Measuring the gut microbiome in birds: comparison of faecal and cloacal sampling

Authors

Elin Videvall[1], Maria Strandh[1], Anel Engelbrecht[2], Schalk Cloete[2], Charlie K. Cornwallis[1]

Affiliation

[1]Department of Biology, Lund University, Sölvegatan 37, SE-22362 Lund, Sweden

[2]Directorate Animal Sciences, Western Cape Department of Agriculture, Private Bag X1, Elsenburg, South Africa

Corresponding author

Elin Videvall (elin.videvall@biol.lu.se)

1

Abstract

The gut microbiomes of birds and other animals are increasingly being studied in ecological and evolutionary contexts. While methods for preserving samples and processing high-throughput sequence data to characterise bacterial communities have received considerable attention, there has been little evaluation of non-invasive sampling methods. Numerous studies on birds and reptiles have made inferences about gut microbiota using cloacal sampling, however, it is not known whether the bacterial community of the cloaca provides an accurate representation of the avian gut microbiome. We examined the accuracy with which cloacal swabs and faecal samples measure the microbiota in three different parts of the gastrointestinal tract (ileum, caecum, and colon) using a case study on juvenile ostriches, *Struthio camelus*, and high-throughput 16S rRNA sequencing. We found that faeces were significantly better than cloacal swabs in representing the bacterial community of the colon. Cloacal samples had a higher abundance of Gammaproteobacteria and fewer Clostridia relative to the gut and faecal samples. However, both faecal and cloacal samples were poor representatives of the microbial communities in the caecum and ileum. Furthermore, the accuracy of the sampling methods in measuring the abundance of different bacterial taxa was highly variable: Bacteroidetes was the most highly correlated phylum between all three gut sections and both methods, whereas colonic Actinobacteria correlated strongly only with faecal samples. This study demonstrates that sampling methods can have significant effects on the inferred gut microbiome in studies of birds. Based on our results, we recommend sampling faeces, whenever possible, as this provides the most accurate assessment of the colon microbiome. The fact that neither sampling technique portrayed the bacterial community of the ileum or the caecum illustrates the difficulty in non-invasively monitoring gut bacteria located further up in the gastrointestinal tract. These results have important implications for the interpretation of avian gut microbiome studies.

Introduction

The community of bacteria harboured within the gastrointestinal tract of animals – 'the gut microbiome' – has been established as an important determinant of host health and physiology (Sekirov *et al.* 2010). Although research has largely focused on humans and model organisms, it is becoming increasingly recognised that the gut microbiome may play an important role in a variety of ecological and evolutionary processes, as it has been associated with disease resistance, behaviour, mate selection, longevity, and adaptation (Sharon *et al.* 2010; Koch & Schmid-Hempel 2011; Muegge *et al.* 2011; Ezenwa *et al.* 2012; Brooks *et al.* 2016; Smith *et al.* 2017). As a result, it is necessary that accurate methods for monitoring the gut microbiome in ecologically relevant contexts are developed. To date, multiple studies have focused on the reliability of methods for storing and preserving samples, as well as techniques for processing data from high-throughput sequencing (see e.g., Debelius *et al.* 2016; Song *et al.* 2016). However, it remains unclear whether different sampling techniques accurately represent the bacterial communities in different parts of the gastrointestinal tract.

A large number of studies investigating the gut microbiome of birds and reptiles have sampled bacteria from the cloaca (Bowman & Jacobson 1980; Cooper *et al.* 1985; Lombardo *et al.* 1996; D'Aloia *et al.* 1996; Mills *et al.* 1999; Dickinson *et al.* 2001; Lamberski *et al.* 2003; Moreno *et al.* 2003; Lucas & Heeb 2005; Maul *et al.* 2005; Santoro *et al.* 2006; Hoar *et al.* 2007; Klomp *et al.* 2008; Ruiz-Rodríguez *et al.* 2009a; b; Martin *et al.* 2010; Xenoulis *et al.* 2010; Santos *et al.* 2012; Charruau *et al.* 2012; Dewar *et al.* 2013, 2014; van Dongen *et al.* 2013; Stenkat *et al.* 2014; Allegretti *et al.* 2014; Matson *et al.* 2015; Stanley *et al.* 2015; Kreisinger *et al.* 2015; Barbosa *et al.* 2016; Merkeviciene *et al.* 2017; Lobato *et al.* 2017; Ganz *et al.* 2017). Cloacal sampling is widely used because it is straightforward to perform, allows repeated sampling of individuals, and the possibility of reliably obtaining samples from all individuals at the same time. This can provide practical advantages over faecal sampling, which may be unreliable and provides potential difficulties in identifying sample ownership and time of defecation.

It is, however, not known if the microbiota of the cloaca provides an accurate reflection of the bacterial community in the gut, and whether cloacal sampling is a good alternative method to faecal sampling. From a theoretical point of view, there are reasons to believe that the bacterial community of the cloaca is not simply seeded with bacteria from faeces. The cloaca constitutes the single posterior opening for the digestive, reproductive, and urinary tract in birds, reptiles, amphibians, sharks, rays, and a few mammals, and as such represents an important barrier to foreign bodies, including pathogens. For example, during copulation, many bird species engage in a so-called "cloacal kiss", where they exchange not only sperm, but also cloacal microbes (Kulkarni & Heeb 2007; White *et al.* 2010). In fact, the avian cloaca has a specialised immune organ, the bursa of Fabricius, that is involved in the development of B lymphocytes and antibody production (Warner & Szenberg 1964), and enables contact between cloacal microbes and the lymphoid system (Schaffner *et al.* 1974). Furthermore, the cloacal mucosa likely constitutes an environment that is mostly aerobic compared to the anaerobic environment of the gastrointestinal lumen, as this is the case for the mammalian rectum (Albenberg *et al.* 2014; De Weirdt & Van de Wiele 2015). Taken together, the

3

83   proximity of the mucosal cloacal microbiome to both the external environment and host tissue,

84   including secreted mucus with immune cells and antimicrobial molecules, likely results in a microbial

85   environment different from that of the gut, and potentially therefore structural differences in

86   microbiota. Nevertheless, several studies investigating bacterial gut composition in birds directly refer

87   to cloacal swabs as faecal samples, with the assumption that they are equivalent (Dewar *et al.* 2013,

88   2014; Allegretti *et al.* 2014; Stanley *et al.* 2015).

89   In line with the idea that the cloaca may accommodate different bacteria, two studies

90   evaluating cloacal swabs with caecal samples in chickens found large differences in bacterial

91   communities (Stanley *et al.* 2015; Zhang *et al.* 2017). It has been argued, however, that cloacal

92   samples may still reflect the presence of the vast majority of caecal bacteria if they are sequenced

93   deep enough (Stanley *et al.* 2015), and it is unclear whether faecal sampling would provide a more

94   accurate picture. This raises the issue of whether particular sampling techniques are superior at

95   measuring specific groups of bacteria in the gut microbiome. For example, certain bacterial taxa may

96   be more widely distributed along the gastrointestinal tract and hence easier to monitor, while other

97   taxa may be confined to specific locations in the gut and thus not well represented by any sampling

98   method. Uncovering what attributes of the gut microbiome different types of sampling methods are

99   able to measure, and what they can infer about the microbial communities present in the different

100   sections of the intestinal tract will be essential to advance our understanding of host microbiomes.

