1　**ORIGINAL ARTICLE**

2

3　**Development and evaluation of a culture-free microbiota profiling platform (MYcrobiota) for**

4　**clinical diagnostics**

5

6　Stefan A. Boers[1, *], Saskia D. Hiltemann[2, *], Andrew P. Stubbs[2], Ruud Jansen[3] and John P. Hays[1, #]

7

8　[1]Department of Medical Microbiology and Infectious Diseases, Erasmus University Medical Centre

9　Rotterdam (Erasmus MC), Rotterdam, the Netherlands

10　[2]Department of Bioinformatics, Erasmus University Medical Centre Rotterdam (Erasmus MC),

11　Rotterdam, the Netherlands

12　[3]Department of Molecular Biology, Regional Laboratory of Public Health Kennemerland, Haarlem, the

13　Netherlands

14

15

16　# Corresponding author:

17　Dr. J.P. Hays, Department of Medical Microbiology and Infectious Diseases, Erasmus University

18　Medical Centre Rotterdam, Wytemaweg 80, 3015 CN, Rotterdam, the Netherlands. Telephone: +31

19　107032177, fax: + 31 107033875, email: j.hays@erasmusmc.nl

20

21　* These authors contributed equally to this work.

ABSTRACT

Microbiota profiling has the potential to greatly impact on routine clinical diagnostics by detecting DNA derived from live, fastidious and dead bacterial cells present within clinical samples. Such results could potentially be used to benefit patients by influencing antibiotic prescribing practices, or to generate new classical-based diagnostic methods e.g. culture or PCR. However, technical flaws in 16S rRNA gene next-generation sequencing (NGS) protocols, together with the requirement for access to bioinformatics, currently hinders the introduction of microbiota analysis into clinical diagnostics. Here we report on the development and evaluation of an 'end-to-end' microbiota profiling platform (MYcrobiota), which combines our previously validated micelle PCR/NGS (micPCR/NGS) methodology with an easy-to-use, dedicated bioinformatics pipeline. The newly designed bioinformatics pipeline processes micPCR/NGS data automatically and summarizes the results in interactive, but simple web reports. In order to explore the utility of MYcrobiota in clinical diagnostics, 47 clinical samples (40 'damaged skin' samples and 7 synovial fluids) were investigated using routine bacterial culture as comparator. MYcrobiota confirmed the presence of bacterial DNA in 37/37 culture-positive samples and detected bacterial taxa in 2/10 culture-negative samples. Moreover, 36/38 potentially relevant aerobic bacterial taxa and 3/3 mixtures of anaerobic bacteria were identified using culture and MYcrobiota, with the sensitivity and specificity being 95%. Interestingly, the majority of the 448 bacterial taxa identified using MYcrobiota were not identified using culture, which could potentially have an impact on clinical decision-making. Taken together, the development of MYcrobiota is a promising step towards the introduction of microbiota analysis into clinical diagnostic laboratories.

INTRODUCTION

The detection, identification and further characterization of pathogenic microorganisms is the major step in establishing appropriate (antibiotic) treatment for infectious diseases. However, the causative microorganism of an infection may not always be detected using current 'gold standard' culturing techniques. Further, most molecular-based detection methods e.g. PCR, require *a priori* knowledge of

2

the potential pathogen before a test is performed. To overcome these limitations, the bacterial composition can be defined and genera identified using a culture-free, broad-range PCR strategy that targets the prokaryotic 16S rRNA gene followed by next-generation sequencing (NGS) [1]. However, to date, 16S rRNA gene NGS methods to profile microbial compositions have been focussed on research questions mostly, with only a few studies having evaluated the utility of 16S rRNA gene NGS methods for clinical microbiology [2, 3]. Currently, the utilisation of 16S rRNA gene NGS methods within routine clinical diagnostics has been hindered by issues relating to the generation of PCR artefacts (e.g. chimera formation and PCR competition) and the susceptibility of 16S rRNA gene NGS methods to DNA contamination that is derived from the laboratory environment and/or the reagents/consumables used. These limitations hinder the standardization of current 16S rRNA gene NGS methods to such an extent that non-identical microbiota results may be obtained when repeatedly analysing the same sample [4].

Recently, the authors published a micelle PCR/NGS (micPCR/NGS) methodology that limits the formation of chimeric sequences and prevents PCR competition via the clonal amplification of targeted 16S rRNA gene molecules [5]. In addition, the micPCR/NGS methodology allows for the utilization of an internal calibrator (IC) to calculate the number of 16S rRNA gene copies for each individual operational taxonomic unit (OTU) present within a (clinical) sample, which conveniently enables the subtraction of contaminating bacterial DNA via the quantification of 16S rRNA gene copies within negative extraction control (NEC) samples. The authors showed that the microbiota results obtained using micPCR/NGS possess a much higher accuracy (precision and trueness) compared to traditional 16S rRNA gene NGS protocols and that the ability to determine and subtract contaminating 16S rRNA gene copies, results in contamination-free quantitative microbiota profiles – with a limit of detection (LOD) of only 25 16S rRNA gene copies per OTU [6]. This low LOD allows for the detection of bacterial OTUs at very low abundances, or can confirm the absence of 16S rRNA gene copies in culture-negative results. Based on these findings, the authors suggested that the micPCR/NGS protocol could possess distinct advantages when processing clinical samples for microbiota profiling compared to traditional (semi-quantitative) 16S rRNA gene NGS methods that remain vulnerable to false-positive results (e.g. chimeric sequences or contaminant DNA) and inaccurate measurements of the OTU relative abundances in polymicrobial clinical samples due to template-specific variations in PCR efficiencies (i.e. PCR competition). However, the analysis of 16S rRNA gene NGS data depends on the use of bioinformatics

