1

2

Impact of Sample Type and DNA Isolation Procedure on Genomic

Inference of Microbiome Composition

5

6

Berith E. Knudsen[a], Lasse Bergmark[a*], Patrick Munk[a], Oksana Lukjancenko[a], Anders

Priemé[2], Frank M. Aarestrup[a], Sünje J. Pamp[a#]

9

[a]National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark

[b]Department of Biology, University of Copenhagen, Copenhagen, Denmark

12

B.E.K and L.B. contributed equally to this work.

[*]Present address: Novo Nordisk, Bagsværdvej, 2880 Bagsværd, Denmark

[#]Address correspondence to: Sünje J. Pamp, sjpa@food.dtu.dk

16

17

18

19

Running title: DNA Isolation Methodology for Microbiome Genomics

Keywords: microbial ecology, microbiome, DNA isolation, 16S rDNA gene sequencing,

metagenomics

23 **ABSTRACT**

24

25 Explorations of complex microbiomes using genomics greatly enhance our

26 understanding about their diversity, biogeography, and function. The isolation of DNA

27 from microbiome specimens is a key prerequisite for such examinations, but challenges

28 remain in obtaining sufficient DNA quantities required for certain sequencing

29 approaches, achieving accurate genomic inference of microbiome composition, and

30 facilitating comparability of findings across specimen types and sequencing projects.

31 These aspects are particularly relevant for the genomics-based global surveillance of

32 infectious agents and antimicrobial resistance from different reservoirs. Here, we

33 compare in a stepwise approach a total of eight commercially available DNA extraction

34 kits and 16 procedures based on these for three specimen types (human feces, pig

35 feces, and hospital sewage). We assess DNA extraction using spike-in controls, and

36 different types of beads for bead-beating facilitating cell lysis. We evaluate DNA

37 concentration, purity, and stability, and microbial community composition using 16S

38 rRNA gene sequencing and for selected samples using shotgun metagenomic

39 sequencing. Our results suggest that inferred community composition was dependent on

40 inherent specimen properties as well as DNA extraction method. We further show that

41 bead-beating or enzymatic treatment can increase the extraction of DNA from Gram-

42 positive bacteria. Final DNA quantities could be increased by isolating DNA from a

43 larger volume of cell lysate compared to standard protocols. Based on this insight, we

44 designed an improved DNA isolation procedure optimized for microbiome genomics that

45 can be used for the three examined specimen types and potentially also for other

46   biological specimens. A standard operating procedure is available from:

47   https://dx.doi.org/10.6084/m9.figshare.3475406.

48

49   **IMPORTANCE**

50

51   Sequencing-based analyses of microbiomes may lead to a breakthrough in our

52   understanding of the microbial world associate with humans, animals, and the

53   environment. Such insight could further the development of innovative ecosystem

54   management approaches for the protection of our natural resources, and the design of

55   more effective and sustainable solutions to prevent and control infectious diseases.

56   Genome sequence information is an organism- (pathogen-) independent language that

57   can be used across sectors, space, and time. Harmonized standards, protocols, and

58   workflows for sample processing and analysis can facilitate the generation of such

59   actionable information. In this study, we assessed several procedures for the isolation of

60   DNA for next-generation sequencing. Our study highlights several important aspects to

61   consider in the design and conduction of sequence-based analysis of microbiomes. We

62   provide a standard operating procedure for the isolation of DNA from a range of

63   biological specimens particularly relevant in clinical diagnostics and epidemiology.

64 **INTRODUCTION**

65

66 Microbial communities fulfill central roles in biological systems, such as in human,

67 animal, and environmental ecosystems. Genomics-based interrogations of these

68 communities can provide unprecedented insight into their composition and function, and

69 reveal general principles and rules about their ecology and evolution (1-4).

70 Genomics-based microbiome analyses can also have important practical implications,

71 such as for the diagnosis and management of infectious diseases. Together with

72 relevant metadata, attribute data, and appropriate bioinformatics and statistical

73 approaches, genomic sequencing data could enable the global surveillance of emerging

74 and re-emerging infectious diseases, and teach us about the reservoirs and

75 transmission pathways of pathogens (5-7). Ultimately, genomics-based information

76 about infectious disease epidemiology may help us to predict, prevent, and control

77 infectious diseases faster, more precisely, and more sustainably.

78 In order to facilitate large-scale microbiome analyses, harmonized standards for sample

79 handling and data analysis need to be ensured. To be able to establish pathogen

80 reservoirs and transmission pathways, specimens from different sources, such as from

81 humans, animals, and the environment, will need to be examined. For genomics

82 analysis, the DNA needs to be isolated from the specimens for DNA sequencing.

83 However, DNA isolation methods are often only evaluated and established in the context

84 of specimens from an individual source (e.g. human fecal specimens), and seldom

85 across a variety of specimen types (8-12), which is addressed in the present study.

86 Current sequencing technologies, such as Illumina MiSeq and Hiseq, PacBio,

87 IonTorrent, and Nanopore systems, still require large initial DNA template quantities,

4

88   particularly from the perspective of PCR-free metagenomics-based analysis. In contrast,

89   16S rRNA gene profiling can reveal a bacterial and archaeal composition for samples

90   with low initial DNA template quantities. In metagenomics, low quantities of input DNA

91   can result in low sequencing data output, and impact the inferred microbial community

92   composition (13). Hence, modified DNA isolation protocols for increasing DNA quantities

93   obtained from different types of specimens are desirable.

94   Here, we examine three specimen types (human feces, animal feces, and sewage), a

95   total of eight commercially available DNA isolation kits, and a number of protocol

96   modifications in regard to output DNA (quantity, purity, stability) and microbiome

97   composition (16S rRNA gene profiling, metagenomics). Our results suggest that both,

98   the specimen itself as well as the DNA isolation procedure, can affect DNA quantity and

99   quality, and inferred microbiome composition. Based on the insight gained, we have

100   developed an improved laboratory protocol that can be used for DNA isolations from a

101   variety of biological specimens.

102   **RESULTS**

103

104   **DNA concentration, purity, and stability depend on the type of specimen and DNA**

105   **isolation method.** We extracted DNA from human feces, pig feces, and hospital

106   sewage, using seven commonly used DNA isolation kits and determined DNA

107   concentration, purity, and stability of the isolated DNA (Fig. 1A and Table 1). The DNA

108   concentrations varied greatly (Fig. 1B; see also Table S1A in the supplemental

109   material). For human feces, the highest DNA concentrations were obtained using the

110   EasyDNA, MagNAPure, and QIAStool procedure, for pig feces using the EasyDNA,

111   QIAStool, and QIAStool+BB procedures, and for sewage using the MagNAPure and

112   EasyDNA procedure, while for three methods the DNA concentration from sewage was

113   below the detection limit. On average across the three types of specimen, the highest

114   DNA concentrations were obtained using EasyDNA (44.96 ng/µl +/- 20.99 SEM) and

115   QIAStool 27.88 ng/µl +/- 2.55 SEM), and the lowest using the PowerSoil.HMP (1.55

116   ng/µl +/-0.31 SEM) and InnuPURE (7.77 ng/µl +/- 5.54 SEM) methods.

117   With regard to DNA purity, the best results for human and pig feces were obtained using

118   the EasyDNA, QIAStool, and QIAStool+BB procedure (see Table S1A in the

119   supplemental material). The DNA was generally stable for at least 7 days when stored at

120   room temperature (22°C) with some exceptions (see Table S1A in the supplemental

121   material). A decrease in DNA concentration over time was observed for example for the

122   human feces when extracted with EasyDNA (57% decease in DNA concentration) or

123   MagNAPure (21% decrease in DNA concentration), suggesting the presence of DNases

124   in these extracts. In some cases, an increase in DNA concentration over time was

125   observed, such as for the pig feces when extracted with EasyDNA (32% increase in

6

126   DNA concentration). An increase in DNA concentration over time at room temperature

127   was previously shown to be related to the hyperchromicity of DNA, and dependent on

128   the DNA concentration and ionic strength of the solution (14).

129

130   **Microbial richness and diversity are influenced by DNA isolation procedure.** For

131   the human fecal specimen, the highest bacterial Operational Taxonomic Unit (OTU)

132   richness and diversity were detected using the QIAStool+BB and FastDNA methods,

133   followed by InnuPURE and PowerSoil.HMP as assessed by 16S rRNA gene profiling

134   (Fig. 1C; see also Table S1B in the supplemental material). In comparison, the

135   determined richness and diversity for the EasyDNA method was low, and the relative

136   abundance of Ruminococcaceae and Bifidobacteriaceae dominated the composition

137   compared to the extracts from the other methods (Fig. 1C; see also Fig. S1A in the

138   supplemental material). Thirty-nine samples (human feces, pig feces, and sewage) with

139   high DNA concentration were selected and examined using metagenomic sequencing.

140   In this assessment, the species richness and diversity for human feces was highest for

141   the EasyDNA procedure, and a high relative abundance of Ruminococcaceae and

142   Bifidobacteriaceae was apparent in this analysis as well (see Fig. S1A in the

143   supplemental material).

144   For the pig fecal specimen, the highest bacterial richness and diversity were detected

145   using the PowerSoil.HMP and MagNAPure methods, followed by QIAStool+BB (Fig. 1C;

146   see also Table S1B in the supplemental material). Similarly, richness and diversity were

147   highest using the MagNAPure and EasyDNA methods when assessed using

148   metagenomics. Based on 16S rRNA gene profiling, the richness and diversity for the

149   FastDNA method were lower compared to all other methods, and the relative

7

150   abundance of Clostridiaceae and Turicibacteraceae was higher and the abundance of

151   Prevotellaceae and Ruminococcaceae lower using this method, compared to the other

152   methods (Fig. 1C; see also Fig. S1A in the supplemental material).

153   For the sewage specimen, the highest bacterial richness and diversity was detected

154   using the InnuPURE method, followed by PowerSoil.HMP and QIAStool+BB, and similar

155   levels were achieved using the other methods (Fig. 1C; see also Table S1B in the

156   supplemental material). The relative abundance of Clostridiaceae was highest in the

157   samples extracted using EasyDNA, and the abundance of Enterobacteriales highest in

158   the samples extracted using PowerSoil.HMP.

159   Overall, the relative abundance of predicted Gram-positive bacteria was highest in the

160   human and sewage specimens when extracted with the EasyDNA method, and highest

161   in the pig specimen when extracted using the FastDNA method (see Fig. S2 in the

162   supplemental material). The abundance of predicted Gram-positive bacteria was lowest

163   using MagNAPure and QIAStool, the two methods that did neither include a bead-

164   beating step nor specific enzymatic cell-wall digestion.

165

166   **Microbial community composition depends on the choice of DNA isolation**

167   **procedure.** The microbial communities from the three types of specimen clustered

168   separately according to specimen type when examined in PCoA Bray-Curtis ordination,

169   and not according to DNA isolation procedure (see Fig. S3 in the supplemental

170   material), indicating that the largest differences between these samples are driven by

171   the inherent microbiota composition. Bray-Curtis dissimilarity distance analysis carried

172   out separately for each of the three specimens revealed that the samples largely

173   clustered according to DNA isolation procedure (Fig. 2A-C). For the human fecal

8

174   specimen, the bacterial community composition derived from the EasyDNA isolation

175   differed from the communities obtained using all other methods (Fig. 2A), which is in

176   agreement with the observations on microbial richness (above). The Bray-Curtis

177   distances between the samples from InnuPURE, MagNAPure, FastDNA,

178   PowerSoil.HMP, QIAStool, and QIAStool+BB DNA isolations were on average 0.337 +/-

179   0.012 SEM, whereas the distances between these and the ones derived from the

180   EasyDNA procedure were on average 0.825 +/- 0.014 SEM.

181   For the pig fecal specimen, the bacterial communities derived from the FastDNA

182   isolation differed from all other communities (Fig. 2B). The average Bray-Curtis distance

183   between the samples originating from all but the FastDNA procedure was on average

184   0.473 +/- 0.008 SEM, whereas the distance between these and the ones derived from

185   the FastDNA procedure was on average 0.877 +/- 0.007 SEM.

