The mosquito microbiome includes habitat-specific but rare symbionts

Hans Schrieke, Loïs Maignien, Florentin Constancias, Florian Trigodet, Sarah Chakloute, Ignace Rakotoarivony, Albane Marie, Gregory L'Ambert, Patrick Makoundou, Nonito Pages, et al.

Computational and Structural Biotechnology Journal 20 (2022) 410–420

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

The mosquito microbiome includes habitat-specific but rare symbionts

Hans Schrieke [a], Loïs Maignien [b], Florentin Constancias [c], Florian Trigodet [d], Sarah Chakloute [e], Ignace Rakotoarivony [e], Albane Marie [f], Gregory L'Ambert [f], Patrick Makoundou [g], Nonito Pages [e,h], A. Murat Eren [d,i], Mylène Weill [g], Mathieu Sicard [g], Julie Reveillaud [a,*]

[a] MIVEGEC, University of Montpellier, INRAE, CNRS, IRD, Montpellier, France
[b] Univ Brest, CNRS, IFREMER, Microbiology of Extreme Environments Laboratory, Plouzané, France
[c] CIRAD, UMR Qualisud, Montpellier, France
[d] Department of Medicine, University of Chicago, Chicago, IL, USA
[e] ASTRE, University of Montpellier, CIRAD, INRAE, Montpellier, France
[f] EID Méditerranée, 165 Avenue Paul Rimbaud, 34184 Montpellier, France
[g] ISEM, University of Montpellier, CNRS, EPHE, IRD, Montpellier, France
[h] CIRAD, UMR ASTRE, Guadeloupe, France
[i] Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA

ARTICLE INFO

Article history:
Received 23 September 2021
Received in revised form 11 December 2021
Accepted 12 December 2021
Available online 23 December 2021

Keywords:
Microbiome
Mosquitoes
Symbionts
Microdiversity
Oligotyping

ABSTRACT

Microbial communities are known to influence mosquito lifestyles by modifying essential metabolic and behavioral processes that affect reproduction, development, immunity, digestion, egg survival, and the ability to transmit pathogens. Many studies have used 16S rRNA gene amplicons to characterize mosquito microbiota and investigate factors that influence host-microbiota dynamics. However, a relatively low taxonomic resolution due to clustering methods based on arbitrary threshold and the overall dominance of *Wolbachia* or *Asaia* symbionts obscured the investigation of rare members of mosquito microbiota in previous studies. Here, we used high resolution Shannon entropy-based oligotyping approach to analyze the microbiota of *Culex pipiens*, *Culex quinquefasciatus* and *Aedes* individuals from continental Southern France and overseas Guadeloupe as well as from laboratories with or without antibiotics treatment. Our experimental design that resulted in a series of mosquito samples with a gradient of *Wolbachia* density and relative abundance along with high-resolution analyses of amplicon sequences enabled the recovery of a robust signal from typically less accessible bacterial taxa. Our data confirm species-specific mosquito-bacteria associations with geography as a primary factor that influences bacterial community structure. But interestingly, they also reveal co-occurring symbiotic bacterial variants within single individuals for both *Elizabethkingia* and *Erwinia* genera, distinct and specific *Asaia* and *Chryseobacterium* in continental and overseas territories, and a putative rare *Wolbachia* variant. Overall, our study reveals the presence of previously overlooked microdiversity and multiple closely related symbiotic strains within mosquito individuals with a remarkable habitat-specificity.

1. Introduction

In our changing world, global warming, habitat reduction and increase of transport impact host-parasite interactions and support a global emergence of wildlife zoonotic diseases [1–3]. In this context, mosquitoes represent a significant threat to global public health because of their ability to transmit multiple pathogens including eucaryotes, bacteria, and viruses. Among the most well-known arthropod-borne viruses or arboviruses, dengue, West Nile, Zika and chikungunya are responsible for several hundreds of thousands of deaths and billions of cases every year [4]. Because of a widespread insecticide resistance [5–6] and in the absence of efficient vaccines or therapy for now, manipulation of the microbiota, and in particular of mosquito bacterial symbionts is emerging as a novel vector control strategies [7–10].

The maternally inherited endosymbiotic *Wolbachia,* which is able to modify the host reproduction and limit the transmission of pathogens, can be an effective biological weapon against arbovirus transmission by mosquitoes. Introduction of infected individuals in a local population is a cornerstone for all the

* Corresponding author.
E-mail address: reveillaud.j@gmail.com (J. Reveillaud).

https://doi.org/10.1016/j.csbj.2021.12.019

techniques using *Wolbachia* either to block the pathogens or sterilizing the females [11]. It is therefore critical to understand the evolution of *Wolbachia* in natural populations before deploying these methods at larger scales and believing in their sustainability. For this, fine-tuned studies of *Wolbachia* population diversity variants are required [12]. Likewise, other bacterial symbionts can interfere with mosquito-pathogen interactions and offer great potential for paratransgenic applications, *i.e*, the use of engineered symbionts to express anti-pathogenic molecules [13]. For example, *Asaia* sp. can activate the expression of antimicrobial peptides in *Anopheles stephensi* post *Plasmodium* infection [14]. On the one hand, the intestinal bacterium *Serratia marcescens* enhances the vectorial susceptibility to arbovirus in *Aedes* mosquitoes by secreting the SmEnhancin protein that damages the physical barrier to the dengue viruses, therefore facilitating the infection and spread of the virus [15]. On the other hand, the strain Y1 from the same intestinal bacterium can inhibit the development of *Plasmodium* by activation of the immune system in *Anopheles stephensi* mosquitoes [16]. The authors highlighted that the 16S rRNA gene sequence of isolated *Serratia* strains Y1 showed 99% similarity to *Serratia marcescens*. These studies highlight the importance of a resolution at the strain level during microbiota analysis.

In addition, some of these mosquito bacterial symbionts and other microbial communities also have key roles in the development and physiology of mosquitoes. Of note, the development of axenic mosquitoes shows that some mosquito life history traits may be independent of a microbiome or at least that the relationships can be complex [17–18]. Nevertheless, bacterial communities can have major impacts on mosquito growth and larval development [19–20], blood and sugar digestion [21], immunity [22–23] and egg production [21]. Therefore, a comprehensive understanding of the diversity, potential diverse functions and possible interactions between the different microbiota members, as well as between the microbiota, pathogens and their hosts is key. Critical to the manipulation of mosquito microbiota as a novel vector control approach is an *in-depth* knowledge of their microbial communities.

