Critical To Quality pre-analytical factors and their impact on microbiome analysis

Evgueni Doukhanine[1], Lisa Gamwell[1], Anne Bouevitch[1], Caroline Elgoff[1], Denise Lynch[1], Christopher Smith[1], Aaron Del Duca[1], Roxana Odouli[2], Jeannette Ferber[2], De-Kun Li[2], Carlos Merino[1]

[1] DNA Genotek Inc., Ottawa, Ontario, Canada

[2] Kaiser Foundation Research Institute, Kaiser Permanente, Oakland, California, USA

Introduction

Since the completion of the Human Microbiome Project (HMP), the number of microbiome focused research programs, peer-reviewed publications and patents has grown exponentially. As a result, much progress has been made in understanding the complex relationship between host and microbiome. However, lack of standard procedures, reference materials and quality control metrics limits reproducibility and comparability of published results, which can lead to discordant data interpretation and false inferences. The microbiome field can only reach its translational potential when hypotheses are rigorously tested with optimized methods.

Quality management in metagenomics begins with a systematic identification of workflow steps that can influence factors critical to the quality or analytical validity of the study. Collectively, these factors are known as CTQs (factors Critical To Quality, defined in the table). The identification of processes that impact these CTQs enables risk mitigation through effective experimental design. Surprisingly, the impact of pre-analytical workflow steps like sample collection and quality were only superficially studied in phase 1 of the Microbiome Quality Control Consortium (MBQC).

The goal of this study was to define and quantify the impact of the 3 stages of the pre-analytical metagenomic workflow on microbiome CTQs. These work flow stages are: 1. collection and stabilization methodology, 2. sample transport and 3. nucleic acid extraction method.

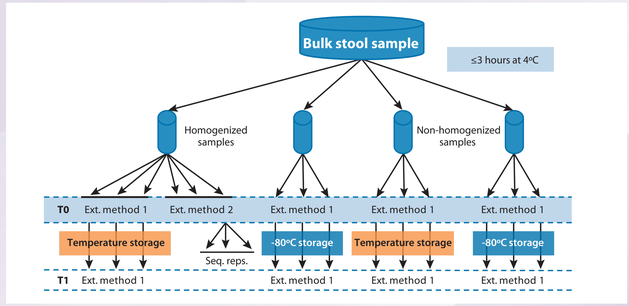
Microbiome CTQ	Definition
Neutrality	A shift in microbial community composition relative to control induced by the stabilization agent, noticeable soon after mixing sample and stabilizer.
Stability	A shift in microbial community composition relative to control, induced by shipping and/or storage conditions and accumulated over time.
Bias	A shift in microbial community composition relative to control induced by extraction methodology.
DNA yield and integrity	The amount of high molecular weight DNA extracted from a microbiome sample.
Repeatability	The variation in measurements taken on the same sample, under the same conditions, and in a short period of time.
Reproducibility	The ability of an entire experiment or study to be duplicated, either by the same researcher or by someone else working independently.

Our results demonstrate that the pre-analytical workflow, including collection, stabilization, sample transport and nucleic acid extraction account for most of the variation observed from sampling to sequencing. For example, we found that inadequate microbial DNA stabilization during transport could account for 40% of dissimilarity observed between samples while inconsistent nucleic acid extraction can contribute 34% dissimilarity. The introduction of such noise could result in an increase in the number of samples required to achieve statistical significance and a reduction in the impact and reproducibility of any biological interpretations.

Methods

Sample collection, stabilization, DNA extraction and sample storage

We designed a series of studies to evaluate CTQs for the pre-analytical workflow, including collection method, neutrality, stability and DNA extraction. Detailed description of each study is included in the results and discussion section. Briefly, a control sample was collected (fresh feces collected in a sterile container, transported as per HMP protocol and extracted within 3 hours of production). In addition, paired samples were collected using OMNIgene®•GUT kits according to the standard instructions or using other stabilization methods (per manufacturer instructions). When neutrality was assessed, all samples were extracted within 3 hours of stabilization. When stability was assessed, samples were subjected to transport conditions as described in the results section. For all extractions, 50 mg of stool or equivalent was extracted using the PowerFecal® DNA Isolation Kit (MoBIO, as per HMP), PowerMag kit (MoBIO) or Repeat Bead Beating (RBB, as per IHMS). When possible, three technical replicates from paired samples were used for the comparisons. The sample size required for proper power was based on a preliminary evaluation of effect size for each CTQ to be tested and ranged from 6 to 30 donors. Confounders were controlled through distribution of experimental conditions within a singular study, such as evaluating extraction methodologies at T0.



