Title: Taxon appearance from extraction and amplification steps demonstrates the value of multiple controls in tick microbiome analysis

Emilie Lejal [1], Agustín Estrada-Peña[2], Maud Marsot[3], Jean-François Cosson[1], Olivier Rué[4], Mahendra Mariadassou[4], Cédric Midoux[4,5], Muriel Vayssier-Taussat[6] and Thomas Pollet[1].

[1] UMR BIPAR, Animal Health Laboratory, INRA, ANSES, Ecole Nationale Vétérinaire d'Alfort, Université Paris-Est, Maisons-Alfort, France

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

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

[4] MaIAGE, INRA, Université Paris-Saclay, 78350, Jouy-en-Josas, France

[5] Irstea, UR PROSE, 92761 Antony, France

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

Contacts:

emilie.lejal@vet-alfort.fr

thomas.pollet@inra.fr

muriel.vayssier@inra.fr

Abstract

- **Background**

The development of high throughput sequencing (HTS) technologies has substantially improved analysis of bacterial community diversity, composition, and functions. Over the last decade, HTS 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 blanks and 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

50 **Background**

51 Analysis of microbial community composition by 16S tag sequencing represents a powerful tool
52 commonly used to evaluate the prokaryotic diversity in many ecosystems. Thanks to high-throughput
53 sequencing (HTS) approaches, it is now feasible to identify microbiota with high coverage and reveal
54 the exceptional and complex microbial diversity of environments ranging from the wide ocean (1,2) to
55 small arthropods (3,4). With the constant improvement of sensitivity associated with its decreasing
56 cost, 16S tag sequencing has spread to many areas, including those dealing with low biomass
57 communities such as tick microbiota.

58 Despite the limitation induced by low biomass, a high number of tick microbiota studies are
59 focusing on individual ticks (4–12) or even on individual or pooled tick organs (9,13–19). The results
60 obtained from these studies have improved our knowledge of tick microbiota, while demonstrating
61 links between its composition and several factors, such as tick stages and gender (5,8,12,18), organ of
62 origin (18), living environment (4,18,20), host blood meal (12), or tick engorgement (15,21,22). The
63 presence of tick-borne pathogens has also been demonstrated to be related to differences in tick
64 microbiota composition and diversity (9,15,19). The continuous improvements in our knowledge, and
65 in particular the assessment of potential links between tick-borne pathogens and particular microbiota
66 members, require us to continue performing analyses on individual ticks, or even better, on individual
67 tick organs.

68 Several potential biases, inherent to the PCR amplification (chimeric sequences) or sequencing
69 steps (random sequencing errors), are now well known (23), and pipelines have been developed to
70 detect and remove these errors during the sequence analysis. However, several studies have begun to
71 warn scientists about the potential distortion of results due to contamination from the
72 extraction/amplification steps or from the environment during sample processing (23–28). This
73 potential contamination is able to significantly modify results concerning microbe diversity and
74 composition, particularly in the case of low biomass samples, and can have serious consequences on
75 the interpretation of results. For example, while investigating 16s rRNA gene profiles from

76 nasopharyngeal swabs regularly sampled from a cohort of children from 1 to 24 months, Salter *et al*.
77 (24) demonstrated that the apparent clustering of samples by age, observed by principal coordinate
78 analysis, in fact corresponded to the commercial kit used to perform the DNA extraction of these
79 samples. While studying the influence of laboratory processing on diluted samples, these authors also
80 observed that the lower the initial biomass of samples, the higher the proportion of contaminants
81 associated with the kits or the environment. In the context of tick microbiota studies, performing using
82 blanks or negative controls would therefore be essential to detect potential contamination, and
83 efficiently distinguish this contamination from genuine tick microbial communities. On the basis of an
84 exhaustive review of the scientific literature, it is very clear that routine multiple control analyses
85 (extraction and PCR negative controls) are not considered at their true value, and are only rarely
86 implemented in tick microbial community analyses (4,6,7). In this context, we performed a 16s rRNA
87 gene sequencing analysis on individual ticks (females, males and nymphs) considering concomitantly
88 several negative controls at each step of the process, prior to the sequencing run (tick crushing, DNA
89 extraction and amplification).

90 **Methods**

91 ***Tick collection and sample selection***

92 Questing *Ixodes ricinus* nymphs and adults were collected for three years by dragging (from April
93 2014 to May 2017) in the Sénart forest in the south of Paris. More details on the sampling location and
94 design, and tick collection, are available in Lejal *et al*. (29). After morphological identification, ticks were
95 stored at -80°C until further analysis.