101   In this study we evaluate the accuracy of two commonly used microbiome sampling

102   techniques for birds: cloacal swabs and faecal samples. We test the similarity of the cloacal and faecal

103   microbiomes to three parts of the gastrointestinal tract: ileum, caecum, and colon. For this purpose, 20

104   juvenile ostriches between four to six weeks old were used as a case study.

4

Materials and methods

Study species

We used the ostrich, *Struthio camelus*, as case study species, kept under controlled conditions at the Western Cape Department of Agriculture ostrich research facility in Oudtshoorn, South Africa. The samples in this study were obtained in 2014 from a total of 20 juveniles, which included ten individuals four weeks old and ten individuals six weeks old. Ostrich chicks can easily be maintained and handled in an experimental setting, and this specific age group is ideal in size and temperament for both faecal sampling and dissection, allowing us to efficiently retrieve all necessary samples in a standardised way. The chicks were housed and reared with their contemporaries in four separate groups in indoor pens in the same building, containing approximately 35-40 individuals in each group at the time of sampling. During the daytime they had access to outside enclosures where they could peck freely in soil, and were given ad libitum access to fresh water and food throughout the trial.

Sample collection

Faecal samples were collected from all chicks one day before scheduled euthanization and dissection, by placing sterile plasters over their cloaca and retrieving the collected fresh faeces approximately one hour later. Two to three chicks were randomly selected from each group for gut sampling, totalling ten individuals per sampling event, one at four weeks of age and one at six weeks. Before dissection, the 20 randomly selected chicks were euthanized by a licensed veterinarian who severed the carotid artery. All procedures were approved by the Departmental Ethics Committee for Research on Animals (DECRA) of the Western Cape Department of Agriculture, reference number R13/90. During the dissection we collected four samples from each individual: cloacal swabs and samples from the ileum, caecum, and colon. Cloacal samples were collected by using sterile cotton swabs that were briefly moistened in phosphate-buffered saline (PBS), and the tip carefully inserted in the cloaca of the birds and gently rotated.

To minimize contamination between samples and individuals, a number of precautions were taken. Lab benches and surfaces were routinely sterilized with 70% ethanol, and equipment used during the dissection was first cleaned with hot water, then rinsed with 70% ethanol and subsequently placed in the open flame of a Bunsen burner between each sample collection for sterilization. Control swabs were collected during both dissection events and during the faecal sampling. The control swabs followed the same initial procedure as the cloacal swabs (dipping sterile cotton swabs in PBS), but instead of sampling the bird, they were exposed to potential microbes in the air by waving the wet swab around in the dissection/sampling room. All samples were collected in plastic 2 ml micro tubes (Sarstedt, cat no. 72.693) between October 28 and November 12, 2014, and stored at -20 °C within two hours of collection. They were subsequently transported on ice to a laboratory and stored at -20 °C.

145

146 *DNA isolation, library preparation, and amplicon sequencing*

147 We prepared sample slurries for all sample types with guidance from Flores *et al.* (2012) and

148 subsequently extracted DNA using the PowerSoil-htp 96 well soil DNA isolation kit (MoBio

149 Laboratories, cat no. 12955-4) as recommended by the Earth Microbiome Project

150 (www.earthmicrobiome.org) (for full details please see Supplementary Methods available online).

151 Libraries for sequencing of the 16S rRNA V3 and V4 regions were prepared using the primers

152 Bakt_341F and Bakt_805R (Herlemann *et al.* 2011) according to the Illumina 16S Metagenomic

153 Sequencing Library Preparation Guide (Part # 15044223 Rev.B). All samples in this study (Table S1)

154 were sequenced in one 300-bp paired end run on an Illumina MiSeq platform at the DNA Sequencing

155 Facility, Department of Biology, Lund University, Sweden. In a subsequent run, we sequenced blank

156 samples and additional control samples that were collected during the trial for a related project. These

157 control samples were not essential for this particular study, but were included to increase the number

158 of controls. As a result, a total of 117 different samples plus 54 sample replicates (see Supplementary

159 Methods) were part of this study.

160

161

162 *Data processing*

163 The 16S amplicon sequences were quality controlled using FastQC (v. 0.11.5) (Andrews 2010)

164 together with MultiQC (Ewels *et al.* 2016). Primers were removed from the sequences using

165 Trimmomatic (v. 0.35) (Bolger *et al.* 2014) and the forward reads were retained for analyses. Quality

166 filtering of the reads were executed using the script multiple_split_libraries_fastq.py from QIIME (v.

167 1.9.1) (Caporaso *et al.* 2010). All bases with a Phred score < 25 at the 3' end of reads were trimmed

168 and samples were multiplexed into a single high-quality multi-fasta file.

169 Operational taxonomic units (OTUs) were assigned and clustered using Deblur (v. 1.0.0)

170 (Amir *et al.* 2017). Deblur circumvents the problems surrounding clustering of OTUs at an arbitrarily

171 threshold by obtaining single-nucleotide resolution OTUs after correcting for Illumina sequencing

172 errors. This results in exact sequence variants (ESVs), also called amplicon sequence variants (ASVs),

173 oligotypes, and sub-OTU (sOTUs). In order to avoid confusion, we chose to call these units OTUs,

174 but the reader should be aware that they differ from the traditional 97% clustering approach (Callahan

175 *et al.* 2017). The minimum reads-option was set to 0 to disable filtering inside Deblur, and all

176 sequences were trimmed to 220 bp. We used the biom table produced after both positive and negative

177 filtering, which by default removes any reads which contain PhiX or adapter sequences, and only

178 retains sequences matching known 16S sequences. Additionally, PCR-originating chimeras were

179 filtered from reads inside Deblur (Amir *et al.* 2017).

180 Taxonomic assignment of OTUs was performed using the Greengenes database (DeSantis *et*

181 *al.* 2006). We filtered all samples on a minimum read count of 1000 sequences, resulting in three out

182 of 171 samples being excluded (one ileum and two control samples). We further filtered all OTUs that

183 only appeared in one sample, resulting in the removal of 8,846 OTUs, with 3,015 remaining. The

184 samples with technical replicates (two control samples and seven individuals with replicates for all

185 sample types; see Supplementary Methods) had the replicates merged within their respective sample

6

186   type (i.e. ileum.rep1 + ileum.rep2) to increase the amount of valuable sequence information. The

187   analyses were evaluated with both rarefied and non-rarefied data, which produced extremely similar

188   and comparable results. We therefore present the results from the non-rarefied data in this study, as

189   recommended by McMurdie & Holmes (2014).

190

191

192   *Data analyses*

193   All analyses were performed in R (v. 3.3.2) (R Core Team 2017). We calculated alpha diversity using

194   the Shannon measure with absolute abundance of reads, and distance measures with the Bray-Curtis

195   distance method on relative read abundances in phyloseq (v. 1.19.1) (McMurdie & Holmes 2013).

