# Quantitative PCR provides a simple and accessible method for quantitative microbiome profiling

Running title: qPCR for quantitative NGS

Ching Jian[1], Panu Luukkonen[2,3], Hannele Yki-Järvinen[2,3], Anne Salonen[1] and Katri Korpela[1*]

[1]*Immunobiology Research Program, Department of Bacteriology and Immunology, University of Helsinki, Helsinki* 00014*, Finland.*

[2]*Minerva Foundation Institute for Medical Research, Helsinki* 00014*, Finland.*

[3]*Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki* 00014*, Finland.*

**Corresponding author**

[*]Katri Korpela, katri.korpela@helsinki.fi, Haartmaninkatu 3, PO box 21, FI-00014, University of Helsinki, Helsinki, Finland

**Contributions**

C.J. performed the laboratory analyses and drafted the manuscript.

A.S. supervised the laboratory work, reviewed and edited the manuscript.

H.Y.-J. and P.L. designed and conducted the clinical trial from where the fecal samples originated.

K.K. conceived the study, performed the statistical analyses, reviewed and edited the manuscript.

25 ## Abstract

26 The use of relative next generation sequencing (NGS) abundance data can lead to
27 misinterpretations of microbial community structures as the increase of one taxon leads to
28 concurrent decrease of the other(s). To overcome compositionality, we provide a quantitative
29 NGS solution, which is achieved by adjusting the relative 16S rRNA gene amplicon NGS data
30 with quantitative PCR (qPCR-based) total bacterial counts. By comparing the enumeration of
31 dominant bacterial groups on different taxonomic levels in human fecal samples using taxon-
32 specific 16S rRNA gene-targeted qPCR we show that quantitative NGS is able to estimate
33 absolute bacterial abundances accurately. We also observed a higher degree of correspondence
34 in the estimated microbe-metabolite relationship when quantitative NGS was applied. Being
35 conceptually and methodologically analogous to amplicon-based NGS, our qPCR-based
36 method can be readily incorporated into the standard, high-throughput NGS sample processing
37 pipeline for more accurate description of interactions within and between the microbes and host.
38

39 ## Main text

40 Microbiota datasets arising from NGS are compositional in nature: the relative abundances of
41 the taxa always sum up to 1 (1). Since the changes of components are mutually dependent,
42 erroneous conclusions may occur if data are analyzed using traditional statistical methods.
43 Compositionality particularly hampers the interpretation of microbiota changes in longitudinal
44 studies, such as interventions. If a single taxon increases or decreases in relative abundance, the
45 relative abundances of other taxa will show a corresponding, opposite change. It becomes
46 impossible to determine which taxon was truly affected by an intervention (Supplementary Fig.
47 S1). Contrary to the speculation that compositionality is dismissible in high complexity
48 environments (2), our simulations revealed that the compositionality effects may lead to
49 extensive false positive findings in samples containing complex microbial communities (e.g.,

50 gut) as well as samples with low diversity (e.g., vaginal microbiome) (Supplementary Fig. S2
51 and S3). Complex analytical methods have been developed in an effort to mitigate the effect of
52 mutual dependence of component changes in compositional microbiota sequencing datasets (2,
53 3). However, in addition to the technical problems of compositional data analysis, there is
54 reason to believe that the absolute abundances of bacteria are a biologically meaningful measure
55 and relying solely on relative abundances results in the omission of important information on
56 the interactions of different taxa with each other and the host (4). Absolute quantification of
57 microbial abundances based on flow cytometry has been applied to complement amplicon
58 sequencing and shown to more precisely describe the temporal dynamics of the microbial
59 community in an engineered freshwater ecosystem (5). Vandeputte et al. (6) developed an
60 integrative workflow that combines flow cytometry-based bacterial enumeration and NGS for
61 fecal samples. Also spike-in bacteria (7) or DNA (8) has been used for the purpose of
62 quantification of microbiota NGS profiles. For fungi, also qPCR has been employed to
63 transform the relative abundance data from pyrosequencing to absolute values (9).
64 Notwithstanding the increased awareness of compositionality and the recent advances in
65 quantitative microbiome profiling, the majority of human microbiome studies to date have not
66 adopted any countermeasures. We believe the practicability and accessibility of many
67 previously proposed solutions represent obstacles for the wider adaptation of quantitative
68 microbiome profiling in human microbiome research. In particular, the flow cytometry-based
69 approach requires considerable expertise for reproducible results, considering flow cytometric
70 enumeration of microbial cells was initially restricted to pure cultures (10) and still remains
71 challenging when performed in complex matrices (11).

