Application Note

Speed meets accuracy: Streamline your microbiome NGS using QIAseq® FX DNA Library Kits with integrated library normalization

Markus Storbeck[1], Dylan Barbera[2], Kerstin Goebels[1], Eric Lader[2] and Peter Hahn[1]

[1] QIAGEN GmbH, Hilden, Germany; [2] QIAGEN Sciences Inc. Frederick, MD, USA

Introduction

Knowledge of the human microbiome and its impact on our health and disease pathogenesis is paving the way for new treatment modalities for various diseases (1). Modern approaches to studying microbiomes no longer involve microbial cultivation but rely on metagenomics, where you analyze the combined genetic material in a sample via next-generation sequencing (NGS).

Metagenomics avoids biases inherent to cultivation regimens, as nucleic acids from bacteria, fungi, viruses and other single-cell organisms representing the microbiome composition are directly isolated and analyzed by NGS.

Yet, microbiome profiling using NGS is prone to biases introduced at multiple points in the workflow, such as ▷

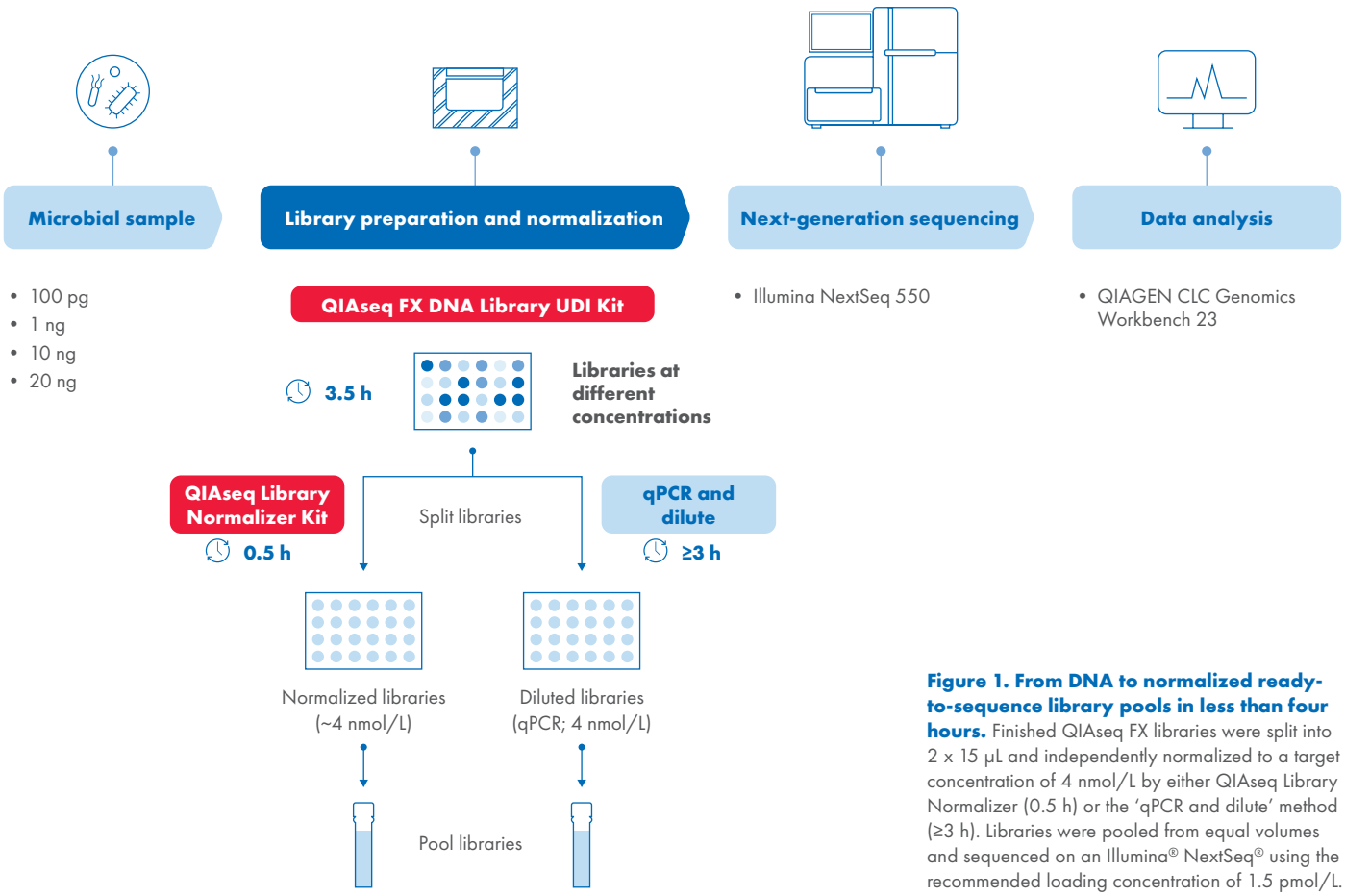


Figure 1. From DNA to normalized ready-to-sequence library pools in less than four hours. Finished QIAseq FX libraries were split into 2 x 15 µL and independently normalized to a target concentration of 4 nmol/L by either QIAseq Library Normalizer (0.5 h) or the 'qPCR and dilute' method (≥3 h). Libraries were pooled from equal volumes and sequenced on an Illumina® NextSeq® using the recommended loading concentration of 1.5 pmol/L.

sample collection and storage (2,3), nucleic acid isolation (4,5), NGS library preparation (6,7), sequencing, and bioinformatic analysis (8). For example, low accuracy of DNA quantification, high GC bias, and high intra- and inter-laboratory variabilities threaten the suitability of NGS for microbiome profiling.

Here we describe the QIAseq FX library preparation workflow with integrated library normalization for accurate and reproducible microbiome NGS (Figure 1). With only 30 minutes of added workflow time, the QIAseq Library Normalizer Kit improves reproducibility by achieving homogenous read representation without requiring qPCR-based library quantification. Using a mock microbial community that includes species with vastly different GC content and genome sizes in a streamlined NGS workflow, we demonstrate how you can achieve highly accurate taxonomic profiling without compromising on speed.