196   Differences between the microbiota of cloacal and faecal samples to the microbiota of each gut

197   section were examined using permutational multivariate analysis of variances (PERMANOVA) on

198   Bray-Curtis distances using the Adonis function in vegan (v. 2.4-2) with 1000 permutations (Oksanen

199   *et al.* 2017). To analyse if there were differences in the variance (dispersion) between sample types,

200   we used the multivariate homogeneity of group dispersions test (betadisper) in vegan (Oksanen *et al.*

201   2017), followed by the Tukey's 'Honest Significant Difference' method. Blank and control samples

202   showed highly dissimilar microbial composition to all other sample types (see Figures S1, S2, S3, and

203   S4) and were not included in any further analyses.

204   To evaluate bacterial abundances, we first filtered out all OTUs with less than 10 sequence

205   reads and then, using DESeq2 (v. 1.14.1), counts were modelled with a local dispersion model and

206   normalised per sample using the geometric mean (see the DESeq2 manual) (Love *et al.* 2014). We

207   examined the strength of the correlations between the abundance of bacteria (normalised OTU

208   abundance), both at the level of phylum and class, in the three parts of the gut in relation to the

209   abundances in both cloacal swabs and faecal samples. Two sets of correlations were performed, one

210   where each data point represented the mean number of OTUs in that bacterial taxon averaged across

211   the 20 individuals ('correlations across bacteria', n = number of OTUs per bacteria phylum or class)

212   and one set of correlations where each data point represented the abundance of a bacterial taxon per

213   individual ('correlations across individuals', n = 20). We used Spearman's rank-order correlation and

214   tested the differences between correlations obtained for cloacal samples and those from faecal

215   samples using cocor (v. 1.1-3) (Diedenhofen & Musch 2015).

216   Differential abundances between sample types were subsequently tested in DESeq2 with a

217   negative binomial Wald test using individual ID as factor and with the beta prior set to false (Love *et*

218   *al.* 2014). The results for specific comparisons were extracted (e.g. faeces versus ileum) and p-values

219   were corrected with the Benjamini and Hochberg false discovery rate for multiple testing. OTUs were

220   labelled significant if they had a corrected p-value (q-value) < 0.01. Plots were made using phyloseq

221   (McMurdie & Holmes 2013) and ggplot2 (Wickham 2009).

7

Results

Overall microbiome composition in different sample types

First, we evaluated the overall pattern of the microbial community reflected by the two sampling techniques (cloacal swabs and faeces) and the three different sections of the avian gastrointestinal tract (Figure 1). The abundance of bacterial taxa in the microbiomes of the caecum, colon, and faeces showed large overall similarities (Figures 1C, 1D), especially the faecal and colon samples which closely clustered in the network plot (Figure 1B). These three sample types also had the highest and most similar alpha diversity values (colon mean Shannon's diversity index H = 4.47, faeces H = 4.25, and caecum H = 4.16; Figure 1A). Bacteria from the classes Clostridia (phylum: Firmicutes) and Bacteroidia (phylum: Bacteroidetes) mainly dominated in the caecum, colon, and faeces (Figures 1C, 1D). In contrast, the cloacal and ileum samples showed large overall taxa dissimilarities in microbiota composition compared to the other samples types (Figure 1). The microbiome of the cloaca had significantly lower alpha diversity compared to the caecum, colon, and faeces (H = 3.40, paired Wilcoxon signed rank test against caecum: V = 37, p = 0.009; against colon: V = 4, p < 0.0001, and against faeces: V = 17, p = 0.0004), and so did the ileum (H = 2.50, pairwise comparisons against caecum, colon, and faeces: V = 0, p < 0.0001). The cloaca showed a distinct microbial community from all other samples at the class level with a high relative abundance of Gammaproteobacteria and Bacilli, and a lower abundance of Clostridia (Figures 1C, 1D). The ileum also showed higher abundance of Bacilli and lower abundance of Clostridia, but was overall dissimilar to all other samples with a high representation of Betaproteobacteria and very few Bacteroidia (Figures 1C, 1D).

Distances between the microbiomes of the cloaca and faeces to the gut sections

Second, to evaluate overall microbiota distance dissimilarities between the two sample methods to the gut samples, we conducted multivariate analyses of variance (Adonis). All comparisons were significantly different from each other (PERMANOVA: p < 0.001), indicating that each sample type harbours a unique microbiome. This was due to differences in mean distances between communities, not differences in variances, as there was no difference in dispersion between sample types (multivariate homogeneity test of group dispersions: adjusted p > 0.152). The two most similar sample types were the faeces and colon, which resulted in a low R^2 (0.069), whereas the cloaca and colon were more dissimilar ($R^2 = 0.099$). Both sampling methods reflected greater dissimilarities to the gut sections further up in the gastrointestinal tract, with faecal samples being more distant to the caecum ($R^2 = 0.191$) and the ileum ($R^2 = 0.160$), as were cloacal samples (caecum: $R^2 = 0.136$, ileum: $R^2 = 0.145$).

To directly test how well cloacal swabs and faecal samples represented the microbiota in the gut, we calculated community Bray-Curtis distances between the faecal and cloacal samples to each of the three sections of the gut for each individual (Figure 2). Neither sampling technique was particularly good at measuring the microbiome of the ileum (cloacal mean distance = 0.87, faecal

8

262   mean distance = 0.84, paired Wilcoxon signed rank test: V = 136, p = 0.104) or the caecum (cloacal

263   mean distance = 0.82, faecal mean distance = 0.84, V = 88, p = 0.546). However, the faecal samples

264   were significantly closer in distance to the colon than the cloacal samples were (cloacal mean distance

265   = 0.74, faecal mean distance = 0.63, V = 164, p = 0.027) (Figure 2).

266

267

268   *Correlation of bacterial abundances in the cloaca and faeces with the gut sections*

269   We further evaluated how accurately the sampling techniques represented the abundance of

270   all OTUs in the ileum, caecum, and colon, and found that the correlations of both faecal and cloacal

271   samples with the ileum and caecum were weak (r = 0.045–0.268; Figure 3). Conversely, the

272   correlations with the colon were strong, especially for the faecal samples (r = 0.558 versus r = 0.476

273   for cloacal swabs; Figure 3). When analysing the abundances of different bacterial phyla separately,

274   we again found that the correlations between the sampling methods and the ileum were weak for all

275   six phyla (r < 0.275; Table S2). The phyla abundance correlations were stronger for the colon (r =

276   0.246–0.803; Table S2), but highly variable for the caecum (r = -0.127–0.633; Table S2). Similar

277   patterns of correlation were also found when analysing abundances across different bacterial classes

278   (Table S3). More specifically, the phylum Bacteriodetes had the strongest correlations between both

279   sampling methods and each of the three gut sections (Table S2), and at a lower taxonomic level, the

280   two classes Bacteriodia (phylum: Bacteroidetes) and Coriobacteriia (phylum: Actinobacteria)

281   displayed strong correlations between each of the two sampling techniques to both the caecum and

282   colon (Table S3). Overall, the correlations between faecal samples and cloacal swabs to the different

283   parts of the gut were similar with a few exceptions. For example, the abundance of Actinobacteria in

284   the colon and caecum appeared to be better represented in faeces, whereas the abundance of

285   Tenericutes and Betaproteobacteria in the same intestinal regions appeared to be better represented in

286   cloacal swabs.

