



HAL
open science

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

Emilie Lejal, Agustin Estrada-Peña, Maud Marsot, Jean-Francois J.-F. Cosson, Olivier Rué, Mahendra Mariadassou, Cedric Midoux, Muriel Vayssier Taussat, Thomas Pollet

► To cite this version:

Emilie Lejal, Agustin Estrada-Peña, Maud Marsot, Jean-Francois J.-F. Cosson, Olivier Rué, et al.. Taxon appearance from extraction and amplification steps demonstrates the value of multiple controls in tick microbiome analysis. *Frontiers in Microbiology*, 2020, 11, 10.3389/fmicb.2020.01093 . hal-02790980

HAL Id: hal-02790980

<https://hal.inrae.fr/hal-02790980>

Submitted on 15 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Taxon Appearance From Extraction and Amplification Steps Demonstrates the Value of Multiple Controls in Tick Microbiota Analysis

Emilie Lejal¹, Agustín Estrada-Peña², Maud Marsot³, Jean-François Cosson¹, Olivier Rué^{4,5}, Mahendra Mariadassou^{4,5}, Cédric Midoux^{4,5,6}, Muriel Vayssier-Taussat⁷ and Thomas Pollet^{1,8*}

¹ UMR BIPAR, Animal Health Laboratory, INRAE, ANSES, Ecole Nationale Vétérinaire d'Alfort, Université Paris-Est, Maisons-Alfort, France, ² Faculty of Veterinary Medicine, University of Zaragoza, Zaragoza, Spain, ³ Laboratory for Animal Health, Epidemiology Unit, ANSES, University Paris-Est, Maisons-Alfort, France, ⁴ INRAE, MaAGE, Université Paris-Saclay, Jouy-en-Josas, France, ⁵ INRAE, Bioinformatics, MIGALE Bbioinformatics Facility, Université Paris-Saclay, Jouy-en-Josas, France, ⁶ INRAE, PROSE, Université Paris-Saclay, Antony, France, ⁷ Animal Health Department, INRAE, Nouzilly, France, ⁸ UMR ASTRE, CIRAD, INRAE, Montpellier, France

OPEN ACCESS

Edited by:

Guillermina Hernandez-Raquet,
Institut National de la Recherche
Agronomique et de l'Environnement
(INRAE), France

Reviewed by:

Charlotte Louise Oskam,
Murdoch University, Australia
Michelle Power,
Macquarie University, Australia
Jingze Liu,
Hebei Normal University, China

*Correspondence:

Thomas Pollet
thomas.pollet@inrae.fr

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 20 August 2019

Accepted: 01 May 2020

Published: 09 June 2020

Citation:

Lejal E, Estrada-Peña A, Marsot M, Cosson J-F, Rué O, Mariadassou M, Midoux C, Vayssier-Taussat M and Pollet T (2020) Taxon Appearance From Extraction and Amplification Steps Demonstrates the Value of Multiple Controls in Tick Microbiota Analysis. *Front. Microbiol.* 11:1093. doi: 10.3389/fmicb.2020.01093

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: homogenization, DNA extraction, and DNA amplification.

Results: Controls yielded a significant number of sequences (1,126–13,198 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 instar and sex: 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 homogenization and extraction reagent controls, highlighting the importance of carefully controlling these steps.

Conclusion: Here, we showed that contaminant OTUs from sample laboratory processing 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 metabarcoding represents a powerful tool commonly used to evaluate the prokaryotic diversity in many ecosystems. High-throughput sequencing approaches allow 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 metabarcoding 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; René-Martellet et al., 2017; Swei and Kwan, 2017; Estrada-Peña et al., 2018; Thapa et al., 2019) 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 instar and sex (Lalzar et al., 2012; Van Treuren et al., 2015; Zolnik et al., 2016; René-Martellet et al., 2017; Swei and Kwan, 2017; Thapa et al., 2019), 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 extraction/amplification steps or from the environment during sample processing (Salter et al., 2014; Strong et al., 2014; Weiss et al., 2014; Narasimhan and Fikrig, 2015; Galan et al., 2016; Glassing et al., 2016; Greay et al., 2018; Knight et al., 2018). This potential contamination is able to significantly skew results concerning microbe diversity and composition, particularly in the case of low biomass samples, and can have serious consequences on the interpretation of results. For example, while investigating 16S rRNA gene profiles from nasopharyngeal swabs regularly sampled from a cohort of children from 1 to 24 months, Salter et al. (2014) demonstrated that the apparent clustering of samples by age, observed by principal coordinate analysis, in fact corresponded to the commercial kit used to perform the DNA extraction of these samples. While studying the influence of laboratory processing on diluted samples, these authors also observed that the lower the initial biomass of samples, the higher the proportion of contaminants associated with the kits or the environment. In the context of tick microbiota studies, performing negative controls would therefore be essential to detect potential contamination, and efficiently distinguish this contamination from genuine tick microbial communities. On the basis of a large review of the scientific literature, it is clear that routine multiple control analyses (extraction and PCR negative controls) are not considered at their true value, and are only rarely implemented (i.e., sequenced and used to filter contaminants from the dataset) in tick microbial community analyses (Hawlena et al., 2013; Gofton et al., 2015a,b; Rynkiewicz et al., 2015; René-Martellet et al., 2017; Sperling et al., 2017; Estrada-Peña et al., 2018; Thapa et al., 2019). In this context, we performed a 16S rRNA gene sequencing analysis on individual ticks (females, males, and nymphs) considering concomitantly several negative controls at each step of the process, prior to the sequencing run (tick homogenization, DNA extraction, and amplification).

MATERIALS AND METHODS

Tick Collection and Sample Selection

Questing *Ixodes ricinus* nymphs and adults were collected for 3 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.

Abbreviations: AC, amplification control; EC, extraction control; HC, homogenization control; OTU, operational taxonomic unit.