96

97 ***Tick washing, crushing, and DNA extraction***

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

107

108 ***DNA amplification and multiplexing***

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

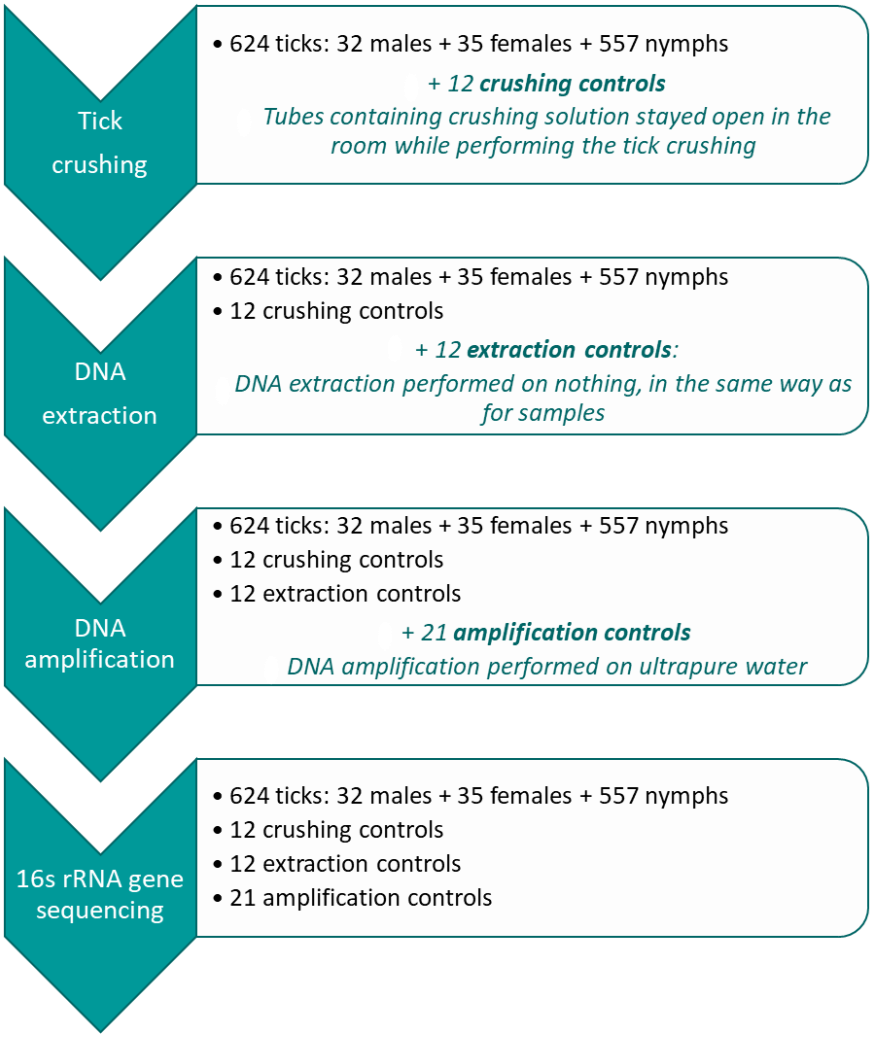
116 compared to prokaryotic DNA. An 8 bp-index was added to the forward and reverse primers. The
117 individual amplification of each sample was performed using a combination of 22 and 32 index-tagged
118 forward and reverse primers, allowing for amplification and multiplexing of a maximum of 704 samples
119 that can be loaded together on an Illumina MiSeq flow cell. Amplification and multiplexing steps were
120 performed under a sterile hood, decontaminated before and after use with DNA remover and UVs,
121 using tube strips with individual caps to prevent cross-contamination. All the PCR amplifications were
122 carried out using the Phusion® High-Fidelity DNA Polymerase amplification kit (Thermo Scientific,
123 Lithuania). For each sample, 5 µL of DNA extract were amplified in a 50 µL final reaction volume,
124 containing 1X Phusion HF buffer, 0.2 µM of dNTPs, 0.2 U/mL of Phusion DNA polymerase, and 0.35 µM
125 of forward and reverse primer. The following thermal cycling procedure was used: initial denaturation
126 at 98°C for 30 s, 35 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s, followed by
127 extension at 72°C for 30 s. The final extension was carried out at 72°C for 10 minutes. Amplification
128 step was performed on a peqSTAR 2x thermocycler (PEQLAB, Germany). For each PCR run, negative
129 controls were performed by using the reaction mixture without template, but using ultrapure water
130 (Invitrogen, United Kingdom). PCR products were checked on 1.5% agarose gels.
131
132 ***DNA purification, quantification, and pooling***
133 Amplicons were cleaned and quantified using a SequalPrep normalization plate kit (Invitrogen,
134 USA), and a Qubit dsDNA High Sensitivity Assay kit (Invitrogen, USA), respectively. Amplicons and
135 controls were pooled at equimolar concentrations and sent to the sequencing platform (GenoScreen,
136 France).
137
138 ***Controls***
139 In total, 45 negative controls were included and analyzed using the same process as for tick
140 samples (Figure 1).

141 **12 Crushing controls (CCs):** Tube containing crushing solution, remaining open in the room while
142 putting ticks inside other tubes between the cleaning and crushing steps. DNA extraction and
143 amplification were then performed on the crushing control in the same conditions as for any other
144 sample.

145 **12 DNA extraction controls (ECs):** For these controls, the different DNA extraction steps were
146 performed under the same conditions as for any other samples, but without adding sample. DNA
147 amplification was then performed on the extraction control in the same conditions as for any other
148 sample.

149 **21 DNA amplification controls (ACs)**: The DNA amplification step was performed on ultrapure
150 water (Invitrogen, United Kingdom), in the same conditions as for any other sample.