82 tools that are complex for non-bioinformatics educated technicians/clinicians to utilize and the required

83 bioinformatics skills are nowadays mostly absent in clinical diagnostic laboratories.

84

85 In this publication, we designed an easy-to-use bioinformatics pipeline to determine bacterial taxa from

86 16S rRNA gene sequences that together with the micPCR/NGS strategy is part of an 'end-to-end'

87 microbiota profiling platform (MYcrobiota). The bioinformatics pipeline enables the full analyses of the

88 NGS data obtained, from raw sequence files to final web reports that summarizes the quantitative

89 microbiota results, without the knowledge of command-line scripts that would normally be required by

90 16S rRNA gene NGS users. As a proof of principle, we explored the utility of MYcrobiota for use in the

91 clinical diagnostic laboratory by processing a total of 47 clinical samples and then comparing the results

92 to conventional 'gold standard' culture results. The samples tested included 40 specimens that were

93 obtained from a variety of damaged skin conditions for which a polymicrobial biomass was expected,

94 and an additional 7 specimens, obtained from patients who were suspected of having (prosthetic) joint

95 infections, for which a low bacterial biomass was expected.

96

97 **MATERIALS AND METHODS**

98 **Ethics statement**

99 An acknowledged national ethics committee from the Netherlands (Medisch Ethische

100 Toetsingscommissie Noord-Holland, http://www.metc.nl) approved the study protocol (M015–021) and

101 all experiments were performed on leftover material of the included clinical samples in accordance with

102 the relevant guidelines and regulations. The national ethics committee waived the need for participant

103 consent as all data were anonymized and analysed retrospectively under code.

104

105 **Sample collection and study design**

106 This study was performed retrospectively using 47 clinical samples obtained from 47 subjects. The

107 results obtained by routine bacterial culturing methods had been used to guide patient treatment and

108 care. In this study, we re-analysed these samples using MYcrobiota and compared the results to the

109 initial outcome of the culture results. The 47 samples included in this study were derived from wounds

110 (22), ulcers (10), abscesses (5), puss (1), erysipelas (1), erythema (1), and 7 synovial fluids obtained

111 from patients with suspected (prosthetic) joint infections.

112

Routine bacterial culture

All samples were cultured according to standard laboratory protocols performed in our laboratory and stored at −80 °C for subsequent MYcrobiota analysis. The routine bacterial culture methods included a 48h incubation at 35 °C on tryptic soy agar plates with 5% sheep blood (TSASB, Oxoid), colistin aztreonam blood agar plates (CAP, Oxoid) and cystine lactose electrolyte deficient agar plates (CLED, Oxoid) under aerobic conditions, a 48h incubation at 35 °C on chocolate agar with Vitox supplement (CHOCV, Oxoid) under 5% CO_2 conditions, and a 48h incubation at 35 °C on TSASB under anaerobic conditions. All Gram-negative rods, beta-haemolytic streptococci, *Staphylococcus aureus*, *Staphylococcus lugdunensis* and anaerobic bacteria cultured were reported as potentially relevant bacteria, of which the identification of aerobic bacteria was obtained using MALDI-TOF mass spectrometry (Bruker). Note that in this study we did not focus on optimizing culturing methods to increase the sensitivity of the culture results and the routine bacterial culture methods used may not be 100% efficient for culturing the bacteria that were detected with MYcrobiota.

Micelle PCR and NGS

DNA was extracted from all 47 samples using the High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's instructions. In addition, DNA from the accompanying elution buffer was extracted as a NEC at the same time in order to allow the subtraction of contaminating bacterial DNA after NGS processing. The total number of 16S rRNA gene copies within each DNA extract was measured using a 16S rRNA gene quantitative PCR (qPCR) according to Yang *et al*. [7], after which each DNA extract was normalized to contain either 10,000, or <1,000 16S rRNA gene copies per microliter. A synthetic microbial community (SMC) sample, containing 10,000 16S rRNA gene copies of *Moraxella catarrhalis* (ATCC 25240), *Staphylococcus aureus* (ATCC 43300), *Haemophilus influenzae* (ATCC 10211), and *Clostridium perfringens* (ATCC 12915), was processed with each batch of clinical samples as a positive control (PC) sample. Prior to amplification by micPCR, 1,000, or 100 16S rRNA gene copies of *Synechococcus* DNA were added respectively as IC to the normalized DNA extracts containing 10,000, or <1,000 16S rRNA gene copies per microliter. One hundred 16S rRNA gene copies of *Synechococcus* DNA were also added to the NEC DNA extract. The IC was used to express the

141 resulting OTUs as a measure of 16S rRNA gene copies by the use of a correction factor (sample OTU

142 copies = sample OTU reads x (initial IC copies/IC OTU reads)) as previously validated elsewhere [6].