Tick Washing, Homogenization, and DNA Extraction

Ticks were first washed once in ethanol 70% for 5 min and rinsed twice in sterile MilliQ water for 5 min each time. Ticks were then individually homogenized in 375 μ L of Dulbecco's Modified Eagle Medium with decomplexed fetal calf serum (10%) and six steel beads using the homogenizer Precellys®24 Dual (Bertin, France) at 5,500 rpm for 20 s. DNA extraction was performed on 100 μ L of tick homogenate, 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, United States) before and after each use. Each time a tick homogenization or DNA extraction was performed, a homogenization control (HC) or extraction control (EC) was added for the corresponding step. Three positives controls, corresponding to a microbial community standard (D6300 ZymoBIOMICS Microbial Community Standard - Zymo Research, United States), were also extracted in the same way.

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 compared to prokaryotic DNA. An 8 bp-index was added to the forward and reverse primers. The individual amplification of each sample was performed using a combination of 22 and 32 index-tagged forward and reverse primers, allowing for amplification and multiplexing of a maximum of 704 samples that can be loaded together on an Illumina MiSeq flow cell. Amplification and multiplexing steps were performed under a sterile hood, decontaminated before and after use with DNA remover (Molecular BioProducts, San Diego, CA, United States) and UVs, using tube strips with individual caps to prevent cross-contamination. All the PCR amplifications were carried out using the Phusion High-Fidelity® DNA Polymerase amplification kit (Thermo Fisher Scientific, Lithuania). For each sample, 5 μ L of DNA extract were amplified in a 50 μ L final reaction volume, containing 1 \times Phusion HF buffer, 0.2 μ M of dNTPs, 0.2 U/mL of Phusion DNA polymerase, and 0.35 μ M of forward and reverse primer. The following thermal cycling procedure was used: initial denaturation 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 extension at 72°C for 30 s. The final extension was carried out at 72°C for 10 min. Amplification step was performed on a peqSTAR 2 \times thermocycler (PEQLAB, Germany). For

each PCR run, negative controls were performed by using the reaction mixture without template, but using ultrapure water (Invitrogen, United Kingdom). PCR products were checked on 1.5% agarose gels.

DNA Purification, Quantification, and Pooling

Amplicons were purified and quantified using a SequelPrep normalization plate kit (Invitrogen, United States), and a Qubit dsDNA High Sensitivity Assay kit (Invitrogen, United States), respectively. Amplicons and controls were pooled at equimolar concentrations and sent to the sequencing platform (GenoScreen, France).

Negative Controls

In total, 45 negative controls were included and analyzed using the same process as for tick samples (**Figure 1**).

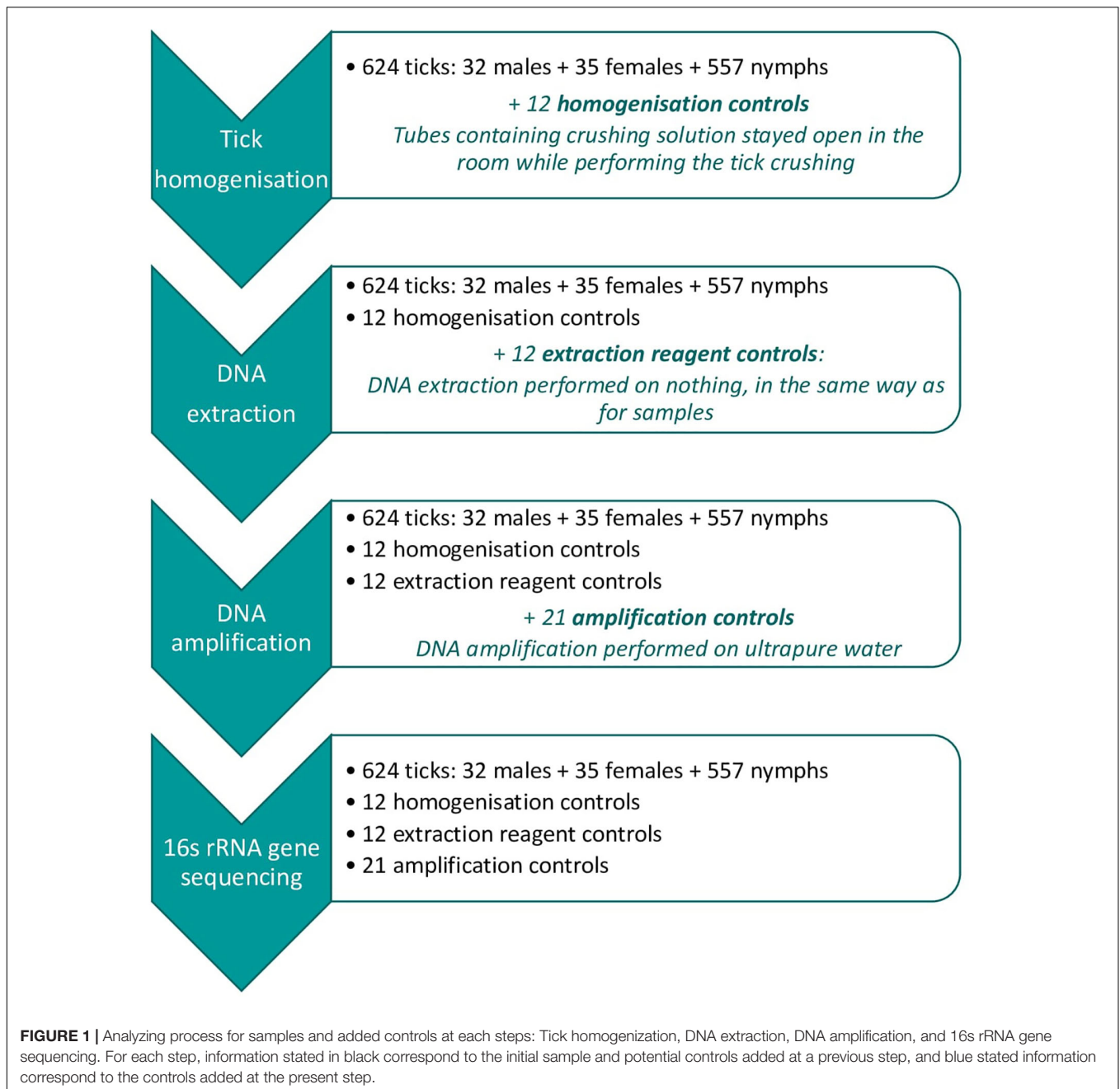
Twelve Homogenization controls (HCs): Tube containing homogenization solution, remained open in the room while putting ticks inside other tubes between the cleaning and homogenization steps. DNA extraction and amplification were then performed on the homogenization control in the same conditions as for any other sample.