186   For the hospital sewage specimen, the bacterial communities originating from the

187   EasyDNA method differed from all others (average Bray-Curtis distance 0.600 +/- 0.006

188   SEM) (Fig. 2C), similar to the human fecal matrix (Fig. 2A). In addition, the communities

189   originating from the QIAStool DNA isolation differed from all others (average Bray-Curtis

190   distance 0.514 +/- 0.009 SEM), whereas the average Bray-Curtis distance between all

191   but the QIAStool and EasyDNA samples was 0.460 +/-0.11 SEM on average.

192

193   **Distinct taxa account for the differences observed between DNA isolation**

194   **methods.** To quantify the effect of DNA isolation method on microbial community

195   composition we tested for differential abundance of taxa between the communities

196   derived from the different DNA isolation methods using DESeq2 analyses. In pairwise

197   comparisons, significant differences between the DNA isolation methods were observed

198   (Fig. 2D-F; see also Table S2 in the supplemental material).

199   The most abundant family on average in the human fecal specimen was Prevotellaceae

200   (Bacteroidetes), and its abundance was significantly lower in the samples extracted with

201   EasyDNA as compared to all other methods (e.g. 18.3-fold lower in EasyDNA vs.

202   QIAStool, adjusted p-value 1.91^{-6}) (Fig. 2D; see also Table S2 in the supplemental

203   material). Similarly, the abundance of Bacteroidaceae (Bacteroidetes),

204   Porphyromonadaceae (Bacteroidetes), Alcaligenaceae (β-Proteobacteria), and

205   Pasteurellaceae (γ-Proteobacteria) was lower in the samples from the EasyDNA

206   isolation compared to the other methods. In contrast, the abundance of

207   Bifidobacteriaceae (Actinobacteria) was higher in the samples originating from the

208   EasyDNA procedure compared to all other methods (e.g. 770-fold higher in EasyDNA

209   vs. QIAStool, adjusted p-value 7.49^{-57}). The abundance of Verrucomicrobiaceae

210   (Verrucomicrobia) was significantly lower in the samples from the QIAStool+BB and

211   PowerSoil.HMP DNA isolations (e.g. 4.15-fold lower in QIAStool+BB vs. QIAStool,

212   adjusted p-value 0.001).

213   The most abundant family on average in the pig fecal specimen was Prevotellaceae

214   (Bacteroidetes), and its abundance differed significantly between the DNA isolation

215   procedures (e.g. 2.3-fold lower in EasyDNA vs. PowerSoil.HMP, adjusted p-value 1.28^{-5})

216   (Fig. 2E; see also Table S2 in the supplemental material). The abundance of

217   Clostridiaceae (Clostridia), the on average fourth most abundant family in the pig feces,

218   was significantly higher in the samples extracted by the FastDNA method (e.g. 166-fold

219   higher in FastDNA vs. EasyDNA, adjusted p-value 7.35^{-110}).

10

220   Moraxellaceae (γ-Proteobacteria) was the most abundant family on average in the

221   hospital sewage, and its abundance was significantly higher in the samples from the

222   EasyDNA isolation compared to other DNA isolation methods (e.g. 2.6-fold higher in

223   EasyDNA vs. PowerSoil.HMP, adjusted p-value 3.82^{-5}) (Fig. 2F; see also Table S2 in

224   the supplemental material). Ruminococcaceae (Clostridia), the on average third most

225   abundant family in sewage, were also significantly more abundant in the samples from

226   the EasyDNA isolation compared to other DNA isolation procedures (e.g. 7.3-fold higher

227   in EasyDNA vs. FastDNA, adjusted p-value 4.28^{-17}).

228

229   **DNA isolation procedure affects the abundance of taxa differently across**

230   **specimens.** Given that differential taxa abundances were observed for the different

231   DNA isolation procedures for the three specimen types, we investigated whether the

232   abundance differed in the same way between DNA isolation procedures across

233   specimens. For example, we were asking: If taxon A is observed at a higher abundance

234   upon DNA isolation with method X compared to method Y in specimen type 1, is this

235   taxon also observed at a higher abundance upon DNA isolation with method X

236   compared to method Y in specimen type 2? We examined taxa that were detected in all

237   three specimen types, and selected representative families from different phyla (Fig. 3).

238   Similar patterns of differential abundance were observed for certain taxa across

239   specimen types, with exceptions, including two families from the Bacteroidetes phylum.

240   The abundance of Prevotellaceae and Bacteroidaceae was significantly lower when

241   human fecal specimen were extracted with EasyDNA compared to other methods. In

242   contrast, these two families were observed at a significantly higher abundance when

243   sewage was extracted with EasyDNA compared to other methods (Fig. 3).

11

244   Likewise, Ruminococcaceae of the phylum Clostridia were observed at a significantly

245   higher abundance in human fecal and hospital sewage samples but not in pig fecal

246   samples when extracted with the EasyDNA method compared to other methods. The

247   same pattern was however not observed for all families of the phylum Clostridia.

248   Clostridiaceae abundance appeared higher in human and pig feces when extracted with

249   FastDNA compared to other methods, and Clostridiaceae abundance appeared higher

250   in sewage when extracted using the EasyDNA method compared to other methods (Fig.

251   3).

252   Thus, we found significant differences in the abundance of certain families according to

253   specimen type, which sometimes depend on the DNA isolation procedure. Some of the

254   differential abundance patterns were similar across the three types of specimens, while

255   others differed.

256

257   **Detection of spiked bacteria is dependent on DNA isolation procedure and**

258   **specimen type.** In order to quantify DNA isolation efficiency, we spiked the three

259   specimen with known numbers of two bacterial representatives, namely *Salmonella*

260   *enterica* serotype Typhimurium DT104 (Gram-negative) and *Staphylococcus aureus*

261   ST398 (Gram-positive) in a CFU ratio of 1.02. Both, *S. enterica* and *S. aureus* were

262   present in negligible numbers in the three specimens before spiking. DNA was isolated

263   from these samples using the seven different DNA isolation methods, and the

264   abundance of the two strains determined using 16S rRNA gene profiling, and for some

265   samples also using metagenomics. Based on 16S rRNA gene profiling, the spiked

266   organisms accounted for an average abundance of 1.0% (+/-0.29 SEM)

267   Enterobacteriaceae, and 0.29% (+/-0.11 SEM) Staphylococcaceae across the three

268   types of specimen.

269   Using QIAStool, a DNA isolation method that does not involve a bead-beating step, the

270   abundance of Enterobacteriaceae was higher in the spiked human fecal specimen than

271   expected, with an Enterobacteriaceae/Staphylococcaceae (E/S) ratio of 13.9 (Fig. 4A).

272   This ratio was lower in the spiked human fecal specimen using InnuPURE, FastDNA,

273   PowerSoil.HMP, and QIAStool+BB, which are all methods that involve a bead-beating

274   step (E/S ratio range 0.3-2.3). The EasyDNA method involves an additional enzymatic

275   lysis step, and using this method the determined E/S ratio was 3.7. Using the

276   MagNAPure method no or lower read numbers assigned to Staphylococcaceae were

277   detected in the spiked samples compared to not spiked samples in the human fecal

278   specimen, and hence the ratio resulted in negative values (Fig. 4A). A similar result was

279   obtained when the samples were examined using metagenomics (see Fig. S4 in the

280   supplemental material).

281   Overall, most DNA isolation methods exhibited a similar tendency across the three types

282   of specimen. For example, for all three specimen types, the E/S ratio was higher using

283   the QIAStool method, compared to the other methods (except MagNAPure for sewage).

284   However, when the strain mix, composed of *S. enterica* and *S. aureus* only, was

285   extracted using the seven DNA isolation procedures, their determined E/S ratio was in

286   almost all cases similar to the expected ratio of 1.02, including the QIAStool method.

287

288   **Protocol modifications for increasing DNA concentration.** One goal in genomics is

289   to obtain a predicted pattern of microbial community composition that closely resembles

290   the actual composition of microorganisms in a particular environment. Another challenge

13

291   is to obtain sufficient DNA for metagenome sequencing. To address this aspect, we

292   examined the effect of modifications to standard protocols on output DNA concentration

293   (modifications are described in detail in the Supplemental Materials & Methods section).

294   We chose the QIAStool method as a starting point, as we obtained DNA extracts using

295   this method that were of high purity and stability (see Table S1A in the supplemental

296   material). Another concern is processing time and costs for DNA isolation procedures,

297   particularly for large-scale microbiome projects. The protocol of the QIAamp Fast DNA

298   Stool Mini kit (QIAFast), a kit that became available at the time the present study was

299   carried out, suggested reduced processing time compared to the QIAStool method.

300   When we compared the QIAStool and QIAFast methods using metagenomic

301   sequencing, we obtained a similar richness, diversity, and microbial community

302   composition with these two methods (see Fig. S5 in the supplemental material).

303   Furthermore, given that our previous results suggested that including a bead-beading

304   step might result in a predicted community composition that was more similar to the

305   community of known composition than without this step (Fig. 4), we included a bead-

306   beating step and examined the effect of beads of differing types and cost (Table 1). We

307   obtained a higher DNA concentration using pig feces and the QIAStool kit, when bead

308   beating was applied and the double amount of volume after cell lysis was transferred

309   (Fig. 5A). Similarly, for the QIAFast method, we obtained an on average 2.6-fold higher

310   DNA concentration by including a bead beading step and transferring the double amount

311   of volume after cell lysis, compared to DNA isolations without these modifications (Fig.

312   5A). Both, DNA purity and stability were in the expected range (see Table S3 in the

313   supplemental material). Even though the DNA concentration was higher with these

14

314   protocol modifications, the richness, diversity and community composition did not

315   significantly differ when assessed by 16S rRNA gene profiling (Fig. 5A).

316   A particular DNA isolation method did not however lead to the highest DNA

317   concentrations for each of the three types of specimen. Whereas the highest DNA

318   concentration for sewage was achieved using the QIAFast+BB.GBT+2Trans method

319   (27.30 ng/ul +/- 4.5 SEM), the highest DNA concentration for human feces was obtained

320   using the QIAStool+BB.LMA method (22.50 ng/ul +/- 4.77 SEM) (Fig. 5B). For pig feces,

321   the highest DNA concentrations were obtained using the QIAStool+BB.LMA (15.43 ng/ul

322   +/-3.48 SEM) and QIAStool (14.57 ng/ul +/-3.62 SEM) methods. On average across the

323   three types of specimen, the highest DNA concentrations were obtained using the

324   QIAFast+BB.GBT+2Trans (17.66 ng/ul +/- 4.82 SEM) and QIAStool+BB.LMA (17.46

325   ng/ul +/- 2.54 SEM) methods.

326

327   **DISCUSSION**

328

329   Genomics-based investigations of complex microbiomes greatly enhance our

330   understanding about microbial community composition and function relevant to human,

331   animal, and plant health, infectious diseases, environmental pollution, agriculture, and

332   food safety. One current ambitious goal is to establish a global surveillance system for

333   infectious agents and antimicrobial resistance based on next-generation DNA

334   sequencing approaches (15). Given that infectious agents occupy various ecological

335   habitats, DNA needs to be extracted from various types of specimen using standardized

336   approaches in a time- and cost-efficient manner. It is advantageous, if a range of

337   different specimens can be processed using the same standard operating procedure. In

15

338 light of these considerations, we compared eight commercially available DNA isolation

339 kits (a total of 16 protocols), and based on the findings developed an improved protocol

340 using the QIAamp® Fast DNA Stool Mini kit.

341 Overall, the amounts of DNA obtained from each DNA isolation method differed greatly,

342 and there was no significant correlation between increasing DNA amount and increase

343 in community diversity or richness. The taxonomic microbiome composition appeared to

344 be dependent on both, the specimen and DNA isolation method. For example, the

345 EasyDNA procedure preferentially extracted DNA from Gram-positive bacteria from the

346 human feces and hospital sewage, while the FastDNA procedure preferentially extracted

347 DNA from Gram-positive bacteria from pig feces. Methods that did not include a bead-

348 beating or enzymatic treatment step generally extracted less DNA from Gram-positive

349 bacteria. Furthermore, the results from our experiment that included the detection of

350 spiked bacteria (Gram-negative and Gram-positive) suggests that quantification of

351 distinct organisms from complex specimens is more challenging when the organisms

352 are present at lower abundance levels. Inherent specimen properties may influence the

353 DNA isolation efficiency leading to a biased pattern of microbial community composition.

354 When using a particular procedure we found some similar abundance patterns of

355 specific bacterial families between the three specimen types. However, we also

356 observed several differences (e.g. Fig. 2 and Fig. 3). Hence, one cannot conclude that

357 the DNA from a particular bacterial family will be extracted preferentially using one

358 specific DNA isolation method across different types of specimens. This could be due to

359 different inherent cellular properties of the taxa belonging to a specific family, affecting

360 mechanical and enzymatic cell lysis. Moreover, the chemical and physical composition

361 of the specimen could influence DNA isolation and downstream procedures. For

16

362   example, it is well known that certain compounds, such as humic acid, polysaccharides,

363   and bilirubin can affect PCR (16). Furthermore, fecal sample consistency, reflecting

364   differences in water content and activity, can impact on microbial community

365   composition (17).