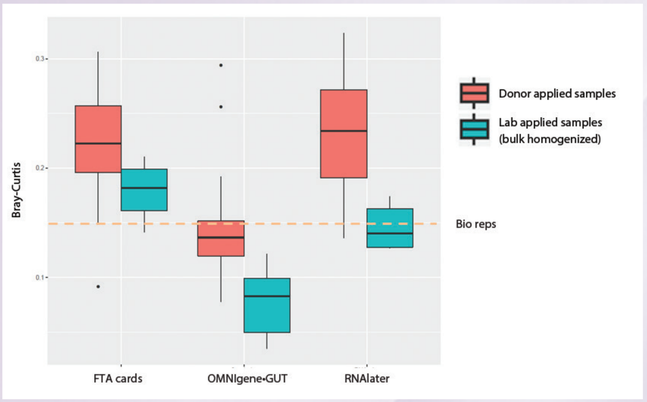
Sequencing, bioinformatics and biostatistics

16S rRNA sequencing library preparation, sequencing (Illumina® MiSeq®) and bioinformatics were conducted by Diversigen, Microbiome Discovery Service, using V4 hypervariable region paired-end amplicon sequencing or by GenoFIND™ Genomic Services (DNA Genotek) using V3-V4 hypervariable region paired-end amplicon sequencing. Sequences were quality filtered using QIIME and custom scripts. Paired-end reads were assembled and compared to the Greengenes database, clustered at 97% by UCLUST. After data normalization, alpha diversity was measured with Chao1 and observed OTU counts. Beta diversity (Bray-Curtis distances) was measured using pair-wise normalization by dividing the sum of differences by the sum of all detected OTU abundances. In all Bray-Curtis measurements, a donor matched fresh sample that had been extracted shortly after collection (control) was used as one side of the pair-wise comparison. Analysis of significant difference between Bray-Curtis Dissimilarities was performed using the Mann-Whitney test. For tracking unique donor features during stabilization, a pairwise statistical comparison (Negative binomial and Fisher's exact test) between test condition and control sample was performed at the phylum through genus levels. This was done for 6 artificial groups, each consisting of 5 donors, to establish differentially abundant taxa that represent uniqueness of donor grouping. Linear Discriminant Analysis (LDA) scores and differential OTU abundances were calculated using the LEfSe (LDA with Effect Size) algorithm[1].

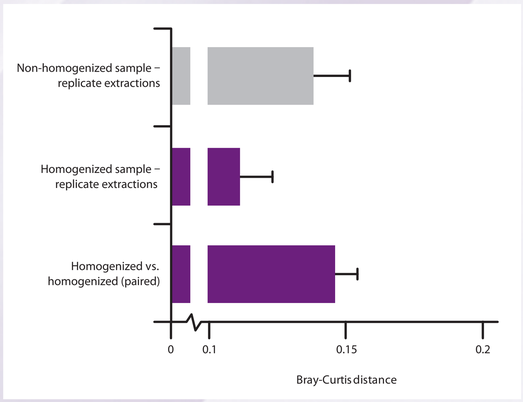
Results and discussion

1. Immediate impact of collection and stabilization methodology on microbial community composition

Sample stabilization, homogenization and donor self-collection has a significant impact on neutrality, repeatability and reproducibility CTQs