72 To render quantitative microbiome profiling from NGS data easy and accessible, we devised
73 and validated a 16S rRNA gene-targeted qPCR-based quantification method. qPCR has been
74 widely used for accurate and targeted quantification of specific bacterial groups and species

75 especially until the era of NGS. Here, we first quantified total bacteria using universal bacterial
76 primers (12) by qPCR (Supplementary Methods) in 114 adult fecal DNA samples that have
77 been analyzed for microbiota composition using Illumina MiSeq for 16S rRNA gene amplicon
78 sequencing (13). The qPCR threshold cycle (Ct) values were converted to the estimates of
79 bacterial genomes present in 1 g of feces (total bacterial counts), which were used to estimate
80 absolute abundance of the NGS-detected taxa (Fig. 1) by a simple equation:

81 $$Relative\ abundance\ of\ taxon\ A\ (MiSeq) \times Total\ bacterial\ counts\ (qPCR)$$
82 $$=\ Estimated\ quantitative\ abundance\ of\ taxon\ A$$

83 Next, we validated the estimated quantitative abundances of four representative taxa by qPCR
84 using the taxon-specific primers (Supplementary Table 1) for the phylum Bacteroidetes,
85 *Clostridium* cluster XIVa (family *Lachnospiraceae*), genus *Bifidobacterium* and *Escherichia*
86 *coli* species. Absolute quantification was achieved by using standard curves. We found near-
87 perfect correlations between the estimated quantitative abundances and qPCR abundances in
88 all tested taxa (Fig. 1b), indicating that qPCR-adjusted NGS (hereafter termed quantitative NGS)
89 profiles are able to reflect absolute bacterial abundances in feces as accurately as taxon-specific
90 qPCR. The correspondence decreases at the very low end of the abundance range, likely due to
91 the relatively lower PCR amplification efficiency and increased stochasticity of low abundance
92 taxa (14). The applied library preparation method (dual index TruSeq-tailed 1-step
93 amplification (15)) causes a slight underestimation of Bacteroidetes abundance (unpublished
94 data), explaining the underestimation observed for this phylum compared to qPCR (Fig. 1b).
95

96 To determine whether the quantitative NGS data can be used to more precisely estimate the
97 abundance of specific gut microbiota functions, we correlated quantitative NGS abundance of
98 butyrate-producing bacteria to the abundance of the butyryl-CoA:acetate CoA-transferase gene
99 determined by qPCR (16). As expected, quantitative NGS data explained a much larger

100 proportion of the variation in butyryl-CoA gene abundance than the relative abundance of the
101 butyrate producers (variance explained 0.47 versus 0.23, Fig. 2).
102
103 We contend that our qPCR-based quantitative NGS enjoys conceptual and practical benefits
104 over other alternatives for quantification of the bacterial load in the NGS samples. First and
105 foremost, the same DNA extract serves as the starting material both for qPCR and NGS, making
106 qPCR easy to implement in the workflow for high-throughput analysis of up to thousands of
107 microbiome samples. Second, unlike flow cytometry that counts cells, qPCR and NGS both
108 target bacterial DNA, including extracellular DNA derived from lysed bacteria. Extracellular
109 DNA can be intrinsic or result from the differential lysis of Gram-positive and negative bacteria
110 during the common freeze-thawing prior to fecal DNA extraction. As the 16S profiles from the
111 gut appear very different for intracellular and extracellular DNA (17), qPCR is expected to
112 reflect the NGS-targeted community structure both quantitatively and qualitatively more
113 closely than flow cytometry. Third, qPCR is cost-effective and more accessible as the
114 laboratory settings, machinery and reagents are essentially the same as those needed for
115 preparing the NGS libraries. While PCR-based methods can be potentially biased e.g. due to
116 inadequate DNA extraction or primer coverage, these factors play a similar role in the NGS
117 itself (18). Simplicity is another strength, as no spikes or other exogenous controls are needed.
118 Lastly, our quantitative microbiome profiling approach is compatible with any downstream
119 bioinformatic pipelines, since it requires no complicated transformation or computation.
120
121 In conclusion, we caution against the analysis of microbiome NGS data solely relying on
122 relative abundances, since compositionality may skew biological inferences from microbiota
123 studies *per* our simulation data as well as the previously published studies. Although relative
124 taxon abundance can be indicative at times, absolute quantification is necessary for obtaining a