However, mosquito microbiota are often dominated by a symbiont, that is, *Wolbachia* or *Asaia* in *Culex* and *Aedes*, respectively, hampering a holistic characterization of the microbiome, and notably of the less abundant but possibly important rare taxa. The rare members of the microbiota are also important reservoirs of genetic and functional diversity with key ecological roles, as shown in the marine and the terrestrial environment as well as in the plant and the human microbiome [24–26]. Because most studies of high-throughput sequencing marker genes datasets use clustering methods, often based on an arbitrary threshold of 97 to 98%, they can overlook closely related ecologically relevant sequence variants. The recently developed Minimum Entropy Decomposition (MED) and oligotyping methods use the Shannon entropy to identify ecologically relevant units at the whole community or at a taxa level (from phylum to species level), in an unsupervised and supervised manner, respectively [27–28]. These information-theory based approaches have allowed unveiling previously undetected ecological patterns for microbial communities from diverse free-living and host-associated environments including sewages, the sponge and coral microbiome, as well as the oral and plaque microbiota among others [27,29–31]. Recently, Coon and colleagues [32] suggested that bacterial communities differed substantially in *Aedes* and *Culex* larvae from different collection sites in the United States using oligotyping, although they focused on pools of larvae. Overall, the single-nucleotide resolution of MED and oligotypes make it possible to reveal novel bacterial variants, with high level of host specificity, and to characterize bacterial populations in a microdiversity-aware manner.

To gain deeper insights into typically overlooked rare microbial symbionts in mosquitoes, here we used a high-resolution sampling and analysis strategy to study their microbial community compositions. We investigated microbial communities of *Aedes aegypti*, *Culex pipiens* and *Culex quinquefasciatus* whole mosquito individuals and their reproductive organs from field (continental and overseas regions of Southern France and Guadeloupe) and insectaries (including some that were antibiotically-treated) to obtain a gradient of *Wolbachia* relative abundance. Together with fine scale entropy-based MED and oligotyping approaches as well as metagenome screening, we aimed to get access to less abundant yet possibly functionally important mosquito-associated bacterial taxa.

2. Materials and methods

2.1. Specimen collection and dissection

We collected field mosquito individuals using a carbon dioxide mosquito trap in Languedoc, Herault, France (Bosc Viel in Mauguio and Camping l'Europe Vic La Gardiole, with the help of Entente Interdépartementale de Démoustication Méditerranée EID) and in Guadeloupe (Prise d'eau, Commune de Petit Bourg) in 2017 and 2018. We transported specimens alive to the laboratory immediately upon recovery. Laboratory specimen *Culex pipiens* from Lavar (St Etienne, France) and *Culex quinquefasciatus* (Slab treated by Tetracycline for five generations before our experiments, originally from Southern California, also called Slab TC) were reared at Institute of Evolutionary Science of Montpellier insectarium (ISEM, Montpellier, France) as detailed in Duron et al [33].

We afforded mosquito manipulation and dissection as described in Reveillaud et al [34]. Briefly, we anesthetized adult females for 4 min at −20 °C and surface-sterilized insects through gentle vortexing with cold (4 °C) 96% ethanol. We transferred them into a sterile cold (4 °C) phosphate-buffered saline (PBS) 1 × solution and then onto a sterile microscope slide with sterile PBS 1 × on top of a cold plate prior to dissection of ovaries using sterilized tweezers. We stored whole or dissected mosquito specimens at −80 °C until further processing.

2.2. DNA extraction, V3-V4 polymerase chain reaction (PCR) and Illumina tag sequencing

We extracted total genomic DNA in a sterile hood with the DNeasy Blood and Tissue (Qiagen) following manufacturer instructions. We systematically added one extraction blank (negative control, hereafter blank), corresponding to an extraction without sample tissue, to each of a series of 10 DNA extraction microtubes. We conducted all pre-PCR laboratory manipulations with filtered tips under a sterile environment in a DNA-free room dedicated to the preparation of PCR mix and equipped with hoods that are kept free of DNA by UV irradiation. We used primers and the dual-index method of Kozich [35] to amplify a 429-bp portion of the V3-V4 region of the 16S rRNA gene with the primers V3F (CCTACGGGAGGCAGCAG) / V4R (GGACTACHVGGGTWTCTAAT) in triplicate for each sample. PCR was carried out in a 25-μL reaction volume using 2 μl of genomic DNA mixed with 23 μl of reactional mix containing 0,1 μl of Platinum™ *Taq* DNA Polymerase High Fidelity (Invitrogen). PCR was run with an initial denaturation step of 94 °C for 2 min, followed by 35 cycles of denaturation at 94 °C for 20 s, annealing at 55 °C for 15 s and extension at 72 °C for 5′, followed by a final extension step at 72 °C for 10 min. After pooling, we purified PCR products using Ampure XP® kit. We quantified DNA with a QUBIT 2.0 Fluorometer. We created two pools per run, grouping samples based on how well they amplified. These pools had a concentration of 1 ng/μl and 30 ng/μl. We diluted over-concentrated samples (9 from run 2 exclusively, see Supplemen-

tary Table 1) to match these concentrations. This procedure attempted generating libraries with equimass amounts of DNA for each run. The combination of primers including different 8-bp indices (8 i5-indexed forward primers and 12 i7-indexed reverse primers) allowed multiplexing 96 samples onto three MiSeq flow cells. Sequencing was done at the Genotyping Platform from UMR AGAP (Montpellier, France).

We analyzed a total of 136 samples including 71 whole specimens and 50 dissected ovaries (where vertically transmitted symbionts can be more specifically found) together with 15 blanks (see Supplementary Table 1). Whole specimen samples consisted of 43 *Culex pipiens* (7 individuals from Camping Europe, 14 from Bosc and 22 from Lavar), 18 *Culex quinquefasciatus* samples (4 from Guadeloupe, and 14 Slab TC) and 10 *Aedes aegypti* from Guadeloupe. Among the ovaries, we counted 44 *Culex pipiens* samples (21 from Camping Europe, 10 from Bosc and 13 from Lavar), 3 *Culex quinquefasciatus* and 3 *Aedes aegypti* samples, both from Guadeloupe.