287   In addition, we examined how well the abundances of different bacteria correlated between

288   samples from the same host individuals (Figure S4). Overall OTU abundance in the ileum was weakly

289   correlated with faeces (r = 0.162), but more strongly with cloacal swabs (r = 0.493). In contrast,

290   individual faecal samples showed extremely high correlations to both the caecum (r = 0.872) and the

291   colon (r = 0.893), whereas cloacal samples only showed intermediate correlations to the caecum (r =

292   0.509) and colon (r = 0.538) (Figure S4).

293

294

295   *Differences in abundance of specific OTUs in the cloaca and faeces versus the gut sections*

296   Next, we analysed whether specific OTUs were more or less abundant when using either of the two

297   sampling techniques by testing for significant differences (q < 0.01) in OTU abundance in the cloacal

298   and faecal samples compared to the three gut sections (Figure 4; Tables S4-S9). Consistent with our

299   previous analyses, we found the highest number of significantly different OTUs when comparing the

300   ileum to both the faecal (n = 307) and cloacal samples (n = 250), followed by the comparisons with

301   the caecum (144 significant OTUs for faeces versus 123 for cloacal swabs). The colon showed the

302   least differences in abundance to both sampling methods, but the cloacal samples had twice as many

9

303   significant OTUs (n = 64) compared to faecal samples (n = 32), indicating substantially more

304   differences between cloaca-colon than faeces-colon (Tables S8-S9).

305   We further evaluated the taxa that showed significantly different abundances across the six

306   sample comparisons. Relative to the ileum, a large number of OTUs in the phylum Firmicutes were

307   significantly more abundant in both faeces and the cloaca (Figure 4; Tables S4-S5). The most

308   significant Firmicutes families included Ruminococcaceae, Lachnospiraceae, Erysipelotrichaceae,

309   Clostridiaceae, and Christensenellaceae (Tables S4-S5). The Enterobacteriaceae (Proteobacteria), the

310   Verrucomicrobiaceae (Verrucomicrobia), and several families within the Bacteroidetes were also

311   significantly more abundant in both faeces and the cloaca compared to the ileum. When comparing

312   sampling methods against the caecum, several Firmicutes bacteria showed significantly different

313   abundances in both directions (Figure 4). The caecum showed, however, a significantly higher

314   abundance of Bacteroidetes relative to both the cloaca and faeces, with one exception: an OTU within

315   the Rikenellaceae, which was completely absent in the caecum samples but present in both sampling

316   methods. Interestingly, the cloaca had a lot more significantly different Proteobacteria OTUs (n = 19)

317   than faeces did (n = 2) in the comparison with the caecum, and 94.7% of those were more abundant in

318   the cloacal samples (Figure 4; Tables S6-S7). Finally, the comparison between the colon and faeces

319   only resulted in 13 significantly different bacterial families within five phyla, while the difference

320   between the colon and cloaca was much larger and phylogenetically broader, representing 28

321   significantly different families from 11 phyla (Figure 4; Tables S8-S9).

Discussion

322

323

324

325 Measuring the gut microbiome of birds and other animals is becoming increasingly important for
326 ecologists and evolutionary biologists due to its potential implications for host fitness. Numerous
327 studies sample either the cloacae or faeces of birds as a proxy for estimating the bacterial community
328 in the gut, however, it has remained untested whether cloacal or faecal sampling constitute accurate
329 ways of measuring avian gut bacteria. In this study we examined the microbiota of cloacal swabs and
330 faeces and compared them to the microbiota in three different sections of the gastrointestinal tract. We
331 found that cloacal swabs were less accurate at representing the microbiome of the colon relative to
332 faecal samples, which had more similar community composition and abundances of bacteria. Neither
333 faeces nor cloacal swabs could, however, accurately estimate the bacterial communities of the ileum
334 and the caecum. These results have important implications for the interpretation of bird gut
335 microbiomes, and we hope they will aid researchers in the planning of future studies.

336   The different sections of the gastrointestinal tract were associated with spatial heterogeneity
337 in their bacterial composition, which is largely expected given their different physiological functions.
338 The ileum is the final part of the small intestine and has a primary role of absorbing nutrients from
339 food while maintaining a neutral pH. In our study, the ileal microbiome had the lowest alpha
340 diversity, which is consistent with other studies investigating the small intestine of birds and reptiles
341 (Bjerrum *et al.* 2006; Danzeisen *et al.* 2015; Kohl *et al.* 2017). It also had the highest relative
342 abundance of Bacilli and Betaproteobacteria compared to the other sample types. The second sample
343 site of the gastrointestinal tract, the caecum, provides important functions by breaking down plant and
344 fibrous material, and birds typically have two caeca, located between the small and large intestines.
345 Although the caecal samples in our study were dissimilar to other sample types, they most closely
346 clustered with samples from the colon, at least at higher taxonomic levels. Both of these intestinal
347 regions had high abundances of Clostridia and Bacteroidia. In comparison to both faecal and cloacal
348 samples, the caecum had a significantly higher abundance of several Bacteroidetes, similar to
349 previous research on the chicken caecum (Stanley *et al.* 2015). The final part of the intestinal tract, the
350 colon, has a primary function to absorb water and salt from ingested material. The colon samples in
351 our study had the highest alpha diversity of all sample types and the strongest taxa correlations to both
352 sampling methods, although significantly better to faeces than to the cloaca.

353   The similarities of both the cloacal and faecal microbiota to that of the gut increased the
354 further down the gastrointestinal tract we sampled, as perhaps expected given the proximity to the
355 cloaca. Nevertheless, there was substantial variation in how well bacterial abundances in different
356 parts of the gut correlated with faecal and cloacal samples across different taxonomic groups, both
357 when examining across individual OTUs and across individual hosts. This variation does not appear
358 to be simply explained by differences in the total abundance of different bacteria (e.g. more abundant
359 bacteria might be more widely distributed in the gut and so more strongly correlated across samples),
360 as some classes of bacteria had high numbers of OTUs, but were poorly correlated and vice versa
361 (Table S3). The strength of correlations between different sample types may potentially reflect the
362 fact that different bacteria vary markedly in the environmental conditions they can tolerate, and hence
363 the breadth of their spatial distribution in the gut. The causes underlying this variation require further

364   investigation, but by presenting effect sizes of the strength of associations we hope to provide useful

365   information on which bacteria can reliably be monitored in different locations of the gastrointestinal

366   tract (Tables S2-S3).

367   A common goal of microbiome studies, particularly in ecological contexts, is to understand

368   how gut bacteria relate to phenotypic variation. Because it is not feasible to collect intestinal samples

369   in wild birds without highly invasive techniques, faecal or cloacal samples are often the only option,

370   especially if repeated sampling is required. Our results suggest that the bacterial communities in the

371   upper and middle gastrointestinal tract are distinct from those recovered by non-invasive sampling

372   methods, and as such, any inferences made about the gut microbiome and its relationship to

373   phenotypic variation may only be possible for processes occurring in the colon. Further studies are

374   needed to investigate if the results of this study hold true for other animals with cloacae, such as frogs,

375   lizards, and egg-laying mammals. Most mammals possess a rectum instead of a cloaca, which differs

376   in both function and physiology, and rectal swabs are therefore likely to differ substantially to cloacal

377   swabs in the degree to which they are useful for monitoring gut microbiomes. The current evidence as

378   to whether rectal swabs constitute a representative sampling method of the gut microbiome of

379   mammals is conflicting and suffers from low sample sizes, thus warranting additional evaluation

380   (Budding *et al.* 2014; Alfano *et al.* 2015; Bassis *et al.* 2017).