125 comprehensive understanding of the dynamics and interactions of the microbiome. To facilitate
126 the standard use of quantitative microbiome data, we propose a qPCR-based quantitative NGS
127 approach that entails a single additional amplification reaction and can be integrated into an
128 NGS workflow seamlessly. While our proposed approach does not solve the limitations and
129 biases inherent to the NGS technology, we believe our reductionist solution to compositionality
130 offers researchers improved biological insight from their microbiota sequencing datasets.

131

132 **Acknowledgements**

133 We thank Professor Willem de Vos for comments on the manuscript.

134

135 **Competing interests**

136 The authors declare no competing financial interests.

137

138 ## References

139 1. Aitchison J. The Statistical Analysis of Compositional Data. J R Stat Soc Series B Stat
140 Methodol. 1982;44(2):139-77.
141 2. Tsilimigras MC, Fodor AA. Compositional data analysis of the microbiome:
142 fundamentals, tools, and challenges. Ann Epidemiol. 2016;26(5):330-5.
143 3. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are
144 Compositional: And This Is Not Optional. Front Microbiol. 2017;8:2224.
145 4. Lewis ZT, Totten SM, Smilowitz JT, Popovic M, Parker E, Lemay DG, et al. Maternal
146 fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants.
147 Microbiome. 2015;3:13.
148 5. Props R, Kerckhof FM, Rubbens P, De Vrieze J, Hernandez Sanabria E, Waegeman W,
149 et al. Absolute quantification of microbial taxon abundances. ISME J. 2017;11(2):584-7.

6. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. Nature. 2017;551(7681):507-11.

7. Stammler F, Glasner J, Hiergeist A, Holler E, Weber D, Oefner PJ, et al. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome. 2016;4(1):28.

8. Tkacz A, Hortala M, Poole PS. Absolute quantitation of microbiota abundance in environmental samples. Microbiome. 2018;6(1):110.

9. Dannemiller KC, Lang-Yona N, Yamamoto N, Rudich Y, Peccia J. Combining real-time PCR and next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations. Atmospheric Environment. 2014;84:113-21.

10. Czechowska K, Johnson DR, van der Meer JR. Use of flow cytometric methods for single-cell analysis in environmental microbiology. Curr Opin Microbiol. 2008;11(3):205-12.

11. Frossard A, Hammes F, Gessner MO. Flow Cytometric Assessment of Bacterial Abundance in Soils, Sediments and Sludge. Front Microbiol. 2016;7:903.

12. Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology. 2002;148(Pt 1):257-66.

13. Luukkonen PK, Sadevirta S, Zhou Y, Kayser B, Ali A, Ahonen L, et al. Saturated Fat Is More Metabolically Harmful for the Human Liver Than Unsaturated Fat or Simple Sugars. Diabetes Care. 2018;41(8):1732-9.

14. Gonzalez JM, Portillo MC, Belda-Ferre P, Mira A. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. PLoS One. 2012;7(1):e29973.

174 15. Raju SC, Lagstrom S, Ellonen P, de Vos WM, Eriksson JG, Weiderpass E, et al.
175 Reproducibility and repeatability of six high-throughput 16S rDNA sequencing protocols for
176 microbiota profiling. J Microbiol Methods. 2018;147:76-86.

177 16. Louis P, Flint HJ. Development of a semiquantitative degenerate real-time pcr-based
178 assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in
179 complex bacterial samples. Appl Environ Microbiol. 2007;73(6):2009-12.

180 17. Qi C, Li Y, Yu RQ, Zhou SL, Wang XG, Le GW, et al. Composition and immuno-
181 stimulatory properties of extracellular DNA from mouse gut flora. World J Gastroenterol.
182 2017;23(44):7830-9.