2.3. Identification of oligotypes, filtration of contaminants, and taxonomic assignment

We merged and quality-filtered raw sequences using the Illumina-Utils pipeline version 2.8 [36] with stringent settings (min-overlap-size 30 --max-num-mismatches 0 --enforce-Q30). We used the algorithm minimum entropy decomposition (MED) [27] with default parameters to identify sequence variants in high-throughput sequencing of 16S rRNA gene amplicons. MED is an algorithm to partition a given set of sequences into discrete sequence groups (terminal MED nodes, hereafter unsupervised oligotypes (UO)) by minimizing the total entropy in the dataset [27]. We identified and removed contaminant UO that could come from several sources (laboratory manipulation, DNA extraction kit or PCR reagents kitomes) [37] using the prevalence method from the R package Decontam version 1.6.0 [38]. We computed a statistical score based on a presence / absence table of UO counts in true samples and blanks. We used a threshold of 0.6 to remove contaminants, which consisted of UO with higher prevalence in blanks than in samples (Supplementary Fig. 1). We assigned taxonomy to the representative sequence of each retained UO using the *assignTaxonomy()* function (minBoot = 80) based on the RDP Naive Bayesian Classifier algorithm [39] and the *addSpecies()* function from the R package Dada2 version 1.14.1 [40] using the silva138 reference database [41]. We checked for the presence of mitochondria and eukaryotic contamination in our sequences and did not retrieve any.

In cases where maximum sensitivity was essential, we further investigated genus-level UO generated from our data by manually identifying nucleotide positions of high entropy using the supervised oligotyping method [28], which enables supervised input into the curation of final results. Briefly, we concatenated the sequences from all UO assigned to a specific genus and ran the oligotyping pipeline. Supervised oligotypes (SO) are defined as the concatenation of the high-information nucleotide positions identified by Shannon entropy with a threshold value of minimum 0.2. For this, we used the default component (-c) to identify the nucleotide positions with the highest Shannon entropy. We set the minimum substantive abundance threshold for a supervised oligotype (-M) to 10. We further confirmed these results for M = 100 (data not shown). Random sequencing errors are reported to generate entropy values lower than 0.2 for Illumina data (https://meren-lab.org/2013/11/04/oligotyping-best-practices/).

2.4. Statistical analyses

We imported raw unsupervised oligotypes count table, taxonomic assignments and metadata in the Phyloseq package for data handling [42] using R version 3.6 [43]. We analyzed microbial communities using alpha diversity metrics including UO observed richness, Chao1's estimated richness and Shannon's diversity index on unnormalized count data. Both Observed and Chao1 indexes account for the number of species (richness) with the latter giving with more weight to rare species using a singletons to doubletons ratio while Shannon accounts for both abundance and evenness. Community similarity (*i.e.,* beta-diversity) was quantified using Bray-Curtis distance on normalized count table (relative abundance) as in McKnight et al [44] and visualized using NMDS ordination. Samples with a sequencing depth lower than 1000 reads were discarded (13/136) to allow an exhaustive description of bacterial communities with some exceptions (five samples from Guadeloupe NP20, NP29, NP30, NP34, NP36, see Supplementary Note 1 for details). We tested differences in alpha and beta-diversity between species or location using ANOVA (Analysis of variance, *aov()* function from the basic package stats from R version 3.6 [43] and PERMANOVA statistical tests (*adonis()* function from the vegan R package version 2.5–6 [45], respectively. In order to test if sequencing depth influenced differences in alpha-diversity between groups, the latter was added in the model (alpha-diversity $\sim$ Group * Sequencing depth after rarefaction of the dataset to 100 and 1000). A significant interaction between Group and Sequencing depth was detected only for the Shannon index (see Supplementary Table 2), indicating that sequencing depth influenced Shannon values among groups. Therefore, statistical tests for the Shannon index were performed on the rarefied dataset as well (using the *rarefy_even_depth*() function from the Phyloseq package; 100 reads). Multiple group comparisons were performed using Tukey's test (*glht()* function from multcomp R package version 1.4 [46]) for alpha diversity. We also checked the heteroscedasticity in order to test for eventual biases of heterogeneity of variance for PERMANOVA results [47]. To do that, we i) computed the Bray-Curtis distance from the normalized count table using the *vegdist()* function from the vegan R package version 2.5-6 [45], ii) used the distance matrix to compute the multivariate dispersion with the function *betadisper()* from the vegan R package version 2.5-6 [45], and iii) performed ANOVA on dispersion groups using the *anova()* function from the stats R package from R version 3.6. We then generated rarefaction curves using the *ggrare()* function from the R package ranacapa [48] to investigate sequencing depth efforts per sample.

2.5. Wolbachia relative abundance and density in Culex specimen

We investigated the *Wolbachia* relative abundance from the high-throughput sequencing *Culex* dataset using a Hierarchical Cluster Analysis (or HCA) that limits the arbitrary choice of a threshold value to split data. For that, we used three functions from the basic stats package from R 3.6 [43]: *dist()* to convert the percentage of *Wolbachia* per sample into an Euclidean distance matrix, *hclust()* to perform the HCA using the "ward.D2" method and *cutree()* to cut the tree resulting from the HCA into three groups (k = 3, hereafter referred to "Low infection", "Medium infection" and "High infection" groups). Using this approach, we ended up splitting samples with *Wolbachia* relative abundance between 0 and 5%, 10 to 69% and 69 to 100%, respectively (Supplementary Fig. 2 and Supplementary Table 3).

In addition, we used real-time quantitative PCR (qPCR) on a subset of representative whole body samples (n = 34) in order to estimate the *Wolbachia* density and investigate its correspondence between the *Wolbachia* infection groups obtained with HCA. Briefly, we used *Wolbachia* wsp (using wolpipdir and wolpiprev primers as in Berticat et al [49] and *Culex Ace-2* genes (using acequantidir and acequantirev) as in Weill et al [50] to estimate *Wolbachia* infection level. We assumed that both genes are present in a

single copy per host and symbiont haploid genome, and therefore that the ratio between both gene concentrations provided the number of *Wolbachia* genomes relative to the mosquito genomes. Standard curves were obtained by diluting a plasmid containing one copy of each gene. We used 0,5 μl (2 ng) of genomic DNA mixed with 1,5 μl of reactional mix containing 0,6 μM of each primer and 0,75 μl of master mix (Sensifast SYBR No-ROX mix 2x, Meridian Bioscience from Bioline). Each PCR reaction was done in triplicate. PCR was run with 95 °C for 3 min followed by 45 cycles defined by 95 °C for 10 sec, 60 °C for 10 sec and 72 °C for 15 sec. Melting curves were generated by a post-amplification melting step between 70 °C and 95 °C, for Tm analysis. Analysis was done with the Advanced Relative Quantification method of the LightCycler 480 software v.1.5.0. Efficiency was confirmed as doubling at each cycle for each gene. Samples with a Ct (cycle threshold) value greater than 33 were excluded to avoid uncertain low titer quantification.