381   In conclusion, for gut microbiome sampling of birds, we recommend faecal samples

382   whenever possible, as this sampling procedure best captured the bacterial community of the colon.

Acknowledgements

Author contributions

398
399
400   Study design was planned by E.V. and C.K.C. Sample collection was performed by E.V., C.K.C., and
401   A.E. Animal facilities were provided by S.C., and supervised by A.E. The laboratory work was
402   planned and performed by M.S. The analyses were performed by E.V. with C.K.C. assisting. E.V.
403   wrote the paper with input from all authors.

References

Albenberg L, Esipova T V, Judge CP *et al.* (2014) Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology*, **147**, 1055–1063.e8.

Alfano N, Courtiol A, Vielgrader H *et al.* (2015) Variation in koala microbiomes within and between individuals: effect of body region and captivity status. *Scientific Reports*, **5**, 10189.

Allegretti L, Revolledo L, Astolfi-Ferreira CS *et al.* (2014) Isolation and molecular identification of lactic acid bacteria and *Bifidobacterium* spp. from faeces of the blue-fronted Amazon parrot in Brazil. *Beneficial Microbes*, **5**, 497–503.

Amir A, McDonald D, Navas-Molina JA *et al.* (2017) Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*, **2**, e00191-16.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Barbosa A, Balagué V, Valera F *et al.* (2016) Age-Related Differences in the Gastrointestinal Microbiota of Chinstrap Penguins (*Pygoscelis antarctica*). *PLOS ONE*, **11**, e0153215.

Bassis CM, Moore NM, Lolans K *et al.* (2017) Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. *BMC Microbiology*, **17**, 78.

Bjerrum L, Engberg RM, Leser TD *et al.* (2006) Microbial Community Composition of the Ileum and Cecum of Broiler Chickens as Revealed by Molecular and Culture-Based Techniques. *Poultry Science*, **85**, 1151–1164.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bowman TA, Jacobson ER (1980) Cloacal Flora of Clinically Normal Captive Psittacine Birds. *Journal of Zoo Animal Medicine*, **11**, 81–85.

Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR (2016) Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biology*, **14**, e2000225.

Budding AE, Grasman ME, Eck A *et al.* (2014) Rectal Swabs for Analysis of the Intestinal Microbiota. *PLOS ONE*, **9**, e101344.

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker gene data analysis. *bioRxiv*. doi: 10.1101/113597.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Charruau P, Pérez-Flores J, Pérez-Juárez JG, Cedeão-Vázquez JR, Rosas-Carmona R (2012) Oral and cloacal microflora of wild crocodiles *Crocodylus acutus* and *C. moreletii* in the Mexican Caribbean. *Diseases of Aquatic Organisms*, **98**, 27–39.

Cooper JE, Needham JR, Lawrence K (1985) Studies on the cloacal flora of three species of free-living British reptile. *Journal of Zoology*, **207**, 521–525.

D'Aloia MA, Bailey TA, Samour JH, Naldo J, Howlett JC (1996) Bacterial flora of captive houbara (Chlamydotis undulata), kori (Ardeotis kori) and rufous-crested (Eupodotis ruficrista) bustards. *Avian Pathology*, **25**, 459–468.

446 Danzeisen JL, Clayton JB, Huang H *et al.* (2015) Temporal Relationships Exist Between Cecum,
447   Ileum, and Litter Bacterial Microbiomes in a Commercial Turkey Flock, and Subtherapeutic
448   Penicillin Treatment Impacts Ileum Bacterial Community Establishment. *Frontiers in Veterinary*
449   *Science*, **2**, 56.
450 Debelius J, Song SJ, Vazquez-Baeza Y *et al.* (2016) Tiny microbes, enormous impacts: what matters
451   in gut microbiome studies? *Genome Biology*, **17**, 217.
452 DeSantis TZ, Hugenholtz P, Larsen N *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene
453   database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**,
454   5069–5072.
455 Dewar ML, Arnould JPY, Dann P *et al.* (2013) Interspecific variations in the gastrointestinal
456   microbiota in penguins. *MicrobiologyOpen*, **2**, 195–204.
457 Dewar ML, Arnould JPY, Krause L, Dann P, Smith SC (2014) Interspecific variations in the faecal
458   microbiota of *Procellariiform* seabirds. *FEMS Microbiology Ecology*, **89**, 47–55.
459 Dickinson VM, Duck T, Schwalbe CR, Jarchow JL, Trueblood MH (2001) Nasal and cloacal bacteria
460   in free-ranging desert tortoises from the western United States. *Journal of Wildlife Diseases*, **37**,
461   252–257.
462 Diedenhofen B, Musch J (2015) Cocor: A comprehensive solution for the statistical comparison of
463   correlations. *PLOS ONE*, **10**, 1–12.
464 van Dongen WF, White J, Brandl HB *et al.* (2013) Age-related differences in the cloacal microbiota
465   of a wild bird species. *BMC Ecology*, **13**, 11.
466 Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple
467   tools and samples in a single report. *Bioinformatics*, **32**, 3047–3048.
468 Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB (2012) Animal Behavior and the
469   Microbiome. *Science*, **338**, 198–199.
470 Flores GE, Henley JB, Fierer N (2012) A Direct PCR Approach to Accelerate Analyses of Human-
471   Associated Microbial Communities. *PLOS ONE*, **7**, 1–11.
472 Ganz HH, Doroud L, Firl AJ *et al.* (2017) Community-Level Differences in the Microbiome of
473   Healthy Wild Mallards and Those Infected by Influenza A Viruses. *mSystems*, **2**, e00188-16.
474 Herlemann DPR, Labrenz M, Jürgens K *et al.* (2011) Transitions in bacterial communities along the
475   2000 km salinity gradient of the Baltic Sea. *The ISME Journal*, **5**, 1571–1579.
476 Hoar BM, Whiteside DP, Ward L, Inglis G, Morck DW (2007) Evaluation of the enteric microflora of
477   captive whooping cranes (*Grus americana*) and sandhill cranes (*Grus canadensis*). *Zoo Biology*,
478   **26**, 141–153.
479 Klomp JE, Murphy MT, Smith SB *et al.* (2008) Cloacal microbial communities of female spotted
480   towhees *Pipilo maculatus*: Microgeographic variation and individual sources of variability.
481   *Journal of Avian Biology*, **39**, 530–538.
482 Koch H, Schmid-Hempel P (2011) Socially transmitted gut microbiota protect bumble bees against an
483   intestinal parasite. *Proceedings of the National Academy of Sciences of the United States of*
484   *America*, **108**, 19288–92.
485 Kohl KD, Brun A, Magallanes M *et al.* (2017) Gut microbial ecology of lizards: insights into diversity
486   in the wild, effects of captivity, variation across gut regions and transmission. *Molecular*
487   *Ecology*, **26**, 1175–1189.

