1 **Taxon appearance from extraction and amplification steps demonstrates the value**

2 **of multiple controls in tick microbiota analysis**

3 Emilie Lejal[1], Agustín Estrada-Peña[2], Maud Marsot[3], Jean-François Cosson[1], Olivier Rué[4], Mahendra

4 Mariadassou[4], Cédric Midoux[4,5], Muriel Vayssier-Taussat[6] and Thomas Pollet[1]*

5

6 **[1]**UMR BIPAR, Animal Health Laboratory, INRA, ANSES, Ecole Nationale Vétérinaire d'Alfort, Université

7 Paris-Est, Maisons-Alfort, France

8 **[2]**Faculty of Veterinary Medicine, University of Zaragoza, Zaragoza, Spain

9 **[3]**ANSES, University Paris-Est, Laboratory for Animal Health, Epidemiology Unit, Maisons-Alfort, France

10 [4]MaIAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France

11 [5]Irstea, UR PROSE, Antony, France

12 [6]INRA, Animal health department, Nouzilly, France

13 *** Correspondence:**

14 Dr Thomas Pollet

15 thomas.pollet@inra.fr

Abstract

Background: The development of high throughput sequencing technologies has substantially improved analysis of bacterial community diversity, composition, and functions. Over the last decade, high throughput sequencing has been used extensively to identify the diversity and composition of tick microbial communities. However, a growing number of studies are warning about the impact of contamination brought along the different steps of the analytical process, from DNA extraction to amplification. In low biomass samples, e.g. individual tick samples, these contaminants may represent a large part of the obtained sequences, and thus generate considerable errors in downstream analyses and in the interpretation of results. Most studies of tick microbiota either do not mention the inclusion of controls during the DNA extraction or amplification steps, or consider the lack of an electrophoresis signal as an absence of contamination. In this context, we aimed to assess the proportion of contaminant sequences resulting from these steps. We analyzed the microbiota of individual *Ixodes ricinus* ticks by including several categories of controls throughout the analytical process: crushing, DNA extraction, and DNA amplification.

Results: Controls yielded a significant number of sequences (1,126 to 13,198 mean sequences, depending on the control category). Some operational taxonomic units (OTUs) detected in these controls belong to genera reported in previous tick microbiota studies. In this study, these OTUs accounted for 50.9% of the total number of sequences in our samples, and were considered contaminants. Contamination levels (i.e. the percentage of sequences belonging to OTUs identified as contaminants) varied with tick stage and gender: 76.3% of nymphs and 75% of males demonstrated contamination over 50%, while most females (65.7%) had rates lower than 20%. Contamination mainly corresponded to OTUs detected in crushing and DNA extraction controls, highlighting the importance of carefully controlling these steps.

Conclusion: Here, we showed that contaminant OTUs from extraction and amplification steps can represent more than half the total sequence yield in sequencing runs, and lead to unreliable results when characterizing tick microbial communities. We thus strongly advise the routine use of negative controls in tick microbiota studies, and more generally in studies involving low biomass samples.

Keywords: Contaminant, Low biomass sample, Tick, Microbiota, High-throughput sequencing.

Introduction

Analysis of microbial community composition by 16S tag sequencing represents a powerful tool commonly used to evaluate the prokaryotic diversity in many ecosystems. Thanks to high-throughput sequencing approaches, it is now feasible to identify microbiota with high coverage and reveal the exceptional and complex microbial diversity of environments ranging from the wide ocean (Sogin *et al.*, 2006; Hamdan *et al.*, 2013) to small arthropods (Degli Esposti and Martinez Romero, 2017; Estrada-Peña *et al.*, 2018). With the constant improvement of sensitivity associated with its decreasing cost, 16S tag sequencing has spread to many areas, including those dealing with low biomass communities such as tick microbiota.

Despite the limitation induced by low biomass, a high number of tick microbiota studies are focusing on individual ticks (Lalzar *et al.*, 2012; Hawlena *et al.*, 2013; Rynkiewicz *et al.*, 2015; Van Treuren *et al.*, 2015; Gall *et al.*, 2016; Khoo *et al.*, 2016; Kwan *et al.*, 2017; Swei and Kwan, 2017; Estrada-Peña *et al.*, 2018) or even on individual or pooled tick organs (Andreotti *et al.*, 2011; Budachetri *et al.*, 2014; Narasimhan *et al.*, 2014; Qiu *et al.*, 2014; Clayton *et al.*, 2015; Gall *et al.*, 2016; Zolnik *et al.*, 2016; Abraham *et al.*, 2017). The results obtained from these studies have improved our knowledge of tick microbiota, while demonstrating links between its composition and several factors, such as tick stages and gender (Lalzar *et al.*, 2012; Van Treuren *et al.*, 2015; Zolnik *et al.*, 2016; Swei and Kwan, 2017), organ of origin (Zolnik *et al.*, 2016), living environment (Carpi *et al.*, 2011; Zolnik *et al.*, 2016; Estrada-Peña *et al.*, 2018), host blood meal (Swei and Kwan, 2017), or tick engorgement (Moreno *et al.*, 2006; Heise *et al.*, 2010; Narasimhan *et al.*, 2014). The presence of tick-borne pathogens has also been demonstrated to be related to differences in tick microbiota composition and diversity (Narasimhan *et al.*, 2014; Gall *et al.*, 2016; Abraham *et al.*, 2017). The continuous improvements in our knowledge, and in particular the assessment of potential links between tick-borne pathogens and particular microbiota members, require us to continue performing analyses on individual ticks, or even better, on individual tick organs.

Several potential biases, inherent to the PCR amplification (chimeric sequences) or sequencing steps (random sequencing errors), are now well known (Galan *et al.*, 2016), and pipelines have been developed to detect and remove these errors during the sequence analysis. However, several studies have begun to warn scientists about the potential distortion of results due to contamination from the

3

73   extraction/amplification steps or from the environment during sample processing (Salter *et al.*, 2014;

74   Weiss *et al.*, 2014; Narasimhan and Fikrig, 2015; Galan *et al.*, 2016; Glassing *et al.*, 2016; Knight *et al.*,

75   2018). This potential contamination is able to significantly modify results concerning microbe

76   diversity and composition, particularly in the case of low biomass samples, and can have serious

77   consequences on the interpretation of results. For example, while investigating 16s rRNA gene

78   profiles from nasopharyngeal swabs regularly sampled from a cohort of children from 1 to 24 months,

79   Salter *et al*. (2014) demonstrated that the apparent clustering of samples by age, observed by

80   principal coordinate analysis, in fact corresponded to the commercial kit used to perform the DNA

81   extraction of these samples. While studying the influence of laboratory processing on diluted

82   samples, these authors also observed that the lower the initial biomass of samples, the higher the

83   proportion of contaminants associated with the kits or the environment. In the context of tick

84   microbiota studies, performing using negative controls would therefore be essential to detect

85   potential contamination, and efficiently distinguish this contamination from genuine tick microbial

86   communities. On the basis of an exhaustive review of the scientific literature, it is very clear that

87   routine multiple control analyses (extraction and PCR negative controls) are not considered at their

88   true value, and are only rarely implemented in tick microbial community analyses (Hawlena *et al.*,

89   2013; Gofton *et al.*, 2015; Rynkiewicz *et al.*, 2015; René-Martellet *et al.*, 2017; Sperling *et al.*, 2017;

90   Estrada-Peña *et al.*, 2018; Thapa *et al.*, 2019). In this context, we performed a 16s rRNA gene

91   sequencing analysis on individual ticks (females, males and nymphs) considering concomitantly

92   several negative controls at each step of the process, prior to the sequencing run (tick crushing, DNA

93   extraction and amplification).