143

144 16S rRNA gene amplicon library preparation using micPCR was performed as previously published [6],

145 but we utilized a different micPCR primer set that made it possible to replace the former Roche 454

146 NGS platform with the Illumina MiniSeq platform. In this study, micPCR amplification was performed

147 using modified 515F (5′-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG TGY CAG CMG

148 CCG CGG TAA-3′) and 806R (5′-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA

149 CNV GGG TWT CTA AT-3′) primers that amplified the V4 regions of 16S rRNA genes as recommended

150 for Illumina NGS and which incorporated universal sequence tails at their 5′ ends to allow for a two-step

151 amplification strategy. During the second round of amplification, dual indices and Illumina sequencing

152 adapters were attached using the Nextera XT Index kit (Illumina). Paired-end sequencing of the 16S

153 rRNA gene amplicon library was performed using the MiniSeq system in combination with the 2x150 bp

154 MiniSeq System High-Output Kit (Illumina), after which FASTQ-formatted sequences were extracted

155 from the MiniSeq machine for downstream analysis. We utilized the micPCR/NGS approach to process

156 all samples, including the NEC and the PC, in triplicate in order to increase accuracy and to correct for

157 contaminating bacteria DNA derived from the laboratory environment as previously described [6].

158

159 **Bioinformatics pipeline**

160 The bioinformatics pipeline designed during this study consists of 23 well-established mothur tools

161 (v.1.36) [8] and an additional 9 custom-made tools developed by the authors that have been integrated

162 and combined in Galaxy as a full analysis service to deliver 16S rRNA gene analysis for micPCR/NGS

163 experiments. Essentially, we have incorporated the functionality of mothur in Galaxy, which is a project

164 dedicated to simplify the use of complex command-line bioinformatics tools (such as mothur) using a

165 user-friendly web interface [9-11], and added new calculator tools to allow for a completely automatic

166 processing of quantitative micPCR/NGS data. Importantly, the bioinformatics pipeline presents the

167 microbiota results together with an extensive overview of the quality control measurements performed

168 during the micPCR/NGS data analysis, to the user in an organized fashion via an interactive web report.

169 The complete workflow of the bioinformatics pipeline is visualized in Fig. 1. All the tools required for the

170 bioinformatics pipeline can be found in Galaxy's Tool Shed (https://toolshed.g2.bx.psu.edu/). A workflow

171 definition   file   can   be   downloaded   from   GitHub   (https://github.com/ErasmusMC-

172 Bioinformatics/MYcrobiota) and may be imported to any Galaxy platform, thereby offering the required

173 set of bioinformatics tools. For more information on how to install and use this pipeline, please refer to

174 the documentation in GitHub (https://github.com/ErasmusMC-Bioinformatics/MYcrobiota).

175

176 **Quantitative PCR methods**

177 The total bacterial biomass within each DNA extract was measured using a 16S rRNA gene quantitative

178 PCR (qPCR) that targets the 16S rRNA gene V5-V7 region, which is a different region of the 16S rRNA

179 gene compared to MYcrobiota [7]. Therefore, the 16S rRNA gene qPCR is a complementary technique

180 that enables the validation of the MYcrobiota process when determining the total number of 16S rRNA

181 gene copies. For this, CT-values were related to a serial dilution of the previous calibrated and

182 normalized SMC sample that contained mixed and equimolar concentrations of four bacterial species

183 and ranged from a total of 100 to 10,000 16S rRNA gene copies per PCR. In addition, the *S. aureus*

184 specific biomass was assessed within each DNA extract using a *S. aureus* qPCR that employs a *S.*

185 *aureus* specific marker as described by Martineau *et al*. [12]. Here, CT-values were related to a serial

186 dilution of only the calibrated *S. aureus* (ATCC 43300) DNA stock that ranged from a total of 10 to

187 10,000 copy numbers of the Martineau fragment. The PCRs were performed in 10 µL reaction volumes

188 using the LightCycler 480 Probes Master (Roche) with the addition of 0.5 µM and 1.0 µM of each PCR

189 primer for the 16S rRNA gene and *S. aureus* qPCRs respectively. Also, 0.25 µM of a Fam-labelled probe

190 was added for the real-time detection of the 16S rRNA gene amplification and 1x Resolight Dye (Roche)

191 was added to the *S. aureus* qPCR in order to measure the *S. aureus* DNA amplification. All PCRs were

192 performed using the following conditions: initial denaturation at 95 °C for 5 minutes followed by 45 cycles

193 of PCR, with cycling conditions of 5 seconds at 95 °C, 10 seconds at 55 °C, and 30 seconds at 72 °C.