We explored the impact of three stabilization methods (FTA cards, OMNIgene•GUT and RNAlater), bulk sample homogenization and donor self-collection on the neutrality CTQ. In one study, bulk stool was homogenized in the lab by a technician prior to application to each collection method (lab applied sample)[2]. This study was independently replicated to assess the impact of donor self-collection and bulk sample homogenization on the results (donors self-collected samples at home with no bulk sample homogenization). DNA from all samples, in both studies, was extracted within 3 hours of collection (PowerFecal kit). In both studies, higher dissimilarity (Bray-Curtis) in the profile was observed in samples stabilized with FTA cards and RNAlater when compare with control samples, indicating the introduction of bias or "non-neutral stabilization". This dissimilarity was greater than that observed between biological replicates (Bray-Curtis distance of 0.15). In both studies, samples collected in OMNIgene•GUT showed the lowest dissimilarity compared to control samples, with the least dissimilarity observed in the lab applied condition, followed by the donor applied OMNIgene•GUT samples. When comparing the magnitude of dissimilarity in both studies, homogenization of the bulk sample prior to application on the stabilization method reduced the dissimilarity from control.



Post-collection sample homogenization can impact aliquot to aliquot reproducibility, a key CTQ for successful replication of results. To characterise this, fecal samples were collected from 6 donors. Paired homogenized (OMNIgene•GUT) or non-homogenized samples were extracted in triplicate within 3 hours of sample collection. Microbiome change visualized by Bray-Curtis distance was calculated by comparing replicates to each other or between methods of homogenization. A trend towards higher dissimilarity was observed in non-homogenized samples when compared with homogenized samples[3].

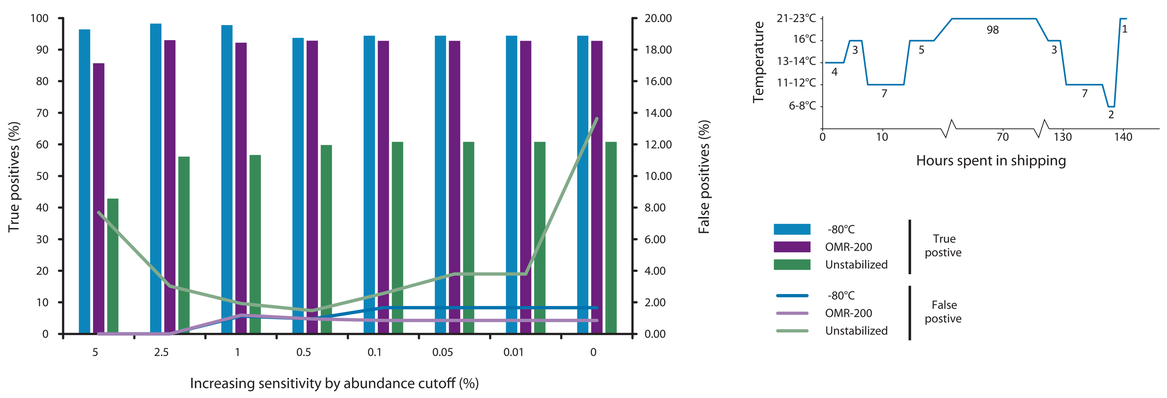
CTQ impact summary

The degree of bias introduced by each stabilization method differed. Among them, OMNIgene•GUT showed the lowest dissimilarity when compared to fresh and remained comparable to the dissimilarity expected between biological replicates. The neutrality of a stabilization method should be evaluated with naïve donors and should always include a comparison with a fresh, immediately extracted sample.

Bulk sample homogenization prior to collection improved neutrality in all stabilization methods tested. Sample homogenization prior to extraction increased the reproducibility of the microbiome profile. As bulk sample homogenization prior to collection is not well accepted by donors and is difficult to standardize outside of a laboratory setting, we do not recommend that bulk sample homogenization be included in sample collection SOPs and instead recommend a sample collection SOP that includes sample homogenization at the time of collection.