Materials and Methods

Sample

For taxonomic profiling, we used a 20 Strain Staggered Mix Genomic Material (MSA-1003™ from ATCC®) composed of 20 different microbial species at variable abundances ranging from 0.02–18.0%. The genome sizes varied from 1.6–6.3 megabases, with relative GC contents ranging from 29.9–69.1%.

Library preparation and normalization

Whole genome shotgun libraries were prepared using QIAseq FX DNA Library UDI Kits (QIAGEN; cat. no.: 180479–180482). The kit uses a single-tube combined enzymatic DNA fragmentation and end-repair reaction enabling library preparation in only 3.5 hours. To benchmark the detection limit for less abundant microbes, libraries were prepared from as little starting material as 100 pg (Figure 1).

In the enrichment PCR, adapter-ligated libraries were amplified using QIAGEN's HiFi PCR Master Mix included in the QIAseq FX kits while using QIAseq Normalizer Primer Mix from the QIAseq Library Normalizer Kits (QIAGEN; cat. no.: 180603/180605). Half of the finished library volume (15 of 30 µL) was normalized to a concentration of 4 nmol/L using the fast 30-minute QIAseq Library Normalizer protocol. The remaining half of the finished library was quantified by qPCR and diluted to 4 nmol/L using the longer 3-hour protocol ('qpCR and dilute' method).

NGS

The two normalized library pools generated using either method were sequenced on an Illumina NextSeq 550 instrument (mid-output flow cell, 2 x 150 bp paired-end sequencing).

Data analysis

Sequence data were mapped to reference genomes of microbes present in the mock community using QIAGEN® CLC Genomics Workbench 23. The relative abundance of any species in the mock community was inferred from the fraction of reads mapping to the respective genome and compared to the expected microbial abundance within the MSA-1003 mock community.