Methods

Tick collection and sample selection

Questing *Ixodes ricinus* nymphs and adults were collected for three years by dragging (from April 2014 to May 2017) in the Sénart forest in the south of Paris. More details on the sampling location and design, and tick collection, are available in Lejal *et al*. (2019). After morphological identification, ticks were stored at -80°C until further analysis.

Tick washing, crushing, and DNA extraction

Ticks were first washed once in ethanol 70% for 5 minutes and rinsed twice in sterile MilliQ water for 5 minutes each time. Ticks were then individually crushed in 375 µL of Dulbecco's Modified Eagle Medium with decomplemented fetal calf serum (10%) and six steel beads using the homogenizer Precellys®24 Dual (Bertin, France) at 5,500 rpm for 20 seconds. DNA extraction was performed on 100 µL of tick crushing, using the NucleoSpin® Tissue DNA extraction kit (Macherey-Nagel, Germany). For these previous steps, sterile tubes and filter tips were used on a dedicated area for DNA extraction. Bench and materials were decontaminated with DNA remover (Molecular BioProducts, San Diego, CA, USA) before and after each use. Each time a tick crushing or DNA extraction was performed, a crushing or extraction control was added for the corresponding step.

DNA amplification and multiplexing

In total, 32 males, 35 females, and 557 nymphs were subjected to sequencing analysis. DNA amplifications were performed on the V4 region of the 16s rRNA gene using the primer pair used by Galan *et al*. (2016) (16S-V4F: 5'-GTGCCAGCMGCCGCGGTAA-3' and 16S-V4R: 5'-GGACTACHVGGGTWTCTAATCC-3'), producing a 251 bp amplicon. This primer pair was chosen for its ability to specifically amplify prokaryotic DNA (86% or 95% of referenced bacterial DNA in the Silva database, and 52.3% or 93% of Archaeal DNA, when considering 0 or 1 mismatch, respectively). This characteristic is crucial in the context of tick microbiota analysis, due to the high amount of tick DNA

120   compared to prokaryotic DNA. An 8 bp-index was added to the forward and reverse primers. The

121   individual amplification of each sample was performed using a combination of 22 and 32 index-tagged

122   forward and reverse primers, allowing for amplification and multiplexing of a maximum of 704

123   samples that can be loaded together on an Illumina MiSeq flow cell. Amplification and multiplexing

124   steps were performed under a sterile hood, decontaminated before and after use with DNA remover

125   and UVs, using tube strips with individual caps to prevent cross-contamination. All the PCR

126   amplifications were carried out using the Phusion® High-Fidelity DNA Polymerase amplification kit

127   (Thermo Scientific, Lithuania). For each sample, 5 µL of DNA extract were amplified in a 50 µL final

128   reaction volume, containing 1X Phusion HF buffer, 0.2 µM of dNTPs, 0.2 U/mL of Phusion DNA

129   polymerase, and 0.35 µM of forward and reverse primer. The following thermal cycling procedure

130   was used: initial denaturation at 98°C for 30 s, 35 cycles of denaturation at 98°C for 10 s, annealing

131   at 55°C for 30 s, followed by extension at 72°C for 30 s. The final extension was carried out at 72°C

132   for 10 minutes. Amplification step was performed on a peqSTAR 2x thermocycler (PEQLAB, Germany).

133   For each PCR run, negative controls were performed by using the reaction mixture without template,

134   but using ultrapure water (Invitrogen, United Kingdom). PCR products were checked on 1.5% agarose

135   gels.

136

137   **DNA purification, quantification, and pooling**

138   Amplicons were cleaned and quantified using a SequalPrep normalization plate kit (Invitrogen,

139   USA), and a Qubit dsDNA High Sensitivity Assay kit (Invitrogen, USA), respectively. Amplicons and

140   controls were pooled at equimolar concentrations and sent to the sequencing platform (GenoScreen,

141   France).

142

143   **Controls**

144   In total, 45 negative controls were included and analyzed using the same process as for tick

145   samples (**Figure 1**).

146   **12 Crushing controls (CCs):** Tube containing crushing solution, remaining open in the room while

147   putting ticks inside other tubes between the cleaning and crushing steps. DNA extraction and

148   amplification were then performed on the crushing control in the same conditions as for any other

149   sample.

150   **12 DNA extraction controls (ECs):** For these controls, the different DNA extraction steps were

151   performed under the same conditions as for any other samples, but without adding sample. DNA

152   amplification was then performed on the extraction control in the same conditions as for any other

153   sample.

154   **21 DNA amplification controls (ACs):** The DNA amplification step was performed on ultrapure water