2.6. Screening ovary shotgun metagenomes for the presence of potentially distinct Wolbachia variants

To further validate the occurrence of putative distinct *Wolbachia* variants within the reproductive organs, we screened available ovary shotgun metagenomes with the identified SO. We extracted fragments of 30 bp from the SO sequences including the positions with highest entropy that were used to discriminate the different *Wolbachia* variants, and used the anvi'o [51]) program "anvi-script-get-primer-matches" to search for metagenomic sequences that include these fragments in four ovaries metagenomes of *Culex pipiens* mosquitoes from Southern France [34] (European Nucleotide Archive ENA accession numbers: ERS2407346, Culex O03; ERS2407347, Culex O07; ERS2407348, Culex O11; ERS2407349, Culex O12) as well as egg-rafts from three *C. pipiens* isofemale lines from North Africa [52] (SRR5810516, Pipiens_MGx_Istanbul; SRR5810518, Pipiens_MGx_Tunis; SRR5810517 Pipiens_MGx_Harash). We quality-filtered raw sequences using the anvi'o [51] "iu-filter-quality-minoche" command with default parameters before the metagenomic screening.

3. Results

3.1. Estimated bacterial diversity

We obtained a total of 6,452,623 high-quality reads distributed into 67 unsupervised oligotypes (UO) from 113 samples, including 67 whole specimens (3,402,384 reads) and 46 ovary samples (3,050,239 reads) after stringent quality filtering of Illumina reads and decontamination (see Supplementary Note 1 for the detailed bioinformatic analyses). Supplementary Fig. 3 and Supplementary Table 4 recapitulates the numbers of reads for each of the 113 samples and UO per group of samples, respectively. Most rarefaction curves plateaued (except for the three *Culex quinquefasciatus* whole individuals and the two *Aedes aegypti* ovary samples that showed much lower number of reads, as expected) indicating that the sequencing depth was sufficient to describe most of bacterial taxa (Supplementary Note 2 and Supplementary Fig. 4).

Whole mosquito bacterial community richness appeared significantly higher for *Aedes aegypti* samples compared to the *Culex* specimens (Fig. 1, see Supplementary Note 3 and Supplementary Table 5 for statistical data). In addition, bacterial richness appeared greater for *Culex quinquefasciatus* Slab TC *(Wolbachia*-) compared to the *Culex quinquefasciatus* from the field (*Wolbachia*+) samples. Similarly, bacterial community richness was slightly higher for *Culex pipiens* specimens from the laboratory (Lavar) than from field locations using all three Observed, Chao1 and Shannon diversity indexes. In the case of *Culex* specimen, all samples were positive for *Wolbachia* (Supplementary Note 4), with distinct relative abundance of this endosymbiont, suggesting a higher bacterial diversity for laboratory specimens as compared to the field ones in this study independent of any previous antibiotic treatment. Noteworthy, *Culex pipiens* samples from the laboratory (Lavar) showed a lower *Wolbachia* relative abundance (Supplementary Fig. 5A) yet a higher *Wolbachia* density as compared to field specimen from Camping Europe (Supplementary Fig. 5B). These data show differences for *Wolbachia* relative abundance and densities, as confirmed by our statistical analyses (Supplementary Note 4, Supplementary Fig. 6).

We further investigated the relative influence of species, geography, insectary conditions as well as *Wolbachia* relative abundance on mosquito bacterial communities taking into account eventual biases of heteroscedasticity and sample size. Data are provided in Supplementary Note 5, Supplementary Fig. 7 & 8.

3.2. Mosquitoes harbor multiple symbiotic strains

We identified amplicon sequence variants in our data using unsupervised oligotyping by running it automatically and in some cases manually refining the results for maximum quality assurance (see Materials and Methods). Overall, the 67 UO generated by MED were affiliated to a small number of bacteria belonging to 12 different genera: *Wolbachia*, *Asaia*, *Legionella*, *Elizabethkingia*, *Chryseobacterium*, *Erwinia*, *Morganella*, *Pseudomonas*, *Delftia*, *Methylobacterium-Methylorubrum*, *Serratia* and *Coetzeea*. One UO (N0318) was only assigned at the order level (Lactobacillales) and represented 0.13% of the total number of reads. A representation of the number of reads per genus is available in Supplementary Fig. 9A.

Whole *Aedes aegypti* individuals, which are mostly known as *Wolbachia* free mosquitoes, were dominated by *Asaia* (85.8%), *Pseudomonas* (10%) and *Chryseobacterium* (3.4%) (Supplementary Fig. 9B) while *Culex* spp were dominated by *Wolbachia*. We detected the presence of *Wolbachia* in *Aedes aegypti* samples (ranging from 1.21e-05 to 0.02 %), in agreement with recent studies reporting the detection of *Wolbachia* infection of *A. aegypti* at low frequency through molecular detection, and possibly not representing active infections [53–54]. The three most abundant genera in wild *Culex quinquefasciatus* were *Wolbachia* (74.7%), *Asaia* (21.1%) and *Pseudomonas* (2.3%). In *Culex pipiens*, *Wolbachia* represented 99.8%, 99% and 59.6% of total microbiota in Bosc, Camping Europe and Lavar, respectively. *Culex pipiens* Lavar specimens showed a higher bacterial diversity with *Asaia* as the second most abundant genus (12.7%) and *Elizabethkingia* as the third one (10.1%). On the contrary, *Culex quinquefasciatus* Slab TC (*Wolbachia*-) specimens were dominated by *Elizabethkingia* (35.9%), *Legionella* (29.4%) and *Erwinia* (15%). Surprisingly we observed *Wolbachia* sequences in the latter samples, representing 0 to 10.74% (*ie.*, in sample CTC14, Supplementary Table 3), suggesting potential *Wolbachia* DNA remnants that are not detected through qPCR. As for the reproductive organs, we observed *Wolbachia* dominated both *Culex pipiens* and *Culex quinquefasciatus* samples with 100% relative abundance.