488   Kreisinger J, Čížková D, Kropáčková L, Albrecht T (2015) Cloacal Microbiome Structure in a Long-
489   Distance Migratory Bird Assessed Using Deep 16sRNA Pyrosequencing. *PLOS ONE*, **10**,
490   e0137401.
491   Kulkarni S, Heeb P (2007) Social and sexual behaviours aid transmission of bacteria in birds.
492   *Behavioural Processes*, **74**, 88–92.
493   Lamberski N, Hull AC, Fish AM, Beckmen K, Morishita TY (2003) A survey of the choanal and
494   cloacal aerobic bacterial flora in free-living and captive red-tailed hawks (*Buteo jamaicensis*)
495   and Cooper's hawks (*Accipiter cooperii*). *Journal of Avian Medicine and Surgery*, **17**, 131–135.
496   Lobato E, Geraldes M, Melo M, Doutrelant C, Covas R (2017) Diversity and composition of
497   cultivable gut bacteria in an endemic island bird and its mainland sister species. *Symbiosis*, **71**,
498   155–164.
499   Lombardo MP, Thorpe PA, Cichewicz R *et al.* (1996) Communities of Cloacal Bacteria in Tree
500   Swallow Families. *The Condor*, **98**, 167–172.
501   Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-
502   seq data with DESeq2. *Genome Biology*, **15**, 550.
503   Lucas FS, Heeb P (2005) Environmental factors shape cloacal bacterial assemblages in great tit *Parus*
504   *major* and blue tit *P. caeruleus* nestlings. *Journal of Avian Biology*, **36**, 510–516.
505   Martin MO, Gilman FR, Weiss SL (2010) Sex-specific asymmetry within the cloacal microbiota of
506   the striped plateau lizard, *Sceloporus virgatus*. *Symbiosis*, **51**, 97–105.
507   Matson KD, Versteegh MA, van der Velde M, Tieleman BI (2015) Effects of immune
508   supplementation and immune challenge on bacterial assemblages in the avian cloaca. *Journal of*
509   *Ornithology*, **156**, 805–810.
510   Maul JD, Gandhi JP, Farris JL (2005) Community-Level Physiological Profiles of Cloacal Microbes
511   in Songbirds (Order: Passeriformes): Variation Due to Host Species, Host Diet, and Habitat.
512   *Microbial Ecology*, **50**, 19–28.
513   McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and
514   Graphics of Microbiome Census Data. *PLOS ONE*, **8**, e61217.
515   McMurdie PJ, Holmes S (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is
516   Inadmissible. *PLOS Computational Biology*, **10**, e1003531.
517   Merkeviciene L, Ruzauskaite N, Klimiene I *et al.* (2017) Microbiome and antimicrobial resistance
518   genes in microbiota of cloacal samples from European herring gulls (*Larus argentatus*). *Journal*
519   *of Veterinary Research*, **61**, 27–35.
520   Mills TK, Lombardo MP, Thorpe PA (1999) Microbial colonization of the cloacae of nestling Tree
521   Swallows. *The Auk*, **116**, 947–956.
522   Moreno J, Briones V, Merino S *et al.* (2003) Beneficial Effects of Cloacal Bacteria on Growth and
523   Fledging Size in Nestling Pied Flycatchers (*Ficedula Hypoleuca*) in Spain. *The Auk*, **120**, 784–
524   790.
525   Muegge BD, Kuczynski J, Knights D *et al.* (2011) Diet Drives Convergence in Gut Microbiome
526   Functions Across Mammalian Phylogeny and Within Humans. *Science*, **332**, 970–974.
527   Oksanen J, Blanchet FG, Friendly M *et al.* (2017) Vegan: Community Ecology Package. *R package*
528   *version 2.4-2*. https://cran.r-project.org/package=vegan
529   R Core Team (2017) R: A language and environment for statistical computing. *R Foundation for*

530   *Statistical Computing, Vienna, Austria.*

531   Ruiz-Rodríguez M, Lucas FS, Heeb P, Soler JJ (2009a) Differences in intestinal microbiota between

532   avian brood parasites and their hosts. *Biological Journal of the Linnean Society*, **96**, 406–414.

533   Ruiz-Rodríguez M, Soler JJ, Lucas FS *et al.* (2009b) Bacterial diversity at the cloaca relates to an

534   immune response in magpie *Pica pica* and to body condition of great spotted cuckoo *Clamator*

535   *glandarius* nestlings. *Journal of Avian Biology*, **40**, 42–48.

536   Santoro M, Hernández G, Caballero M (2006) Aerobic bacterial flora of nesting green turtles

537   (*Chelonia mydas*) from Tortuguero National Park, Costa Rica. *Journal of Zoo and Wildlife*

538   *Medicine*, **37**, 549–552.

539   Santos SS, Pardal S, Proença DN *et al.* (2012) Diversity of cloacal microbial community in migratory

540   shorebirds that use the Tagus estuary as stopover habitat and their potential to harbor and

541   disperse pathogenic microorganisms. *FEMS Microbiology Ecology*, **82**, 63–74.

542   Schaffner T, Mueller J, Hess MW *et al.* (1974) The Bursa of Fabricius: A Central Organ Providing for

543   Contact between the Lymphoid System and Intestinal Content. *Cellular Immunology*, **13**, 304–

544   312.

545   Sekirov I, Russell S, Antunes L (2010) Gut microbiota in health and disease. *Physiological Reviews*,

546   **90**, 859–904.

547   Sharon G, Segal D, Ringo JM *et al.* (2010) Commensal bacteria play a role in mating prefence of

548   Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States*

549   *of America*, **107**, 20051–20056.

550   Smith P, Willemsen D, Popkes M *et al.* (2017) Regulation of Life Span by the Gut Microbiota in The

551   Short-Lived African Turquoise Killifish. *bioRxiv*. doi: 10.1101/120980.

552   Song SJ, Amir A, Metcalf JL *et al.* (2016) Preservation Methods Differ in Fecal Microbiome

553   Stability, Affecting Suitability for Field Studies. *mSystems*, **1**, e00021-16.

554   Stanley D, Geier MS, Chen H, Hughes RJ, Moore RJ (2015) Comparison of fecal and cecal

555   microbiotas reveals qualitative similarities but quantitative differences. *BMC Microbiology*, **15**,

556   51.

557   Stenkat J, Krautwald-Junghanns ME, Schmitz Ornés A, Eilers A, Schmidt V (2014) Aerobic cloacal

558   and pharyngeal bacterial flora in six species of free-living birds. *Journal of Applied*

559   *Microbiology*, **117**, 1564–1571.

560   Warner NL, Szenberg A (1964) The Immunological Function of the Bursa of Fabricius in the

561   Chicken. *Annual Review of Microbiology*, **18**, 253–266.

562   De Weirdt R, Van de Wiele T (2015) Micromanagement in the gut: microenvironmental factors

563   govern colon mucosal biofilm structure and functionality. *npj Biofilms and Microbiomes*, **1**,

564   15026.

565   White J, Mirleau P, Danchin E *et al.* (2010) Sexually transmitted bacteria affect female cloacal

566   assemblages in a wild bird. *Ecology Letters*, **13**, 1515–1524.

567   Wickham H (2009) *ggplot2: elegant graphics for data analysis.* New York: Springer.

568   Xenoulis PG, Gray PL, Brightsmith D *et al.* (2010) Molecular characterization of the cloacal

569   microbiota of wild and captive parrots. *Veterinary Microbiology*, **146**, 320–325.