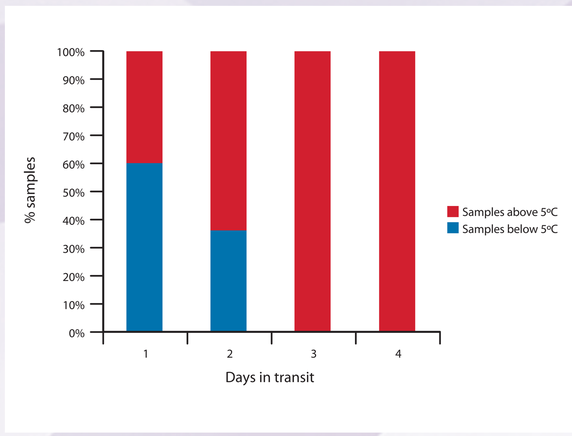
2. Effect of temperature and time during sample transport on microbial overgrowth and degradation

Ineffective stabilization during sample transport impacts stability and reproducibility CTQs causing microbiome profile drift and a reduction in statistical power



To understand the impact of temperature and time exposure during transit, stool samples were shipped at ambient temperature across Canada (Ottawa to Vancouver and back, approx. 6 days) by UPS. Stool samples were collected from 30 adult donors and an aliquot of each sample was immediately extracted (control), held at -80°C for 6 days or shipped at ambient temperature (stabilized in OMNIgene•GUT or unstabilized). Temperature exposure during transit was measured using an on-board real-time temperature tracker. The observed temperature ranged between 2 and 23 °C, supporting the requirement for sample stabilization (inset). Reference microbiome profiles were established using control samples (OTU characteristic features were identified from read counts using a negative binomial test). Shipped samples were processed upon arrival, at which point the paired -80°C stored samples were also processed and were compared to their corresponding control samples. The impact of temperature and time exposure during shipping on the microbial profile was measured using a Power analysis test (Bray-Curtis Dissimilarly was 0.403, see last figure). The percentage of True Positive (features shared with control) and False Positives (new features identified after shipping) were quantified. Abundance cut-off was based on normalized OTU read counts (0% to a ≥5% abundance). Unstabilized samples showed a significant reduction in the number of true positives and an increase in false positives. No significant differences were observed in samples held at -80°C or shipped while stabilized in OMNIgene•GUT.

The HMP cold chain shipping protocol is ineffective at maintaining optimal sample transport conditions



As it is commonly used as a method for transporting samples, the temperature control efficiency of the HMP stool collection protocol was assessed. In brief, the HMP protocol dictates samples be transported in a Styrofoam container with frozen gel packs and processed within 24 hours. In this study, 23 donors collected fecal samples, and mailed samples to a processing lab using USPS First Class Mail (1-3 business days). The temperature inside the box was measured upon arrival and was greater than 5°C in 40% of the samples after 1 day in the mail, 65% after two days and 100% after 3 or more days in the mail. Of note, 1 box arrived at ambient temperature after only 1 day in transit, which indicates that the donor did not comply with the sample collection instructions and did not completely freeze the icepacks prior to shipping the sample.

CTQ impact summary

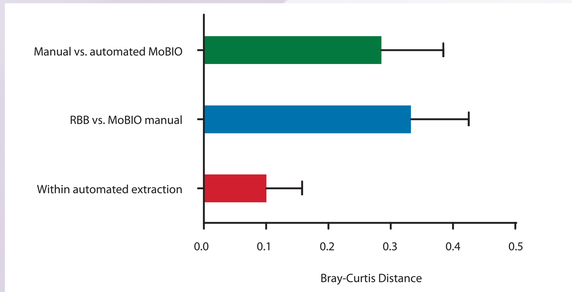
Temperature exposure during transit is highly variable and has a significant impact on the microbial profile. Current cold-chain protocols designed to mitigate these changes are labor intensive for the donor and inadequately address the realities of ambient temperature transport. In addition, they are highly inefficient, expensive and not easily standardized.

Environmental exposure during sample transport is difficult to control for and may lead to false confidence in sample integrity, highlighting the need for donor accepted sample stabilization at the point of collection. Ideally a method that does not really on cold chain, is easy for donors to comply with and is robust enough to withstand shipping delays should be used whenever possible. This is particularly relevant during the execution of epidemiology studies that rely on at home sample collection and centralized processing

We found that improper sample stabilization represents the major source of variability; therefore, we propose that stabilization is the most important pre-analytical CTQ to consider when establishing microbiome sample collection SOPs. As stabilization techniques are often analyte-specific, it is critical to ensure all analytes required for the study at hand are properly stabilized.