Results and Discussion

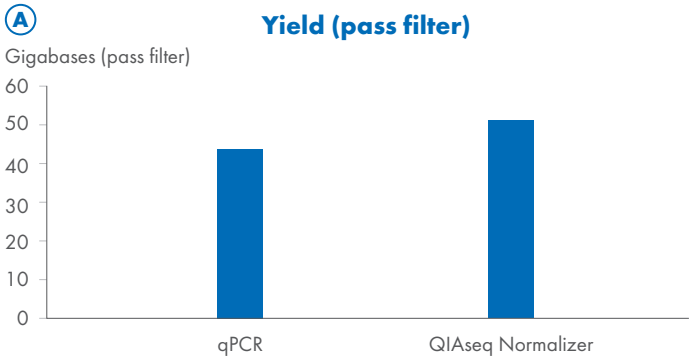
QIAseq Normalizer yielded a more balanced sequence read representation than the 'qPCR & dilute' method

Both QIAseq Normalizer and 'qPCR & dilute' methods exceeded the specified sequence data output of 32–39 Gb (Figure 2A) with more than 75% of bases called of high quality (≥Q30) (Figure 2B). Due to a higher clustering density, QIAseq Normalizer had a slightly smaller fraction of ≥Q30 reads but yielded more usable sequence data in total compared to qPCR-based normalization.

The distribution of sequence reads per library differed considerably between the two normalization methods (Figure 2D). Libraries normalized using the 30-minute QIAseq Library Normalizer protocol were homogeneously distributed across all used library input amounts (CV=15.1%; GINI=0.076). In contrast, qPCR-normalized libraries were unbalanced, with two overrepresented libraries accounting for two-thirds of all reads (CV=86.2%; GINI=0.392) (Figure 2C).

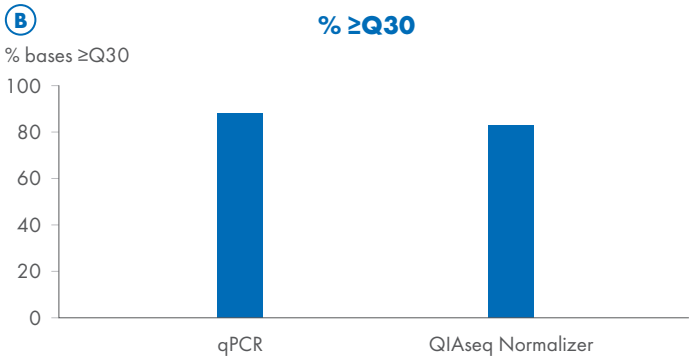
Although widely used, the 'qPCR and dilute' normalization method can be susceptible to errors leading to under- or over-representation of some libraries. For example, the libraries with 10 ng and 20 ng input yielded very high concentrations of 330 and 350 nmol/L, respectively. Either both libraries were quantified as too low in concentration by qPCR, or the dilution from 330/350 nmol/L to 4 nmol/L was not accurate enough. QIAseq Library Normalizer, on the other hand, reduced the highly concentrated libraries to levels similar to other libraries in the sequencing pool (31–82 nmol/L initial concentration).

These results indicate that the QIAseq Library Normalizer Kit is an accurate yet fast and easy-to-use tool to normalize libraries across a broad range of concentrations. The resulting library pools are well-balanced, and the total pool concentration assures optimal flow cell loading and high sequencing quality.