Twelve Extraction reagent controls (ECs): For these controls, the different DNA extraction steps were performed under the same conditions as for any other samples, but without adding sample. DNA amplification was then performed on the extraction control in the same conditions as for any other sample.

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

Sequencing and Data Processing

The equimolar mix was concentrated and sequenced by GenoScreen (Lille, France) using MiSeq Illumina 2 \times 250 bp chemistry. Quality control checks on raw data were performed using FastQC v0.11.3 (Andrews, 2010), and no problems were detected. First, pairs of amplicons were merged using Pear v0.9.11 (Zhang et al., 2014) with a minimum overlap of 20 bp, a minimum assembly length of 100 bp, a maximum assembly length of 480 bp, and an e-value of 0.01. Then, the sequencing orientation of extended sequences was checked by using cutadapt v1.12 (Martin, 2011), with primers in forward and reverse strands (without removing them). Sequences with a reversed orientation were reverse-complemented to obtain a set of sequences with the same orientation (5' to 3'). The sequence analysis was performed using FROGS (Escudie et al., 2018). Primers were removed using cutadapt and resulting sequences were dereplicated. Sequences smaller or longer than 250 bp \pm 20 bp, or containing N-nucleotides were removed from the dataset. Operational taxonomic units (OTUs) were built using Swarm v2.1.12 (Mahé et al., 2015) with parameter $d = 3$, and chimeras were filtered using vsearch v1.4 (Rognes et al., 2016) in a *de novo* mode. Finally, OTUs with low abundance (<0.005% of the total sequence count) were filtered out. OTUs were affiliated by blasting cluster seed



sequences against the Silva database v132 (Quast et al., 2013). OTUs corresponding to mitochondrial/chloroplast DNA or not affiliated to any domain were removed.

Data Processing to Distinguish Contaminant Versus Tick Microbiota OTUs

The control composition was used to distinguish tick microbial OTUs from contaminants. All the OTUs corresponding to more than 1% of the total number of sequences obtained in each category of control were identified as main contaminants.

However, due to possible errors in sequence assignment occurring during sequencing due to the formation of mixed clusters on the flow cell (Kircher et al., 2012), negative controls can also contain a few sequences of tick microbial OTUs, complicating contaminant identification under this 1% threshold. Therefore, for each OTU, the number of sequences expected to spread to other samples due to false assignment (i.e., expected number) was determined by calculating 0.02% of the total number of each OTU in the dataset, corresponding to a false assignment rate estimated previously by Galan et al. (2016). This expected number was compared to the maximum number of obtained sequences in controls (i.e., maximum number).

As this rate was not estimated from our dataset, we decided to allow variability around this threshold, while calculating a 99% confidence interval around the expected number of 0.02%. OTUs whose maximum sequence number was above the upper threshold were considered contaminants. Please note that cross-contaminations from samples to controls were considered negligible. Indeed, no negative controls presented a higher maximum sequence number than the expected number for an OTU belonging to the most abundant genus of tick microbiota in our dataset (Ca. *Midichloria*), which is also the 3rd most abundant OTU in the dataset (266,454 sequences).

Statistical Analyses

The percentage of samples presenting a contamination rate higher than 50% was compared between samples categories (nymphs, males, and females) according to a Chi-square statistical test.

Alpha diversity was estimated before and after contaminating OTUs filtering. Samples containing fewer than 500 sequences after contaminant filtering were removed from the analysis, even if they contained more than 500 sequences before the filtering. Several measures of alpha diversity were calculated. Basically, the observed number of OTUs (species richness) was used as well as the Faith phylogenetic diversity (PD) index, corresponding to the total branch length of the subtree spanned by the community, and also two indexes taking into account the number of sequences: the Shannon index (more influenced by rare OTUs) and Simpson index (more influenced by dominant OTUs). Phylogenetic diversity was estimated using the Picante package (Kembel et al., 2010), while other measures were assessed thanks to Easy 16S,¹ based on the phyloseq package (McMurdie and Holmes, 2013). Alpha diversity measure comparison was performed between contaminant filtered and not filtered groups, as well as between instars, using non-parametric statistical tests (Wilcoxon and Kruskal–Wallis, respectively). Statistical computations were performed in R 3.6.0 (R Core Team, 2018).

RESULTS

After sequencing, demultiplexing and merging amplicons, we obtained a total of 4,767,043 sequences. The impact of the different steps of the data processing: (1) merging filters, (2) FROGS pre-process, (3) Swarm clustering, (4) Chimera removing, and (5) filtering of rare OTUs are described for each sample in the **Supplementary Table S1**. After performing all these steps, we obtained a dataset composed of 608 OTUs for 2,804,545 sequences. After the affiliation step using the SILVA data base, OTUs corresponding to mitochondrial/chloroplast DNA or not affiliated to any domain, have been removed, and we finally obtained a dataset composed of 513 OTUs and 2,695,360 sequences.

Positive controls (microbial community standards) presented a mean number of sequences of 5,702 (± 634) and a mean number of OTUs of 60 (± 9). All the expected micro-organisms were

detected and 98.1% (± 1.1) of the sequences were attributed to the microbial community standard.