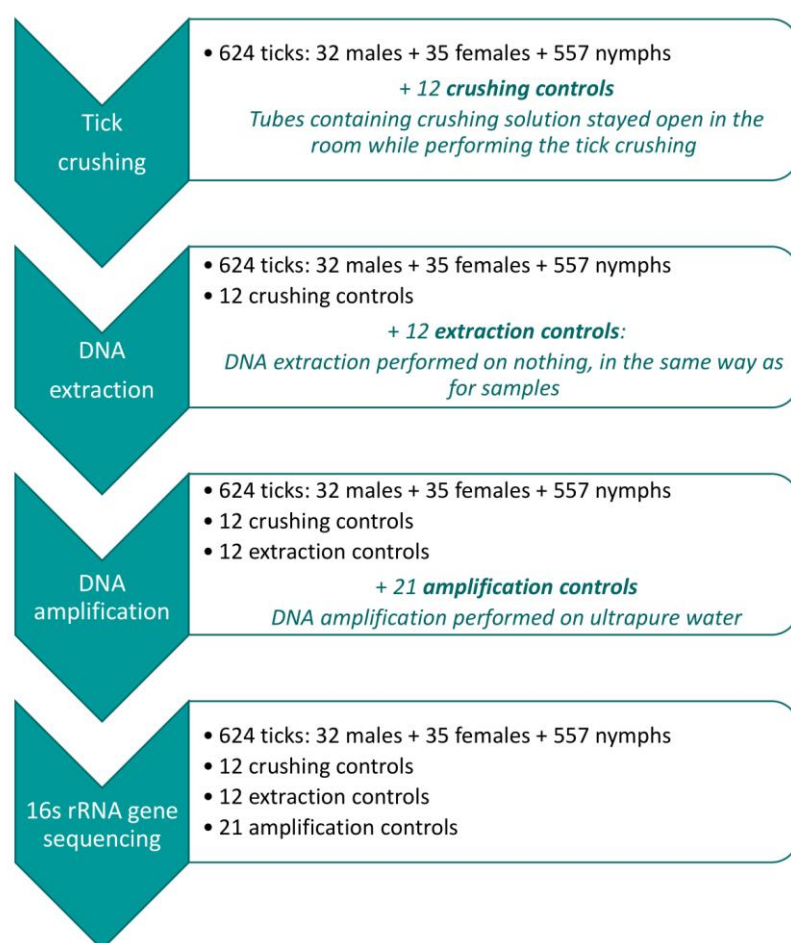155   (Invitrogen, United Kingdom), in the same conditions as for any other sample.

156

157   **Figure 1. Analyzing process for samples and added controls.**

7

158 **Sequencing and data processing**

159   The equimolar mix was concentrated and sequenced by GenoScreen (Lille, France) using
160 MiSeq Illumina 2 x 250 bp chemistry. Quality control checks on raw data were performed using FastQC
161 v0.11.3 (Andrews, 2010), and no problems were detected. First, pairs of amplicons were merged using
162 Pear v0.9.11 (Zhang *et al.*, 2014) with a minimum overlap of 20 bp, a minimum assembly length of
163 100 bp, a maximum assembly length of 480 bp, and an e-value of 0.01. Then, the sequencing
164 orientation of extended sequences was checked by using cutadapt v1.12 (Martin, 2011), with primers
165 in forward and reverse strands (without removing them). Sequences with a reversed orientation were
166 reverse-complemented to obtain a set of sequences with the same orientation (5' to 3'). The
167 sequence analysis was performed using FROGS (Escudié *et al.*, 2018). Primers were removed using
168 cutadapt and resulting sequences were dereplicated. Operational taxonomic units (OTUs) were built
169 using Swarm v2.1.12 (Mahé *et al.*, 2015) with parameter d = 3, and chimeras were filtered using
170 vsearch v1.4 (Rognes *et al.*, 2016) in a de novo mode. Finally, OTUs with low abundance (< 0.005% of
171 the total sequence count) were filtered out, resulting in 608 OTUs (corresponding to 2,835,506
172 sequences). OTUs were affiliated by blasting cluster seed sequences against the Silva database v132
173 (Quast *et al.*, 2013). OTUs corresponding to mitochondrial/chloroplast DNA or not affiliated to any
174 domain were removed, leading to a dataset composed of 513 OTUs and 2,726,287 sequences.

175

176 **Data processing to distinguish contaminant versus tick microbiota OTUs**

177   The control composition was used to distinguish tick microbial OTUs from contaminants. All
178 the OTUs corresponding to more than 1% of the total number of sequences obtained in each category
179 of control were identified as main contaminants. However, due to possible errors in sequence
180 assignment occurring during sequencing due to the formation of mixed clusters on the flow cell
181 (Kircher *et al.*, 2012), negative controls can also contain a few sequences of tick microbial OTUs,
182 complicating contaminant identification under this 1% threshold. Therefore, for each OTU, the
183 number of sequences expected to spread to other samples due to false assignment (i.e. expected
184 number) was determined by calculating 0.02% of the total number of each OTU in the dataset,
185 corresponding to a false assignment rate estimated previously by Galan *et al*. (2016). This expected
186 number was compared to the maximum number of obtained sequences in controls (i.e. maximum

8

187   number). As this rate was not estimated from our dataset, we decided to allow variability around this

188   threshold, while calculating a 99% confidence interval around the expected number of 0.02%. OTUs

189   whose maximum sequence number was above the upper threshold were considered contaminants.

190   Importantly, cross-contamination from samples to controls was considered negligible, as no negative

191   controls presented a higher maximum sequence number than the expected number for an OTU

192   belonging to the most abundant and characteristic genus of tick microbiota in our dataset

193   (Midichloria), which is also the 3rd most abundant OTU in the dataset (266,516 sequences).

194

Statistical analyses

196   The percentage of samples presenting a contamination rate higher than 50% was compared

197   between samples categories (nymphs, males and females) according to a Chi^2 statistical test.

198   Alpha diversity was estimated before and after sample cleaning. Samples containing fewer

199   than 500 sequences after cleaning were removed from the analysis, even if they contained more than

200   500 sequences before cleaning. Several measures of alpha diversity were calculated. Basically, the

201   observed number of OTUs (species richness) was used as well as the Faith Phylogenetic Diversity (PD)

202   index, corresponding to the total branch length of the subtree spanned by the community, and also

203   two indexes taking into account the number of sequences: the Shannon index (more influenced by

204   rare OTUs) and Simpson index (more influenced by dominant OTUs). Phylogenetic diversity was

205   estimated using the Picante package (Kembel *et al.*, 2010), while other measures were assessed

206   thanks to Easy 16S (http://genome.jouy.inra.fr/shiny/easy16S/), based on the phyloseq package

207   (McMurdie and Holmes, 2013). Alpha diversity measure comparison was performed between cleaned

208   and not cleaned groups, as well as between stages, using non-parametric statistical tests (Wilcoxon

209   and Kruskal–Wallis, respectively). Statistical computations were performed in R 3.6.0 (R Core Team,

210   2018).

211

212 **Results**

213   Nymphs, males and females presented a mean number of sequences per sample of 3,721,

214   2,689 and 5,316, respectively. Concerning controls, a mean number of 13,198, 11,235 and 1,126

215   sequences per sample was obtained for crushing, extraction and amplification controls (CCs, ECs and

216   ACs), respectively.

217   The most abundant OTUs, with relative abundance higher than 1% in at least one category of

218   samples or controls, are represented in **Figure 2**. All other OTUs, i.e. those with relative abundance

219   lower than 1% in all categories were grouped together in the "Others" category.