151
152
153
154
155
156
157
158
159
160
161
162
163
164
165



166 **Figure 1. Process followed by samples and added control steps.**

167

168 ***Sequencing and data processing***

169 The equimolar mix was concentrated and sequenced by GenoScreen (Lille, France) using MiSeq
170 Illumina 2 x 250 bp chemistry. Quality control checks on raw data were performed using FastQC v0.11.3
171 (30), and no problems were detected. First, pairs of amplicons were merged using Pear v0.9.11 (31)
172 with a minimum overlap of 20 bp, a minimum assembly length of 100 bp, a maximum assembly length
173 of 480 bp, and an e-value of 0.01. Then, the sequencing orientation of extended sequences was
174 checked by using cutadapt v1.12 (32), with primers in forward and reverse strands (without removing
175 them). Sequences with a reversed orientation were reverse-complemented to obtain a set of
176 sequences with the same orientation (5' to 3'). The sequence analysis was performed using FROGS (33).
177 Primers were removed using cutadapt and resulting sequences were dereplicated. Operational
178 taxonomic units (OTUs) were built using Swarm v2.1.12 (34) with parameter d = 3, and chimeras were
179 filtered using vsearch v1.4 (35) in a *de novo* mode. Finally, OTUs with low abundance (< 0.005% of the
180 total sequence count) were filtered out, resulting in 608 OTUs (corresponding to 2,835,506 sequences).
181 OTUs were affiliated by blasting cluster seed sequences against the Silva database v132 (36). OTUs
182 corresponding to mitochondrial/chloroplast DNA or not affiliated to any domain were removed, leading
183 to a dataset composed of 513 OTUs and 2,726,287 sequences.

184

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

186 The control composition was used to distinguish tick microbial OTUs from contaminants. All the OTUs
187 corresponding to more than 1% of the total number of sequences obtained in each category of control
188 were identified as main contaminants. However, due to possible errors in sequence assignment
189 occurring during sequencing due to the formation of mixed clusters on the flow cell (37), negative
190 controls can also contain a few sequences of tick microbial OTUs, complicating contaminant
191 identification under this 1% threshold. Therefore, for each OTU, the number of sequences expected to
192 spread to other samples due to false assignment (i.e. expected number) was determined by calculating

193 0.02% of the total number of each OTU in the dataset, corresponding to a false assignment rate
194 estimated previously by Galan *et al*. (23). This expected number was compared to the maximum
195 number of obtained sequences in controls (i.e. maximum number). As this rate was not estimated from
196 our dataset, we decided to allow variability around this threshold, while calculating a 99% confidence
197 interval around the expected number of 0.02%. OTUs whose maximum sequence number was above
198 the upper threshold were considered contaminants. Importantly, cross-contamination from samples to
199 controls was considered negligible, as no negative controls presented a higher maximum sequence
200 number than the expected number for an OTU belonging to the most abundant and characteristic
201 genus of tick microbiota in our dataset (Midichloria), which is also the 3rd most abundant OTU in the
202 dataset (266,516 sequences).
203
204 ***Statistical analyses***
205 The percentage of samples presenting a contamination rate higher than 50% was compared
206 between samples categories (nymphs, males and females) according to a Chi^2 statistical test.
207 Alpha diversity was estimated before and after sample cleaning. Samples containing fewer than
208 500 sequences after cleaning were removed from the analysis, even if they contained more than 500
209 sequences before cleaning. Several measures of alpha diversity were calculated. Basically, the observed
210 number of OTUs (species richness) was used as well as the Faith Phylogenetic Diversity (PD) index,
211 corresponding to the total branch length of the subtree spanned by the community, and also two
212 indexes taking into account the number of sequences: the Shannon index (more influenced by rare
213 OTUs) and Simpson index (more influenced by dominant OTUs). Phylogenetic diversity was estimated
214 using the Picante package (38), while other measures were assessed thanks to Easy 16S
215 (http://genome.jouy.inra.fr/shiny/easy16S/), based on the phyloseq package (39). Alpha diversity
216 measure comparison was performed between cleaned and not cleaned groups, as well as between
217 stages, using non-parametric statistical tests (Wilcoxon and Kruskal–Wallis, respectively). Statistical
218 computations were performed in R 3.6.0 (40).