C

Variation	qPCR pool	QIAseq Normalizer pool
Max/Min	6.36	1.48
CV%	86.2%	15.1%
GINI	0.392	0.076



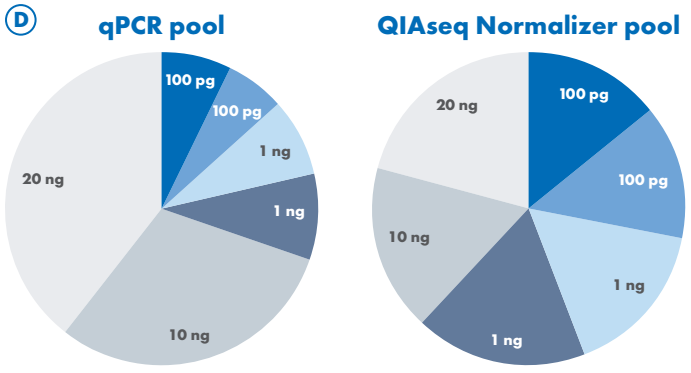


Figure 2. QIAseq Library Normalizer enables balanced read representation and optimal flow cell loading. A Both normalization methods yielded more than 40 Gb of sequence data and clearly exceeded the specified output of 32–39 Gb. **B** More than 80% of bases were called at high quality ≥Q30 (≥75% are specified). **C** Normalization metrics. The GINI coefficient is a value between 0 and 1, where 0 means all libraries have the same read count, and 1 means a single library has all the reads. **D** Distribution of sequence reads after normalization by qPCR and QIAseq Normalizer. Data labels indicate the input DNA amount of the library preparation.

Unbiased taxonomic profiling: QIAseq FX–Normalizer combination accurately characterized the microbial mock community

All 20 microbes present in the MSA-1003 mock community were successfully identified from sequence data and the estimated abundances were highly similar to the values expected for the mock community. In particular, abundance estimates maintain high precision across all reference genomes with variable GC content from 29.9% to 69.1% and allow reliable detection of not just highly abundant genomes but also genomes representing only 0.1% or less of the microbiome (Figure 3). This strongly suggests that the QIAseq FX library preparation workflow effectively fragments DNA irrespective of the GC content and preserves fragments of low-abundant genomes due to its high library conversion rate. Neither enzymatic fragmentation nor PCR-based library amplification appears to introduce a significant GC bias based on the accurate quantification of high- and low-GC genomes.

Since the same libraries were used in both normalization protocols, we could determine whether QIAseq Normalizer introduced any sequencing bias. Both normalization methods yielded identical results, indicating that the QIAseq Normalizer protocol does not introduce bias to the library composition compared to merely diluting libraries. Also, QIAseq Normalizer does not alter the composition of low- and high-input libraries. The standard deviation of microbial abundance estimates between libraries with 100 pg to 20 ng DNA input were identical for all species, regardless of the normalization method.