Further, when considering the different genera, we observed several UO affiliated to *Asaia* (N0939, N1147, N1156), *Elizabethkingia* (N0990, N1160), *Chryseobacterium* (N1034, N1035), *Legionella* (N0635, N1065) and *Erwinia* (N0311, N0798) distributed among most mosquito samples (Fig. 2). Moreover, while unsupervised oligotype N0711 was the most abundant *Wolbachia* UO, the genus also showed additional variants that were not part of the 15 most abundant UO and were therefore grouped into "Others" (see heatmap with all oligotypes assigned to *Wolbachia* in Supplementary Fig. 10). We observed the same *Wolbachia* unsupervised

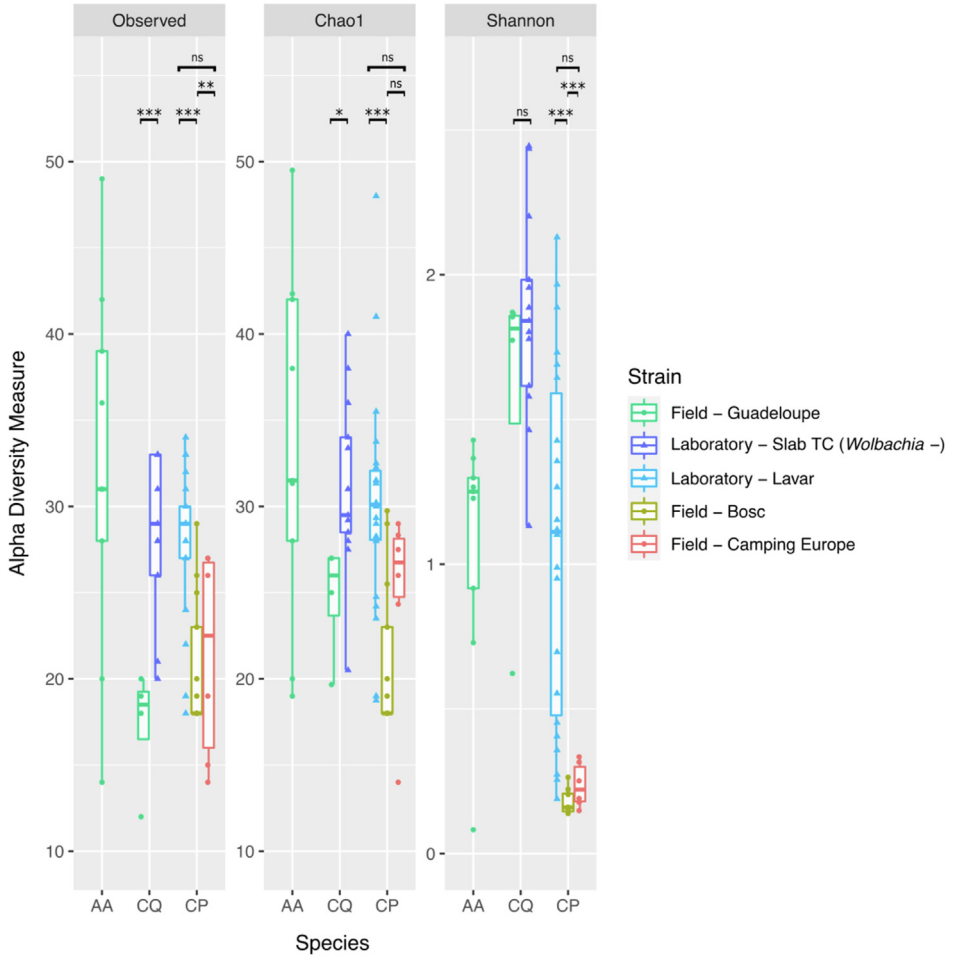

Fig. 1. Alpha diversity measures using the Observed, Chao1 and Shannon diversity indexes in whole *Culex pipiens* (CP), *Culex quinquefasciatus* (CQ) and *Aedes aegypti* (AA). Symbols above brackets represent the significance of the Tukey's test performed in *Culex pipiens* and *Culex quinquefasciatus* samples: 'ns' corresponded to p > 0.05, '*' to p <= 0.05, '**' to p <= 0.01 and '***' to p <= 0.001.

oligotype N0711 dominating the ovaries samples, yet co-occurring with additional putative distinct *Wolbachia* variants in these organs too (Supplementary Fig. 11).

In order to further investigate the occurrence of putative distinct variants within these mosquito symbionts and commensal taxa, we ran the supervised oligotyping pipeline on genus-level UO from genera that showed multiple UO for both whole individuals and ovary samples.

3.3. Co-occurring and habitat-specific supervised oligotypes

3.3.1. Asaia supervised oligotypes

Asaia supervised oligotypes (SO) were identified within the 1,755,102 sequences affiliated to this genus. We were able to manually decompose it further into five SO using five nucleotide positions including three abundant and two rare oligotypes. Oligotypes revealed by the entropy found in the entire dataset showed an abundance ratio of ca. 1:2 between the 3rd (376,837 reads) and the two most abundant *Asaia* SO (699,320 and 661,010 reads, respectively). This 1:2:2 ratio between *Asaia* oligotypes suggests that this variation could be explained by a single strain of *Asaia* with 5 copies of the rRNA operon with subtle differences, consistent with Ribosomal RNA Database [55] query results for this genus (Supplementary Fig. 13, panel C). Yet, breaking down the distribution of SO across individual samples did not support this hypothesis, since the copy number should be preserved across samples. Indeed, we observed some samples with such expected 1:2:2 ratio (*e.g.*, for whole *Aedes* and *Culex quinquefasciatus* mosquito individuals collected in Guadeloupe, a typical sample is NP38) while some other showed distinct ratios like a 5:0:1 (*e.g.*, S50, *Culex pipiens*, Bosc, France), 3:0:1 (*e.g.*, S44, *Culex pipiens*, Camping Europe, France) or 1:0:0 (*e.g.*, S40, *Culex pipiens* isofemelle lines from Lavar laboratory or CTC1, Slab TC, Supplementary Fig. 13 panel A1 and A2). These data instead suggest some distinct *Asaia* populations, with one sequence variant (CATAG) found exclusively in Guadeloupe (in 13 out of 15 samples). Interestingly, we observed the co-occurrence of two to three *Asaia* SO in several individual samples. In addition, specimens from insectaries (both Lavar and Slab TC) shared a similar *Asaia* SO.