219

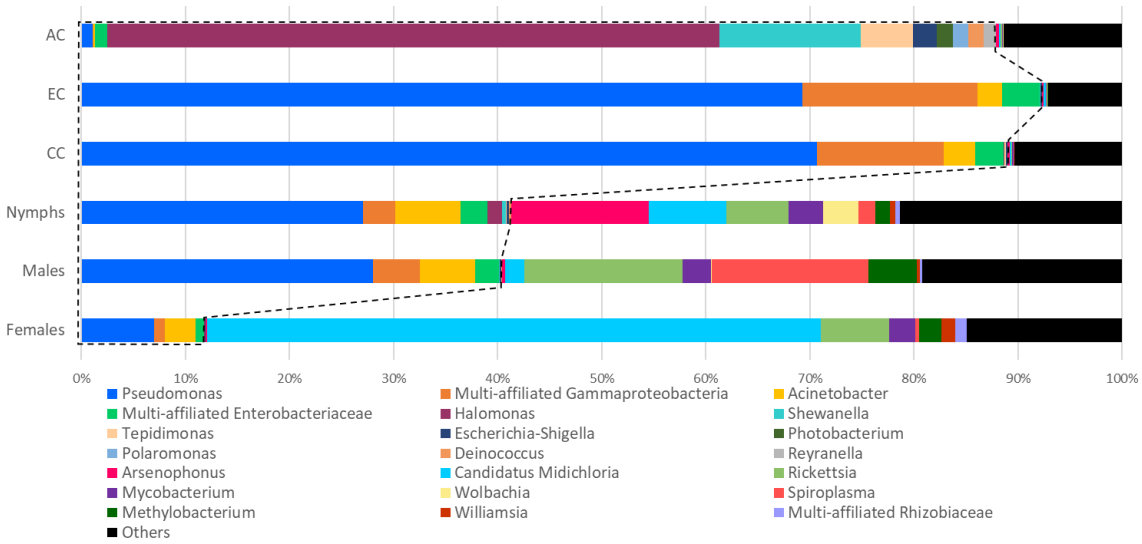
220 **Results**

221 Nymphs, males and females presented a mean number of sequences per sample of 3,721,
222 2,689 and 5,316, respectively. Concerning controls, a mean number of 13,198, 11,235 and 1,126
223 sequences per sample was obtained for crushing, extraction and amplification controls (CCs, ECs and
224 ACs), respectively.

225 The most abundant OTUs, with relative abundance higher than 1% in at least one category of
226 samples or controls, are represented in Figure 2. All other OTUs, i.e. those with relative abundance
227 lower than 1% in all categories were grouped together in the "Others" category.

228 The examination of the at-least-1% OTU composition of the control (CCs, ECs, ACs) allowed to
229 identify OTUs belonging to the genera: *Pseudomonas*, *Acinetobacter*, *Halomonas*, *Shewanella*,
230 *Tepidimonas*, *Esherichia-Shigella*, *Photobacterium*, *Polaromonas*, *Deinococcus* and *Reyranella*, as well
231 as two OTUs belonging to the Enterobacteriaceae family but multi-affiliated at the genus level, and
232 one OTU only affiliated to Gammaproteobacteria at the Class level (Figure 2). OTU composition was
233 contrasted across the control types. While the same OTUs were identified in CCs and ECs and belonged
234 to *Pseudomonas* [70.7% and 69.3%, respectively], multi-affiliated Gammaproteobacteria [12.2% and
235 16.7%, respectively], *Acinetobacter* [3.1% and 2.4%, respectively], and multi-affiliated
236 Enterobacteriaceae [2.7% and 3.6%, respectively]), the microbial community identified in AC controls
237 was dominated by OTUs belonging to *Halomonas* (58.8%), *Shewanella* (13.6%), *Tepidimonas* (5.1%),
238 *Escherichia-Shigella* (2.3%), *Photobacterium* (1.5%), *Polaromonas* (1.5%), *Deinococcus* (1.5%),
239 *Reyranella* (1.2%) and multi-affiliated Enterobacteriaceae (1.2%). These OTUs represented 38.9% of
240 the total number of sequences coming from the tick samples. Focusing on the sample category, these
241 OTUs correspond to 11.9%, 40.4% and 41.3% of sequences detected in females, males and nymphs,
242 respectively (Figure 2). These OTUs were therefore considered to be the main contaminants of the
243 dataset. It is interesting to note that most of these OTUs correspond to those detected in both EC and
244 CC controls. In fact, only 2.8% of sequences detected in the different categories of samples (nymphs,

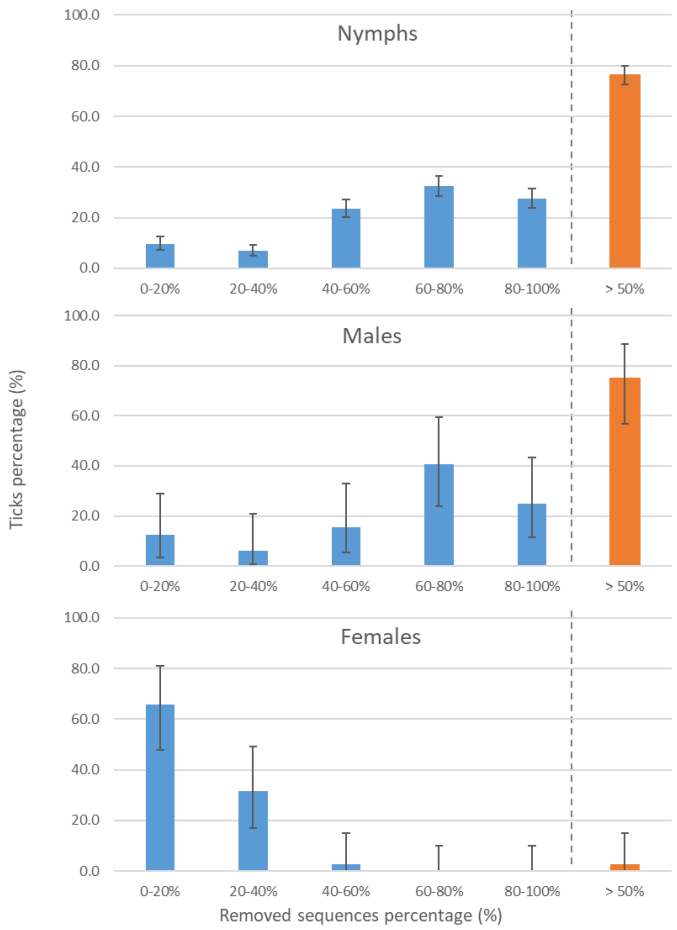
245 females and males) correspond to OTUs detected in the AC controls and identified as main
246 contaminants.
247
248
249
250
251
252