366   Our observations from 16S rRNA gene profiling and metagenomics generally agreed,

367   but the taxonomic patterns also exhibited some differences. One reason could be the

368   known primer biases towards certain taxa in 16S rRNA gene based analysis (18). An

369   additional reason could be differences in the composition of the reference databases

370   used for the two sequence-based strategies. While 16S rRNA gene databases are

371   composed of 16S rRNA gene sequences from a large diversity of taxa, the

372   metagenomic sequence databases are based on whole and draft genome sequences

373   from fewer and less diverse taxa. Both strategies complement each other, and efforts

374   are ongoing in developing harmonized analytical workflows for sequence-based

375   microbial community analysis.

376   Based on the insight gained in this study, we have developed an improved DNA

377   isolation method based on the QIAamp$^{®}$ Fast DNA Stool Mini kit. This procedure

378   includes a bead beading step to obtain DNA from both, Gram-positive and Gram-

379   negative taxa, and a step in which the double amount of cell lysate is transferred to the

380   column to increase the DNA quantity. For aqueous sample types, like sewage,

381   additional modifications are included, such as increasing the input amount and

382   processing aliquots in parallel, as described in the SOP. While there was no single

383   approach among the 16 procedures tested that appeared to completely resolve all

384   challenges, we find the SOP based one the QIAamp$^{®}$ Fast DNA Stool Mini kit useful for

385   a number of reasons, including: 1) DNA extracts contained high amounts of DNA

17

386   (sufficient to permitting PCR-free metagenomic sequencing) with high reproducibility 2)

387   DNA extracts were of high quality in terms of DNA purity and stability, 3) DNA from both,

388   Gram-positive and Gram-negative bacteria were reasonably well extracted (including

389   Bifidobacteria), as determined by 16S rRNA amplicon profiling and metagenomic

390   sequencing of spiked and un-spiked complex samples, 4) the method worked well for all

391   examined sample types based on the DNA quality assessment and inferred microbiota

392   composition, 5) the reagents and materials required were cheaper, and 6) the time

393   needed for carrying out the DNA isolation was shorter, compared to several of the other

394   procedures. A standard operating procedure for this DNA isolation method is available

395   from https://dx.doi.org/10.6084/m9.figshare.3475406, and which can be used for

396   different specimen types, and may be relevant to projects like EFFORT-against-AMR,

397   COMPARE-Europe, the International Microbiome Initiative, and International Human

398   Microbiome Standards.

399   In summary, our findings provide new insight into the effect of different specimen types

400   and DNA isolation methods on DNA quantities and genomic-based inference of

401   microbiome composition. We offer an optimized strategy for the DNA isolation for

402   different sample types providing a representative insight into community composition,

403   and which can be conducted in a time- and cost-efficient manner.

18

404   **MATERIAL AND METHODS**

405

406   **Specimen Collection and Handling**

407   Human fecal specimens were collected from a healthy individual. Pig fecal specimens

408   were collected from animals at a conventional pig production farm in Denmark.

409   Untreated sewage was collected from the sewage inlet of the Herlev hospital waste

410   water treatment plant, Denmark. For details regarding sample handling and processing,

411   see Supplemental Materials and Methods (Text S1).

412

413   **Spiking with strain mix**

414   Subsequent to specimen collection, about half of the aliquots from the human, pig, and

415   sewage were spiked with a representative of Gram-positive and Gram-negative bacteria,

416   namely *Staphylococcus aureus* ST398 (strain S0385) and *Salmonella enterica* serotype

417   Typhimurium DT104. For details regarding the preparation of the strain mix, see

418   Supplemental Materials and Methods (Text S1).

419

420   **DNA isolation**

421   In a first step, seven DNA isolation procedures were examined, namely: InnuPure® C16,

422   Analytic Jena AG (InnuPURE); MagNA Pure LC DNA isolation Kit III, Roche

423   (MagNAPure); Easy-DNA^{TM} gDNA Purification Kit, Invitrogen (EasyDNA); MP

424   FastDNA^{TM} Spin Kit, MP Biomedicals (FastDNA); PowerSoil® DNA Isolation kit, MoBio

425   (PowerSoil.HMP); QIAamp® DNA Stool Mini Kit, Qiagen (QIAStool); QIAamp® DNA

426   Stool Mini Kit +Bead Beating, Qiagen (QIAStool+BB) (see Table 1, and details below).

427   In a second step, a variety of modifications to two Qiagen kits were examined, namely

428   the QIAamp® DNA Stool Mini Kit (QIAStool), and QIAamp® Fast DNA Stool Mini Kit

429   (QIAFast). The standard operating procedure for an improved DNA isolation method (i.e.

430   QIAamp Fast DNA Stool Modified, corresponding to QIAFast+BB.GBT+2Trans

431   described here) can be found at https://dx.doi.org/10.6084/m9.figshare.3475406. For

432   details regarding the individual DNA isolation procedures, see Supplemental Materials

433   and Methods (Text S1).

434

435   **DNA quantitation and quality assessment**

436   Subsequent to DNA isolation, the DNA was portioned into 10-µl aliquots to prevent

437   repeated freeze-thawing cycles, and stored at -20°C. DNA concentrations were

438   measured using Qubit® dsDNA BR Assay Kit on a Qubit® 2.0 Fluorometer (Invitrogen,

439   Carlsbad, CA). As DNA extracts can contain contaminants, such as proteins and other

440   organic molecules that can affect downstream procedures such as DNA amplifications in

441   PCR, we determined the DNA purity by measuring the ratios of absorbance at 260/280

442   and 260/230, respectively, using a NanoDrop 1000 Spectrophotometer (Thermo

443   Scientific, Pittsburgh, USA). DNA extracts with a 260/280 ratio between ~1.7 to ~ 2.0,

444   and 260/230 ration between ~2.0 to ~2.2 are regarded as "pure". The stability of the

445   DNA in the extracts was determined by measuring the DNA concentration after 2 and 7

446   days incubation at 22°C. A decrease in DNA concentration over time can indicate the

447   presence of DNases in the extract.

448

449   **16S rRNA gene profiling**

450   16S rRNA gene amplicon libraries were generated using a two-step protocol similar as

451   described in Part # 15044223 Rev. B by Illumina. In a first PCR, the V4 region of the

452   16S rRNA genes were amplified using the universal primers (515f 5'-

20

453   TGCCAGCAGCCGCGGTAATAC (19) and 806r 5'-GGACTACNNGGGTATCTAAT (20).

454   The samples were pooled in equal concentrations, and concentrated using 'DNA clean

455   and concentrator-5 kit' (Zymo Research, Orange, CA). Paired-end 2 × 250 bp

456   sequencing of barcoded amplicons was performed on a MiSeq machine running v2

457   chemistry (Illumina Inc., San Diego, CA, USA). The sequences were processed using

458   the UPARSE pipeline (21) and a OTU x sample contingency table was created. Using

459   QIIME1.8.0 (22), taxonomy was assigned with uclust using assign_taxonomy.py based

460   on the Greengenes 13.8 reference database. Ecological diversity estimates and

461   microbial community comparisons were performed using the relevant scripts provided by

462   QIIME, phyloseq, and R (22-24). For details regarding the 16S rRNA gene-based

463   microbial community analysis, see Supplemental Materials and Methods (Text S1), and

464   the additional material provided through Figshare,

465   https://figshare.com/projects/DNA_Isolation_Methodology_for_Microbiome_Genomics/1

466   4774.

467

468   **Metagenomics**

469   A subset of the DNA extracts was subjected to metagenomic sequencing. The samples

470   were prepared and sequenced following the Nextera XT DNA Library Preparation Guide

471   for the MiSeq system Part # 15031942 Rev. D, using paired-end v2 2×250bp

472   sequencing. The taxonomic microbiome compositions were determined through the use

473   of the MGmapper pipeline (25). The MGmapper package is available for download at

474   www.cbs.dtu.dk/public/MGmapper/. For details regarding the metagenomics-based

475   microbial community analysis, see Supplemental Materials and Methods (Text S1).

476

477   **Differential abundance analysis**

21

478   In order to test for the differential abundance of taxa that may drive the differences

479   observed between the communities derived from the different DNA isolation procedures,

480   we performed DESeq2 analyses. The read count tables from the 16S rRNA gene

481   profiling and metagenomics sequence analysis, respectively, were aggregated to the

482   family level in R (v. 3.2.3, 64bit) (24) We performed an analysis that allows for varied

483   sequencing depth, similar as suggested previously (26), and carried out two-sided Wald

484   tests as implemented in the DESeq2 (v. 1.10.1) package (27). The size factors were

485   determined by DESeq2 from the read count tables. For details regarding the differential

486   abundance analysis, see Supplemental Materials and Methods (Text S1).

487

488   **Quantification of strain mix**

489   The samples that were spiked with the strain mix composed of *S. enterica* Typhimurium

490   DT104 and *S. aureus* ST398 were extracted, sequenced, and analyzed together with

491   the non-spiked samples. For each type of specimen and isolation method, the

492   abundance of Enterobacteriaceae and Staphylococcaceae for 16S rRNA gene profiling

493   and metagenomics, respectively, were determined. The ratios between

494   Enterobacteriaceae and Staphylococcaceae was determined for each sample matrix

495   and isolation method, and compared to the *S. enterica* Typhimurium DT104 / *S. aureus*

496   ST398 ratio of CFU that were added to the original samples. For details regarding the

497   quantification of the strain mix, see Supplemental Materials and Methods (Text S1).

498

499   **Ethics**

500   The collection of human and pig fecal specimens as well as sewage was non-invasive,

501   and were performed in accordance with the Declaration of Helsinki, and complied with

502   Danish and European directives (86/609/EEC). The collection of specimens was

22

503   conducted in accordance with the act on research ethics of heath research projects as

504   administrated and confirmed by the National Committee on Health Research Ethics of

505   Denmark (Region Hovedstaden), Journal nr. H-14013582.

506

507   **Accession numbers**

508   The 16S rRNA gene sequences are available through the INSDC, such as from the

509   European Nucleotide Archive (ENA) at the European Bioinformatics Institute (EBI) under

510   accession number PRJEB12431, and the metagenomic sequences from ENA at EBI

511   under accession number PRJEB14814.

512

525

526   **AUTHOR CONTRIBUTIONS**

527 B.E.K., L.B., F.M.A., and S.J.P. designed the research; B.E.K., L.B., O.L., and P.M.

528 performed the research; B.E.K., L.B., O.L., P.M., A.P., F.M.A., and S.J.P. contributed

529 analytic tools; B.E.K., L.B., O.L., P.M., and S.J.P. analyzed the data; B.E.K., L.B., and

530 S.J.P. wrote the manuscript; and O.L., P.M., and A.P. edited the manuscript. All authors

531 have read and approved the manuscript as submitted.

REFERENCES

532

533   1.   **Ding T**, **Schloss PD**. 2015. Dynamics and associations of microbial community types

534   across the human body. Nature **509**:357–360.

535   2.   **Schloissnig S**, **Arumugam M**, **Sunagawa S**, **Mitreva M**, **Tap J**, **Zhu A**, **Waller A**,

536   **Mende DR**, **Kultima JR**, **Martin J**, **Kota K**, **Sunyaev SR**, **Weinstock GM**, **Bork P**. 2013.