3.3.2. Elizabethkingia supervised oligotypes

Similarly, we further investigated the *Elizabethkingia* supervised oligotypes by examining the distribution of entropy in 135,187 reads it accounted for. We manually decomposed it further using one nucleotide position into two SO; a dominant supervised oligotype C with 124,095 reads and a less abundant supervised oligotype G with 11,092 reads (Supplementary Fig. 14). The co-occurrence of both *Elizabethkingia* SO in each individual sample also suggests the presence of closely related *Elizabethkingia* variants. In addition, their differential co-occurrence in almost all samples suggest *Elizabethkingia* variants are not specialized to one or several specific groups of mosquitoes.

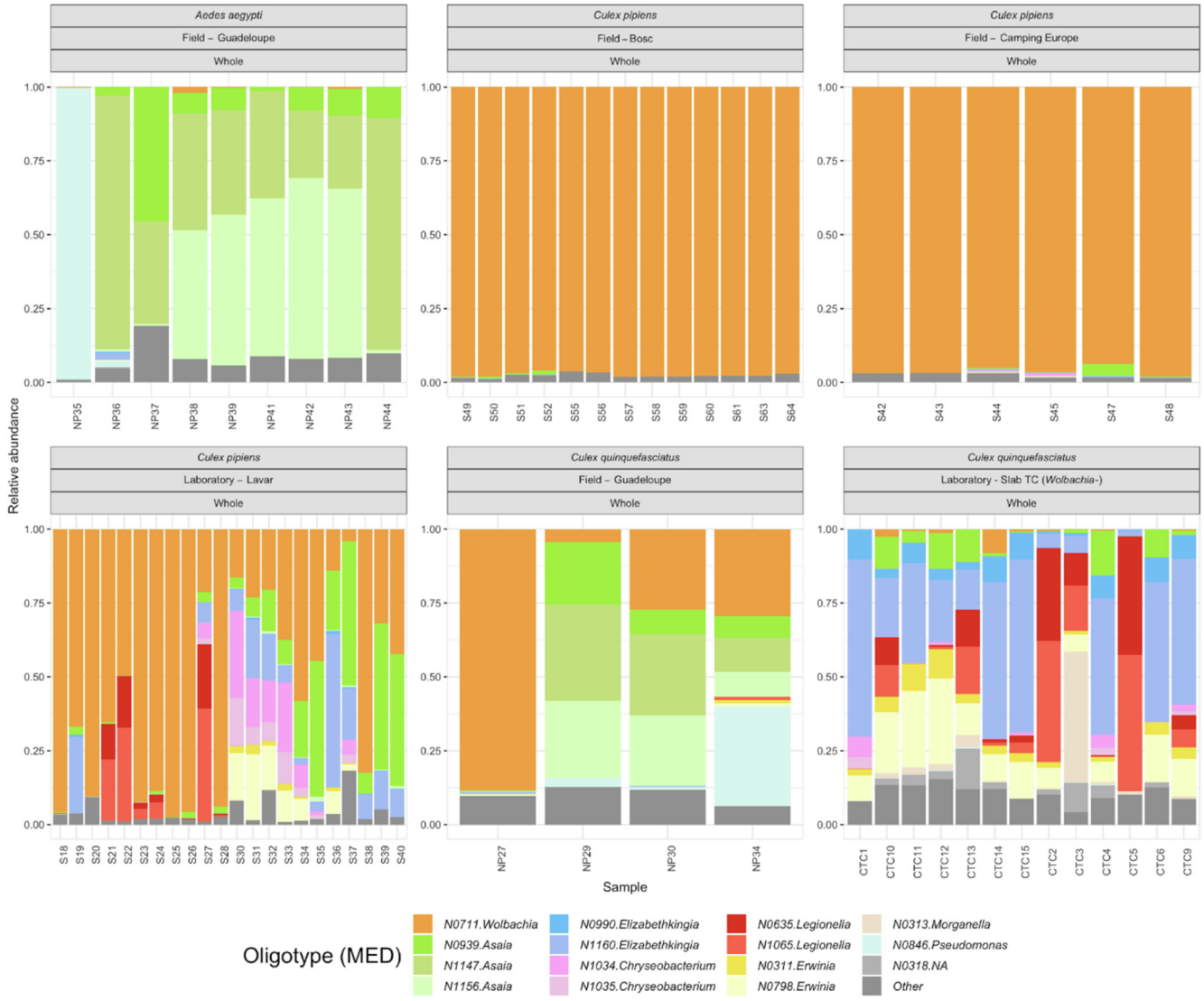

Fig. 2. The 15 most abundant unsupervised oligotypes in whole *Culex* spp and *Aedes* individuals in function of mosquito strain. Each color represents an UO with the taxonomic assignment at the genus level. "Other" groups indicate UO that are not included in the 15 most abundant ones. A representation of all UO using a heatmap is available in Supplementary Fig. 12.

3.3.3. *Erwinia* supervised oligotypes

Oligotyping 51,215 sequences affiliated to *Erwinia* resulted in three supervised oligotypes defined by 8 nucleotide positions. Data showed the presence of one *Erwinia* SO (AAGACTTA; 38,025 reads) dominating all samples, as well as two less abundant *Erwinia* variants (TGAGTCGA; 7,110 reads AAGACTTG; 4,594 reads, respectively, Supplementary Fig. 15). Of note, the Slab TC samples, not dominated by *Wolbachia*, highlighted the co-occurrence of the putative three *Erwinia* variants consistently in all samples, and suggest they could possibly occur in all mosquito samples although they are less or not accessible. These three SO differ from each other by one to 8 nucleotides, representing a minimum of 98,4% percent similarity.

3.3.4. *Chryseobacterium* supervised oligotypes

We further investigated the supervised oligotypes of *Chryseobacterium* by studying the distribution of entropy in 106,099 sequences affiliated to this genus. We decomposed *Chryseobacterium* using 3 nucleotide positions into 3 SO. The most abundant *Chryseobacterium* oligotype CGC (66,891 reads) was found only in mosquito samples from Guadeloupe (*Culex quinquefasciatus* and *Aedes*, in 11/11 samples) while oligotypes TAC (12,689 reads) and oligotype TAT (26,519) were observed in samples from France (from the field and lab mosquitoes) as well as in Slab TC (in 34/34 samples, Supplementary Fig. 16). Of note, a ratio of ca. 1:2 between the second and the third most abundant SO (12,689 for TAC; 26,519 for TAT) highlighted by the entropy analysis of the whole dataset could suggest they represent two copies of the ribosomal RNA operon of a *Chryseobacterium* bacterium. Such ratio was however not retrieved consistently in all samples, probably discarding this hypothesis. Overall, these data suggest specific *Chryseobacterium* variants in mosquitoes from overseas and continental territories in Guadeloupe and Southern France.