220   The examination of the at-least-1% OTU composition of the control (CCs, ECs, ACs) allowed to

221   identify OTUs belonging to the genera: *Pseudomonas*, *Acinetobacter*, *Halomonas*, *Shewanella*,

222   *Tepidimonas*, *Esherichia-Shigella*, *Photobacterium*, *Polaromonas*, *Deinococcus* and *Reyranella*, as well

223   as two OTUs belonging to the Enterobacteriaceae family but multi-affiliated at the genus level, and

224   one OTU only affiliated to Gammaproteobacteria at the Class level (**Figure 2**). OTU composition was

225   contrasted across the control types. While the same OTUs were identified in CCs and ECs and

226   belonged to *Pseudomonas* [70.7% and 69.3%, respectively], multi-affiliated Gammaproteobacteria

227   [12.2% and 16.7%, respectively], *Acinetobacter* [3.1% and 2.4%, respectively], and multi-affiliated

228   Enterobacteriaceae [2.7% and 3.6%, respectively]), the microbial community identified in AC controls

229   was dominated by OTUs belonging to *Halomonas* (58.8%), *Shewanella* (13.6%), *Tepidimonas* (5.1%),

230   *Escherichia-Shigella* (2.3%), *Photobacterium* (1.5%), *Polaromonas* (1.5%), *Deinococcus* (1.5%),

231   *Reyranella* (1.2%) and multi-affiliated Enterobacteriaceae (1.2%). These OTUs represented 38.9% of

232   the total number of sequences coming from the tick samples. Focusing on the sample category, these

233   OTUs correspond to 11.9%, 40.4% and 41.3% of sequences detected in females, males and nymphs,

234   respectively (**Figure 2**). These OTUs were therefore considered to be the main contaminants of the

235   dataset. It is interesting to note that most of these OTUs correspond to those detected in both EC and

236   CC controls. In fact, only 2.8% of sequences detected in the different categories of samples (nymphs,

237   females and males) correspond to OTUs detected in the AC controls and identified as main
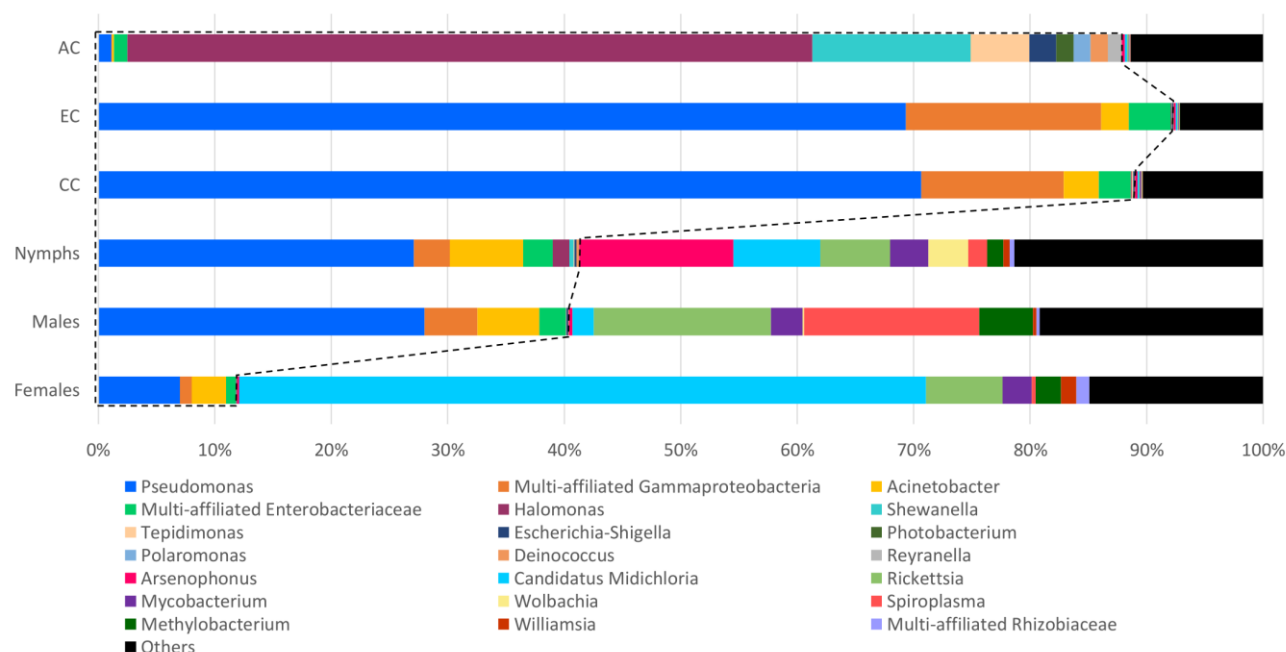
238   contaminants.

10

239

Figure 2. Comparison of the 16S rRNA gene sequencing taxonomic assignments between negative controls and samples. Percentages are represented for the OTUs with proportions corresponding to at least 1% of the total number of sequences assigned to a category of samples or controls. Those corresponding to less than 1% are assigned as "Others". OTUs belonging to the same genus or family (if multi-affiliated at the scale of genus) were assigned under the same name. One OTU was only identified until the class level (multi-affiliated Gammaproteobacteria).