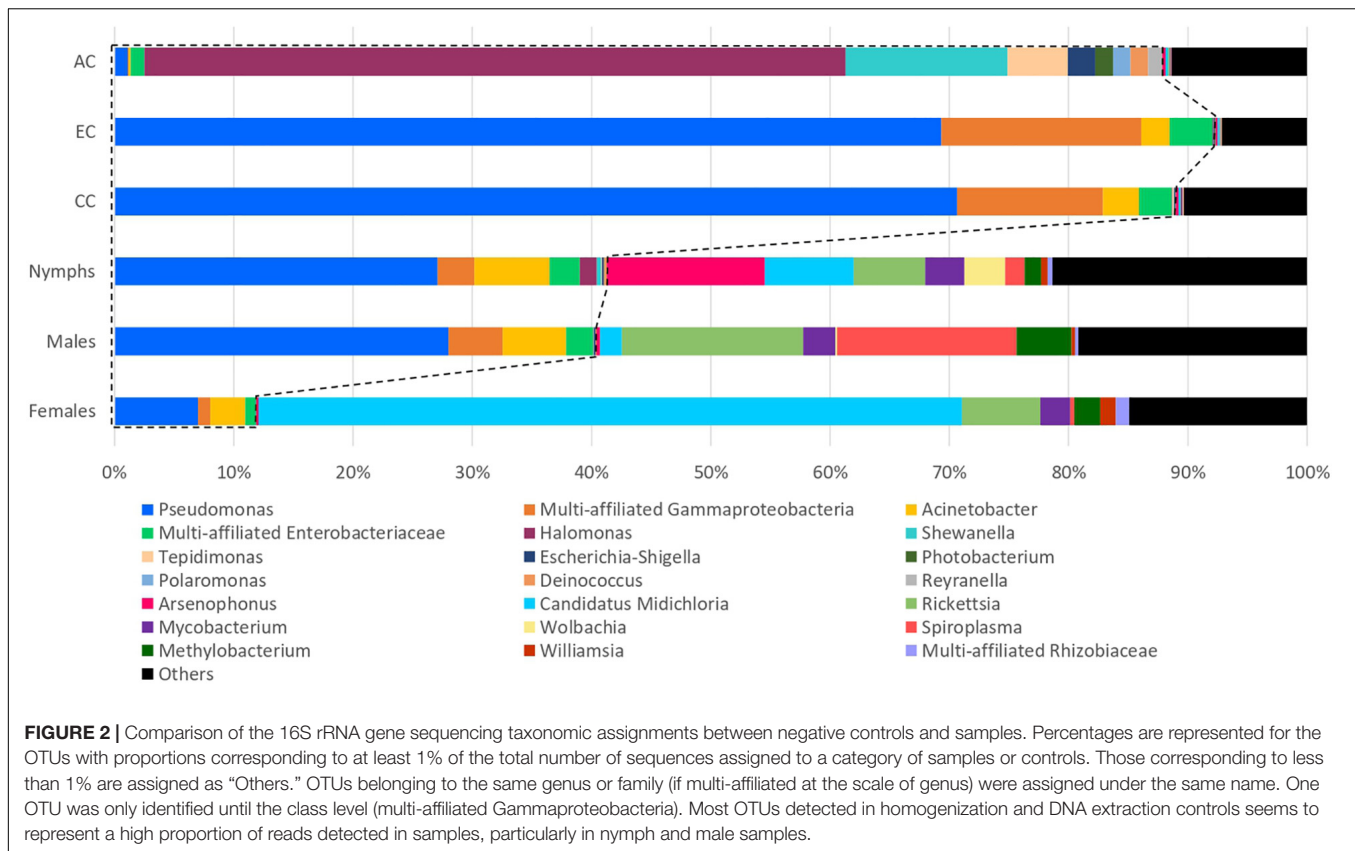
Nymphs, males and females presented a mean number of sequences per sample of 3,721 ($\pm 3,017$), 2,689 ($\pm 1,762$) and 5,316 ($\pm 2,300$), and a mean number of OTUs per samples of 114 (± 30), 107 (± 22), 107 (± 29), respectively. Concerning negative controls, a mean number of 13,198 ($\pm 3,219$), 11,235 ($\pm 2,742$), and 1,126 ($\pm 1,904$) sequences per sample and a mean number of 123 (± 23), 127 (± 14), and 26 (± 11) were obtained for HCs, Ecs, and ACs, respectively.

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

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

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 contaminant filtering steps, 201 samples contained less than 500 sequences compared to 8 before filtering. For 76.6% of nymph and 75% of male samples, contaminant OTUs represented more than 50% of the total

¹<http://genome.jouy.inra.fr/shiny/easy16S/>



sequence count (**Figure 3**). These sample contamination rates are significantly higher than those obtained for females ($p < 0.05$ according to a Chi-square test), for which most of the samples (65.7%) exhibited a contamination rate below 20%. The impact of this contamination on diversity, according to the sample category was estimated (**Figure 4**). As expected, for all the studied diversity measures, contaminant filtering 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 contaminant filtering on alpha diversity comparison between tick instars was been studied. Without filtering, significant differences in alpha diversity were found between tick instars for observed, Shannon and Simpson measures. Concerning filtered data, significant differences were still observed between instars when using Shannon and Simpson indexes, but disappeared when using the observed diversity measure.