253
254
255
256 **Figure 2. 16S rRNA gene sequencing taxonomic assignment comparison between negative controls**
257 **and samples.** Percentages are represented for the OTUs corresponding to at least 1% of the total number of
258 sequences assigned to a category of samples or controls, those corresponding to less than 1% are grouped under
259 the term “Others”. OTUs belonging to the same genus or family (if multi-affiliated at the scale of genus) were
260 grouped under the same name. One OTU was only identified until the class level (multi-affiliated
261 Gammaproteobacteria).
262
263 A more in-depth identification of contaminants, including those corresponding to less than 1%
264 of the total number of sequences was performed, allowing us to identify 160 OTUs as contaminants
265 among the 513 initially identified OTUs in the dataset. This identification increased the proportions of
266 identified contaminants in the samples, which reach 50.9% of the total sequence count from the tick
267 samples, and 20.1%, 51.2% and 53.7% of sequences detected in females, males and nymphs,
268 respectively. Furthermore, after these cleaning steps, 201 samples contained less than 500 sequences
269 compared to 8 before cleaning. For 76.6% of nymph and 75% of male samples, contaminant OTUs
270 represented more than 50% of the total sequence count (Figure 3). These sample contamination rates

271 are significantly higher than those obtained for females ($p > 0.05$ according to a Chi^2 test), for which
272 most of the samples (65.7%) exhibited a contamination rate below 20%.

273
274
275
276
277
278
279
280
281
282
283
284
285
286



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

288

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

Concerning cleaned data, significant differences were still observed between stages when using Shannon and Simpson indexes, but disappeared when using the observed diversity measure.

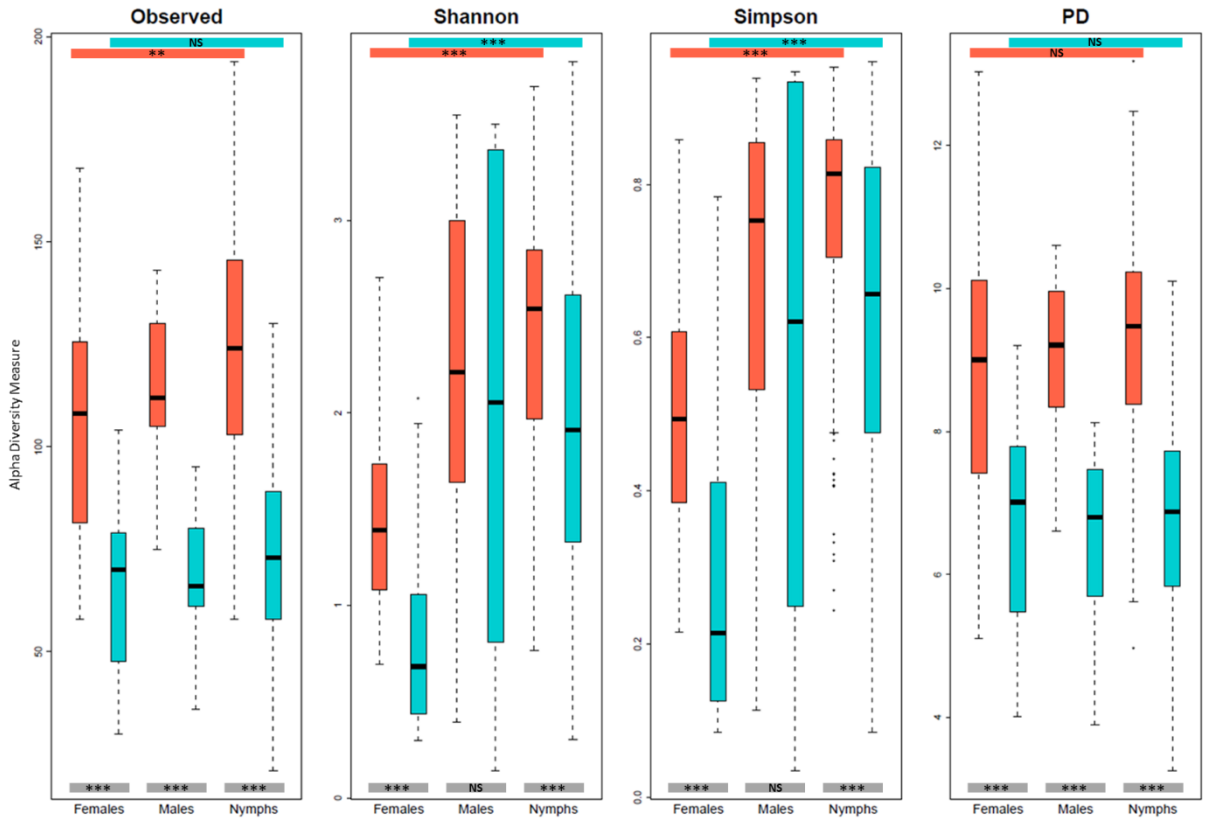


Figure 4. Impact of contamination on alpha diversity estimations according to tick stages. Alpha diversity has been estimated according to the observed number of OTUs, the Shannon and Simpson indexes and the Faith's phylogenetic diversity (PD), for each stages on cleaned (turquoise) and not cleaned (orange-red) datasets. Results of the statistical test comparing alpha diversity measurement between stages, as well as between the cleaned and uncleaned groups, are represented on coloured bands at the top and at the bottom of the frame respectively: NS= Not Significant, ***= $p < 0.001$, **= $p < 0.01$, *= $p < 0.05$.