537   Genomic variation landscape of the human gut microbiome. Nature **493**:45–50.

538   3.   **Guidi L**, **Chaffron S**, **Bittner L**, **Eveillard D**, **Larhlimi A**, **Roux S**, **Darzi Y**, **Audic S**,

539   **Berline L**, **Brum JR**, **Coelho LP**, **Espinoza JCI**, **Malviya S**, **Sunagawa S**, **Dimier C**,

540   **Kandels-Lewis S**, **Picheral M**, **Poulain J**, **Searson S**, **Stemmann L**, **Not F**, **Hingamp P**,

541   **Speich S**, **Follows M**, **Karp-Boss L**, **Boss E**, **Ogata H**, **Pesant S**, **Weissenbach J**,

542   **Wincker P**, **Acinas SG**, **Bork P**, **de Vargas C**, **Iudicone D**, **Sullivan MB**, **Raes J**,

543   **Karsenti E**, **Bowler C**, **Gorsky G**. 2016. Plankton networks driving carbon export in the

544   oligotrophic ocean. Nature **532**:465–470.

545   4.   **Costello EK**, **Stagaman K**, **Dethlefsen L**, **Bohannan BJM**, **Relman DA**. 2012. The

546   application of ecological theory toward an understanding of the human microbiome.

547   Science **336**:1255–1262.

548   5.   **Relman DA**. 2014. Actionable Sequence Data on Infectious Diseases in the Clinical

549   Workplace. Clinical Chemistry **61**:38–40.

550   6.   **Petersen TN**, **Rasmussen S**, **Hasman H**, **Carøe C**, **Bælum J**, **Schultz AC**, **Bergmark L**,

551   **Svendsen CA**, **Lund O**, **Sicheritz-Ponten T**, **Aarestrup FM**. 2015. Meta-genomic

552   analysis of toilet waste from long distance flights; a step towards global surveillance of

553   infectious diseases and antimicrobial resistance. Sci Rep 1–9.

554   7.   **Lipkin WI**. 2013. The changing face of pathogen discovery and surveillance. Nature

555   Reviews Microbiology **11**:133–141.

556 8. **Wesolowska-Andersen A**, **Bahl MI**, **Carvalho V**, **Kristiansen K**, **n TS-P**, **Gupta R**,

557 **Licht TR**. 2014. Choice of bacterial DNA extraction method fromfecal material influences

558 community structure asevaluated by metagenomic analysis. Microbiome **2**:1–11.

559 9. **Mackenzie BW**, **Taylor MW**, **Waite DW**. 2015. Evaluating variation in human gut

560 microbiota profiles dueto DNA extraction method and inter-subject differences 1–11.

561 10. **Henderson G**, **Cox F**, **Kittelmann S**, **Miri VH**, **Zethof M**, **Noel SJ**, **Waghorn GC**,

562 **Janssen PH**. 2013. Effect of DNA Extraction Methods and Sampling Techniques on the

563 Apparent Structure of Cow and Sheep Rumen Microbial Communities. PLoS ONE

564 **8**:e74787.

565 11. **Yuan S**, **Cohen DB**, **Ravel J**, **Abdo Z**, **Forney LJ**. 2012. Evaluation of Methods for the

566 Extraction and Purification of DNA from the Human Microbiome. PLoS ONE **7**:e33865.

567 12. **Albertsen M**, **Karst SM**, **Ziegler AS**, **Kirkegaard RH**, **Nielsen PH**. 2015. Back to Basics

568 – The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of

569 Activated Sludge Communities. PLoS ONE **10**:e0132783.

570 13. **Bowers RM**, **Clum A**, **Tice H**, **Lim J**, **Singh K**, **Ciobanu D**, **Ngan CY**, **Cheng J-F**,

571 **Tringe SG**, **Woyke T**. 2015. Impact of library preparation protocols and template quantity

572 on the metagenomic reconstruction of a mock microbial community. BMC Genomics

573 **16**:1–12.

574 14. **Gupta Das NN**, **Basu S**, **Bagchi B**. 1967. Conformational Changes in Denatured DNA,

575 pp. 663–688. *In* Ramachandran, GN (ed.), Conformation of Biopolymers. Academic

576 Press, London and New York.

577 15. **Aarestrup FM**, **Koopmans MG**. 2016. Sharing Data for Global Infectious Disease

578 Surveillance and Outbreak Detection. Trends in Microbiology **24**:241–245.

26

579   16.   **Radstrom P**, **Knutsson R**, **Wolffs P**, **Lovenklev M**, **Lofstrom C**. 2004. Pre-PCR

580   processing: strategies to generate PCR-compatible samples. Mol Biotechnol **26**:133–146.

581   17.   **Vandeputte D**, **Falony G**, **Vieira-Silva S**, **Tito RY**, **Joossens M**, **Raes J**. 2016. Stool

582   consistency is strongly associated with gut microbiota richness and composition,

583   enterotypes and bacterial growth rates. Gut **65**:57–62.

584   18.   **Klindworth A**, **Pruesse E**, **Schweer T**, **Peplies J**, **Quast C**, **Horn M**, **Glockner FO**.

585   2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-

586   generation sequencing-based diversity studies. Nucleic Acids Research **41**:e1–e1.

587   19.   **Yu Y**, **Lee C**, **Kim J**, **Hwang S**. 2005. Group-specific primer and probe sets to detect

588   methanogenic communities using quantitative real-time polymerase chain reaction.

589   Biotechnol Bioeng **89**:670–679.

590   20.   **Sundberg C**, **Al-Soud WA**, **Larsson M**, **Alm E**, **Yekta SS**, **Svensson BH**, **Sørensen SJ**,

591   **Karlsson A**. 2013. 454 pyrosequencing analyses of bacterial and archaeal richness in 21

592   full-scale biogas digesters. FEMS Microbiology Ecology **85**:612–626.

593   21.   **Edgar RC**. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon

594   reads. Nature Methods **10**:996–998.

595   22.   **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello EK**,

596   **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**, **Knights D**,

597   **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**, **Pirrung M**, **Reeder J**,

598   **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**, **Yatsunenko T**, **Zaneveld J**,

599   **Knight R**. 2010. QIIME allows analysis of high-throughput community sequencing data.

600   Nature Methods **7**:335–336.

601   23.   **McMurdie PJ**, **Holmes S**. 2013. phyloseq: An R Package for Reproducible Interactive

27

602   Analysis and Graphics of Microbiome Census Data. PLoS ONE **8**:e61217.

603   24.   **R Development Core Team**. 2014. R: A Language and Environment for Statistical

604   Computing. R Foundation for Statistical Computing, Vienna, Austria.

605   25.   **Petersen TN**, **Bælum J**, **Lukjancenko O**, **Geertz-Hansen HM**, **Thomsen MCF**,

606   **Sperotto MM**, **Lund O**, **Aarestrup FM**, **Sicheritz-Ponten T**. 2015. MGmapper: an

607   automated pipeline for mapping and stratification of metagenomics sequence data.

608   Submitted.

609   26.   **McMurdie PJ**, **Holmes S**. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data

610   Is Inadmissible. PLOS Computational Biology **10**:e1003531.

611   27.   **Love MI**, **Huber W**, **Anders S**. 2014. Moderated estimation of fold change and dispersion

612   for RNA-seq data with DESeq2. Genome Biology **15**:31.

613

614   **FIGURE LEGENDS**

615

616   **FIG 1. Comparison of DNA extraction methods**. (A) Experimental design. Human

617   feces, pig feces, and hospital sewage were extracted using seven different DNA

618   extraction methods (see also Table 1): InnuPure® C16, MagNA Pure LC DNA isolation

619   Kit III, Easy-DNA™ gDNA Purification Kit, MP FastDNA™ Spin Kit, PowerSoil® DNA

620   Isolation kit, QIAamp® DNA Stool Mini Kit, QIAamp® DNA Stool Mini Kit + Bead Beating

621   (For details see Materials and Methods). DNA concentration, purity, and stability were

622   examined, and microbial community composition determined using 16S rRNA gene

623   profiling and metagenomics (selected samples). (B) DNA from each method was

624   dissolved in 100 ul solution and DNA concentrations were determined using Qubit®

625   dsDNA BR Assay Kit measurements. Values represent averages from duplicate or

626   triplicate DNA extractions (See also Supplemental Table S1A). (C) Ecological richness

627   (Chao 1) and diversity (Shannon index) were determined based on contingency tables

628   from 16S rRNA gene profiling and metagenomic sequencing data at OTU and species

629   levels, respectively (See also Supplemental Table S1B).

630

631   **FIG 2. Microbial community dissimilarity.** The dissimilarity between the microbiotas

632   from the human, pig, and sewage samples based on DNA extraction methods was

633   examined using Principal Coordinates Analysis of Bray-Curtis distances (A-C) and

634   differential abundance analysis using DESeq2 (D-F) from 16S rRNA amplicon data. (A-

635   F) For the PCoA Bray-Curtis ordination analysis only samples with 800 or more reads

636   were included. (D-F) For the differential abundance analysis pairwise testing by DNA

637   extraction method was performed, and bacterial families were considered significantly

638   differentially abundant if their adjusted P-value was <0.1 (see also Table S2 in the

639   supplemental material). Examples for differentially abundant families are shown that are

640   among the top10 most abundant taxa found in the sample, respectively. For each family,

641   the total number of DNA isolation procedures, that exhibit significantly different

642   abundance values compared to a particular DNA isolation procedure, are indicated

643   above the plot, respectively.

644

645

29

646 **FIG 3. Differential abundance of bacterial families.** Pairwise testing by DNA
647 extraction method was performed using DESeq2, and the log2-fold difference displayed
648 (column vs. rows) for selected families present in all sample matrices if their adjusted P-
649 value was <0.1 (see also Table S2 in the supplemental material). The rank abundance
650 position for each family per sample matrix type is noted according to their regularized
651 log abundance. The baseMean (bM) indicates the mean of negative-binominal-based
652 normalized read counts. The pairwise comparisons based on relative abundance
653 normalization (total-sum scaling) of the bacterial families for the different DNA isolation
654 procedures and three sample types is available though figshare at
655 https://dx.doi.org/10.6084/m9.figshare.3811254.

656

657 **FIG 4. Detection of spiked bacteria.** The human fecal (A), pig fecal (B), and hospital
658 sewage (C) samples were spiked with a strain mix composed of *Salmonella enterica*
659 serotype Typhimurium DT104 and *Staphylococcus aureus* ST398 in a CFU ratio of 1.02.
660 The three sample matrices, as well as aliquots of the strain mix (D) were extracted using
661 seven different DNA extraction methods. The two strains were detected by 16S rRNA
662 gene profiling, and their ratios determined. For details, see Materials and Methods. An
663 asterisk in (D) indicates that the values for the particular DNA extraction of the strain mix
664 are based on single measurements. All other values are based on averages from
665 duplicate or triplicate DNA extractions. The dashed line indicates the ratio of the strain
666 mix based on CFU determinations. The x-axis scale is the same for all panels (A-D), and
667 the y-axis scale specific for each sample type.

668

669

670 **FIG 5. Effect of protocol modifications.** A) Pig feces was extracted using standard as
671 well as modified protocols based on the QIAamp® DNA Stool Mini and QIAamp® Fast
672 DNA Stool Mini kits. The modifications included bead beading, pre-treatment of the
673 sample, and transfer of the double amount of volume after cell lysis. In the bead-beating
674 step, different bead types were examined (For details, see Materials and Methods, and
675 Table 1). The alpha diversity (Chao 1 and Shannon index) was determined at OTU-
676 level, and the microbial community composition examined at family-level based on 16S
677 rRNA gene profiling. B) Selected standard and modified DNA extraction protocols were

678   employed to extract DNA from human feces, pig feces, and sewage and their DNA

679   concentration was displayed in a star plot. The values indicate the averages from

680   duplicate extractions.

681

682   **SUPPLEMENTAL MATERIAL LEGENDS**

683

684   **Text S1. Supplemental Materials and Methods.** Details regarding specimen collection

685   and handling, spiking with the strain mix, DNA isolation, DNA quantitation and quality

686   assessment, 16S rRNA gene profiling, metagenomics, differential abundance analysis,

687   and quantification of the strain mix are described.

688

689   **FIG S1. Microbial community composition.** The Top 10 most abundant families for

690   the human fecal (A), pig fecal (B), and hospital sewage (C) samples based on (i) 16S

691   rRNA gene profiling, (ii) metagenomics analysis that include normalization based on

692   reference genome size, and (iii) metagenomics analysis without normalization according

693   to genome size. For details regarding sequence data analysis and normalization see

694   Materials and Methods.

695

696   **FIG S2. Microbial community composition based on predicted Gram-staining.**

697   Gram-positive and Gram-negative affiliations were assigned at the order-level based on

698   information found in the literature. For some taxa the Gram-staining status was

699   unknown.