570   Zhang Y, Simon SE, Johnson JA, Allen MS (2017) Spatial Microbial Composition Along the

571   Gastrointestinal Tract of Captive Attwater's Prairie Chicken. *Microbial Ecology*, **73**, 966–977.

17

572 # **Supporting information**

573

574

575   Supporting information has been made available online.

576

577   **Appendix S1**. Supplementary Methods, Supplementary Figures S1–S4, Supplementary Tables S1–

578   S3, and Supplementary References.

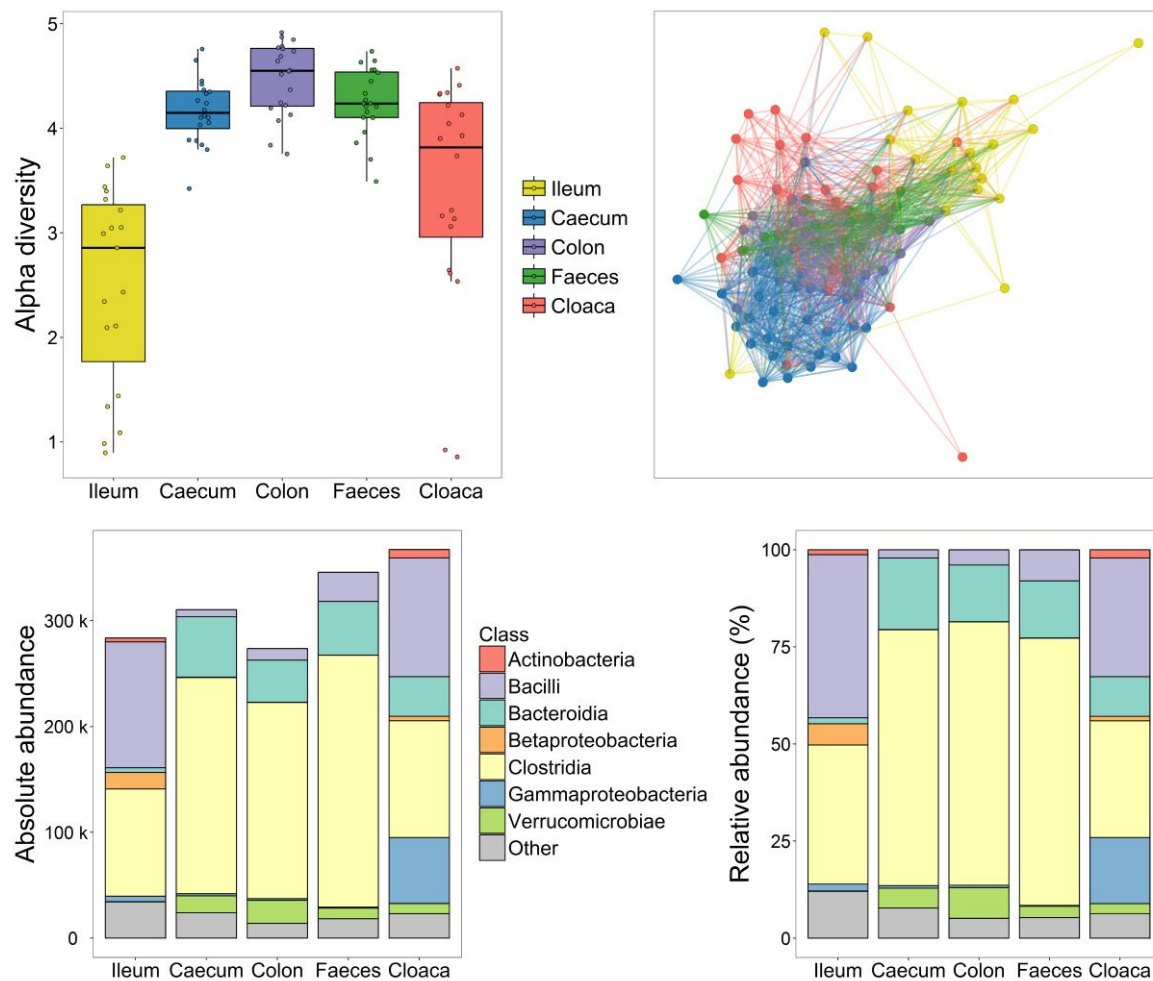579   **Appendix S2**. Supplementary Tables S4–S9.

Figures

Figure 1. Overall microbiota similarities and differences between sample types. (A) Boxplot of Shannon alpha diversity within sample types. (B) Network of Bray-Curtis distances between samples where colours indicate sample type and lines are drawn to nearest neighbours with a maximum distance of 0.85. The barplots show the (C) absolute and (D) relative abundance of all OTUs for each sample type. Taxonomic classifications are coloured at the class level.
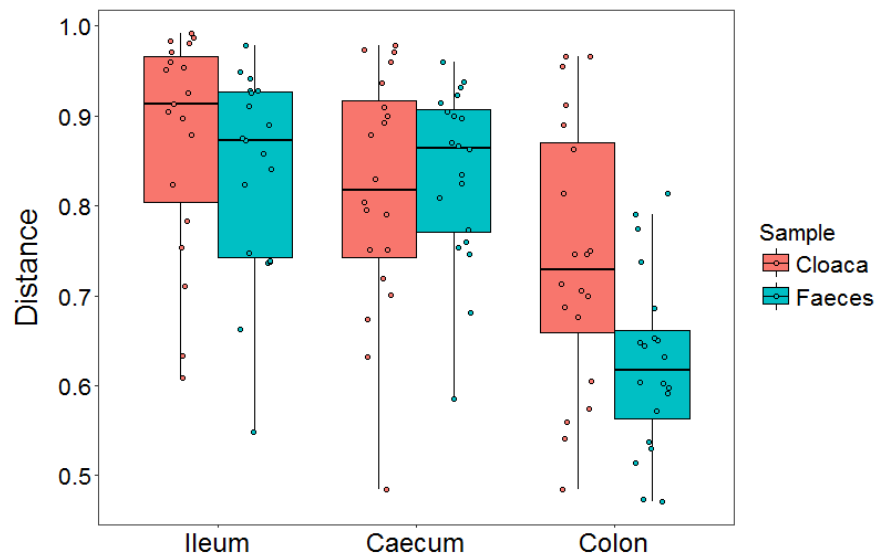
589

590

591 **Figure 2.** Bray-Curtis distances between the microbiota in cloacal and faecal samples compared to the

592 microbiota in three parts of the gut (ileum, caecum, colon) within the same host individual. Higher