194

195 **RESULTS**

196 **Development of an easy-to-use bioinformatics pipeline**

197 In order to analyse 16S rRNA gene NGS data obtained using micPCR/NGS, we designed a Galaxy-

198 based bioinformatics pipeline for use in clinical diagnostics. This workflow is largely based on the well-

199 established standard operating procedure (SOP) defined by the creators of mothur [13]. We have

200 adapted the SOP to our specific use-case by integrating several custom-made tools that allows for the

7

201 subsampling of large datasets, the averaging over multiple technical replicates, converting the number

202 of obtained sequence reads per OTU to 16S rRNA gene copies per OTU via the use of an IC, and

203 correction for contaminating bacterial DNA via the use of NECs. All results are presented to the user as

204 a single, interactive web report in Galaxy using the iReport tool [14]. The iReport was designed to

205 visualize the resultant microbiota profiles using KRONA [15], list quantitative microbiota profiles in OTU

206 tables (with the microbial load per OTU reported as 16S rRNA gene copies), summarize results of

207 diversity calculators, and provide an extensive overview of the quality control measurements during the

208 analysis. Importantly, the iReport is relatively small in size (~ 6 MB per sample for our datasets) that

209 enables easy sharing and storage of 16S rRNA gene NGS results (Fig. 1).

210

211 **Validation of the MYcrobiota process.**

212 As shown in Fig 2, MYcrobiota results obtained from the PC that was profiled in three independent

213 experiments showed a median value of only a 1.3-fold (± 0.2) difference between the measured 16S

214 rRNA gene copies per bacterial species and the expected 10,000 16S rRNA gene copies per bacterial

215 species present in the PC. In addition, comparisons between the measured 16S rRNA gene copies

216 determined in actual clinical samples using MYcrobiota compared to qPCR results revealed an average

217 of only a 1.5-fold (± 0.5) and a 1.3-fold (± 0.4) difference for the total bacterial biomass and the

218 *Staphylococcus* OTU-specific biomass respectively. Of note, 10 of the 47 clinical samples included in

219 this study resulted in culture-negative results and the absence of bacterial DNA in these samples was

220 confirmed with both qPCR and MYcrobiota methods. Also, one discrepant sample was detected that

221 showed a 20-fold higher abundance of staphylococci detected by MYcrobiota compared to qPCR. This

222 result can be explained by the presence of *S. aureus* and *S.* non-*aureus* within this sample. In fact, the

223 *S. aureus* qPCR showed a 100% specificity compared to *S. aureus* culture-positive results and indicates

224 the presence of *S.* non-*aureus* bacteria within 7 additional samples in which the *Staphylococcus* OTU

225 was detected using MYcrobiota but no *S. aureus* could be cultured. Taken together, these data

226 demonstrate the accuracy of the MYcrobiota process and the ability to incorporate quantitative results

227 obtained from additional (species-specific) qPCRs.

228

229 **Comparing MYcrobiota results to routine bacterial culture.**

230 In order to explore the utility of MYcrobiota in the field of clinical diagnostics, we processed a total of 47

clinical samples and compared the results to routine bacterial culture. All bacterial genera detected using culture and MYcrobiota are reported per sample in Table1. Using standard bacterial culture techniques, our laboratory detected a total of 38 potentially relevant aerobic bacterial genera within 25 clinical samples and obtained a positive culture of a mixture of anaerobic bacteria in 3 samples. No bacteria were cultured from 10 samples and an additional 10 samples resulted in the growth of bacteria that were all presumed to be commensal flora. In contrast, using MYcrobiota we detected a total of 448 bacterial operational taxonomic units (OTUs) in 39 samples of which 337 OTUs (75%) could be identified as anaerobic bacterial genera that were detected in 21 samples. No bacterial DNA was measured in 8 out of 10 culture-negative samples. The sensitivity for bacterial culture detection by MYcrobiota was determined at 100% and the specificity at 83% using culture as 'gold standard'.

The majority of bacterial genera identified with culture were also identified using MYcrobiota. As shown in Table 2, MYcrobiota detected 36 of all 38 aerobic bacteria cultured on a genus-level taxonomy and confirmed the growth for anaerobic bacteria in 3 samples (sensitivity 95%; specificity 95%). Important to note, the two discrepant bacterial genera were measured using the micPCR/NGS strategy, but below the technique's LOD of 25 16S rRNA gene copies per OTU. In contrast, the vast majority of bacterial genera identified with MYcrobiota were presumed to belong to the commensal flora using culture or were not cultured at all (Table 1). These additional taxa include potential pathogens such as the *Kingella* OTU that was detected from a synovial fluid sample obtained from a juvenile patient that was not detected using culture and was confirmed using a *Kingella kingae* specific PCR.