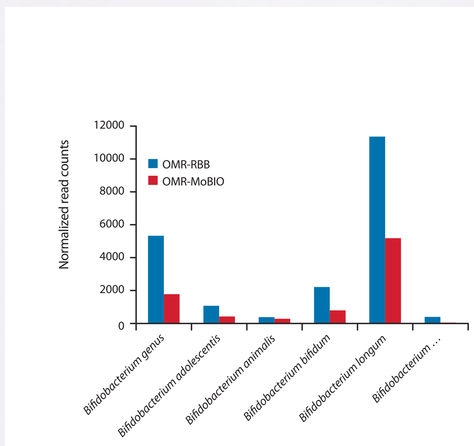
3. The impact of nucleic acid extraction method choice on microbiome profile and nucleic acid quality

Choice of DNA extraction method impacts DNA yield and integrity, bias, repeatability and reproducibility CTQ



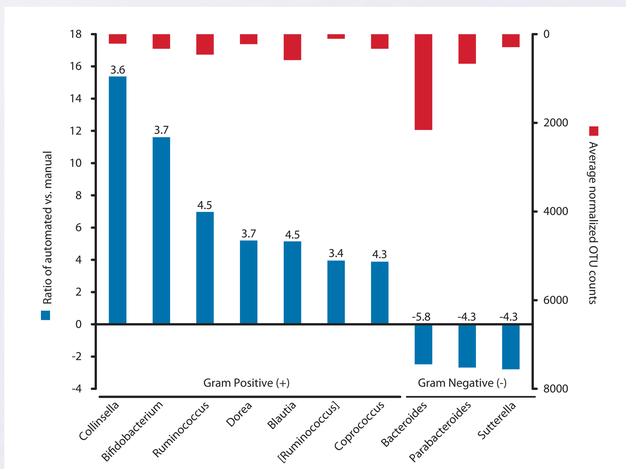
The effect of the DNA extraction method on the microbiome profile was assessed by comparing triplicate extractions using two guanidium-based methods from the same manufacturer (green bar, MoBIO manual PowerFecal vs. automated PowerMag extraction, n=12 adult samples) or manual PowerFecal vs. repeat beat-beating with alcohol precipitation (RBB) (blue bar, n=8 infant samples). Extraction method impacted DNA yield and quality, with PowerMag resulting in a higher DNA yield compared to PowerFecal in the adult samples and RBB returning a higher yield compared to PowerFecal in the infant samples (data not shown). Microbiome profile differences between extraction methods were calculated by Bray-Curtis distance. Significant dissimilarity was observed between paired samples when extracted using different methods. While such differences can be primarily explained by lysis efficiency, they could impact the relative abundance of taxa and therefore the results of a microbiome study. While evaluating the potential bias introduced by each extraction method with a mock bacterial community is the next step, some conclusions can be made from a closer comparison of each method.

Extraction method impacts "Abundance" of *Bifidobacterium* genera in infant samples



Infant fecal samples (n=8) were extracted as above. Paired extractions were performed side by side and showed consistent increase in *Bifidobacterium* taxa when extracted with RBB, highlighting the potential bias introduced by each extraction method. The increase in read counts could have a major impact in the microbiome profile (shift in the relative abundance of all OTUs). This effect is likely caused by more efficient lysis within the RBB protocol, and next steps include evaluating these protocols using mock communities.

Manual and automated extraction return discordant Gram Positive and Gram Negative relative abundances



PowerFecal (manual) and PowerMag (automated) extraction kits, were evaluated using adult fecal samples as above (n=12). Paired extractions were performed on the same day to eliminate differences due to sample storage. Ratio of taxa between the two extractions methods was generated using normalized sequencing data, values above zero indicate an increase in abundance of the taxon with PowerMag compared to PowerFecal. Data was re-analyzed using LEfSe, with LDA scores shown above the bars. Average normalized OTU counts for each taxon is presented in order to evaluate relevance of change. Biological difference between taxa is represented as gram positive and negative stain results. Significant differences were observed between manual and automated extractions; this is likely caused by differential lysis and evaluations of methods using mock communities constitutes the next steps.