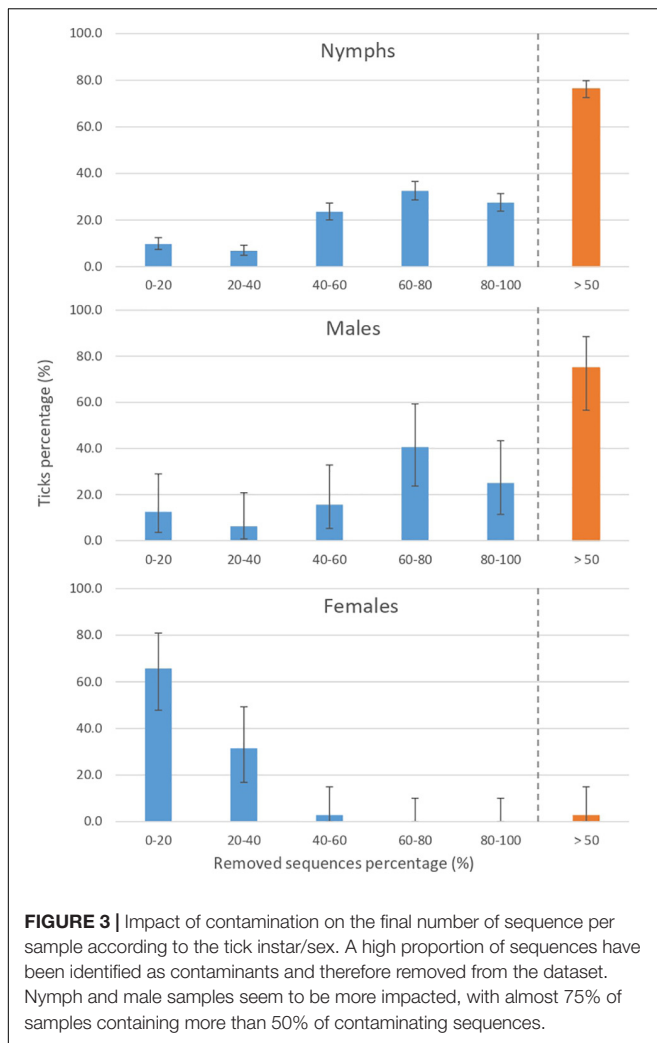
DISCUSSION

In this study, we sequenced the microbiota of 624 individual *I. ricinus* ticks including several categories of negative controls to identify and quantify the contaminating sequences, in particular those produced during the extraction or amplification steps. The

performance of multiple negative controls allowed us to perform in-depth identification of contaminant OTUs. These OTUs represented 50.9% of the total sequence yield from tick samples, and had a significant impact on microbial diversity estimation. This shows the importance of contamination potentially arising from processing steps, and demonstrates that its consequences on data analyses can be significant and should not be neglected.

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

Each processing step leads to a different contamination pattern. Most of the main contaminants do not seem to originate from the DNA amplification step. Furthermore, as the composition of homogenization and extraction controls was the same, while the reagents used were different and homogenization control also underwent the extraction step, the contaminant sequences most probably came from the DNA extraction step, rather than from the homogenization step. The key role of the