DISCUSSION

In this study, we developed and explored the utility of a 'end-to-end' microbiota profiling platform (MYcrobiota) – consisting of our previously published 16S rRNA gene sequencing methodology (micPCR/NGS) in combination with an easy-to-use bioinformatics pipeline – to investigate human samples for the clinical diagnostics laboratories. The bioinformatics pipeline designed during this study allows for a fully automated sequence interpretation of 16S rRNA gene NGS data that is obtained using the validated micPCR/NGS protocol without the need for advanced bioinformatics skills that are often unavailable in the clinical diagnostic laboratories. The MYcrobiota results are presented using (interactive) visualizations and tables, including an overview of all removed sequences during the

9

261 analysis that allows for a manual evaluation of the quality measurements pre-installed within the

262 bioinformatics pipeline. Moreover, connections of OTU representative sequences to the external NCBI

263 database are available and can be used to ensure that the taxonomic identification of bacterial genera

264 is correct [16]. Importantly, the summarizing reports are relatively small in size and storage of these files

265 enables the traceability of patient test results that is required for clinical diagnostic laboratories according

266 to quality requirements.

267

268 Using MYcrobiota, we processed a total of 47 clinical samples and compared the results to routine

269 bacterial culture. Our results showed that the majority of bacteria identified with culture were also

270 identified with MYcrobiota, but the majority of bacterial taxa identified with MYcrobiota were not identified

271 using culture. Many of the additional bacterial taxa identified using MYcrobiota are obligate anaerobes

272 that were commonly detected as a large component of the microbial population in samples obtained

273 from damaged skin sites, which is consistent with previous studies [17, 18]. Indeed, it is well-known that

274 anaerobic bacteria are able to cause serious and life-threatening infections but are often overlooked

275 due to their requirement for appropriate methods of collection, transportation and cultivation [19].

276 Therefore, the culture-free MYcrobiota detection platform can play an important role in the identification

277 of the bacteriological aetiology of anaerobic infections, or any other infections caused by fastidious

278 microorganisms. Of note, it could be argued that the development of extensive culture techniques (so-

279 called 'culturomics') may eventually facilitate the successful culture of supposedly 'non-culturable'

280 microbial isolates [20].

281

282 In addition to the accurate detection and identification of bacterial OTUs within clinical samples,

283 MYcrobiota also provide the relative abundances in combination with the absolute abundances for each

284 detected bacterial OTU. This feature allows clinicians to obtain a comprehensive overview of the

285 microbial composition of the clinical sample so that each quantified bacterial OTU, as well as the

286 bacterial community as a whole, might be taken into account in clinical decision-making. Additionally,

287 MYcrobiota allows for the removal of contaminating DNA from environmental sources in order to

288 accurately and reliably investigate very low bacterial biomass, or no bacterial biomass, clinical samples

289 [6]. For example, MYcrobiota confirmed the absence of 16S rRNA gene copies in 8 of the 10 samples

290 that generated culture-negative results. The two discrepant samples contained either anaerobic bacteria

291   or low amounts of the fastidious *Kingella* bacterium respectively. The ability to confirm culture-negative

292   results improves the reliability of culture-negative diagnostic results. Additionally, the ability of

293   MYcrobiota to detect bacterial OTUs at very low abundances makes MYcrobiota a suitable method to

294   investigate normally sterile body sites, such as synovial fluids, cerebrospinal fluids, blood samples, etc.

295   It should be noted however, that the authors are aware of the fact that the construction of MYcrobiota is

296   only a first step in the transition of microbiota research into actual clinical diagnostics. Extensive clinical

297   and financial validation studies will be needed in order to validate and justify the routine introduction of

298   molecular microbiota profiling methods into clinical diagnostic laboratories.

299

300   In conclusion, the stepwise development of MYcrobiota paves the way to introduce quantitative

301   microbiota profiling into the clinical diagnostic laboratory. The method provides an highly accurate and

302   comprehensive overview of the microbial composition of clinical samples or, alternatively, confirms the

303   absence of 16S rRNA gene copies in culture-negative samples, using a standardized and validated 16S

304   rRNA gene NGS workflow. Despite some shortcomings e.g. lack of species identification and the

305   inability to provide detailed information on antibiotic susceptibility, our data illustrates that MYcrobiota

306   has promising applications in the field of clinical diagnostics and warrants investment in future studies

307   to accurately evaluate the clinical relevance of 16S rRNA gene NGS results in clinical samples.

308

309   **AVAILABILITY OF DATA AND MATERIALS**

310   The datasets generated and analysed during the current study are available in the Sequence Read

311   Archive repository with accession number SRP109023,

312   https://www.ncbi.nlm.nih.gov/sra/?term=SRP109023.