700

701   **FIG S3. Microbial community dissimilarity.** The dissimilarity between the microbiotas

702   from the human, pig, and sewage samples was examined using Principal Coordinates

703   Analysis of Bray-Curtis distances based on the 16S rRNA gene count data. For the

704   PCoA Bray-Curtis ordination analysis only samples with a minimum of 800 reads were

705   included. Additional results regarding community dissimilarity (based on Bray Curtis)

706   and similarity (based on Jaccard similarity coefficient) within and between DNA

707   extraction procedures across sample types as well as for a given sample type, are

708   available through figshare: https://dx.doi.org/10.6084/m9.figshare.3814239.

709

710   **FIG S4. Detection of spiked bacteria using metagenomics.** The human fecal (A), and

711   pig fecal (B) samples were spiked with a strain mix composed of *Salmonella enterica*

712   serotype Typhimurium DT104 and *Staphylococcus aureus* ST398 in a CFU ratio of 1.02.

713   These two sample matrices, as well as aliquots of the strain mix (C) were extracted

714   using three different DNA extraction methods. The two strains were detected by
715   metagenomics analysis, and their ratios determined. For details, see Materials and
716   Methods. An asterisk indicates that the values for the particular DNA extraction of the
717   strain mix (D) are based on single measurements. All other values are based on
718   averages from duplicate or triplicate measurements. The dashed line indicates the ratio
719   of the strain mix based on CFU determinations.

720

721   **FIG S5. Comparison between QIAStool and QIAFast DNA extraction methods by**
722   **metagenomics.** Pig feces was extracted using the QIAamp® DNA Stool Mini and
723   QIAamp® Fast DNA Stool Mini kits, and analyzed using metagenomics. The alpha
724   diversity (Chao 1 and Shannon index) was determined at species-level. The microbial
725   community composition was examined at genus-level and the relative abundance of the
726   Top 10 most abundant taxa are shown here.

727

728   **Table S1. Comparison of DNA extraction methods.** (A) DNA concentration, purity,
729   and stability, and (B) Microbiome richness and diversity.

730

731   **Table S2. Differential abundance of families.** (A) Human fecal microbial community,
732   (B) Pig fecal microbial community, (C) Hospital sewage microbial community.

733

734   **Table S3. Comparison of DNA extraction methods.** DNA concentration, purity, and
735   stability, for different DNA isolation procedures based on the QIAamp® DNA Stool Mini
736   and QIAamp® Fast DNA Stool Mini Kits.

737

TABLE 1. Overview of DNA extraction procedures

Extraction Method	Sample amount (g)	Cell lysis method	Bead type	DNA separation	Cost pr. extraction (€)[a]	Processing time for 20 samples (h)
Step 1: Seven commonly used DNA extraction kits						
InnuPure® C16 (Analytic Jena AG) [A]	0.1	Chemical, Mechanical, Heat	Ceramic	Magnetic beads	7.3	4
MagNA Pure LC DNA isolation Kit III (Roche) [A]	0.25	Chemical, Heat	-	Magnetic beads	2.6[b]	2.5
Easy-DNA™ gDNA Purification Kit (Invitrogen)	0.25	Chemical, Enzymatic	None	Phenol:Chloroform, Precipitation	4.5	8.8
MP FastDNA™ Spin Kit (MP Biomedicals)	0.5	Chemical, Mechanical	Ceramic & Garnet	Silica membrane-based columns	14.1[c]	5
PowerSoil® DNA Isolation kit (MoBio)	0.25	Chemical, Mechanical, Heat	Garnet	Silica membrane-based columns	5.3	5.5
QIAamp® DNA Stool Mini Kit (Qiagen)	0.2	Chemical, Heat	-	Silica membrane-based columns	5.3	4
QIAamp® DNA Stool Mini Kit (Qiagen) +BB (Lysing Matrix A, MP Biomedicals)	0.2	Chemical, Mechanical, Heat	Ceramic & Garnet	Silica membrane-based columns	12.7	4
Step 2: New DNA extraction kit and modified DNA extraction procedures						
QIAamp® DNA Stool Mini Kit (Qiagen) +BB (Garnet Bead Tubes, MoBio)	0.2	Chemical, Mechanical, Heat	Garnet	Silica membrane-based columns	8.5	3
QIAamp Fast DNA Stool Mini	0.2	Chemical, Mechanical, Heat	-	Silica membrane-based columns	6.2	2.6
QIAamp Fast DNA Stool Mini +BB (Lysing Matrix A, MP Biomedicals)	0.2	Chemical, Mechanical, Heat	Ceramic & Garnet	Silica membrane-based columns	13.6	3
QIAamp Fast DNA Stool Mini +BB	0.2	Chemical, Mechanical,	Glass	Silica membrane-based columns	10	3

(Pathogen Lysis Tubes S, Qiagen)		Heat				
QIAamp Fast DNA Stool Mini +BB (Pathogen Lysis Tubes L, Qiagen)	0.2	Chemical, Mechanical, Heat	Glass	Silica membrane-based columns	10	3
QIAamp Fast DNA Stool Mini +BB (Garnet Bead Tubes, MoBio)	0.2	Chemical, Mechanical, Heat	Garnet	Silica membrane-based columns	8.5	3
QIAamp Fast DNA Stool Mini +BB (Bead Beating Tubes, A&A Biotechnology)	0.2	Chemical, Mechanical, Heat	Zirconia / Silica	Silica membrane-based columns	8.2	3

[A] Automated procedure

BB Bead beating

[a]Calculations do not include costs for additional laboratory supply, such as pipette tips and reaction tubes.

[b]Excluding costs for special pipette tips and plastic cartridges required for the robot.

[c]Based on price in the USA, excluding general sales tax that is being added in other countries.

A

Sample matrix

Human   Pig   Sewage

↓

DNA Isolation

| InnuPURE | MagNAPure | EasyDNA | FastDNA | PowerSoil HMP | QIAStool | QIAStool +BB |

↓

Quality control

Nucleic acid measurement   Microbial community analysis

| Concentration | Purity | Stability | 16S rRNA gene profiling | Metagenomics |

B

DNA concentration

Human
Pig
Sewage

C

Richness

16S rRNA gene profiling   metagenomics

Diversity

16S rRNA gene profiling   metagenomics

Prevotellaceae (Bacteroidetes)

Rank 1, bM 18839.0 👶
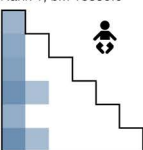
Rank 1, bM 7308.7 🐖
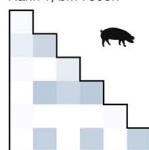
Rank 11, bM 1181.6 🌊
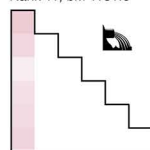
Bacteroidaceae (Bacteroidetes)

Rank 5, bM 3052.9 👶
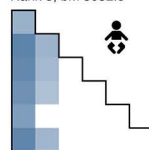
Rank 20, bM 216.5 🐖
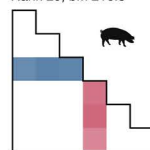
Rank 5, bM 2750.7 🌊
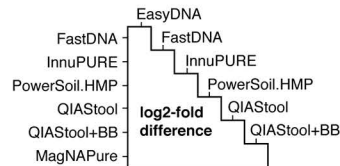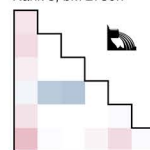
Ruminococcaceae (Clostridia)

Rank 2, bM 14193.0
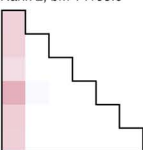
Rank 2, bM 2516.2
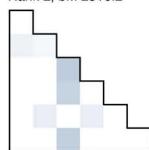
Rank 3, bM 4234.5
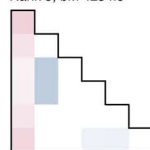
Clostridiaceae (Clostridia)

Rank 11, bM 164.7
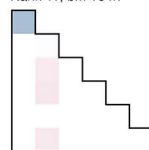
Rank 4, bM 14966.6
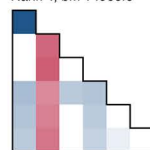
Rank 35, bM 164.7
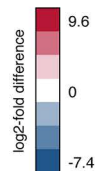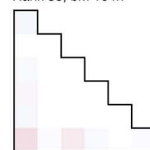
Veillonellaceae (Firmicutes)

Rank 4, bM 3174.4
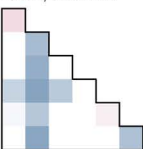
Rank 7, bM 908.7
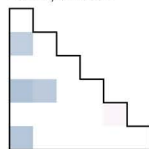
Rank 18, bM 616.0
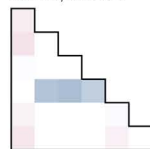
Streptococcaceae (Firmicutes)

Rank 15, bM 118.4
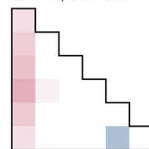
Rank 9, bM 576.8
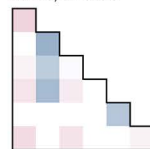
Rank 14, bM 719.1
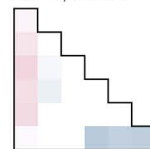
Bifidobacteria (Actinobacteria)