317 **Discussion**

318 In this study, we sequenced the microbiota of 624 individual *I. ricinus* ticks including several
319 categories of negative controls to identify and quantify the contaminating sequences, in particular
320 those produced during the extraction or amplification steps. The performance of multiple negative
321 controls allowed us to perform in-depth identification of contaminant OTUs. These OTUs represented
322 50.9% of the total sequence yield from tick samples, and had a significant impact on microbial diversity
323 estimation. This shows the importance of contamination potentially arising from processing steps, and
324 demonstrates that its consequences on data analyses can be significant and should not be neglected.

325 A high number of sequences in each category of controls was detected, particularly in CCs and
326 ECs, in which the mean number of sequences was even higher than the mean sequence number
327 detected in samples. Importantly, this higher number might be partly overestimated, first due to the
328 presence of tick DNA during the PCR that may have inhibited the reaction on the 16s rRNA gene in
329 samples while not in controls, and second during the measure of DNA concentrations that may have
330 led to a higher estimation of concentrations in samples compared to controls, before the pooling step.
331 Despite this, the high number of sequences in controls raises the issue of contaminant presence, while
332 performing each step of the sample processing.

333 Each processing step leads to a different contamination pattern. Most of the main
334 contaminants do not seem to originate from the DNA amplification step. Furthermore, as the
335 composition of crushing and extraction controls was the same, while the reagents used were different
336 and crushing control also underwent the extraction step, the contaminant sequences most probably
337 came from the DNA extraction step, rather than from the crushing step. The key role of the extraction
338 step in the contamination of low biomass samples has previously been demonstrated by Salter *et al*.
339 (24). Furthermore, three of the four main contaminants identified in our analysis, corresponding to
340 *Pseudomonas*, *Acinetobacter*, and multi-affiliated Enterobacteriaceae, belong to families and genera
341 identified by these authors as contaminants coming from DNA extraction kits. Our observations
342 differed from those obtain by Galan *et al*. (23), who mainly detected contaminants coming from the

343 DNA amplification step, while this step represented only 2% of the main contaminants detected in our
344 samples. In their study of kit contaminants, Salter *et al.* (24) observed that the extent of contamination
345 arising from DNA extraction was different depending on the kit brand used to perform this step. While
346 using the less contaminating kit, they extracted DNA from several 10-fold cascade dilutions of a pure
347 suspension of *Salmonella bongori*, and subsequently assessed 16s rRNA gene concentrations via q-
348 PCR. They demonstrated a linear relationship between the 16s rRNA gene copy number and the
349 dilution fold only for the first dilution points, after which levels started to stabilize and to reach a
350 threshold of 10^5 mean copy number/mL. Consequently, the higher rate of extraction, rather than
351 amplification contaminants in samples, could be due either to variations of contaminating load
352 between our extraction and amplification kits, that are not sourced from the same company, or to the
353 presence of a basal level of extraction contaminant bacterial DNA, potentially limiting the
354 contamination arising from the next step (amplification).

355 Interestingly, we showed that in terms of contamination rates, the influence of contaminant
356 presence was significantly higher in nymph and male samples compared to female ones. Because low
357 biomass samples were shown to be particularly sensitive to contamination (24), we might suppose that
358 the amount of bacterial DNA would be lower in males and nymphs, explaining the higher presence of
359 contaminants in these samples. It would be useful to develop epifluorescence or flow cytometry
360 protocols adapted to tick samples, to assess whether the amount of bacteria composing the female
361 microbiota is effectively higher than in the male and nymph. Additionally, previous studies dealing with
362 tick microbes have identified higher microbiota diversity in males and nymphs compared to females,
363 while investigating richness and PD (8,18) as well as Shannon and Simpson indexes (18). However, no
364 negative controls and data cleaning were reported in these studies, and they could therefore result
365 from contamination. In the present study, we determined that the presence of contaminants can lead
366 to an overestimation of microbial diversity in all the tick stages, and that this is the case for several
367 diversity indexes, especially richness and PD estimations, which are the measures most affected by the
368 cleaning process. Furthermore, the impact of removing contaminants on diversity comparisons

369 between stages was estimated and showed that this cleaning step can influence the final results, and
370 particularly in our case, concerning the observed diversity estimation. Our results therefore call for
371 caution in attempts to draw conclusion about tick stage diversity without considering negative
372 controls.

373 As already mentioned by Narasimhan and Fikrig (26), the understanding of arthropod microbial
374 communities in the context of arthropod survival and pathogen transmission may open new routes in
375 arthropod-borne pathogen control strategies. In this context, many studies have assessed tick
376 microbiota composition in individual ticks or even in tick organs (4–19). However, very few reported
377 the use of negative controls or mentioned the identification and removal of subsequently identified
378 contaminants (4,6,7), thus obscuring the risk that identified micro-organisms as members of the
379 microbiota could correspond to contaminants. Nonetheless, the bias generated by the presence of
380 contaminants can have a considerable influence on the characterization of tick microbiota composition
381 and structure. Here, we showed that contaminants represented almost 38.9% of the total sequence
382 yield of tick samples. Furthermore, several contaminants we identified belong to genera (especially
383 *Pseudomonas* and *Acinetobacter*) previously reported in several tick microbiota studies (reviewed by
384 26). This does not necessarily mean that these OTUs are contaminants, as several OTUs belonging to
385 the same genera may correspond either to tick microbiota members or to contaminants. However,
386 this points out the lack of reliability concerning this crucial point in the understanding of tick microbiota
387 composition, and the need to perform negative controls to clean the data set and to rule out the
388 hypothesis of potential contamination.