313

317

318   **CONFLICT OF INTEREST**

319   The authors declare that they have no conflict of interest.

320

321   **REFERENCES**

322   [1]   Fournier PE, Raoult D (2011) Prospects for the future using genomics and proteomics in clinical

323   microbiology. Annu Rev Microbiol 65:169-188

324   [2]   Rhoads DD, Wolcott, RD, Sun Y, Dowd SE (2012) Comparison of culture and molecular

325   identification of bacteria in chronic wounds. Int J Mol Sci 13:2535-2550

326   [3]   Salipante SJ, Sengupta DJ, Rosenthal C et al (2013) Rapid 16S rRNA next-generation

327   sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. PloS

328   one 8:e65226

329   [4]   Hiergeist A, Reischl U, Priority Program Intestinal Microbiota Consortium/ quality assessment

330   participants, Gessner A (2016) Multicenter quality assessment of 16S ribosomal DNA-

331   sequencing for microbiome analyses reveals high inter-center variability. Int J Med Microbiol

332   306:334-342

333   [5]   Boers SA, Hays JP, Jansen R (2015) Micelle PCR reduces chimera formation in 16S rRNA

334   profiling of complex microbial DNA mixtures. Sci Rep 5:14181

335   [6]   Boers SA, Hays JP, Jansen R (2017) Novel micelle PCR-based method for accurate, sensitive

336   and quantitative microbiota profiling. Sci Rep 7:45536

337   [7]   Yang S, Lin S, Kelen GD, Quinn TC, Dick JD, Gaydos CA, Rothman RE (2002) Quantitative

338   multiprobe PCR assay for simultaneous detection and identification to species level of bacterial

339   pathogens. J Clin Microbiol 40:3449-3454

340   [8]   Schloss PD, Westcott SL, Ryabin T et al (2009) Introducing mothur: open-source, platform-

341   independent, community-supported software for describing and comparing microbial

342   communities. Appl Environ Microbiol 75:7537-7541

343   [9]   Giardine B, Riemer C, Hardison RC et al (2005) Galaxy: a platform for interactive large-scale

344   genome analysis. Genome Res 15:1451-1455

345   [10]   Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A,

346   Taylor J (2010) Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc

347   Mol Biol Chapter 19:Unit 19.10.1-21

348   [11]   Goecks J, Nekrutenko A, Taylor J, Galaxy Team (2010) Galaxy: a comprehensive approach for

349   supporting accessible, reproducible, and transparent computational research in the life

350   sciences. Genome Biol 11:R86

351   [12]   Martineau F, Picard FJ, Roy PH, Ouellette M, Bergeron MG (1998) Species-specific and

352   ubiquitous-DNA-based assays for rapid identification of *Staphylococcus aureus*. J Clin Microbiol

353   36:618-623

354   [13]   Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-

355   index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the

356   MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112-5120

357   [14]   Hiltemann S, Hoogstrate Y, der Spek, Pv, Jenster G, Stubbs A (2014) iReport: a generalised

358   Galaxy solution for integrated experimental reporting. Gigascience 3:19

359   [15]   Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a Web

360   browser. BMC Bioinformatics 12:385

361   [16]   Boers SA, Jansen R, Hays JP (2016) Suddenly everyone is a microbiota specialist. Clin

362   Microbiol Infect 22:581-582

363   [17]   Price LB, Liu CM, Melendez JH et al (2009) Community analysis of chronic wound bacteria

364   using 16S rRNA gene-based pyrosequencing: impact of diabetes and antibiotics on chronic

365   wound microbiota. PloS one 4:e6462

366   [18]   Smith K, Collier A, Townsend EM, O'Donnell LE, Bal AM, Butcher J, Mackay WG, Ramage G,

367   Williams C (2016) One step closer to understanding the role of bacteria in diabetic foot ulcers:

368   characterising the microbiome of ulcers. BMC Microbiol 16:54

369   [19]   Brook I (2002) Clinical review: bacteremia caused by anaerobic bacteria in children. Crit Care

370   6:205-211

371   [20]   Lagier J-C, Hugon P, Khelaifia S, Fournier P-E, La Scola B, Raoult D (2015) The rebirth of

372   culture in microbiology through the example of culturomics to study human gut microbiota. Clin

373   Microbiol Rev 28:237-264

TABLES AND FIGURES

375 **Table 1.** Bacterial genera identified from 47 clinical samples using routine bacterial culture and