A more in-depth identification of contaminants, including those corresponding to less than 1% of the total number of sequences was performed, allowing us to identify 160 OTUs as contaminants among the 513 initially identified OTUs in the dataset. This identification increased the proportions of identified contaminants in the samples, which reach 50.9% of the total sequence count from the tick samples, and 20.1%, 51.2% and 53.7% of sequences detected in females, males and nymphs, respectively. Furthermore, after these cleaning steps, 201 samples contained less than 500 sequences compared to 8 before cleaning. For 76.6% of nymph and 75% of male samples, contaminant OTUs represented more than 50% of the total sequence count (**Figure 3**). These sample contamination rates are significantly higher than those obtained for females ($p < 0.05$ according to a Chi^2 test), for which most of the samples (65.7%) exhibited a contamination rate below 20%.
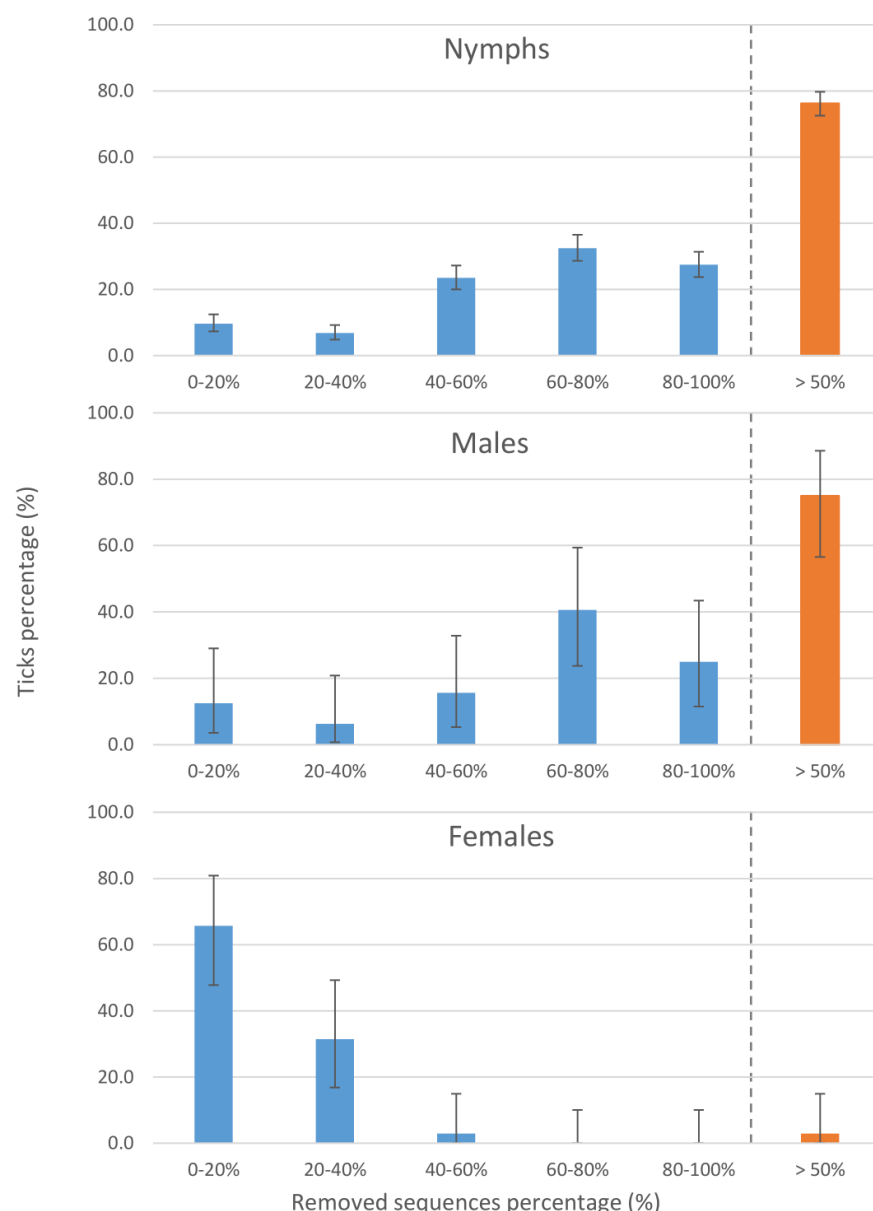
11

Figure 3. Impact of contamination on the final number of sequences per sample according to the tick stage/gender.

The impact of this contamination on diversity, according to the sample category was estimated (**Figure 4**). As expected, for all the studied diversity measures, cleaning led to a decrease in diversity. Concerning nymphs and females, this decrease was significant for all the tested diversity indexes, while for males, it was only significant for the observed and phylogenetic diversity measures, suggesting that for males, contamination concerned mostly low-abundance taxa. The impact of this cleaning on alpha diversity comparison between tick stages was been studied. Without cleaning,

12

265   significant differences in alpha diversity were found between tick stages for observed, Shannon and

266   Simpson measures. Concerning cleaned data, significant differences were still observed between

267   stages when using Shannon and Simpson indexes, but disappeared when using the observed diversity

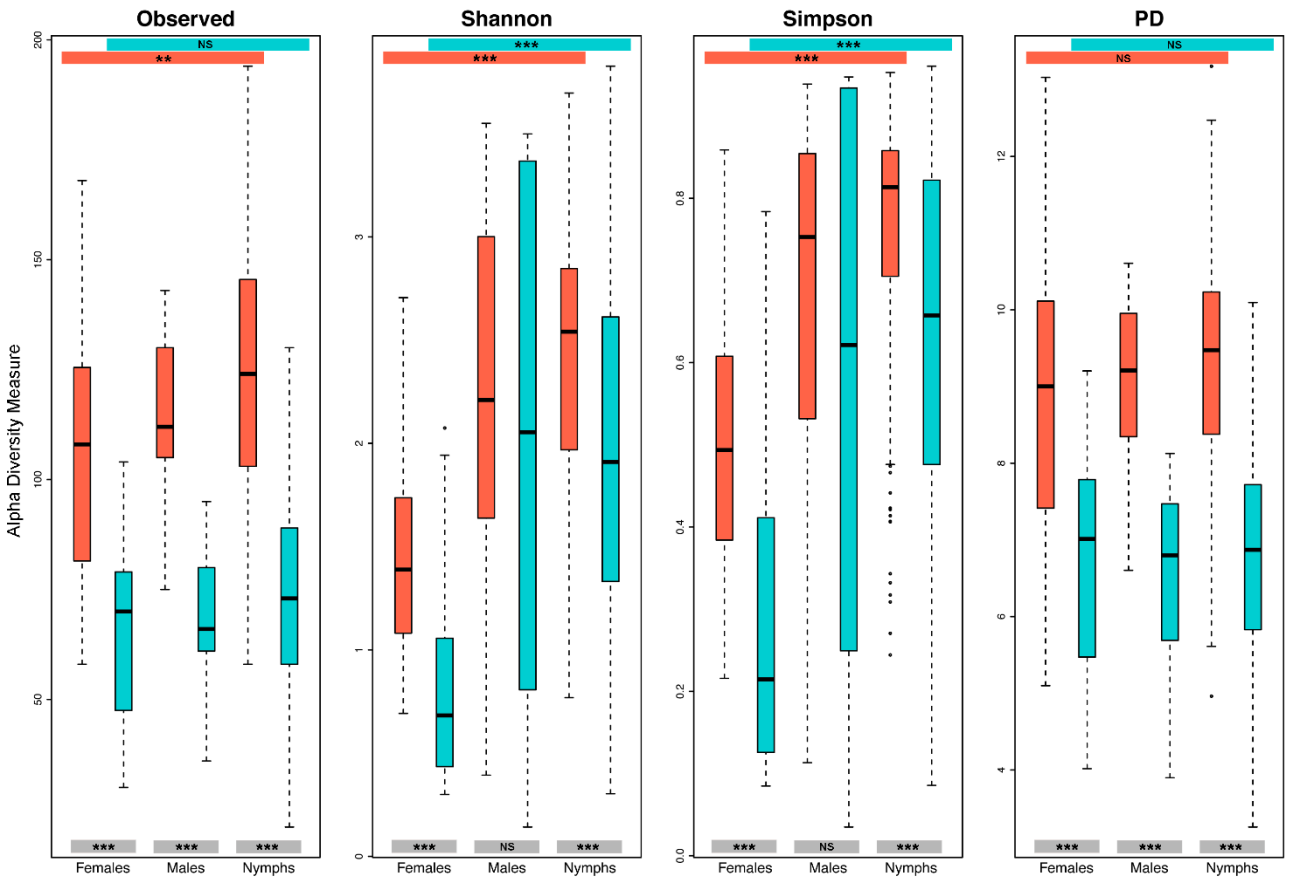268   measure.