389

390 **Conclusions**

391 Because of their ease of use and the fact that they employ safer chemicals than those in
392 conventional methods (i.e. phenol chloroform extraction), the use of commercial DNA extraction kits
393 has become standard, particularly in the context of studies dealing with the identification of tick
394 microbial communities. However, the issue of contaminants coming from DNA extraction kits has

395 recently been raised for microbiota studies dealing with low biomass samples, such as blood, intestinal
396 tissue, mucosal tissue, or nasopharyngeal samples. Our study adds further evidence supporting these
397 important concerns in the field of tick microbiota. Here, our results suggest that many negative
398 controls should be performed at each step of sample processing prior to performing microbiota
399 analyses, allowing us to efficiently identify a potentially serious bias in the study of the actual tick
400 microbial community. While more and more studies try to identify tick microbiota to in fine potentially
401 propose hypothesis for the development of new tick-borne pathogen control strategies, we strongly
402 advise the routine use of negative controls in microbiota studies, particularly in the context of low
403 biomass samples like ticks or other small arthropods. These procedures should be used routinely in
404 order to obtain the most reliable results possible.
405
406

Declarations

Competing interests

No conflicts of interest are declared.

Acknowledgements

The authors would like to thanks Sabine Delannoy and the IdentyPath genomic platform that allowed us to perform part of the experiments in their laboratory. Thank you also to Maxime Galan for simulating discussions and for providing index sequences.

Funding

This research work was funded by the French National Institute of Agronomic Research (INRA) through grants obtained from the Métaprogramme: “Adaptation of Agriculture and Forests to Climate Change” and “Meta-omics and Microbial Ecosystems”. The salary of Emilie Lejal, the PhD student working on this project was funded by the Ile-de-France region.

Authors’ contributions

Conceived and designed the experiments: TP, MVT, JFC, EL. Performed the experiments: EL. Analysed the data: EL, AE, TP, MM, MM, CM, OR. Wrote the paper: EL, AE, MM, JFC, OR, MM, CM, MVT, TP.

References

1. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proc Natl Acad Sci. 2006 Aug 8;103(32):12115–20.

2. Hamdan LJ, Coffin RB, Sikaroodi M, Greinert J, Treude T, Gillevet PM. Ocean currents shape the microbiome of Arctic marine sediments. ISME J. 2013 Apr;7(4):685–96.

434 3. Degli Esposti M, Martinez Romero E. The functional microbiome of arthropods. PLOS ONE. 2017
435 May 5;12(5):e0176573.

436 4. Estrada-Peña A, Cabezas-Cruz A, Pollet T, Vayssier-Taussat M, Cosson J-F. High Throughput
437 Sequencing and Network Analysis Disentangle the Microbial Communities of Ticks and Hosts
438 Within and Between Ecosystems. Front Cell Infect Microbiol [Internet]. 2018 [cited 2019 Mar
439 15];8. Available from: https://www.frontiersin.org/articles/10.3389/fcimb.2018.00236/full

440 5. Lalzar I, Harrus S, Mumcuoglu KY, Gottlieb Y. Composition and Seasonal Variation of
441 Rhipicephalus turanicus and Rhipicephalus sanguineus Bacterial Communities. Appl Env
442 Microbiol. 2012 Jun 15;78(12):4110–6.

443 6. Hawlena H, Rynkiewicz E, Toh E, Alfred A, Durden LA, Hastriter MW, et al. The arthropod, but not
444 the vertebrate host or its environment, dictates bacterial community composition of fleas and
445 ticks. ISME J. 2013 Jan;7(1):221–3.

446 7. Rynkiewicz EC, Hemmerich C, Rusch DB, Fuqua C, Clay K. Concordance of bacterial communities
447 of two tick species and blood of their shared rodent host. Mol Ecol. 2015;24(10):2566–79.

448 8. Van Treuren W, Ponnusamy L, Brinkerhoff RJ, Gonzalez A, Parobek CM, Juliano JJ, et al. Variation
449 in the Microbiota of Ixodes Ticks with Regard to Geography, Species, and Sex. Appl Env Microbiol.
450 2015 Sep 15;81(18):6200–9.

451 9. Gall CA, Reif KE, Scoles GA, Mason KL, Mousel M, Noh SM, et al. The bacterial microbiome of
452 Dermacentor andersoni ticks influences pathogen susceptibility. ISME J. 2016 Aug;10(8):1846–
453 55.

454 10. Khoo J-J, Chen F, Kho KL, Ahmad Shanizza AI, Lim F-S, Tan K-K, et al. Bacterial community in
455 Haemaphysalis ticks of domesticated animals from the Orang Asli communities in Malaysia. Ticks
456 Tick-Borne Dis. 2016 Jul 1;7(5):929–37.

457 11. Kwan JY, Griggs R, Chicana B, Miller C, Swei A. Vertical vs. horizontal transmission of the
458 microbiome in a key disease vector, Ixodes pacificus. Mol Ecol. 2017;26(23):6578–89.

459 12. Swei A, Kwan JY. Tick microbiome and pathogen acquisition altered by host blood meal. ISME J.
460 2017 Mar;11(3):813–6.