593 distance measures indicate higher dissimilarity, where 1 = completely dissimilar bacterial community.
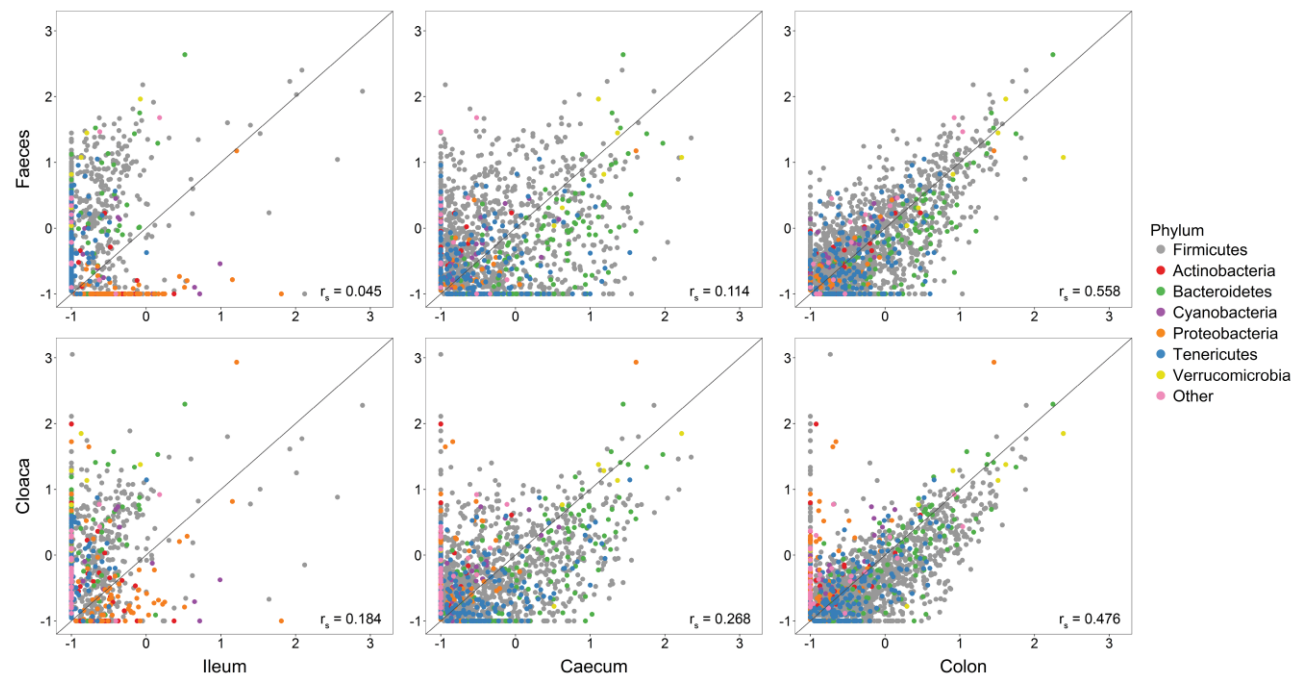
594

595

596

Figure 3. Correlations in OTU abundances between each of two sampling methods (cloaca and faeces) compared to three sections of the gut (ileum, caecum, and colon). OTU abundances were normalised according to the DESeq2 method and log-transformed (+0.1) for graphical purposes. The straight line shows the 1:1 relationship between the samples, and the r_s values display the Spearman rank correlation across all OTUs.
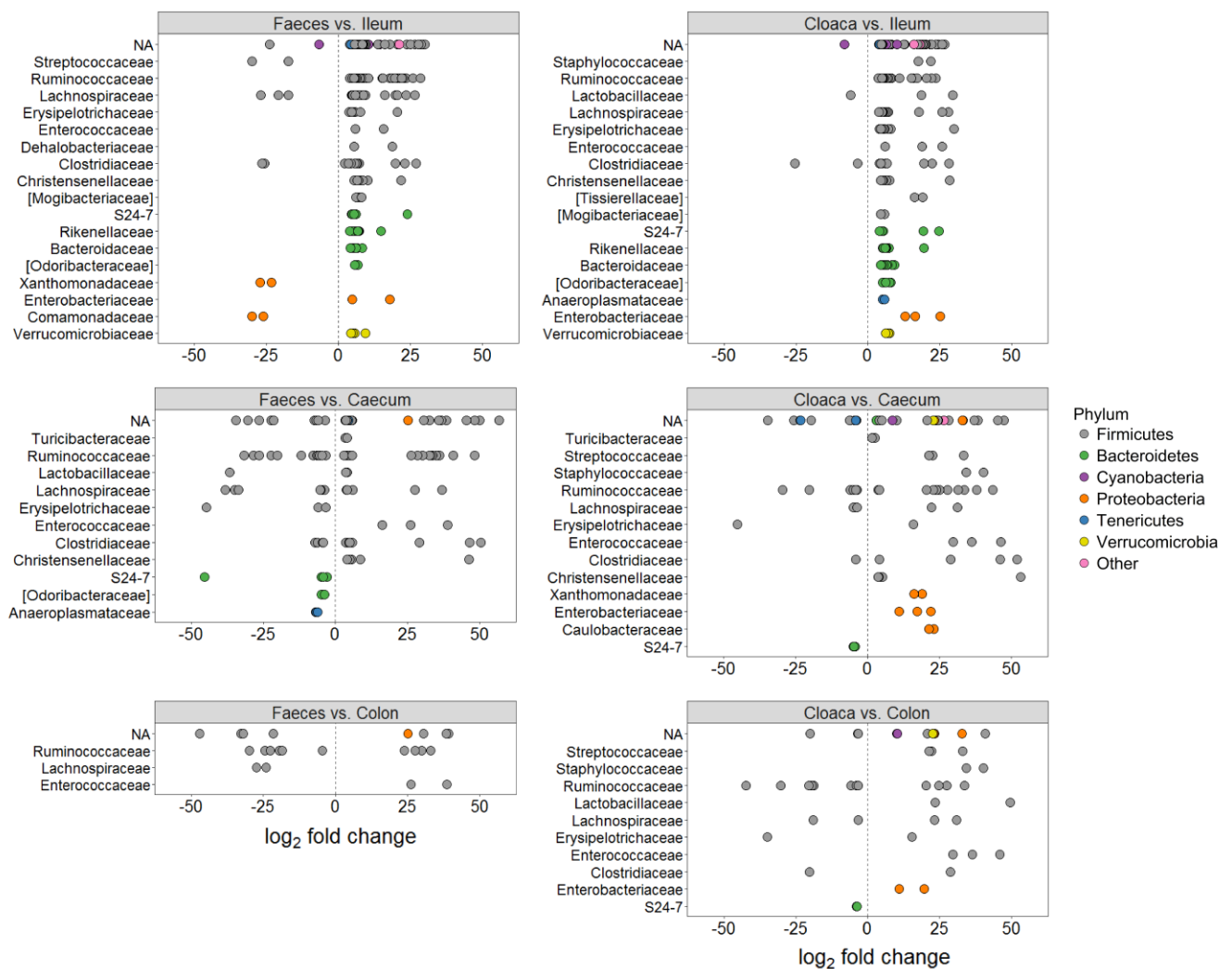
602

603

604 **Figure 4.** OTUs that were significantly different in abundance in the sampling methods (faeces and

605 cloaca) relative to three parts of the gut (ileum, caecum, and faeces). The y-axes show taxonomic

606 families and all OTUs have been coloured within their respective phylum. Positive $\log_2$ fold changes

607 signify increased OTU abundance in either faeces (left column) or cloacal swabs (right column), and

608 negative $\log_2$ fold changes display increased abundance in one of the gut sections (ileum, caecum, or

609 colon). Families with only a single significant OTU are not shown due to space limitations; complete

610 dataset can be found at Tables S4-S9. Family names in brackets are proposed taxonomies by

611 Greengenes.