376 MYcrobiota.

Sample	Routine bacterial culture	MYcrobiota
01_U	Commensal flora (1+)	**Anaerobic bacteria (346,300)**, *Corynebacterium* (10,725)
02_U	Commensal flora (2+)	*Staphylococcus* (941)
03_W	Commensal flora (1+)	**Anaerobic bacteria (263)**, *Streptococcus* (33), *Staphylococcus* (25)
04_U	*Pseudomonas* (3+), *Staphylococcus* (2+)	*Pseudomonas* (4,706), *Staphylococcus* (848), ***Enterococcus (135)***, **Anaerobic bacteria (102)**
05_U	*Proteus* (2+), *Enterobacteriaceae** (2+), *Streptococcus* (1+), Commensal flora (1+)	**Anaerobic bacteria (8,271)**, *Proteus* (3,510), *Streptococcus* (632), *Enterobacteriaceae** (333)
06_U	Commensal flora (1+)	*Moraxella* (8,947), *Corynebacterium* (734)
07_W	*Enterobacteriaceae* (1+)	*Enterobacteriaceae** (5,386), ***Bacillus (44)***
08_W	Negative	Negative
09_W	Commensal flora (1+)	**Anaerobic bacteria (523)**, *Staphylococcus* (31)
10_A	Anaerobic bacteria (2+), *Pasteurella* (2+), *Streptococcus* (2+)	Anaerobic bacteria (3,704,750), *Pasteurella* (242,250), *Streptococcus* (28,625)
11_W	*Enterobacteriaceae** (3+), *Staphylococcus* (3+)	*Enterobacteriaceae** (3,420,786), ***Acinetobacter (1,126,632)***, *Staphylococcus* (32,760)
12_A	*Enterobacteriaceae** (2+), *Streptococcus* (2+)	*Enterobacteriaceae** (18,046), *Streptococcus* (6,409), ***Enterococcus (67)***
13_Es	Commensal flora (1+)	*Staphylococcus* (344), **Anaerobic bacteria (150)**, *Dermabacteraceae** (93), *Haemophilus* (64), *Corynebacterium* (53)
14_W	Commensal flora (1+)	*Staphylococcus* (31)
15_U	*Staphylococcus* (4+)	*Staphylococcus* (17,035)
16_U	*Enterobacteriaceae** (3+), *Stenotrophomonas* (2+), Commensal flora (2+), *Proteus* (1+), ***Pseudomonas* (1+)**	*Enterobacteriaceae** (828,310), *Proteus* (250,670), *Stenotrophomonas* (11,760)
17_W	*Staphylococcus* (1+), Commensal flora (1+)	*Staphylococcus* (4,886)
18_W	*Staphylococcus* (3+), Commensal flora (2+)	*Staphylococcus* (141,120), *Corynebacterium* (4,959)
19_W	*Streptococcus* (2+), *Staphylococcus* (1+)	*Streptococcus* (114,257), *Staphylococcus* (44,772), ***Corynebacterium (8,749)***, **Anaerobic bacteria (897)**
20_W	*Enterobacteriaceae** (3+), ***Staphylococcus* (2+)**	*Enterobacteriaceae** (4,574,310)
21_U	Commensal flora (2+)	*Moraxella* (1,066,608), *Acinetobacter* (142,155), *Pseudomonas* (30,051), **Anaerobic bacteria (30,051)**, *Corynebacterium* (23,976), *Alkanindiges* (2,187)
22_W	*Staphylococcus* (2+), Commensal flora (1+)	*Staphylococcus* (105,648)
23_Et	*Staphylococcus* (2+), Commensal flora (2+)	*Staphylococcus* (14,803), *Corynebacterium* (66)
24_U	*Staphylococcus* (3+), *Streptococcus* (3+), Commensal flora (2+)	*Staphylococcus* (231,756), **Anaerobic bacteria (96,740)**, *Streptococcus* (15,904), *Enterococcus* (1,680)
25_W	*Staphylococcus* (3+), Commensal flora (2+)	*Staphylococcus* (23,175), *Corynebacterium* (15,488), **Anaerobic bacteria (1,271)**
26_W	Commensal flora (1+)	*Staphylococcus* (4,142), **Anaerobic bacteria (101)**, *Corynebacterium* (94), *Streptococcus* (47)
27_A	*Streptococcus* (1+)	**Anaerobic bacteria (1,062,060)**, *Streptococcus* (5,490), ***Treponema (3,435)***, ***Gemella (1,425)***, ***Mycoplasma (870)***, ***Tannerella (720)***

28_W	Commensal flora (2+), *Streptococcus* (2+)	**Anaerobic bacteria (114,004)**, *Streptococcus* (43,208)
29_A	*Streptococcus* (1+)	*Streptococcus* (12,225), **Anaerobic bacteria (3,384)**, ***Gemella* (299)**, ***Enterococcus* (295)**, ***Haemophilus* (221)**, ***Capnocytophaga* (156)**, ***Granulicatella* (122)**, ***Neisseria* (119)**, ***Rothia* (52)**, ***Lautropia* (35)**
30_W	Negative	Negative
31_U	*Acinetobacter* (2+), *Enterobacteriaceae** (2+), Commensal flora (2+)	*Acinetobacter* (518,396), *Stenotrophomonas* (423,320), *Enterobacteriaceae** (12,046), *Corynebacterium* (5,928), *Bordetella* (4,636), *Brevibacterium* (988)
32_W	*Staphylococcus* (2+), Commensal flora (1+)	**Anaerobic bacteria (251,692)**, *Streptococcus* (30,408), *Staphylococcus* (8,960)
33_W	*Staphylococcus* (1+), Commensal flora (1+)	*Staphylococcus* (466), **Anaerobic bacteria (171)**, *Streptococcus* (105), *Acinetobacter* (84), *Corynebacterium* (41)
34_W	*Staphylococcus* (3+)	*Staphylococcus* (218,141)
35_W	Commensal flora (2+)	*Staphylococcus* (5,121), **Anaerobic bacteria (769)**, *Roseomonas* (40)
36_W	Anaerobic bacteria (3+), Commensal flora (1+)	Anaerobic bacteria (493,183), *Streptococcus* (1,045)
37_W	*Streptococcus* (2+)	*Streptococcus* (11,457)
38_W	Anaerobic bacteria (3+)	Anaerobic bacteria (830,531)
39_P	*Streptococcus* (2+)	*Streptococcus* (10,277,376)
40_A	Negative	Anaerobic bacteria (94,633), *Enterobacteriaceae** (2,944), *Streptococcus* (44), *Thalassospira* (36)
41_S	Negative	Negative
42_S	Negative	Negative
43_S	Negative	Negative
44_S	Negative	Negative
45_S	Negative	Negative
46_S	Negative	Negative
47_S	Negative	***Kingella* (25)**