269

270   **Figure 4. Impact of contamination on alpha diversity estimations according to the tick stages.** Alpha diversity

271   has been estimated according to the observed number of OTUs, the Shannon and Simpson indexes and the

272   Faith's phylogenetic diversity (PD), for each stage on cleaned (turquoise) and not cleaned (orange-red)

273   datasets. Results of the statistical test comparing alpha diversity measurement between stages, as well as

274   between the cleaned and uncleaned groups, are represented on coloured bands at both the top and the

275   bottom of the frame respectively: NS= Not Significant, ***= p < 0.001, **= p < 0.01.

13

276   **Discussion**

277   In this study, we sequenced the microbiota of 624 individual *I. ricinus* ticks including several

278   categories of negative controls to identify and quantify the contaminating sequences, in particular

279   those produced during the extraction or amplification steps. The performance of multiple negative

280   controls allowed us to perform in-depth identification of contaminant OTUs. These OTUs represented

281   50.9% of the total sequence yield from tick samples, and had a significant impact on microbial

282   diversity estimation. This shows the importance of contamination potentially arising from processing

283   steps, and demonstrates that its consequences on data analyses can be significant and should not be

284   neglected.

285   A high number of sequences in each category of controls was detected, particularly in CCs and

286   ECs, in which the mean number of sequences was even higher than the mean sequence number

287   detected in samples. Importantly, this higher number might be partly overestimated, first due to the

288   presence of tick DNA during the PCR that may have inhibited the reaction on the 16s rRNA gene in

289   samples while not in controls, and second during the measure of DNA concentrations that may have

290   led to a higher estimation of concentrations in samples compared to controls, before the pooling

291   step. Despite this, the high number of sequences in controls raises the issue of contaminant presence,

292   while performing each step of the sample processing.

293   Each processing step leads to a different contamination pattern. Most of the main

294   contaminants do not seem to originate from the DNA amplification step. Furthermore, as the

295   composition of crushing and extraction controls was the same, while the reagents used were different

296   and crushing control also underwent the extraction step, the contaminant sequences most probably

297   came from the DNA extraction step, rather than from the crushing step. The key role of the extraction

298   step in the contamination of low biomass samples has previously been demonstrated by Salter *et al*.

299   (2014). Furthermore, three of the four main contaminants identified in our analysis, corresponding

300   to *Pseudomonas*, *Acinetobacter*, and multi-affiliated Enterobacteriaceae, belong to families and

301   genera identified by these authors as contaminants coming from DNA extraction kits. Our

302   observations differed from those obtain by Galan *et al*. (2016), who mainly detected contaminants

303   coming from the DNA amplification step, while this step represented only 2% of the main

304   contaminants detected in our samples. In their study of kit contaminants, Salter *et al*. (2014) observed

14

305 that the extent of contamination arising from DNA extraction was different depending on the kit

306 brand used to perform this step. While using the less contaminating kit, they extracted DNA from

307 several 10-fold cascade dilutions of a pure suspension of *Salmonella bongori*, and subsequently

308 assessed 16s rRNA gene concentrations *via* q-PCR. They demonstrated a linear relationship between

309 the 16s rRNA gene copy number and the dilution fold only for the first dilution points, after which

310 levels started to stabilize and to reach a threshold of 10^5 mean copy number/mL. Consequently, the

311 higher rate of extraction, rather than amplification contaminants in samples, could be due either to

312 variations of contaminating load between our extraction and amplification kits, that are not sourced

313 from the same company, or to the presence of a basal level of extraction contaminant bacterial DNA,

314 potentially limiting the contamination arising from the next step (amplification).

315   Interestingly, we showed that in terms of contamination rates, the influence of contaminant

316 presence was significantly higher in nymph and male samples compared to female ones. Because low

317 biomass samples were shown to be particularly sensitive to contamination (Salter *et al.*, 2014), we

318 might suppose that the amount of bacterial DNA would be lower in males and nymphs, explaining

319 the higher presence of contaminants in these samples. It would be useful to develop epifluorescence

320 or flow cytometry protocols adapted to tick samples, to assess whether the amount of bacteria

321 composing the female microbiota is effectively higher than in the male and nymph. Additionally,

322 previous studies dealing with tick microbes have identified higher microbiota diversity in males and

323 nymphs compared to females, while investigating richness and PD (Van Treuren *et al.*, 2015; Zolnik

324 *et al.*, 2016) as well as Shannon and Simpson indexes (Zolnik *et al.*, 2016). However, no negative

325 controls and data cleaning were reported in these studies, and they could therefore result from

326 contamination. In the present study, we determined that the presence of contaminants can lead to

327 an overestimation of microbial diversity in all the tick stages, and that this is the case for several

328 diversity indexes, especially richness and PD estimations, which are the measures most affected by

329 the cleaning process. Furthermore, the impact of removing contaminants on diversity comparisons

330 between stages was estimated and showed that this cleaning step can influence the final results, and

331 particularly in our case, concerning the observed diversity estimation. Our results therefore call for

332 caution in attempts to draw conclusion about tick stage diversity without considering negative

333 controls.

334 As already mentioned by Narasimhan and Fikrig (2015), the understanding of arthropod

335 microbial communities in the context of arthropod survival and pathogen transmission may open new

336 routes in arthropod-borne pathogen control strategies. In this context, many studies have assessed

337 tick microbiota composition in individual ticks or even in tick organs (Andreotti *et al.*, 2011; Lalzar *et*

338 *al.*, 2012; Hawlena *et al.*, 2013; Budachetri *et al.*, 2014; Narasimhan *et al.*, 2014; Qiu *et al.*, 2014;

339 Clayton *et al.*, 2015; Rynkiewicz *et al.*, 2015; Van Treuren *et al.*, 2015; Gall *et al.*, 2016; Khoo *et al.*,

340 2016; Zolnik *et al.*, 2016; Abraham *et al.*, 2017; Kwan *et al.*, 2017; Swei and Kwan, 2017; Estrada-Peña

341 *et al.*, 2018). However, very few reported the use of negative controls or mentioned the identification

342 and removal of subsequently identified contaminants (Hawlena *et al.*, 2013; Rynkiewicz *et al.*, 2015;

343 Estrada-Peña *et al.*, 2018), thus obscuring the risk that identified micro-organisms as members of the

344 microbiota could correspond to contaminants. Nonetheless, the bias generated by the presence of

345 contaminants can have a considerable influence on the characterization of tick microbiota

346 composition and structure. Here, we showed that contaminants represented almost 50 % of the total

347 sequence yield of tick samples. Furthermore, several contaminants we identified belong to genera

348 (especially *Pseudomonas* and *Acinetobacter*) previously reported in several tick microbiota studies

349 (reviewed by Narasimhan and Fikrig, 2015). This does not necessarily mean that these OTUs are

350 contaminants, as several OTUs belonging to the same genera may correspond either to tick

351 microbiota members or to contaminants. However, this points out the lack of reliability concerning

352 this crucial point in the understanding of tick microbiota composition, and the need to perform

353 negative controls to clean the data set and to rule out the hypothesis of potential contamination.