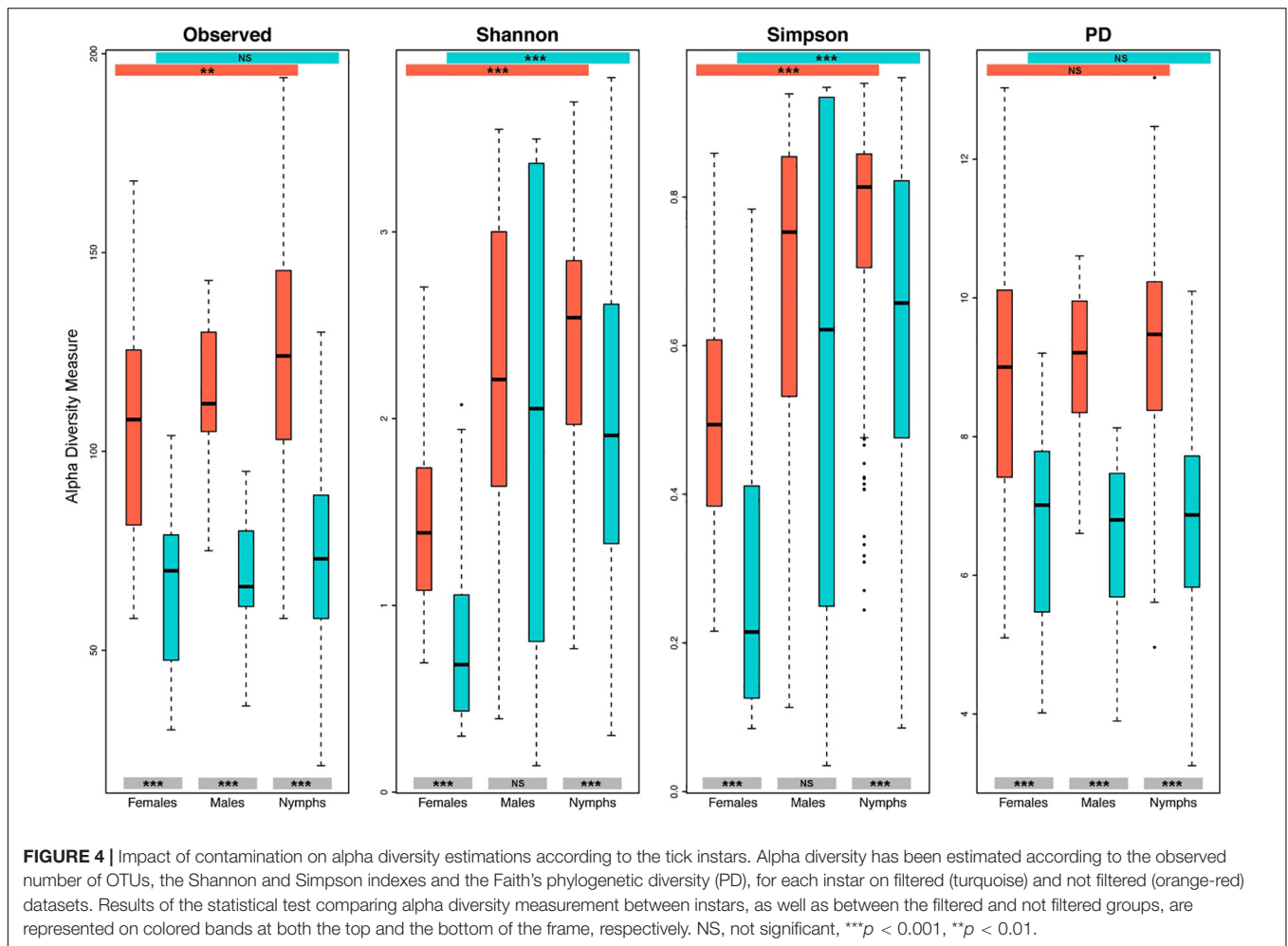


extraction step in the contamination of low biomass samples has previously been demonstrated by Salter et al. (2014). Furthermore, three of the four main contaminants identified in our analysis, corresponding to *Pseudomonas*, *Acinetobacter*, and multi-affiliated Enterobacteriaceae, belong to families and genera identified by these authors as contaminants coming from DNA extraction kits. Our observations differed from those obtained by Galan et al. (2016), who mainly detected contaminants coming from the DNA amplification step, while this step represented only 2% of the main contaminants detected in our samples. In their study of kit contaminants, Salter et al. (2014) observed that the extent of contamination arising from DNA extraction was different depending on the kit brand used to perform this step. While using the less contaminating kit, they extracted DNA from several 10-fold cascade dilutions of a pure suspension of *Salmonella bongori*, and subsequently assessed 16S rRNA gene concentrations via q-PCR. They demonstrated a linear relationship between the 16S rRNA gene copy number and the dilution fold only for the first dilution points, after which levels started to stabilize and to reach a threshold of 10^5 mean copy number/mL. Consequently, the higher rate of extraction, rather

than amplification contaminants in samples, could be due either to variations of contaminating load between our extraction and amplification kits, that are not sourced from the same company, or to the presence of a basal level of extraction contaminant bacterial DNA, potentially limiting the contamination arising from the next step (amplification). These types of contamination seem to be the main in our dataset. However, some other sources of external contamination, such as those potentially coming from the dragged material or inherent to 16s metabarcoding (i.e., cross talk during the sequencing run, leading to false assignment of sequences between samples) could be considered. False assignment rate can be assessed through the use of an alien sequence (Galan et al., 2016). Such a sequence was not used in our dataset. Nonetheless, if false assignments would have highly affected our data set, we probably would have detected high proportion of *Ca. Midichloria* (the most abundant known tick microbial member) sequences in our negative controls, that would have led to its identification as a contaminant. Even if we are aware that dragged material or false assignments might be another source of contamination, we hypothesize, based on our results, that they would be probably much weaker than those linked to the extraction steps and would represent a low percentage of the total number of sequences.

Interestingly, we showed that in terms of contamination rates, the influence of contaminant presence was significantly higher in nymph and male samples compared to female ones. Because low biomass samples were shown to be particularly sensitive to contamination (Salter et al., 2014), we might suppose that the amount of bacterial DNA would be lower in males and nymphs, explaining the higher presence of contaminants in these samples. It would be useful to develop epifluorescence or flow cytometry protocols adapted to tick samples, to assess whether the amount of bacteria composing the female microbiota is effectively higher than in the male and nymph. Additionally, previous studies dealing with tick microbes have identified higher microbiota diversity in males and nymphs compared to females, while investigating richness and PD (Van Treuren et al., 2015; Zolnik et al., 2016) as well as Shannon and Simpson indexes (Zolnik et al., 2016). However, no negative controls and data filtering from contaminants were reported in these studies, and they could therefore be impacted by contamination. In the present study, we determined that the presence of contaminants can lead to an overestimation of microbial diversity in all the tick instars, and that this is the case for several diversity indexes, especially richness and PD estimations, which are the measures most affected by the filtering process. Furthermore, the impact of removing contaminants on diversity comparisons between instars was estimated and showed that this filtering step can influence the final results, and particularly in our case, concerning the observed diversity estimation. Our results therefore call for caution in attempts to draw conclusion about tick instar diversity without considering negative controls.

As already mentioned by Narasimhan and Fikrig (2015), the understanding of arthropod microbial communities in the context of arthropod survival and pathogen transmission may open new routes in arthropod-borne pathogen control strategies. In this context, many studies have assessed tick microbiota



composition in individual ticks or even in tick organs (Andreotti et al., 2011; Lalar et al., 2012; Hawlena et al., 2013; Budachetri et al., 2014; Narasimhan et al., 2014; Qiu et al., 2014; Clayton et al., 2015; Rynkiewicz et al., 2015; Van Treuren et al., 2015; Gall et al., 2016; Khoo et al., 2016; Zolnik et al., 2016; Abraham et al., 2017; Kwan et al., 2017; Swei and Kwan, 2017; Estrada-Peña et al., 2018). However, very few reported the use of negative controls or mentioned the identification and removal of subsequently identified contaminants (Hawlena et al., 2013; Gofton et al., 2015a,b; Rynkiewicz et al., 2015; René-Martellet et al., 2017; Sperling et al., 2017; Estrada-Peña et al., 2018; Thapa et al., 2019), thus obscuring the risk that identified micro-organisms as members of the microbiota could correspond to contaminants. Nonetheless, the bias generated by the presence of contaminants can have a considerable influence on the characterization of tick microbiota composition and structure. Here, we showed that contaminants represented almost 50.9% of the total sequence yield of tick samples. Furthermore, several contaminants we identified belong to genera (especially *Pseudomonas* and *Acinetobacter*) previously reported in several tick microbiota studies (reviewed by Narasimhan and Fikrig, 2015). This does not necessarily mean that all OTUs belonging

to these genera are contaminants, as several OTUs belonging to the same genera may correspond either to tick microbiota members or to contaminants. However, this points out the lack of reliability concerning this crucial point in the understanding of tick microbiota composition, and the need to perform negative controls to filter contaminants from the data set and to rule out the hypothesis of potential contamination.

CONCLUSION

The use of commercial DNA extraction kits has become standard, particularly in the context of studies dealing with the identification of tick microbial communities, notably because of their ease of use and the fact that they employ safer chemicals than those in conventional methods (i.e., phenol chloroform extraction). However, the issue of contaminants coming from DNA extraction kits has recently been raised for microbiota studies dealing with low biomass samples, such as blood, intestinal tissue, mucosal tissue, or nasopharyngeal samples. Our study adds further evidence supporting these important concerns in the field of tick microbiota. Here, our results suggest that many

negative controls should be performed at each step of sample processing prior to performing microbiota analyses, allowing us to efficiently identify a potentially serious bias in the study of the actual tick microbial community. While more and more studies try to identify tick microbiota to *in fine* potentially propose hypothesis for the development of new tick-borne pathogen control strategies, we strongly advise the routine use of negative controls in microbiota studies, particularly in the context of low biomass samples like ticks or other small arthropods. These procedures should be used routinely in order to obtain the most reliable results possible.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the European Nucleotide Archive. Project accession number: PRJEB36903 (ERP120162) – Sample accession numbers: ERS4353953–ERS4354625 – Run accession numbers: ERR3956669–ERR3957340.