183 18. Bonk F, Popp D, Harms H, Centler F. PCR-based quantification of taxa-specific
184 abundances in microbial communities: Quantifying and avoiding common pitfalls. J Microbiol
185 Methods. 2018;153:139-47.

186

187

188

189

190

191

192

193

194

195

196

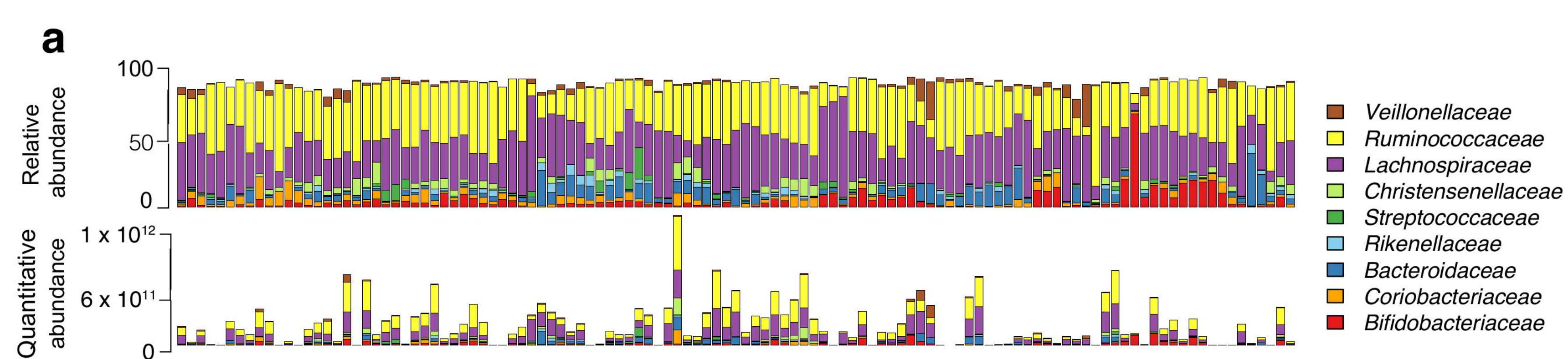
197

198 **Figure Legends**

199 **Figure 1: Relative microbiota profiles translated into quantitative microbiota profiles**
200 **using quantitative NGS.**
201 **a**, Comparison of relative abundance and quantitative sequencing abundance of dominant
202 bacterial families in 114 fecal samples. **b**, Correlation between the qPCR abundance (16S rRNA
203 gene copies per g feces) and the quantitative sequencing abundance of four taxa representing
204 the species, genus, family and phylum levels. The dashed line shows the expected 1:1
205 correspondence.

206

207 **Figure 2: Association between the abundance of the butyryl-CoA:acetate CoA-transferase**
208 **gene, measured by qPCR, and the abundance of butyrate producers.**
209 **a**, Sequencing-based quantitative abundance. **b**, Relative abundance. The variance explained
210 and the p-value are from a linear regression.




b

E. coli correlation 0.95
Bifidobacterium correlation 0.93
Lachnospiraceae correlation 0.97
Bacteroidetes correlation 0.89

Quantitative Sequencing abundance

$1.1 \times 10^{13}$
$7.2 \times 10^{10}$
$4.9 \times 10^{8}$
$3.3 \times 10^{6}$
$2.2 \times 10^{4}$

$2.2 \times 10^{4}$ $3.3 \times 10^{6}$ $4.9 \times 10^{8}$ $7.2 \times 10^{10}$ $1.1 \times 10^{13}$

qPCR abundance

**a**

Variance explained 47%

$p = 1.63 \times 10^{-11}$

Butyryl–CoA

0.025 0.030 0.035 0.040

$8.9 \times 10^{6}$ $1.8 \times 10^{8}$ $3.6 \times 10^{9}$ $7.2 \times 10^{10}$

Quantitative abundance of butyrate producers

**b**

Variance explained 23%

$p = 1.08 \times 10^{-5}$

Butyryl–CoA

0.025 0.030 0.035 0.040

0.1 0.2 0.3 0.4 0.5 0.6

Relative abundance of butyrate producers