354

Conclusions

356 Because of their ease of use and the fact that they employ safer chemicals than those in

357 conventional methods (i.e. phenol chloroform extraction), the use of commercial DNA extraction kits

358 has become standard, particularly in the context of studies dealing with the identification of tick

359 microbial communities. However, the issue of contaminants coming from DNA extraction kits has

360 recently been raised for microbiota studies dealing with low biomass samples, such as blood,

361 intestinal tissue, mucosal tissue, or nasopharyngeal samples. Our study adds further evidence

362 supporting these important concerns in the field of tick microbiota. Here, our results suggest that

363　many negative controls should be performed at each step of sample processing prior to performing

364　microbiota analyses, allowing us to efficiently identify a potentially serious bias in the study of the

365　actual tick microbial community. While more and more studies try to identify tick microbiota to *in*

366　*fine* potentially propose hypothesis for the development of new tick-borne pathogen control

367　strategies, we strongly advise the routine use of negative controls in microbiota studies, particularly

368　in the context of low biomass samples like ticks or other small arthropods. These procedures should

369　be used routinely in order to obtain the most reliable results possible.

370

371　**List of abbreviations**

372　OTU: Operational Taxonomic Unit

373　*I. ricinus*: *Ixodes ricinus*

374　CC: Crushing Control

375　EC: Extraction Control

376　AC: Amplification Control

377

378　**Author Contributions**

379　Conceived and designed the experiments: TP, MVT, JFC, EL. Performed the experiments: EL. Analysed

380　the data: EL, AE, TP, MM, MM, CM, OR. Wrote the paper: EL, AE, MM, JFC, OR, MM, CM, MVT, TP. All

381　the authors read and approved the final manuscript.

382

383　**Funding**

387 The salary of Emilie Lejal, the PhD student working on this project was funded by the Ile-de-France

388 region.

Acknowledgments

390   The authors would like to thanks Sabine Delannoy and the IdentyPath genomic platform that

391 allowed us to perform part of the experiments in their laboratory. We also thank Maxime Galan for

392 providing index sequences and stimulating discussions. Thank you as well to Ladislav Šimo for his

393 precious help in figure improvement.

References

395 Abraham, N.M., Liu, L., Jutras, B.L., Yadav, A.K., Narasimhan, S., Gopalakrishnan, V., et al. (2017)
396   Pathogen-mediated manipulation of arthropod microbiota to promote infection. *Proc Natl*
397   *Acad Sci* **114**: E781–E790.

398 Andreotti, R., Pérez de León, A.A., Dowd, S.E., Guerrero, F.D., Bendele, K.G., and Scoles, G.A. (2011)
399   Assessment of bacterial diversity in the cattle tick Rhipicephalus (Boophilus) microplus
400   through tag-encoded pyrosequencing. *BMC Microbiol* **11**: 6.

401 Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data.

402 Budachetri, K., Browning, R.E., Adamson, S.W., Dowd, S.E., Chao, C.-C., Ching, W.-M., and Karim, S.
403   (2014) An Insight Into the Microbiome of the Amblyomma maculatum (Acari: Ixodidae). *J*
404   *Med Entomol* **51**: 119–129.

405 Carpi, G., Cagnacci, F., Wittekindt, N.E., Zhao, F., Qi, J., Tomsho, L.P., et al. (2011) Metagenomic
406   Profile of the Bacterial Communities Associated with Ixodes ricinus Ticks. *PLOS ONE* **6**:
407   e25604.

408 Clayton, K.A., Gall, C.A., Mason, K.L., Scoles, G.A., and Brayton, K.A. (2015) The characterization and
409   manipulation of the bacterial microbiome of the Rocky Mountain wood tick, Dermacentor
410   andersoni. *Parasit Vectors* **8**: 632.

411 Degli Esposti, M. and Martinez Romero, E. (2017) The functional microbiome of arthropods. *PLOS*
412   *ONE* **12**: e0176573.

413 Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., et al. (2018) FROGS: Find,
414   Rapidly, OTUs with Galaxy Solution. *Bioinformatics* **34**: 1287–1294.

415 Estrada-Peña, A., Cabezas-Cruz, A., Pollet, T., Vayssier-Taussat, M., and Cosson, J.-F. (2018) High
416   Throughput Sequencing and Network Analysis Disentangle the Microbial Communities of
417   Ticks and Hosts Within and Between Ecosystems. *Front Cell Infect Microbiol* **8**: 236.

418 Galan, M., Razzauti, M., Bard, E., Bernard, M., Brouat, C., Charbonnel, N., et al. (2016) 16S rRNA
419   Amplicon Sequencing for Epidemiological Surveys of Bacteria in Wildlife. *mSystems* **1**:
420   e00032-16.

18

421   Gall, C.A., Reif, K.E., Scoles, G.A., Mason, K.L., Mousel, M., Noh, S.M., and Brayton, K.A. (2016) The
422   bacterial microbiome of Dermacentor andersoni ticks influences pathogen susceptibility.
423   *ISME J* **10**: 1846–1855.

424   Glassing, A., Dowd, S.E., Galandiuk, S., Davis, B., and Chiodini, R.J. (2016) Inherent bacterial DNA
425   contamination of extraction and sequencing reagents may affect interpretation of
426   microbiota in low bacterial biomass samples. *Gut Pathog* **8**: 24.

427   Gofton, A.W., Oskam, C.L., Lo, N., Beninati, T., Wei, H., McCarl, V., et al. (2015) Inhibition of the
428   endosymbiont "Candidatus Midichloria mitochondrii" during 16S rRNA gene profiling reveals
429   potential pathogens in Ixodes ticks from Australia. *Parasit Vectors* **8**: 345.

430   Hamdan, L.J., Coffin, R.B., Sikaroodi, M., Greinert, J., Treude, T., and Gillevet, P.M. (2013) Ocean
431   currents shape the microbiome of Arctic marine sediments. *ISME J* **7**: 685–696.