Samples were derived from wounds (W), ulcers (U), abscesses (A), puss (P), erysipelas (Es), erythema (Et), and suspected joint infections (S). Cultured bacteria other than Gram-negative rods, beta-hemolytic streptococci, *S. aureus*, *S. lugdunensis* and anaerobic bacteria were reported as commensal flora. The semi-quantitative culture results are presented as 1+, 2+, 3+, or 4+, depending on which quadrants demonstrate bacterial growth. The presence of anaerobic bacteria was reported as either a positive or negative result. Bacterial species and OTUs detected using culture and MYcrobiota respectively are grouped at the genus level to compare results. Red shades indicate bacterial genera that were only identified by culture and blue shades indicate bacterial genera that were only identified by MYcrobiota (with 'commensal flora' culture results representing a positive detection signal for any kind of aerobic bacterial OTU identified by MYcrobiota). The number of 16S rRNA genes measured using MYcrobiota are indicated in parentheses. (*) Several bacterial genera that belong to the *Enterobacteriaceae* and *Dermabacteraceae* families could not be differentiated at a 97% similarity level using MYcrobiota.

389

390 **Table 2.** Comparison of the cultured bacterial taxa to MYcrobiota results.

Bacterial taxa	Number of positive samples		Sensitivity	Specificity
	Routine bacterial culture	MYcrobiota		
Acinetobacter	1	4	100%	98%
*Enterobacteriaceae**	7	8	100%	98%
Pasteurella	1	1	100%	100%
Proteus	2	2	100%	100%
Pseudomonas	2	2	67%	100%
Staphylococcus	14	20	93%	100%
Stenotrophomonas	1	2	100%	100%
Streptococcus	10	16	100%	97%
Anaerobic bacteria	3	21	100%	71%
Total	**41**	**76**	**95%**	**95%**

391 The culture results are restricted to genus-level classifications in order to compare the OTUs detected

392 using MYcrobiota to the culture-based results. The presence of anaerobic bacteria was reported as

393 either a positive or negative result. 'Commensal flora' culture results were interpreted as a positive

394 detection signal for any kind of aerobic bacterial OTU identified by MYcrobiota to perform specificity

395 calculations. (*) Several bacterial genera that belong to the *Enterobacteriaceae* family could not be

396 differentiated at a 97% similarity level using MYcrobiota.

Fig. 1. Schematical overview of the bioinformatics pipeline.

FASTQ-formatted sequences obtained from triplicate experiments using micPCR/NGS (R1, R2 and R3) are automatically processed via the use of 32 (mothur) tools that have been integrated and combined in Galaxy as an 'end-to-end' analysis service. The results obtained per sample (average of triplicate results) are presented to the user in a single, interactive iReport that consist of 3 tabs. The taxonomy tab visualizes and lists the resultant microbiota profiles. The diversity tab summarizes the results of 3 diversity calculators (Chao1, Shannon and Simpson). The quality control tab provides an extensive overview of the quality control measurements during the analysis.

Fig. 2. Accuracy of 16S rRNA gene copy determination using MYcrobiota.

The expected number of 16S rRNA gene copies within the positive control (PC) was compared to the measured number of 16S rRNA gene copies using MYcrobiota (green dots). The PC contained 10,000 16S rRNA gene copies of four different bacterial species and was processed in three independent MYcrobiota experiments. The indirect estimation of the total bacterial biomass within 37 clinical samples using MYcrobiota was compared to the total 16S rRNA gene copies measured directly using a 16S rRNA gene qPCR (blue dots). The *Staphylococcus* OTU-specific biomass from 13 *S. aureus* culture-positive samples was compared to the *S. aureus* biomass detected directly using a *S. aureus*-specific qPCR (yellow dots). In order to compare the number of *S. aureus* genome copies estimated using qPCR to the number of 16S rRNA gene copies detected using MYcrobiota, the estimated *S. aureus* genome copies were first multiplied by a factor of 6 to correct for differences in copy numbers of the Martineau fragment and the 16S rRNA gene present on the *S. aureus* genome. The calculated differences between methods were plotted using a binary logarithmic scale.