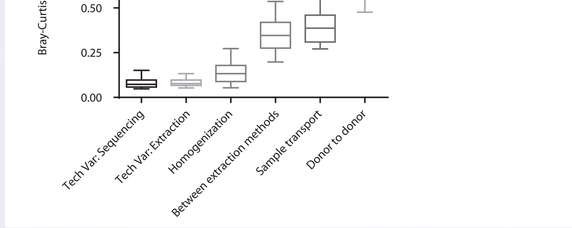
CTQ impact summary

DNA extraction methodology has a significant impact on the microbiome profile (relative abundance and diversity) in both high and low diversity samples (adult and infant samples, respectively). In both cases, the relative abundance of key taxa differed between methods.

DNA extraction methodology also impacts DNA yield and integrity, (data not shown). While low yield is not a common problem in samples from healthy adult donors it can present a considerable challenge in infant donors.

Bias introduced by the extraction method could cause inappropriate data interpretation, increased signal to noise ratio and limited reproducibility. Ideally extraction methodology would remain consistent throughout a study. Particular caution should be taken when combining DNA samples or data sets from studies where extraction methodology differs, or is not well documented.

4. Quantification of the impact of pre-analytical factors



The effect of different pre-analytical factors on the microbiome profile was analyzed. The graph shows microbiome profile differences observed between extraction methods (aliquots of sample extracted with two different kits, n=12), changes during sample shipping (control sample compared to paired unstabilized, shipped sample, n=30). Donor to donor variability was included as a control, showing maximum expected difference between microbiome profiles (n=30). Similarly, technical variability (tech var) associated with sequencing (same DNA re-sequenced in triplicates, n=6), extraction (aliquots of same sample extracted multiple times with same kit, n=12) was included to demonstrate minimum expected variability.

Conclusions

Of the workflow stages assessed, our findings indicate that sample stabilization and DNA extraction method are the most important pre-analytical processes to assess when establishing SOPs. A summary of our findings, with suggested mitigation to decrease noise and variability introduced by pre-analytical processing can be found in the following table.

Pre-analytical workflow steps	CTQ	Mitigation
Sample stabilization	Neutrality, DNA yield and integrity (data not shown), repeatability, reproducibility	Confirm stabilization method is neutral by comparing to fresh, immediately extracted paired sample.
Sample homogenization (bulk and sub sample)	Neutrality, repeatability, reproducibility	Homogenize sample post collection following a reproducible, standardized and donor accept manner.
Sample transport	Stability, bias, repeatability, reproducibility	Choose a donor accepted stabilization method that does not rely on cold chain and is robust enough to withstand shipping delays.
Storage temperature, including freeze-thaw cycles (data not shown)	Stability, bias, repeatability, reproducibility, DNA yield and integrity	Unless stabilization method provides protection, store at -80°C and limit number of freeze-thaws.
Nucleic acid extraction	DNA yield and integrity (data not shown), bias, repeatability, reproducibility	Ensure extraction method does not introduce bias and provides enough high molecular weight DNA for down stream processing; extraction method should be consistent across all samples in a study.

References

[1] N. Segata, J. Izard, L. Waldron, D. Gervers, L. Miropolsky, W. Garrett and C. Huttenhower, "Metagenomic biomarker discovery and explanation," *Genome Biology*, vol. 12, p. R60, 2011.

[2] S. J. Song, A. Amir, J. L. Metcalf, K. R. Amato, Z. Z. Zu, G. Humphrey and R. Knight, "Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies," *mSystems*, vol. 1, no. 3, pp. e00021-16, 2016.

[3] E. Doukhanine, A. Bouevitch, L. Pozza and C. Merino, "OMNIgene•GUT enables reliable collection of high quality fecal samples," DNA Genotek., pp. PD-WP-00040, 2014.

DNA genotek

Superior samples
Proven performance

www.dnagenotek.com • support@dnagenotek.com