3.3.5. *Serratia* and *Legionella* supervised oligotypes

Following the supervised analysis of *Serratia* (8,511 reads), an abundance ratio of ca. 1:2 between the 1st (A; 6,013 reads) and the second most abundant oligotypes (G; 2,498 reads) in almost all samples could suggest that mosquito-associated *Serratia* genomes harbored 2 copies of the rRNA operon (Supplementary Fig. 17). However, a ratio of ca. 1:2 between the two most abundant *Legionella* supervised oligotypes CGGA (56,507 reads) and

AAAA (34,561 reads) out of 100,761 *Legionella* reads as shown by the entropy found in the entire dataset was not confirmed in all individual samples, suggesting distinct *Legionella* variants (Supplementary Fig. 18). Overall, the presence and abundance of *Serratia* and *Legionella* SO varied between samples with no obvious consistent pattern.

3.3.6. Wolbachia supervised oligotypes

We then further studied the SO *Wolbachia* (4,081,319 sequences), leading to a decomposition into seven oligotypes mainly based on two high Shannon entropy positions 260 and 268 (Supplementary Fig. 19, panel C). Therefore, each SO differed by as little as one to two nucleotides (Fig. 3). Of note, supervised oligotype CC presented a unique entropy profile with important entropy value for more than 250 positions (data not shown). This particularity can be explained by a deletion at position 192 as we can see in the sequence alignment on Fig. 3. *Wolbachia* SO showed a different size (number of reads), with AT accounting for the higher number of reads (3,890,567) and AC accounting for the smallest one (2,504, Fig. 3). In relative abundance, data showed a dominant *Wolbachia* variants (supervised oligotype AT corresponding to the unsupervised oligotype N0711), accounting for 47 to 100% and rare "variants" that represented between 0 and 26% for supervised oligotype AG, 0 to 21% for supervised oligotype GT, 0 to 16% for supervised oligotype TT, 0 to 0.5% for supervised oligotype CC, 0 to 3.6% for supervised oligotype GG, and 0 to 2.5% for supervised oligotype AC in the different studied samples (Supplementary Table 6 – Sheet 1).

When studying the distribution of these SO, we observed the co-occurrence of one to four different SO in a single sample like in S42 or CTC6, respectively (Supplementary Fig. 19, panel A1 and A2). We also noted multiple co-existing SO in samples with low (Slab TC) to medium or high (most of *Culex pipiens* from Bosc locality for instance) *Wolbachia* relative abundance, suggesting the presence of these SO is not linked to the *Wolbachia* predominance.

We eventually blasted these SO in order to investigate the presence of these putative variants in the available databases. On the one hand, we observed the dominant supervised oligotype AT had a match with 100% identity with *Wolbachia pipientis* isolates including TuFIVIA19m (NCBI Accession number MN123175.1) while GT, AG, CC, TT, and GG had a match with 99.5% to 99.75% identity to the same sequence. On the other hand, supervised oligotype AC had a match with 100% identity to three distinct *Wolbachia* strains including *Wolbachia* endosymbiont of *Acrocephalus palustris* isolate Acr172 (NCBI Accession number MF374624.1), *Wolbachia* endosymbiont of *Ectropis obliqua* clone WX-03 (NCBI Accession number KU058642.1), *Wolbachia* endosymbiont of *Medythia nigrobilineata* strain Mni (NCBI Accession number GU236941.1). While the sequence variants that occur in a small number of positions throughout the amplicons give credence to true biological diversity, it is difficult to exclude PCR amplification-associated systematic errors to result in variants with high relative abundance, especially in the presence of a remarkably numerous template. Of note, we observed a homopolymer region "TTTTAAA" before the first high entropy peak (position 260) that could favor a possible error (sequencing or polymerase) while the second entropy peak (268) appeared as independent of this region (Fig. 3).

3.4. Metagenome screening of Wolbachia supervised oligotypes

Since amplification-free shotgun metagenomes are less prone to PCR errors, we further investigated the presence of additional *Wolbachia* variants suggested by the highly-resolved analysis of amplicon sequences in single individuals using metagenomic datasets generated independently. Using variable regions of SO sequences as 'primers', we searched high-quality metagenomic short reads in available *Culex pipiens* ovary metagenomes [34] and egg-rafts metagenomes [52] from France and different locations in Northern Africa generated using the Illumina sequencing technology. For that purpose, we extracted small nucleotide fragments (Supplementary Fig. 20) including the two highest entropy positions and searched for their presence in the *Culex* metagenome datasets. As expected, supervised oligotypes AT counted a huge number of hits within the four *Culex pipiens* ovary metagenome samples, that is between 89 and 322 hits (Supplementary Fig. 21). In addition, other SO showed single hits (Supplementary Fig. 21) which could further suggest putative multiple *Wolbachia* infection in *Culex* samples, although these remain based on very small read numbers and therefore do represent ambiguous data.

4. Discussion

In this study, we realized a fine-scale 16S rRNA gene analysis of the bacterial communities of *Culex pipiens*, *Culex quinquefasciatus* and *Aedes aegypti* mosquito individuals collected in continental and overseas regions of France. We examined the bacterial assemblages in the whole bodies as well as dissected ovaries from specimen collected in the wild or in insectaries (Lavar *C. pipiens* and antibiotically-treated *C. quinquefasciatus* Slab TC) as well as with low to high *Wolbachia* relative abundance and densities. We used Minimum Entropy Decomposition (MED) and oligotyping that use the Shannon entropy to identify ecologically relevant, and to discriminate taxa that differ by as little as a single nucleotide, in the sequenced region of these datasets [27–28].

4.1. The influence of host species, geography, environmental conditions, symbiont relative and absolute abundance on mosquito microbiota

Bacterial community structure differed significantly between our *Culex pipiens*, *Culex quinquefasciatus* and *Aedes aegypti* samples, confirming different mosquito microbiota for each species.