Rank 6, bM 8087.8
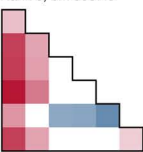
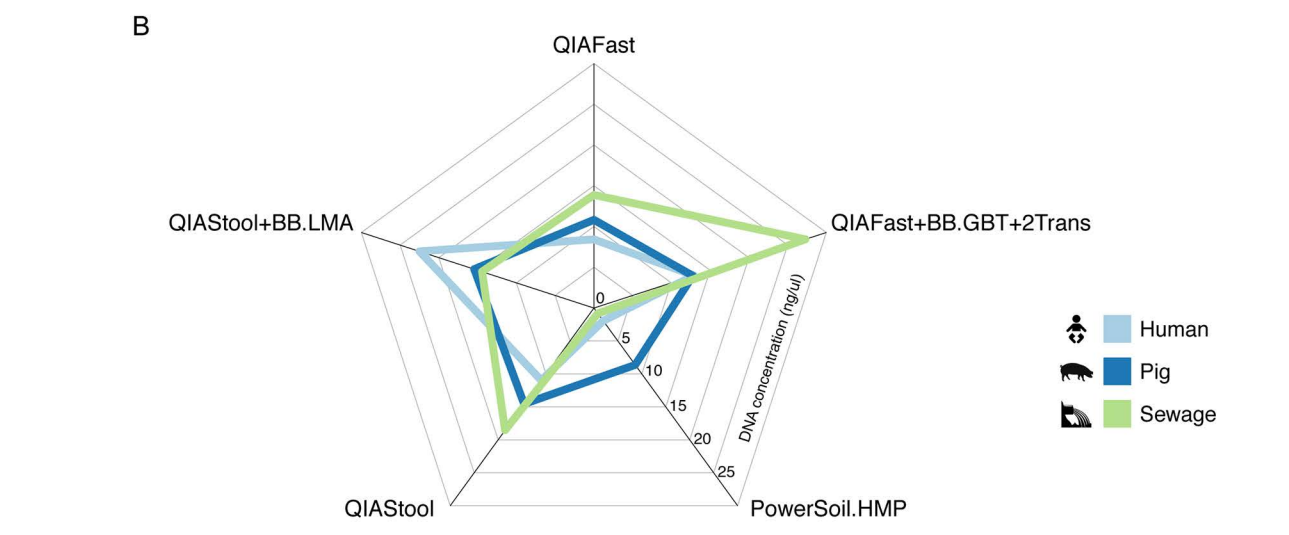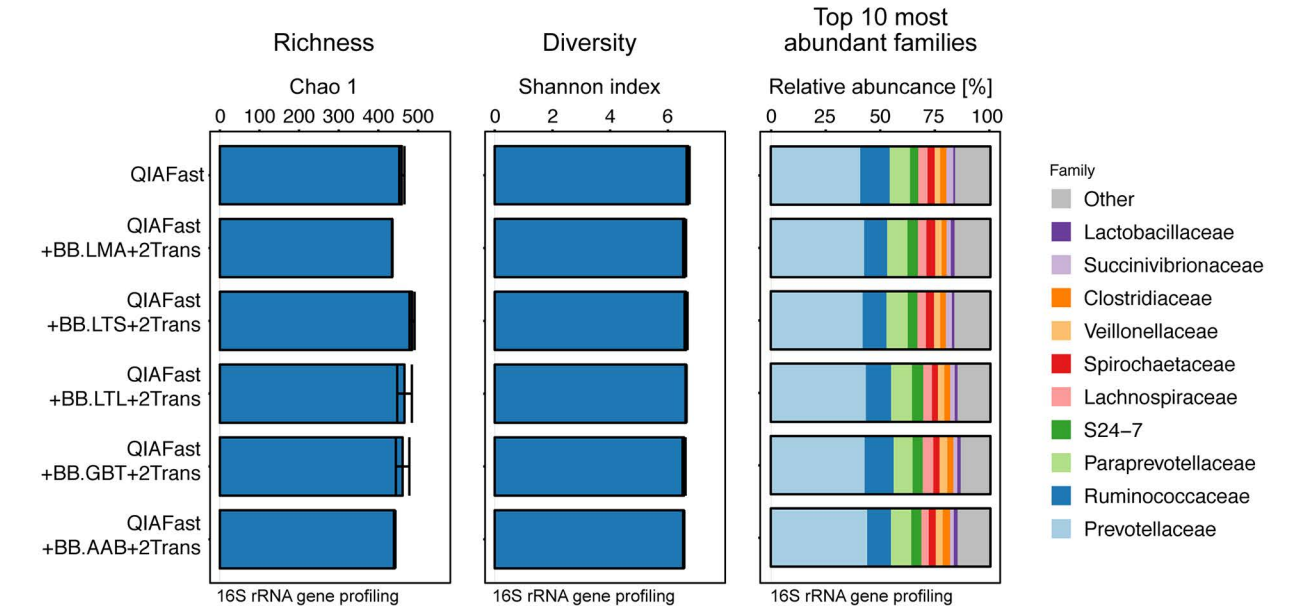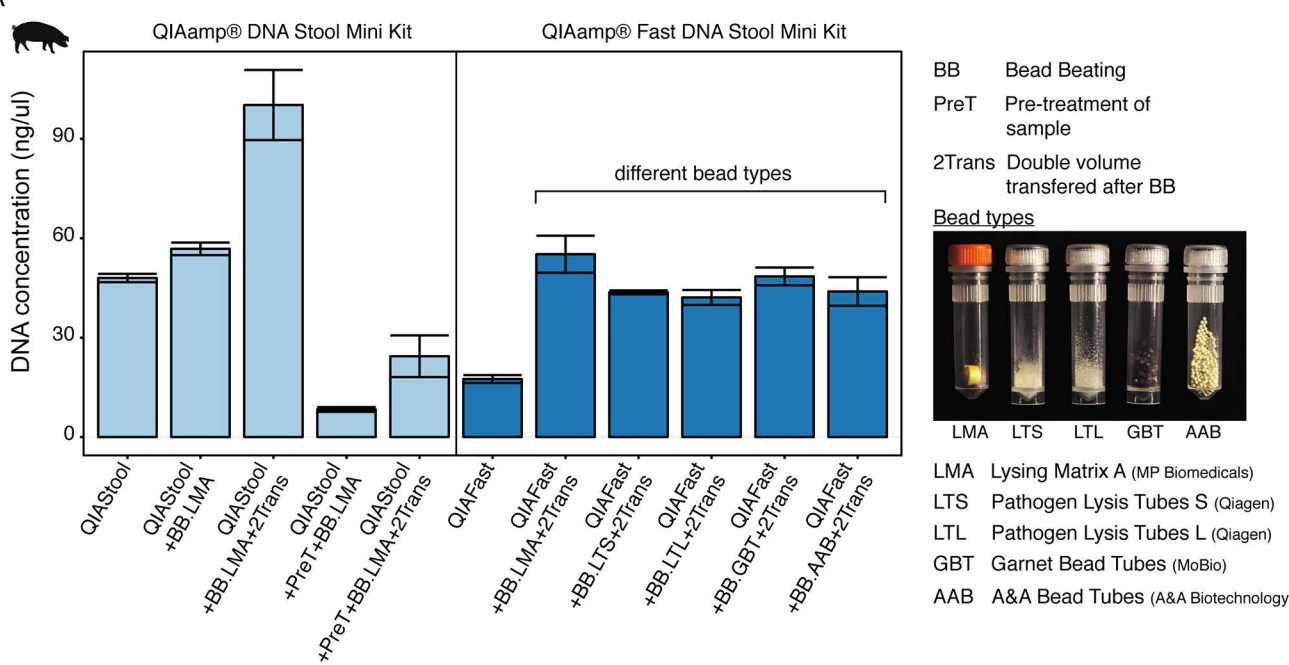Rank 25, bM 42.2
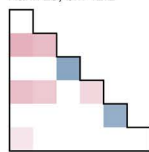
Rank 31, bM 649.2
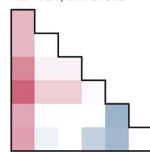
Enterobacteriaceae (γ-Proteobacteria)

Rank 21, bM 33.4
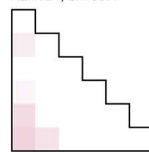
Rank 19, bM 62.3

Rank 19, bM 736.5

EasyDNA
FastDNA FastDNA
InnuPURE InnuPURE
PowerSoil.HMP PowerSoil.HMP
QIAStool QIAStool
QIAStool+BB QIAStool+BB
MagNAPure

log2-fold difference

log2-fold difference

9.6

0

-7.4

👶 Human

🐖 Pig

🌊 Sewage

A

B

C

D

A

QIAamp® DNA Stool Mini Kit / QIAamp® Fast DNA Stool Mini Kit

DNA concentration (ng/ul)

BB — Bead Beating
PreT — Pre-treatment of sample
2Trans — Double volume transferred after BB

Bead types

LMA — Lysing Matrix A (MP Biomedicals)
LTS — Pathogen Lysis Tubes S (Qiagen)
LTL — Pathogen Lysis Tubes L (Qiagen)
GBT — Garnet Bead Tubes (MoBio)
AAB — A&A Bead Tubes (A&A Biotechnology)

Richness — Chao 1
Diversity — Shannon index
Top 10 most abundant families — Relative abunance [%]

16S rRNA gene profiling

Family:
Other
Lactobacillaceae
Succinivibrionaceae
Clostridiaceae
Veillonellaceae
Spirochaetaceae
Lachnospiraceae
S24–7
Paraprevotellaceae
Ruminococcaceae
Prevotellaceae

B

Human
Pig
Sewage

A Human

16S rRNA gene profiling

metagenomics

B Pig

16S rRNA gene profiling
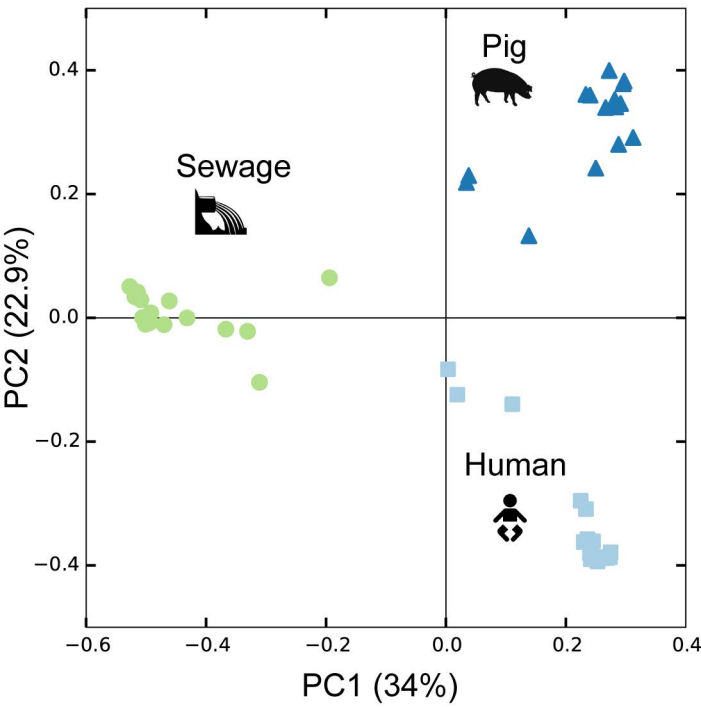
metagenomics

C Sewage

16S rRNA gene profiling

☐ Unknown
☐ Gram - (predicted)
■ Gram + (predicted)

PCoA - Bray Curtis

Richness

Diversity

Top 10 Genera

Other
Roseburia
Mitsuokella
Selenomonas
Megasphaera
Faecalibacterium
Eubacterium
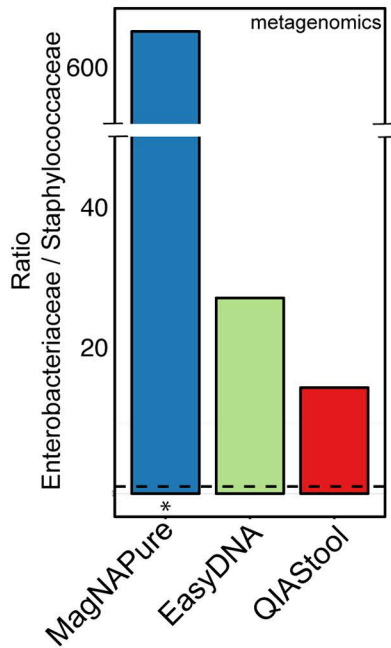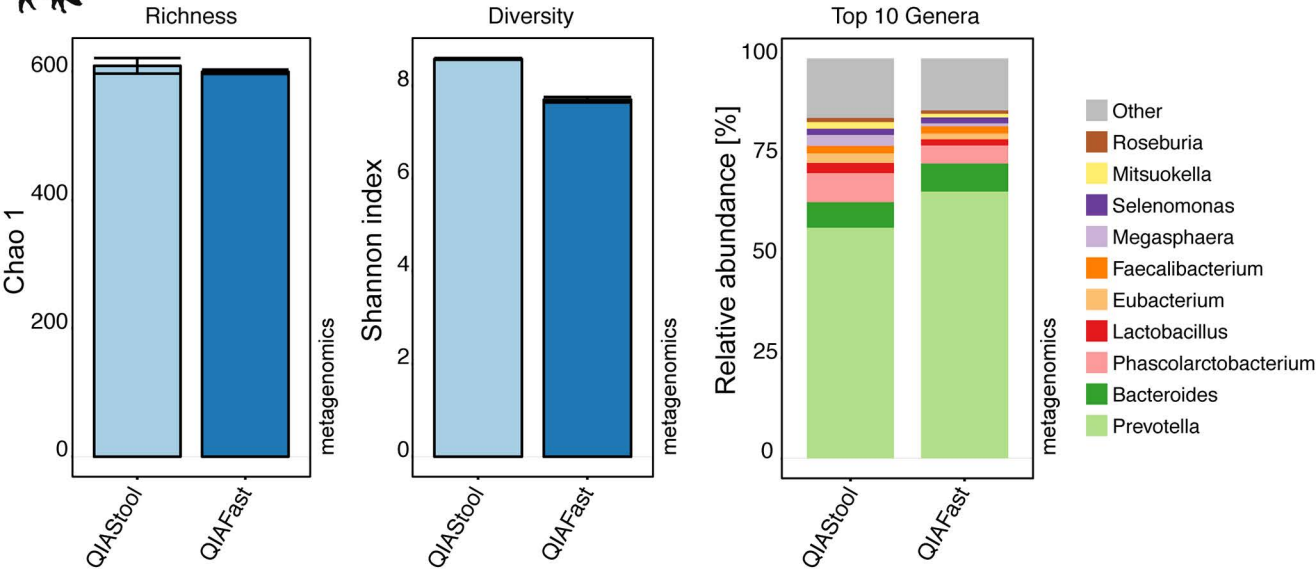Lactobacillus
Phascolarctobacterium
Bacteroides
Prevotella

SUPPLEMENTAL MATERIAL AND METHODS

Specimen Collection and Handling

Human fecal specimens were collected from a healthy individual at three time points over a single day. The specimens were kept at 4°C, and transported to the laboratory within 24 hours. Upon arrival, the three samples were pooled and homogenized. For this study, fecal specimens from an infant were chosen, as infant fecal samples often contain a high proportion of Actinobacteria (e.g. Bifidobacteria), from which genomic DNA can be difficult to isolate. Pig fecal specimens were collected from animals at a conventional pig production farm in Denmark. Samples from individual animals were obtained directly after defecation, stored in a cooling box, and transported to the laboratory within four hours. Upon arrival, three random samples were pooled and homogenized. Untreated sewage was collected from the sewage inlet of the Herlev hospital waste water treatment plant, Denmark. Specimens were stored in a cooling box and transported to the laboratory within two hours. Upon arrival 24 x 40 ml sewage samples were sedimented for 10 minutes at 8000xg in an Eppendorf 5810R centrifuge. The sewage pellets were pooled and homogenized. For all three types of specimen (human feces, pig feces, sewage), the homogenized samples were separated into 0.5 g aliquots, respectively. A subset of aliquots for each specimen type was spiked with two bacterial strains (see details below). The individual sample aliquots with and without strain mix were stored at -80°C until further processing.