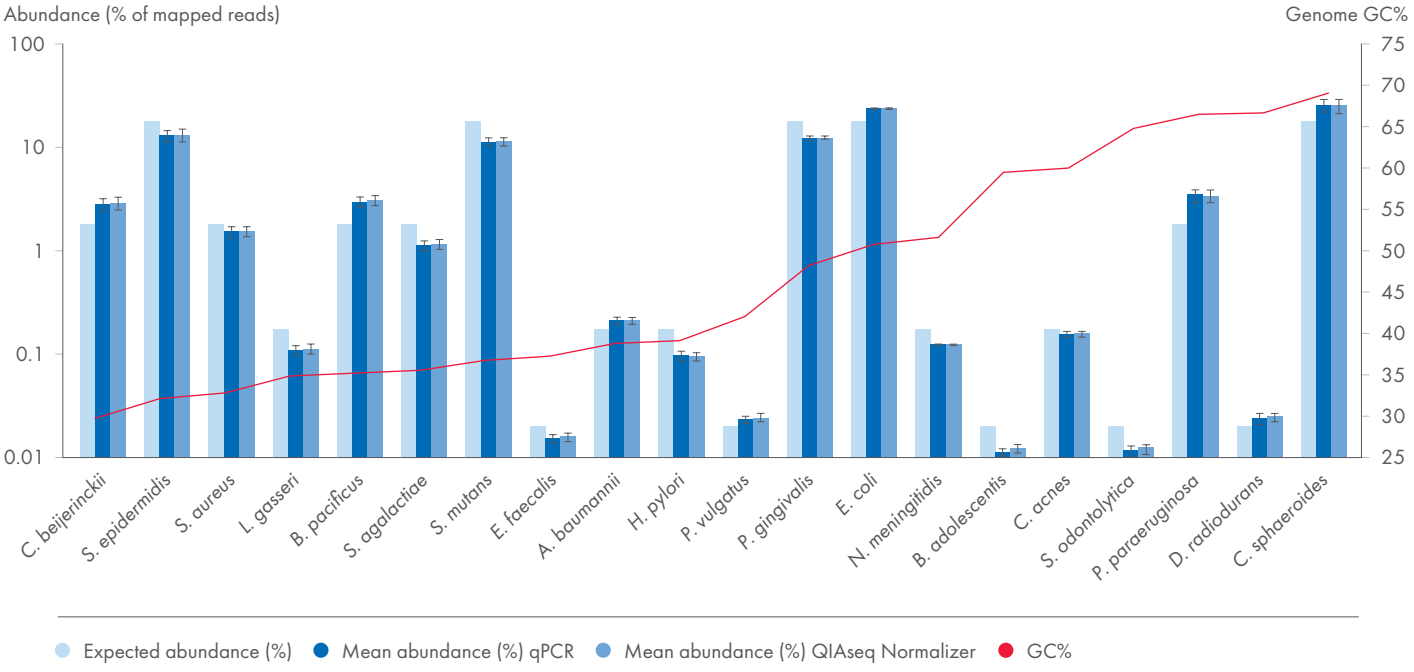


Figure 3. Microbial profile of the MSA-1003 (ATCC) mock community. Light blue bars indicate the expected microbial abundance for the MSA-1003 mock community (log scale). Blue (qPCR) and middle blue (QIAseq Normalizer) bars show the mean abundances calculated from sequence data of libraries generated from 100 pg, 1 ng, 10 ng and 20 ng input DNA. Error bars indicate the standard deviation between libraries. The 20 mock community species are shown in order of increasing genome GC content.

Conclusion

NGS has become the go-to method for microbial profiling. Unbiased taxonomic profiling can be rapidly carried out in a high-throughput manner while eliminating biases inherent to microbial cultivation. This study evaluated QIAseq FX DNA Library Kits with integrated library normalization for taxonomic profiling of a mock microbial community.

This study corroborates the independent assessment of Tourlousse and colleagues, who recommended the QIAseq FX workflow for taxonomic profiling based on its high accuracy, low GC bias and excellent transferability across laboratories owing to the use of enzymatic DNA fragmentation (7). Adding the QIAseq Normalizer protocol to the QIAseq FX workflow provides homogeneous ready-to-sequence libraries without tedious qPCR-based normalization. With just 30 minutes of added workflow time, the QIAseq FX–Normalizer workflow offers a faster and more economical way to profile microbial communities accurately.

References

1. Sorbara MT, Pamer EG. Microbiome-based therapeutics. *Nat Rev Microbiol.* 2022;20(6):365–380.
2. Choo JM, Leong LE, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep.* 2015;5:16350.
3. Watson EJ, Giles J, Scherer BL, Blatchford P. Human faecal collection methods demonstrate a bias in microbiome composition by cell wall structure. *Sci Rep.* 2019;9(1):16831.
4. 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.
5. Yang F, et al. Assessment of fecal DNA extraction protocols for metagenomic studies. *Gigascience.* 2020;9(7):giaa071.
6. Clooney AG, et al. Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis. *PLoS One.* 2016;11(2):e0148028.
7. Tourlousse DM, et al. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome.* 2021;9(1):95.
8. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic classification. *Cell.* 2019;178(4):779–794.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at **www.qiagen.com** or can be requested from QIAGEN Technical Services or your local distributor.

Scan the QR codes or use the links below to learn more.

NGS library preparation for metagenomics: **www.qiagen.com/QIAseq-FX**

NGS library normalization: **www.qiagen.com/QIAseq-Normalizer**