461 13. Andreotti R, Pérez de León AA, Dowd SE, Guerrero FD, Bendele KG, Scoles GA. Assessment of
462 bacterial diversity in the cattle tick Rhipicephalus (Boophilus) microplusthrough tag-encoded
463 pyrosequencing. BMC Microbiol. 2011 Jan 6;11(1):6.

464 14. Budachetri K, Browning RE, Adamson SW, Dowd SE, Chao C-C, Ching W-M, et al. An Insight Into
465 the Microbiome of the Amblyomma maculatum (Acari: Ixodidae). J Med Entomol. 2014 Jan
466 1;51(1):119–29.

467 15. Narasimhan S, Rajeevan N, Liu L, Zhao YO, Heisig J, Pan J, et al. Gut Microbiota of the Tick Vector
468 Ixodes scapularis Modulate Colonization of the Lyme Disease Spirochete. Cell Host Microbe. 2014
469 Jan 15;15(1):58–71.

470 16. Qiu Y, Nakao R, Ohnuma A, Kawamori F, Sugimoto C. Microbial Population Analysis of the Salivary
471 Glands of Ticks; A Possible Strategy for the Surveillance of Bacterial Pathogens. PLOS ONE. 2014
472 Aug 4;9(8):e103961.

473 17. Clayton KA, Gall CA, Mason KL, Scoles GA, Brayton KA. The characterization and manipulation of
474 the bacterial microbiome of the Rocky Mountain wood tick, Dermacentor andersoni. Parasit
475 Vectors. 2015 Dec 10;8(1):632.

476 18. Zolnik CP, Prill RJ, Falco RC, Daniels TJ, Kolokotronis S-O. Microbiome changes through ontogeny
477 of a tick pathogen vector. Mol Ecol. 2016;25(19):4963–77.

478 19. Abraham NM, Liu L, Jutras BL, Yadav AK, Narasimhan S, Gopalakrishnan V, et al. Pathogen-
479 mediated manipulation of arthropod microbiota to promote infection. Proc Natl Acad Sci. 2017
480 Jan 31;114(5):E781–90.

481 20. Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP, et al. Metagenomic Profile of the
482 Bacterial Communities Associated with Ixodes ricinus Ticks. PLOS ONE. 2011 Oct
483 13;6(10):e25604.

484 21. Moreno CX, Moy F, Daniels TJ, Godfrey HP, Cabello FC. Molecular analysis of microbial
485 communities identified in different developmental stages of Ixodes scapularis ticks from
486 Westchester and Dutchess Counties, New York. Environ Microbiol. 2006 May 1;8(5):761–72.

487 22. Heise SR, Elshahed MS, Little SE. Bacterial Diversity in Amblyomma americanum (Acari: Ixodidae)
488 With a Focus on Members of the Genus Rickettsia. J Med Entomol. 2010 Mar 1;47(2):258–68.

489 23. Galan M, Razzauti M, Bard E, Bernard M, Brouat C, Charbonnel N, et al. 16S rRNA Amplicon
490 Sequencing for Epidemiological Surveys of Bacteria in Wildlife. mSystems. 2016 Aug
491 30;1(4):e00032-16.

492 24. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory
493 contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014 Nov
494 12;12(1):87.

495 25. Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of
496 experimental contamination in microbiome studies. Genome Biol. 2014 Dec 17;15(12):564.

497 26. Narasimhan S, Fikrig E. Tick microbiome: the force within. Trends Parasitol. 2015 Jul 1;31(7):315–
498 23.

499 27. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of
500 extraction and sequencing reagents may affect interpretation of microbiota in low bacterial
501 biomass samples. Gut Pathog. 2016 May 26;8(1):24.

502 28. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for
503 analysing microbiomes. Nat Rev Microbiol. 2018 Jul;16(7):410.

504 29. Lejal E, Marsot M, Chalvet-Monfray K, Cosson J-F, Moutailler S, Vayssier-Taussat M, et al. A three-
505 years assessment of Ixodes ricinus-borne pathogens in a French peri-urban forest. bioRxiv. 2019
506 Apr 5;597013.

507 30. Andrews S. FastQC: a quality control tool for high throughput sequence data. [Internet]. 2010.
508 Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

509 31. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd
510 mergeR. Bioinformatics. 2014 Mar 1;30(5):614–20.

511 32. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
512 EMBnet.journal. 2011 May 2;17(1):10–2.

513 33. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. FROGS: Find, Rapidly, OTUs
514 with Galaxy Solution. Bioinformatics. 2018 Apr 15;34(8):1287–94.

515 34. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm v2: highly-scalable and high-
516 resolution amplicon clustering. PeerJ. 2015 Dec 10;3:e1420.

517 35. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for
518 metagenomics. PeerJ. 2016 Oct 18;4:e2584.

519 36. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene
520 database project: improved data processing and web-based tools. Nucleic Acids Res. 2013 Jan
521 1;41(D1):D590–6.

522 37. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing
523 on the Illumina platform. Nucleic Acids Res. 2012 Jan 1;40(1):e3–e3.

524 38. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools
525 for integrating phylogenies and ecology. Bioinformatics. 2010 Jun 1;26(11):1463–4.

526 39. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and
527 Graphics of Microbiome Census Data. PLOS ONE. 2013 Apr 22;8(4):e61217.

528 40. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna,
529 Austria: R Foundation for Statistical Computing; 2018. Available from: https://www.R-
530 project.org/

531