Spiking with strain mix

Subsequent to specimen collection, about half of the aliquots from the human feces, pig feces, and sewage were spiked with a representative of Gram-positive and Gram-negative bacteria, namely *Staphylococcus aureus* ST398 (strain S0385) and *Salmonella enterica* serotype Typhimurium DT104. The strains were cultivated in Luria-Bertani (LB) broth at 37°C. Cells were harvested when the culture reached late exponential growth phase at OD_{600} ~0.9. The strain mix was prepared by mixing equal volumes of the bacterial cultures. To determine the number of cells of *S. aureus* ST398 and *Salmonella* Typhimurium DT104 in the two cultures, dilutions of these were plated on LB agar, the plates incubated overnight at 37°C, and colony forming units (CFU) determined the following day. The strain mix was added at about 5% of the volume of the aliquot, and the added cell numbers of *S. aureus* and *S. enterica* Thyphimurium were calculated based on the CFU determinations.

DNA isolation

In a first step, seven DNA isolation procedures were examined, namely: InnuPure® C16 (Analytic Jena AG), MagNA Pure LC DNA isolation Kit III (Roche), Easy-DNA™ gDNA Purification Kit (Invitrogen), MP FastDNA™ Spin Kit (MP Biomedicals), PowerSoil® DNA Isolation kit (MoBio), QIAamp® DNA Stool Mini Kit (Qiagen), QIAamp® DNA Stool Mini Kit (Qiagen) +Bead Beating (see Table 1, and details below). These methods were selected because they are

widely used and represent a variety of isolation procedures involving manual or automated DNA isolation, DNA separation using filter-columns or magnetic beads, chemical or mechanical lysis, and phenol/chloroform-based or non-chloroform based isolations. Bead-beating steps were performed in a Qiagen TissueLyser II if not stated otherwise, and centrifugation steps were carried out in an Ole Dich 157.MP Microcentrifuge (Denmark). DNA isolation was performed on duplicate or triplicate aliquots, dependent on specimen availability. One to two isolation controls were included at each round of isolation.

InnuPure® C16, Analytic Jena AG (InnuPURE)

Automatic isolation with the InnuPURE–C16 robot using the InnuPURE Stool DNA Kit–IP–C16 according to the manufacture's instructions. Prior to the automatic isolation, a lysis step was performed according to the protocol for lysis of bacterial DNA from stool samples using a SpeedMILL PLUS provided by the manufacturer. The cell disruption process was carried out two times for 30 sec at 50 Hz (50 s^{-1}). The DNA was eluted in 100 ul of buffer supplied with the kit.

MagNA Pure LC DNA isolation Kit III, Roche (MagNAPure)

Automatic isolation with the MagNA Pure LC instrument using the DNA Isolation Kit III (Bacteria, Fungi) according to the manufacture's instructions. The pre-isolation step for stool samples described in the protocol was performed before transferring the samples to the MagNA Pure LC. The protocol states a starting amount of a peanut-size sample, and in order to ensure consistency across isolations a starting amount of 0.25 g was chosen. The DNA was eluted in 100 ul of buffer supplied with the kit.

Easy-DNA^{TM} gDNA Purification Kit, Invitrogen (EasyDNA)

The DNA isolation was performed according to the manufacturer's instructions with minor modifications. Pretreatment of the samples were performed following the protocol for small amounts of cells, tissues, or plant leaves. Initially, 0.25 g sample aliquots were resuspended in 1.5 ml 0.9% NaCl, respectively. The samples were centrifuged at 600xg for 3 minutes. The supernatant was transferred to new tubes and centrifuged at 8000xg for 10 minutes. After centrifugation, the supernatant was discarded and the pellet was resuspended in 200 µl PBS. 30 µl lysozyme (10 mg/ml) and 15 µl lysostaphin (10 mg/ml) were added and the samples were incubated at 37°C for 20 minutes shaking at 550 rpm, before adding 30 µl 10% SDS. The final pretreatment step included the addition of 15 µl proteinase K (20 mg/ml) and incubation at 37°C for 20 minutes. The final step in the isolation protocol was prolonged to an incubation for 1.5 (instead of 0.5) hours at 37°C. The DNA was eluted in 100 ul of buffer supplied with the kit.

MP FastDNA^{TM} Spin Kit, MP Biomedicals (FastDNA)

The DNA isolation was performed according to the manufacturer's instructions with minor modifications. A centrifugation step at 3000xg for 2 minutes was included to ensure proper settling of the silica matrix. The protocol suggested

eluting the DNA in 50-100 µl DNase/Pyrogen-Free water, and here the DNA was eluted in 100 µl.

PowerSoil® DNA Isolation kit, MoBio Laboratories Inc. (PowerSoil.HMP)
The DNA isolation was performed according to the protocol employed in the Human Microbiome Project (HMP Protocol # 07-001 version 12), with a minor modification to the initial protocol step. The HMP protocol states to resuspend 2 ml fecal sample in 5 ml MoBio lysis buffer. Here, we resuspended 0.5 g sample in 1.25 ml MoBio lysis buffer (i.e. same ratio). Subsequently, the samples were centrifuged according to the HMP protocol and 1 ml of supernatant transferred to a garnet bead tube containing 0.75 ml MoBio buffer. The samples were heated at 65°C for 10 minutes followed by an additional heating step at 95°C for 10 minutes. The samples were processed further according to the HMP protocol including the modification at step 12, where the centrifugation step was prolonged to 2 minutes. The DNA was eluted in 100 ul of buffer supplied with the kit.

QIAamp® DNA Stool Mini Kit, Qiagen (QIAStool)
The DNA isolation was performed according to the manufacturer's protocol for isolation of DNA from stool for pathogen detection with minor modifications. In the lysis step, the samples were first heated at 70°C for 5 minutes and subsequently at 95°C for 5 minutes. The DNA was eluted in 100 µl elution buffer.

QIAamp® DNA Stool Mini Kit, Qiagen +Bead Beating (QIAStool+BB)
The DNA isolation was performed according to the manufacturer's protocol for isolation of DNA from stool for pathogen detection with minor modifications that included a bead-beating step. Sample aliquots of 0.2 g were mixed with 1.4 ml ASL buffer, respectively, and added to Lysing matrix A bead beating tubes (MP Biomedicals) and were briefly homogenized. The samples were treated in a Qiagen TissueLyser II at 30 f/s (Hz) three times for 30 seconds, with placement of the samples on ice in between bead beating steps. Subsequently, the samples were heated at 95°C for 15 minutes. The remaining steps were carried out according to the manufacturer's recommendations, and the DNA eluted in 100 ul elution buffer.

In a second step, a variety of modifications to two Qiagen kits were examined, namely the QIAamp® DNA Stool Mini Kit (QIAStool), and QIAamp® Fast DNA Stool Mini Kit (QIAFast). The latter was released to the market during the course of this study. The main difference between the QIAStool and the QIAFast kits relies in the way inhibitor compounds are being removed. In the QIAStool kit, InhibitEX tablets are being dissolved in the samples that are adsorbing the inhibitors and together are removed via centrifugation. The QIAFast kit contains an InhibitEX buffer to remove inhibitor compounds, and no tablets are required.

QIAamp® DNA Stool Mini Kit, Qiagen and Modifications

Five different protocols based on the QIAamp® DNA Stool Mini Kit were examined. i) QIAStool: see above, ii) QIAStool+BB.LMA: QIAStool+BB procedure using Lysing matrix A tubes (see above), iii) QIAStool+BB.LMA+2Trans: QIAStool+BB procedure using Lysing matrix A tubes (see above) with modifications. To reduce the loss of sample, the double amount of supernatant was transferred to proteinase K (i.e. 400 µl instead of 200 µl). The volumes of Proteinase K, buffer AL and ethanol were doubled, respectively. Due to the increased volume, the passing of the sample through the spin columns is performed in two centrifugation steps. The DNA was washed twice before elution in 100 µl elution buffer. iv) QIAStool+PreT+BB.LMA: QIAStool+BB procedure using Lysing matrix A tubes (see above) with modifications. An increased starting sample amount was used and pre-treated. 0.5g of sample was mixed with 1.5 ml 0.9% NaCl solution. After homogenization by vortexing, the samples were centrifuged at 600 x g for 3 minutes to settle large particles. The supernatant was centrifuged at 8000 x g for 10 minutes to pellet microbial cells. The pellet was resuspended in 200 µl PBS and transferred to Lysing Matrix A bead beating tubes. v) QIAStool+PreT+BB.LMA+2Trans: QIAStool+BB procedure using Lysing matrix A tubes (see above) with modifications described in iii) and iv).

QIAamp® Fast DNA Stool Mini Kit, Qiagen and Modifications
Six different protocols (i–vi) based on the QIAamp® Fast DNA Stool Mini Kit were examined using five different bead types (ii–vi). i) QIAFast: The DNA isolation was performed according to the manufacturer's protocol for isolation of DNA from stool for pathogen detection with minor modifications. In the lysis-step, the samples were first heated at 70°C for 5 minutes and subsequently at 95°C for 5 minutes. The DNA was eluted in 100 µl elution buffer for 2 minutes. ii) QIAFast+BB.LMA+2Trans: The DNA isolation was performed according to the manufacturer's protocol for isolation of DNA from stool for pathogen detection with minor modifications that included a bead-beating step. Sample aliquots of 0.2 g were mixed with 1 ml InhibitEX buffer, respectively, and added to Lysing matrix A bead beating tubes (MP Biomedicals) and are briefly homogenized. The samples are treated in a Qiagen TissueLyser II at 30 f/s (Hz) three times for 30 seconds, with placement of the samples on ice in between bead beating steps. Subsequently, the samples are heated at 95°C for 7 minutes. Similar to the modifications described above, following the bead-beating and heating steps, the double amount of supernatant was transferred to proteinase K (i.e. 400 µl instead of 200 µl). The volumes of proteinase K, Buffer AL and ethanol were also doubled. The passing of the sample through the filter columns were subsequently carried out in two centrifugation steps rather than one, to accommodate the increased sample volume. The remaining steps were carried out according to the manufacturer's recommendations, and the DNA eluted in 100 ul elution buffer. A laboratory protocol for this procedure can be found at https://dx.doi.org/10.6084/m9.figshare.3475406. iii) QIAFast+BB.LTS+2Trans: Same procedure as described in ii) with Pathogen Lysis Tubes S (Qiagen). iv) QIAFast+BB.LTL+2Trans: Same procedure as described in ii) with Pathogen Lysis Tubes L (Qiagen). v) QIAFast+BB.GBT+2Trans: Same procedure as

described in ii) with Garnet Bead Tubes (MoBio). vi) QIAFast+BB.AAB+2Trans: Same procedure as described in ii) with A&A Bead Tubes (A&A Biotechnology, Gdynia, Poland).

Together, the evaluation and improvements of DNA isolation methods were carried out in a step-wise approach. In the first step, seven DNA extraction kits were evaluated using human feces, pig feces, and hospital sewage (Figures 1–4, Supplemental Figures S1-S4, and Supplemental Tables S1+S2). The standard and modified procedures based on the QIAStool and QIAFast methods were tested using a second set of pig fecal samples (Figure 5A, and Supplemental Figure S5, and Supplemental Table S3). Upon evaluation of the different DNA isolation methods, promising procedures were selected and examined using a new set of human feces, pig feces, and hospital sewage (Figure 5B).