AUTHOR CONTRIBUTIONS

TP, MV-T, J-FC, and EL conceived and designed the experiments. EL performed the experiments. EL, AE-P, TP, MMs, MMi, CM, and OR analyzed the data. EL, AE-P, MMs, J-FC, OR, MMi, CM, MV-T, and TP wrote the manuscript. All the authors read and approved the final manuscript.

REFERENCES

- Abraham, N. M., Liu, L., Jutras, B. L., Yadav, A. K., Narasimhan, S., Gopalakrishnan, V., et al. (2017). Pathogen-mediated manipulation of arthropod microbiota to promote infection. *Proc. Natl. Acad. Sci. U.S.A.* 114, E781–E790. doi: 10.1073/pnas.1613422114
- Andreotti, R., Pérez de León, A. A., Dowd, S. E., Guerrero, F. D., Bendele, K. G., and Scoles, G. A. (2011). Assessment of bacterial diversity in the cattle tick *Rhipicephalus (Boophilus) microplus* through tag-encoded pyrosequencing. *BMC Microbiol.* 11:6. doi: 10.1186/1471-2180-11-6
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Berlin: Springer.
- Budachetri, K., Browning, R. E., Adamson, S. W., Dowd, S. E., Chao, C.-C., Ching, W.-M., et al. (2014). An insight into the microbiome of the *Amblyomma maculatum* (Acari: Ixodidae). *J. Med. Entomol.* 51, 119–129. doi: 10.1603/me12223
- Carpi, G., Cagnacci, F., Wittekindt, N. E., Zhao, F., Qi, J., Tomsho, L. P., et al. (2011). Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS One* 6:e25604. doi: 10.1371/journal.pone.0025604
- Clayton, K. A., Gall, C. A., Mason, K. L., Scoles, G. A., and Brayton, K. A. (2015). The characterization and manipulation of the bacterial microbiome of the Rocky Mountain wood tick, *Dermacentor andersoni*. *Parasit Vect.* 8:632. doi: 10.1186/s13071-015-1245-z
- Degli Esposti, M., and Martinez Romero, E. (2017). The functional microbiome of arthropods. *PLoS One* 12:e0176573. doi: 10.1371/journal.pone.0176573
- Escudé, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., et al. (2018). FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics* 34, 1287–1294. doi: 10.1093/bioinformatics/btx791
- Estrada-Peña, A., Cabezas-Cruz, A., Pollet, T., Vayssier-Taussat, M., and Cosson, J.-F. (2018). High throughput sequencing and network analysis disentangle

FUNDING

This work was supported by the Métaprogramme “Metaomics and Microbial Ecosystems” (MEM) and the Métaprogramme “Adaptation of Agriculture and Forests to Climate Change” (ACCAF) granted by the French National Institute for Agricultural Research (France). The salary of EL, the Ph.D. student working on this project was funded by the Ile-de-France region.

ACKNOWLEDGMENTS

The authors would like to thank Sabine Delannoy and the IdentityPath genomic platform that allowed us to perform part of the experiments in their laboratory. The authors also thank Maxime Galan for providing index sequences and stimulating discussions. Thank you as well to Ladislav Šimo for his precious help in figure improvement.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01093/full#supplementary-material>

TABLE S1 | Impact of the different steps of the data processing on each sample.

- the microbial communities of ticks and hosts within and between ecosystems. *Front. Cell Infect. Microbiol.* 8:236. doi: 10.3389/fcimb.2018.00236
- Galan, M., Razzauti, M., Bard, E., Bernard, M., Brouat, C., Charbonnel, N., et al. (2016). 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife. *mSystems* 1:e0032-16. doi: 10.1128/mSystems.00032-16
- Gall, C. A., Reif, K. E., Scoles, G. A., Mason, K. L., Mousel, M., Noh, S. M., et al. (2016). The bacterial microbiome of *Dermacentor andersoni* ticks influences pathogen susceptibility. *ISME J.* 10, 1846–1855. doi: 10.1038/ismej.2015.266
- Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., and Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* 8:24. doi: 10.1186/s13099-016-0103-7
- Gofton, A. W., Doggett, S., Ratchford, A., Oskam, C. L., Paparini, A., Ryan, U., et al. (2015a). Bacterial profiling reveals novel “Ca. Neoehrlichia”, Ehrlichia, and *Anaplasma* species in Australian Human-biting ticks. *PLoS One* 10:e0145449. doi: 10.1371/journal.pone.0145449
- Gofton, A. W., Oskam, C. L., Lo, N., Beninati, T., Wei, H., McCarl, V., et al. (2015b). Inhibition of the endosymbiont “*Candidatus Midichloria mitochondrii*” during 16S rRNA gene profiling reveals potential pathogens in *Ixodes* ticks from Australia. *Parasit Vect.* 8:345. doi: 10.1186/s13071-015-0958-3
- Greay, T. L., Gofton, A. W., Paparini, A., Ryan, U. M., Oskam, C. L., and Irwin, P. J. (2018). Recent insights into the tick microbiome gained through next-generation sequencing. *Parasit Vect.* 11:12. doi: 10.1186/s13071-017-2550-5
- Hamdan, L. J., Coffin, R. B., Sikaroodi, M., Greinert, J., Treude, T., and Gillevet, P. M. (2013). Ocean currents shape the microbiome of Arctic marine sediments. *ISME J.* 7, 685–696. doi: 10.1038/ismej.2012.143
- Hawlena, H., Rynkiewicz, E., Toh, E., Alfred, A., Durden, L. A., Hastriter, M. W., et al. (2013). The arthropod, but not the vertebrate host or its environment, dictates bacterial community composition of fleas and ticks. *ISME J.* 7, 221–223. doi: 10.1038/ismej.2012.71