432   Hawlena, H., Rynkiewicz, E., Toh, E., Alfred, A., Durden, L.A., Hastriter, M.W., et al. (2013) The
433   arthropod, but not the vertebrate host or its environment, dictates bacterial community
434   composition of fleas and ticks. *ISME J* **7**: 221–223.

435   Heise, S.R., Elshahed, M.S., and Little, S.E. (2010) Bacterial Diversity in Amblyomma americanum
436   (Acari: Ixodidae) With a Focus on Members of the Genus Rickettsia. *J Med Entomol* **47**: 258–
437   268.

438   Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., et al. (2010)
439   Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.

440   Khoo, J.-J., Chen, F., Kho, K.L., Ahmad Shanizza, A.I., Lim, F.-S., Tan, K.-K., et al. (2016) Bacterial
441   community in Haemaphysalis ticks of domesticated animals from the Orang Asli
442   communities in Malaysia. *Ticks Tick-Borne Dis* **7**: 929–937.

443   Kircher, M., Sawyer, S., and Meyer, M. (2012) Double indexing overcomes inaccuracies in multiplex
444   sequencing on the Illumina platform. *Nucleic Acids Res* **40**: e3–e3.

445   Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., et al. (2018) Best
446   practices for analysing microbiomes. *Nat Rev Microbiol* **16**: 410.

447   Kwan, J.Y., Griggs, R., Chicana, B., Miller, C., and Swei, A. (2017) Vertical vs. horizontal transmission
448   of the microbiome in a key disease vector, Ixodes pacificus. *Mol Ecol* **26**: 6578–6589.

449   Lalzar, I., Harrus, S., Mumcuoglu, K.Y., and Gottlieb, Y. (2012) Composition and Seasonal Variation of
450   Rhipicephalus turanicus and Rhipicephalus sanguineus Bacterial Communities. *Appl Env*
451   *Microbiol* **78**: 4110–4116.

452   Lejal, E., Marsot, M., Chalvet-Monfray, K., Cosson, J.-F., Moutailler, S., Vayssier-Taussat, M., and
453   Pollet, T. (2019) A three-years assessment of Ixodes ricinus-borne pathogens in a French
454   peri-urban forest. *bioRxiv* 597013.

455   Mahé, F., Rognes, T., Quince, C., Vargas, C. de, and Dunthorn, M. (2015) Swarm v2: highly-scalable
456   and high-resolution amplicon clustering. *PeerJ* **3**: e1420.

457   Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
458   *EMBnet.journal* **17**: 10–12.

459   McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis
460   and Graphics of Microbiome Census Data. *PLOS ONE* **8**: e61217.

461   Moreno, C.X., Moy, F., Daniels, T.J., Godfrey, H.P., and Cabello, F.C. (2006) Molecular analysis of
462   microbial communities identified in different developmental stages of Ixodes scapularis ticks
463   from Westchester and Dutchess Counties, New York. *Environ Microbiol* **8**: 761–772.

464   Narasimhan, S. and Fikrig, E. (2015) Tick microbiome: the force within. *Trends Parasitol* **31**: 315–
465   323.

466   Narasimhan, S., Rajeevan, N., Liu, L., Zhao, Y.O., Heisig, J., Pan, J., et al. (2014) Gut Microbiota of the
467   Tick Vector Ixodes scapularis Modulate Colonization of the Lyme Disease Spirochete. *Cell*
468   *Host Microbe* **15**: 58–71.

469   Qiu, Y., Nakao, R., Ohnuma, A., Kawamori, F., and Sugimoto, C. (2014) Microbial Population Analysis
470   of the Salivary Glands of Ticks; A Possible Strategy for the Surveillance of Bacterial
471   Pathogens. *PLOS ONE* **9**: e103961.

472   Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal
473   RNA gene database project: improved data processing and web-based tools. *Nucleic Acids*
474   *Res* **41**: D590–D596.

475   R Core Team (2018) R: A Language and Environment for Statistical Computing, Vienna, Austria: R
476   Foundation for Statistical Computing.

477   René-Martellet, M., Minard, G., Massot, R., Tran Van, V., Valiente Moro, C., Chabanne, L., and
478   Mavingui, P. (2017) Bacterial microbiota associated with Rhipicephalus sanguineus (s.l.) ticks
479   from France, Senegal and Arizona. *Parasit Vectors* **10**: 416.

480   Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source
481   tool for metagenomics. *PeerJ* **4**: e2584.

482   Rynkiewicz, E.C., Hemmerich, C., Rusch, D.B., Fuqua, C., and Clay, K. (2015) Concordance of bacterial
483   communities of two tick species and blood of their shared rodent host. *Mol Ecol* **24**: 2566–
484   2579.

485   Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., et al. (2014) Reagent
486   and laboratory contamination can critically impact sequence-based microbiome analyses.
487   *BMC Biol* **12**: 87.

488   Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., et al. (2006) Microbial
489   diversity in the deep sea and the underexplored "rare biosphere." *Proc Natl Acad Sci* **103**:
490   12115–12120.

491   Sperling, J.L., Silva-Brandão, K.L., Brandão, M.M., Lloyd, V.K., Dang, S., Davis, C.S., et al. (2017)
492   Comparison of bacterial 16S rRNA variable regions for microbiome surveys of ticks. *Ticks*
493   *Tick-Borne Dis* **8**: 453–461.

494   Swei, A. and Kwan, J.Y. (2017) Tick microbiome and pathogen acquisition altered by host blood
495   meal. *ISME J* **11**: 813–816.

496   Thapa, S., Zhang, Y., and Allen, M.S. (2019) Effects of temperature on bacterial microbiome
497   composition in Ixodes scapularis ticks. *MicrobiologyOpen* **8**: e00719.

498   Van Treuren, W., Ponnusamy, L., Brinkerhoff, R.J., Gonzalez, A., Parobek, C.M., Juliano, J.J., et al.
499   (2015) Variation in the Microbiota of Ixodes Ticks with Regard to Geography, Species, and
500   Sex. *Appl Env Microbiol* **81**: 6200–6209.

501   Weiss, S., Amir, A., Hyde, E.R., Metcalf, J.L., Song, S.J., and Knight, R. (2014) Tracking down the
502   sources of experimental contamination in microbiome studies. *Genome Biol* **15**: 564.

503   Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-
504   End reAd mergeR. *Bioinformatics* **30**: 614–620.

505   Zolnik, C.P., Prill, R.J., Falco, R.C., Daniels, T.J., and Kolokotronis, S.-O. (2016) Microbiome changes
506   through ontogeny of a tick pathogen vector. *Mol Ecol* **25**: 4963–4977.