DNA quantitation and quality assessment

Subsequent to DNA isolation, the DNA was portioned into 10 µl aliquots to prevent repeated freeze-thawing cycles, and stored at -20°C. DNA concentrations were measured using Qubit® dsDNA BR Assay Kit on a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA). As DNA extracts can contain contaminants, such as proteins or other organic molecules that can affect downstream procedures such as DNA amplifications in PCR, we determined the DNA purity by measuring the ratios of absorbance at 260/280 and 260/230, respectively, using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Pittsburgh, USA). DNA extracts with a 260/280 ratio between ~1.7 to ~ 2.0, and 260/230 ration between ~2.0 to ~2.2 are regarded as "pure". The stability of the DNA in the extracts was determined by measuring the DNA concentration after 2 and 7 days incubation at 22°C. A decrease in DNA concentration over time can indicate the presence of DNases in the extract.

16S rRNA gene profiling

16S rRNA amplicon libraries were generated using a two-step protocol similar as described in Part # 15044223 Rev. B by Illumina (http://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). In a first PCR, the V4 region of the 16S rRNA genes were amplified using the universal primers (515f 5'-TGCCAGCAGCCGCGGTAATAC (1) and 806r 5'-GGACTACNNGGGTATCTAAT (2). Each 20-µl PCR reaction contained 2 µl of 10 x AccuPrime PCR Buffer II (15mM MgCl2, Invitrogen), 1 µl (10 µM) of the primers, 0.12 µl AccuPrime Taq DNA polymerase (2 units/µl, Invitrogen), 1 µl template DNA and 14.88 µl ddH2O. PCR conditions: denaturation at 94°C for 2 min; 30 cycles at 94°C for 20 s, 56°C for 20 s, 68°C for 30 s; followed by 68°C for 5 min, and 3 min at 70 °C. Subsequently, the PCR products were placed on ice to prevent hybridization between PCR products and nonspecific amplicons. Samples were quantified using Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsbad, CA) on a Lightcycler 96 (Roche, Mannheim, Germany) and adjusted to equal concentrations. In the second PCR the same conditions were

used as in the first round, except the PCR was reduced to 15 cycles and the primers had a unique adaptor/linker/index sequence per sample (3). The PCR products were purified using Agencourt AmPure XP beads (Beckman Coulter Inc, A63881), and concentrations were measured using Quant-iT™ PicoGreen® dsDNA Assay Kit on a Lightcycler 96. The samples were pooled in equal concentrations, and concentrated using 'DNA clean and concentrator-5 kit' (Zymo Research, Orange, CA). Paired-end 2 × 250 bp sequencing of barcoded amplicons was performed on a MiSeq machine running v2 chemistry (Illumina Inc., San Diego, CA, USA) at University of Copenhagen, Section of Microbiology.

The primer sequences were trimmed, quality filtering performed, and paired sequences assembled using the UPARSE pipeline http://drive5.com/usearch/manual/uparse_pipeline.html (4). Low quality reads were removed with a maximum expected error threshold of 0.5 (maxee). Sequences were barcoded and pooled before dereplication and removal of duplicates (-minseqlength 64). Prior to clustering of the OTUs the dereplicated reads were sorted according to abundance, and singeltons where removed (http://drive5.com/usearch/manual/upp_readprep.html). Chimera filtering was performed using UCHIME (5) with rdp_gold.fa as reference database. The reads were mapped back to OTUs, including singletons, at a 97% identity level and an OTU-table was generated using uc2otutab.py. Using QIIME1.8.0 (6), taxonomy was assigned with uclust using assign_taxonomy.py based on the Greengenes 13.8 reference database. The average number of reads per sample was 192965, and the read length was between 186-251 bp. The average number of reads in the isolation controls was 34063 and the majority of these reads were affiliated with the two strains used for spiking (Enterobacteriaceae and Staphylococcaceae), as well as dominant taxa that were present in the complex samples, such as Ruminococcaceae, Prevotellaceae, and Bacteroidales. Ecological diversity estimates and microbial community comparisons were performed using the relevant scripts provided by QIIME, phyloseq, and R (6-8). For the estimation of bacterial diversity and richness (Fig. 1C, and Fig. 5A), and principal coordinate analysis (Fig. 2A-C, and Fig. S3) the samples were rarefied to 800 reads per sample. The abundance of Gram-positive and Gram-negative bacteria was predicted at order levels based on information from the literature. For some bacteria (mainly Firmicutes), the Gram status could not be assigned at this level, and for those the family level was used instead.

Metagenomics
A subset of thirty-nine DNA extracts was subjected to metagenomic sequencing. The samples were prepared and sequenced following the Nextera XT DNA Library Preparation Guide for the MiSeq system, Part # 15031942 Rev. D (http://support.illumina.com/content/dam/illumina-support/documents/documen tation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-guide-15031942-01.pdf), using paired-end v2 2×250bp sequencing. The taxonomic microbiome compositions were determined through the use of the MGmapper pipeline (9). The MGmapper package is available for download at

www.cbs.dtu.dk/public/MGmapper/. The analysis consisted of three main steps: i) Pre-processing and quality trimming of raw reads, ii) Mapping of reads to reference sequence databases, and iii) Analysis of read count data. In the first step, cutadapt (10) was employed for adapter sequence removal, trimming of low-quality bases from the ends of the reads (-q 30), and removal of reads that were shorter than 30 bp. In a second step, the remaining paired-end reads were mapped in chain-mode to four databases: 1. complete bacterial genomes, 2. draft bacterial genomes, 3. MetaHit Assembly (http://www.sanger.ac.uk/resources/downloads/bacteria/metahit/, July 2014), and 4. Human Microbiome assemby (http://www.hmpdacc.org/resources/data_browser.php, July 2014) using the BWA-MEM algorithm (http://bio-bwa.sourceforge.net). For the analysis in the present study only the reads mapping to the two primary bacterial databases (complete and draft bacterial genomes) were considered. These two databases were composed of 2685 complete and 22224 draft bacterial and archaeal genomes obtained from Genbank on July 2014 and December 2014, respectively. The order by which the databases are specified in chain-mode is important, as reads that exhibit a significant hit to the previous reference database are removed before mapping to the next database. Samtools (10) was used to remove singletons and all reads that did not map as pairs. An alignment of a read pair with a region in a genome was considered a hit only if the sum of the alignment scores (SAS) was higher than any SAS values from other hits in the database. In the third step, the alignments were filtered based on the Fraction of Matches+Mismatches (FMM) threshold, i.e. the fraction of a read that should align. Here, the default FMM threshold of 80% was used. From 96 155 142 raw read pairs, 7 567 574 read pairs mapped genomes in the two reference databases. The final read count table was composed of 9436 bacterial and archaeal reference strains with an average of 69952 mapped read pairs per sample. For each sample, the read counts were normalized according to the genome length of the respective genomes in the database, and for sequencing depth using total sum scaling.

Differential abundance analysis
In order to test for the differential abundance of taxa that may drive the differences observed between the communities derived from the different DNA isolation procedures, we performed DESeq2 analyses. The (unnormalized) read count tables from the 16S rRNA gene profiling and metagenomics sequence analysis, respectively, were aggregated to the family level in R (v. 3.2.3, 64bit) (8). We performed an analysis that allows for varied sequencing depth, similar as suggested previously (12), and carried out two-sided Wald tests as implemented in the DESeq2 package (v. 1.10.1) (13). The size factors were determined by DESeq2 from the read count tables. An example for such an analysis is available from https://dx.doi.org/10.6084/m9.figshare.3811251 (available as .Rmd, .html, and .pdf).
When testing the effect of added strain mix, we included the samples to which the strain mix was added as well as the corresponding samples to which no strain mix was added and accounted for DNA isolation method and sample

matrix type. When testing the effects of the DNA isolation method, we analyzed the data from the three types of fecal specimen separately and extracted results from all two-wise comparisons of DNA isolation methods. For each DESeq2 test, p-values were adjusted for the false-discovery rate (FDR) using the Benjamini-Hochberg procedure (14). As recommended by DESeq2, comparisons with an FDR below 0.1 were considered significant. For the visualization of the data, the read count data were variance-stabilized using the DESeq2 regularized log (rlog) transformation. This transformation also accounts for sequencing depth differences, allowing inter-sample comparisons of taxa.

Quantification of strain mix

The samples that were spiked with the strain mix composed of *S. enterica* Typhimurium DT104 and *S. aureus* ST398 were extracted, sequenced, and analyzed together with the non-spiked samples. For each type of specimen and isolation method, the relative abundance of Enterobacteriaceae and Staphylococcaceae for 16S rRNA gene profiling and metagenomics, respectively, were determined. Our differential abundance analysis using DESeq2 confirmed, that these two strains were present in significantly higher abundance in the spiked samples than in the not spiked samples for 16S rRNA gene profiling: Enterobacteriaceae adjusted P-value 3.08^{-30} and Staphylococcaceae adjusted P-value 2.13^{-10}; and for metagenomics: Enterobacteriaceae adjusted P-value 1.74^{-77} and Staphylococcaceae adjusted P-value 1.07^{-4}. The average relative abundance values from the samples without added strain mix were subtracted from the corresponding samples to which the strain mix was added. Subsequently, the 16S rRNA gene copy numbers of the two added strains were taken into account with 5 for *S. aureus* and 7 for *S. enterica* (for 16S rRNA gene profiling). The ratios between Enterobacteriaceae and Staphylococcaceae were determined for each sample matrix and isolation method, and compared to the *S. enterica* Typhimurium DT104 / *S. aureus* ST398 ratio of CFU that were added to the original samples.

Supplemental References

1.   **Yu Y**, **Lee C**, **Kim J**, **Hwang S**. 2005. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. Biotechnol Bioeng **89**:670–679.

2.   **Sundberg C**, **Al-Soud WA**, **Larsson M**, **Alm E**, **Yekta SS**, **Svensson BH**, **Sørensen SJ**, **Karlsson A**. 2013. 454 pyrosequencing analyses of bacterial and archaeal richness in 21 full-scale biogas digesters. FEMS Microbiology Ecology **85**:612–626.

3.   **Riber L**, **Poulsen PHB**, **Al-Soud WA**, **Skov Hansen LB**, **Bergmark L**, **Brejnrod A**, **Norman A**, **Hansen LH**, **Magid J**, **Sørensen SJ**. 2014. Exploring the immediate and long-term impact on bacterial communities in soil amended with animal and urban organic waste fertilizers using pyrosequencing and screening for horizontal transfer of antibiotic resistance. FEMS Microbiology Ecology **90**:206–224.

4.   **Edgar RC**. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods **10**:996–998.

5.   **Edgar RC**, **Haas BJ**, **Clemente JC**, **Quince C**, **Knight R**. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**:2194–2200.

6.   **Caporaso JG**, **Kuczynski J**, **Stombaugh J**, **Bittinger K**, **Bushman FD**, **Costello EK**, **Fierer N**, **Peña AG**, **Goodrich JK**, **Gordon JI**, **Huttley GA**, **Kelley ST**, **Knights D**, **Koenig JE**, **Ley RE**, **Lozupone CA**, **McDonald D**, **Muegge BD**, **Pirrung M**, **Reeder J**, **Sevinsky JR**, **Turnbaugh PJ**, **Walters WA**, **Widmann J**, **Yatsunenko T**, **Zaneveld J**, **Knight R**. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods **7**:335–336.

7.   **McMurdie PJ**, **Holmes S**. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE **8**:e61217.

8.   **R Development Core Team**. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

9.   **Petersen TN**, **Bælum J**, **Lukjancenko O**, **Geertz-Hansen HM**, **Thomsen MCF**, **Sperotto MM**, **Lund O**, **Aarestrup FM**, **Sicheritz-Ponten T**. 2015. MGmapper: an automated pipeline for mapping and stratification of metagenomics sequence data. Submitted.

10.   **Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal **17**:10–12.

11.   **Li H**, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**, **1000 Genome Project Data Processing Subgroup**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics

25:2078–2079.

12. **McMurdie PJ**, **Holmes S**. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLOS Computational Biology **10**:e1003531.

13. **Love MI**, **Huber W**, **Anders S**. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**:31.

14. **Benjamini Y**, **Hochberg Y**. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) **57**:289–300.