- Heise, S. R., Elshahed, M. S., and Little, S. E. (2010). Bacterial diversity in *Amblyomma americanum* (Acari: Ixodidae) with a focus on members of the genus rickettsia. *J. Med. Entomol.* 47, 258–268. doi: 10.1603/med.09197
- Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464. doi: 10.1093/bioinformatics/btq166
- Khoo, J.-J., Chen, F., Kho, K. L., Ahmad Shanizza, A. I., Lim, F.-S., Tan, K.-K., et al. (2016). Bacterial community in Haemaphysalis ticks of domesticated animals from the Orang Asli communities in Malaysia. *Ticks Tick Borne Dis.* 7, 929–937. doi: 10.1016/j.ttbdis.2016.04.013
- Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40:e3. doi: 10.1093/nar/gkr771
- Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., et al. (2018). Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16:410. doi: 10.1038/s41579-018-0029-9
- Kwan, J. Y., Griggs, R., Chicana, B., Miller, C., and Swei, A. (2017). Vertical vs. horizontal transmission of the microbiome in a key disease vector, *Ixodes pacificus*. *Mol. Ecol.* 26, 6578–6589. doi: 10.1111/mec.14391
- Lalzar, I., Harrus, S., Mumcuoglu, K. Y., and Gottlieb, Y. (2012). Composition and seasonal variation of *Rhipicephalus turanicus* and *Rhipicephalus sanguineus* bacterial communities. *Appl. Env. Microbiol.* 78, 4110–4116. doi: 10.1128/AEM.00323-12
- Lejal, E., Marsot, M., Chalvet-Monfray, K., Cosson, J.-F., Moutailler, S., Vayssier-Taussat, M., et al. (2019). A three-years assessment of Ixodes ricinus-borne pathogens in a French peri-urban forest. *bioRxiv* [Preprint]. doi: 10.1101/597013
- Mahé, F., Rognes, T., Quince, C., Vargas, C., and de Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3:e1420. doi: 10.7717/peerj.1420
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12.
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Moreno, C. X., Moy, F., Daniels, T. J., Godfrey, H. P., and Cabello, F. C. (2006). Molecular analysis of microbial communities identified in different developmental stages of *Ixodes scapularis* ticks from Westchester and Dutchess Counties, New York. *Environ. Microbiol.* 8, 761–772. doi: 10.1111/j.1462-2920.2005.00955.x
- Narasimhan, S., and Fikrig, E. (2015). Tick microbiome: the force within. *Trends Parasitol.* 31, 315–323. doi: 10.1016/j.pt.2015.03.010
- Narasimhan, S., Rajeevan, N., Liu, L., Zhao, Y. O., Heisig, J., Pan, J., et al. (2014). Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the Lyme disease spirochete. *Cell Host Microbe* 15, 58–71. doi: 10.1016/j.chom.2013.12.001
- Qiu, Y., Nakao, R., Ohnuma, A., Kawamori, F., and Sugimoto, C. (2014). Microbial population analysis of the salivary glands of ticks; a possible strategy for the surveillance of bacterial pathogens. *PLoS One* 9:e103961. doi: 10.1371/journal.pone.0103961
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- René-Martellet, M., Minard, G., Massot, R., Tran Van, V., Valiente Moro, C., Chabanne, L., et al. (2017). Bacterial microbiota associated with *Rhipicephalus sanguineus* (s.l.) ticks from France, Senegal and Arizona. *Parasit Vect.* 10:416. doi: 10.1186/s13071-017-2352-9
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584
- Rynkiewicz, E. C., Hemmerich, C., Rusch, D. B., Fuqua, C., and Clay, K. (2015). Concordance of bacterial communities of two tick species and blood of their shared rodent host. *Mol. Ecol.* 24, 2566–2579. doi: 10.1111/mec.13187
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z
- Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.”. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103
- Sperling, J. L., Silva-Brandão, K. L., Brandão, M. M., Lloyd, V. K., Dang, S., Davis, C. S., et al. (2017). Comparison of bacterial 16S rRNA variable regions for microbiome surveys of ticks. *Ticks Tick Borne Dis.* 8, 453–461. doi: 10.1016/j.ttbdis.2017.02.002
- Strong, M. J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., et al. (2014). Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. *PLoS Pathog.* 10:e1004437. doi: 10.1371/journal.ppat.1004437
- Swei, A., and Kwan, J. Y. (2017). Tick microbiome and pathogen acquisition altered by host blood meal. *ISME J.* 11, 813–816. doi: 10.1038/ismej.2016.152
- Thapa, S., Zhang, Y., and Allen, M. S. (2019). Effects of temperature on bacterial microbiome composition in *Ixodes scapularis* ticks. *MicrobiologyOpen* 8:e00719. doi: 10.1002/mbo3.719
- Van Treuren, W., Ponnusamy, L., Brinkerhoff, R. J., Gonzalez, A., Parobek, C. M., Juliano, J. J., et al. (2015). Variation in the microbiota of *Ixodes* ticks with regard to geography, species, and sex. *Appl. Environ. Microbiol.* 81, 6200–6209. doi: 10.1128/AEM.01562-15
- Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J., and Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol.* 15:564. doi: 10.1186/s13059-014-0564-2
- Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620. doi: 10.1093/bioinformatics/btt593
- Zolnik, C. P., Prill, R. J., Falco, R. C., Daniels, T. J., and Kolokotronis, S.-O. (2016). Microbiome changes through ontogeny of a tick pathogen vector. *Mol. Ecol.* 25, 4963–4977. doi: 10.1111/mec.13832

Conflict of Interest: CM was employed by the company Irstea.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lejal, Estrada-Peña, Marsot, Cosson, Rué, Mariadassou, Midoux, Vayssier-Taussat and Pollet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.