each sample) Unifrac distances were calculated between each sample as well as the non-phylogenetic Bray-Curtis dissimilarity measure. These were compared visually using principle component analysis and statistically, by comparing within and between various sample groups.

By every one of these measures, the differences between volunteers was far greater than the differences between fresh and gFOBt samples within those volunteers. These results were more evident for unweighted than weighted analysis, suggesting that the majority populations from each volunteer were mostly the same, but that each person had specific minority populations. These major and minor populations were the same whether they were measured using fresh or gFOBt samples. ICCs for various measures were also calculated. Half had a better (closer to one) score in gFOBt samples, and half had a better score in fresh samples. These results confirm those of Sinha et al [19], who also found that samples from bowel cancer screening card were an acceptable substitute for fresh stool.

The second, and novel aim of this study was to study the effect of long term storage of gFOBt cards on the microbiome detected. If cards can be stored with confidence for long periods, then all of the advantages listed above are increased. Prospective studies would have much more flexibility when operating. Retrospective studies would have more confidence to study older samples, and multiple cards from the same patient over months or even years could be analysed.

Samples from the same volunteer were collected, applied to multiple gFOBt cards, and stored at room temperature for differing times up to three years

before DNA extraction, and examined as before. Diversity analysis, visual inspection of taxa, ICCs, and principle component analysis all demonstrated small differences between samples, but they were not linked to time stored in any way, and appeared to be just random fluctuation. When the time course samples were compared to those from the other volunteers, then again the differences between volunteers dwarfed the minor fluctuations seen amongst the samples from different times.

In summary, the use of gFOBt cards is an acceptable substitute for fresh stool samples when carrying out studies of the faecal microbiome. Storage of cards for up to three years, and probably longer, appears to have no detrimental effect on the data recovered. This has important implications for the ease of carrying out large scale microbiome studies, and possible future benefits for patient screening for not only the NHSBCSP but also screening for other diseases either caused by or associated with the human microbiome.

KEY MESSAGES

- gFOBt cards can be used to store human stool samples prior to microbiome analysis.
- gFOBt cards can be stored for up to three years with no noticeable decline in data quality.

CONTRIBUTION STATEMENT

MT: Study concept and design, technical work, drafting of the manuscript.

HMW: acquisition of data, data analysis and interpretation, drafting of the

manuscript, obtained funding. SPH: supply of faecal occult blood test cards, drafting manuscript and advice. PQ: Study concept and design, drafting of the manuscript, obtained funding

COMPETING INTERESTS

None of the authors have any disclosures

FUNDING

Laboratory consumables supported by a Cancer Research UK development fund. MT, HMW, PQ supported by Yorkshire Cancer Research.

REFERENCES

1. Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. bioRxiv 2016 doi: 10.1101/036103[published Online First: Epub Date]|.
2. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science 2005;**307**(5717):1915-20 doi: 10.1126/science.1104816[published Online First: Epub Date]|.
3. Suau A, Bonnet R, Sutren M, et al. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl Environ Microbiol 1999;**65**(11):4799-807
4. Hayashi H, Sakamoto M, Benno Y. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. Microbiol Immunol 2002;**46**(8):535-48
5. Woo PC, Lau SK, Teng JL, Tse H, Yuen KY. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect 2008;**14**(10):908-34 doi: 10.1111/j.1469-0691.2008.02070.x[published Online First: Epub Date]|.

6. Lazarevic V, Whiteson K, Huse S, et al. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. J Microbiol Methods 2009;**79**(3):266-71 doi: 10.1016/j.mimet.2009.09.012[published Online First: Epub Date]|.

7. Tojo R, Suarez A, Clemente MG, et al. Intestinal microbiota in health and disease: role of bifidobacteria in gut homeostasis. World J Gastroenterol 2014;**20**(41):15163-76 doi: 10.3748/wjg.v20.i41.15163[published Online First: Epub Date]|.

8. Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. Therap Adv Gastroenterol 2013;**6**(4):295-308 doi: 10.1177/1756283X13482996[published Online First: Epub Date]|.

9. Holmes E, Li JV, Athanasiou T, Ashrafian H, Nicholson JK. Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. Trends Microbiol 2011;**19**(7):349-59 doi: 10.1016/j.tim.2011.05.006[published Online First: Epub Date]|.

10. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One 2012;**7**(6):e39743 doi: 10.1371/journal.pone.0039743[published Online First: Epub Date]|.

11. Yang T, Owen JL, Lightfoot YL, Kladde MP, Mohamadzadeh M. Microbiota impact on the epigenetic regulation of colorectal cancer. Trends Mol Med 2013;**19**(12):714-25 doi: 10.1016/j.molmed.2013.08.005[published Online First: Epub Date]|.

12. CRUK. Bowel Cancer Statistics. Secondary Bowel Cancer Statistics 2015.

13. BowelCancer-UK. Bowel Cancer Statistics. Secondary Bowel Cancer Statistics 2016.

14. CRUK. About Bowel Cancer Screening. Secondary About Bowel Cancer Screening 2016.

15. Morris EJ, Whitehouse LE, Farrell T, et al. A retrospective observational study examining the characteristics and outcomes of tumours diagnosed within and without of the English NHS Bowel Cancer Screening Programme. Br J Cancer 2012;**107**(5):757-64 doi: 10.1038/bjc.2012.331[published Online First: Epub Date]|.

16. Rennert G, Kislitsin D, Brenner DE, Rennert HS, Lev Z. Detecting K-ras mutations in stool from fecal occult blood test cards in multiphasic screening for colorectal cancer.

Cancer Lett 2007;**253**(2):258-64 doi: 10.1016/j.canlet.2007.01.023[published Online First: Epub Date]|.

17. Grimes KA, Mohamed JA, Dupont HL, et al. PCR-based assay using occult blood detection cards for detection of diarrheagenic Escherichia coli in specimens from U.S. travelers to Mexico with acute diarrhea. J Clin Microbiol 2008;**46**(7):2227-30 doi: 10.1128/JCM.00073-08[published Online First: Epub Date]|.
18. Dominianni C, Wu J, Hayes RB, Ahn J. Comparison of methods for fecal microbiome biospecimen collection. BMC Microbiol 2014;**14**:103 doi: 10.1186/1471-2180-14-103[published Online First: Epub Date]|.
19. Sinha R, Chen J, Amir A, et al. Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies. Cancer Epidemiol Biomarkers Prev 2016;**25**(2):407-16 doi: 10.1158/1055-9965.EPI-15-0951[published Online First: Epub Date]|.
20. Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta and fecal samples. Biotechniques 2004;**36**(5):808-12
21. Claesson MJ, Wang Q, O'Sullivan O, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res 2010;**38**(22):e200 doi: 10.1093/nar/gkq873[published Online First: Epub Date]|.
22. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 2011;**17**(1)
23. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;**7**(5):335-6 doi: 10.1038/nmeth.f.303[published Online First: Epub Date]|.
24. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;**26**(19):2460-1 doi: 10.1093/bioinformatics/btq461[published Online First: Epub Date]|.
25. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;**72**(7):5069-72 doi: 10.1128/AEM.03006-05[published Online First: Epub Date]|.

26. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 2010;**26**(2):266-7 doi: 10.1093/bioinformatics/btp636[published Online First: Epub Date]|.

27. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007;**73**(16):5261-7 doi: 10.1128/AEM.00062-07[published Online First: Epub Date]|.

28. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;**71**(12):8228-35 doi: 10.1128/AEM.71.12.8228-8235.2005[published Online First: Epub Date]|.

29. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol Monogr 1957;**27**(4):326-49

30. Kinross JM, Darzi AW, Nicholson JK. Gut microbiome-host interactions in health and disease. Genome Med 2011;**3**(3):14 doi: 10.1186/gm228[published Online First: Epub Date]|.

Supplementary Table Legends

Table S1: Sample details, linking volunteer, fresh sample and gFOBT equivalent, or time the sample was collected.

Table S2: Details of samples used. Sample name, volunteer, sample type (FOBT or fresh), original stool sample from which the sample was derived, time stored on FOBT card, and numbers of reads analysed are recorded.

Table S3: Full taxa lists for each volunteer for each sample type.

Table S4: Full taxa lists for each group of samples of the time course experiment.

Table S5: Statistics for weighted and unweighted UniFrac distances, and Bray-Curtis dissimilarity between groups of samples. Distances within and between sample types, and within and between all and individual volunteers are compared.