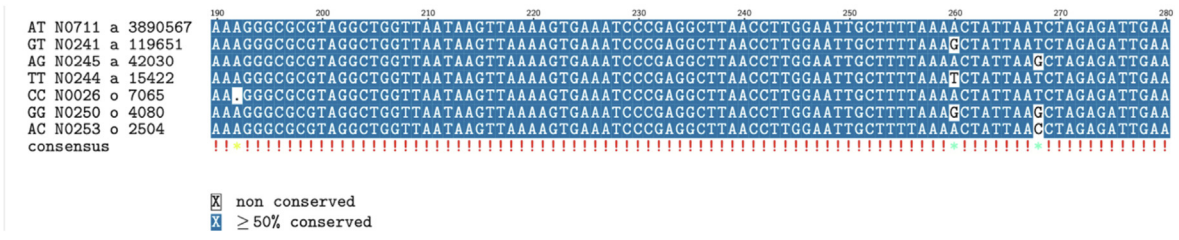

Fig. 3. Alignment of the seven *Wolbachia* supervised oligotypes retrieved in this study. The first letters correspond to the SO (at the 260 and 268 positions that varied), the first number represents the corresponding UO, the following letter indicates abundance information ("a" means that UO is among the 15 most abundant, "o" means that UO is among "Other" in Fig. 2), and the last number indicates the size of SO (number of reads).

Although bacterial communities of wild *Aedes* specimens were dominated by *Asaia* taxa (85.8%), *C. pipiens* and *C. quinquefasciatus* individuals were dominated by *Wolbachia*, *Asaia* and *Elizabethkingia,* yet in different proportions for each species. The predominance of *Wolbachia* was in agreement with previous studies in *Culex pipiens* [56–58] and *Culex quinquefasciatus* [59–60]. Similarly, *Asaia* is one of the most-frequently detected genera in *Aedes aegypti* microbiota [61–65], although it was found in lower abundances than in our study. It is likely that the predominance of *Wolbachia*, *Asaia* and *Elizabethkingia* in specimens from both *Culex* species explains the observed partial overlap in the overall community structure of both species. Geography at a fine scale (such as Bosc and Camping l'Europe sites which are not far from each other) was nevertheless shown to influence bacterial communities within *Culex pipiens*. These data were in agreement with previous studies, showing the site of mosquito collection influences bacterial community structure at a large geographic scale [66].

We then observed clear differences between field and insectaries mosquito microbiota. Drastically distinct bacterial profiles in individuals from *C. quinquefasciatus* lines (*Wolbachia*+ and *Wolbachia*-) which were treated (at least five generations before sampling) and not treated with tetracycline could be due to antibiotics, which eliminate some sensitive resident microbiota from mosquitoes, and consequently may have an effect on the remaining bacterial community. We however cannot ignore some confounding factors that could influence bacterial communities like distinct collection locations for both strains (Guadeloupe vs. California). Nevertheless, our *C. pipiens* specimens reared in the laboratory for several consecutive years (Lavar strain) unexpectedly showed a higher bacterial richness as compared to our wild mosquito individuals. These data differed from previous studies, showing a particularly poor bacterial richness and diversity in laboratory mosquitoes in *Anopheles* species [67]. One hypothesis could be that selection on *Culex* microbiota composition could be stronger in environmental conditions while it might be more relaxed in the lab, where mosquitoes are fed every day and competition is lower allowing the presence of diverse bacteria. Another hypothesis could be that slightly lower *Wolbachia* relative abundance in *Culex* specimen from the lab, as observed herein, could allow for a greater bacterial diversity to be observed (Supplementary Fig. 5A). Nevertheless, *Wolbachia* densities as detected by qPCR were noteworthy higher in *Culex pipiens* from the lab (Supplementary Fig. 5B).

These data underline that high-throughput sequencing approaches provide relative abundances of bacterial taxa, and cannot be directly linked to absolute abundance, in agreement with Williamson and colleagues [68]. In addition, it suggests that higher bacterial diversity in the lab could possibly be associated with higher bacterial loads for the remaining taxon. These different hypotheses might not be exclusive from one another and future studies should investigate these scenarios using more samples and laboratory conditions. In any case, bacterial community differences highlight the importance of lab vs. environmental conditions on microbiota and reminds us to remain cautious and critical about the interpretation of infection outcome when conducting experimental infection using laboratory strains. Mosquito microbiota from lab strains could indeed respond differently to pathogens in experimental conditions and not mimic wild mosquito individuals.

4.2. Closely related 16S rRNA gene sequences highlight multiple symbiont variants

Overall, our combined unsupervised and supervised oligotyping analyses revealed that some genus-level mosquito-associated bacterial taxa contain putative previously undetected co-occurring variants at the scale of single individuals. In particular, it highlighted the predominance of certain variants yet the presence of additional but less abundant ones that would not have been seen using a classical clustering approach.

We noted the co-existence of closely related *Elizabethkingia* and *Erwinia* variants in several individuals, suggesting the possibility of some gene content variation and functional differences, warranting further analyses at the genome scale. The bacterium *Elizabethkingia anophelis* was first isolated from *Anopheles* mosquitoes in 2011 [69] and later detected in *Aedes* dengue fever vector mosquitoes but not in *Culex* [70]. *Elizabethkingia*-like bacteria were identified as interesting candidates for paratransgenic approaches due to some anti-*Plasmodium* activity *in vitro* [71–72] as well as some genetic modification potential [70]. However, a recent study showed a single DNA repair gene disruption in one *Elizabethkingia anophelis* strain was at the origin of an atypical evolutionary rate and a large outbreak of *E. anophelis* human infection in the US from 2015 to 2016 [73]. Although a 1% nucleotide distance separated the outbreak strain from its closest environmental *E. anophelis* relatives at the whole genome level, the authors [73] showed the integration of a mobile genetic element in a specific protein-coding gene within this species could lead to the emergence of a pathogenic strain with serious health implications. In line with this, the co-existence of closely related *Erwinia* variants showing more than 98.4% percent similarity raises questions about the potential phenotypic differences of each strain. Members of the *Erwinia* group are phytopathogenic but also exploit insects both as hosts and vectors [74]. The presence of a dual *Erwinia* infection might have implications for crop management, suggesting some monitoring and further genome-scale analyses are needed.

In addition to different strains coexisting in a single individual, our study showed some variants specific to certain populations of mosquitoes, like for *Asaia* and *Chryseobacterium* in Guadeloupe vs. continental France. We actually observed one *Asaia* variant that seemed to occur in Guadeloupe only, and not in populations from continental France. These data are in agreement with previous studies, showing that different *Asaia* strains are present in different mosquito populations [75]. The authors indicate the importance of uncovering the diversity of *Asaia* strains, proposed as potential antiplasmodial agents, in the context of paratransgenic approaches. Similarly, the observation of specific *Chryseobacterium* sp. strains in different geographic locations is in agreement with their mutualist / commensal characteristics. *Chryseobacterium culicis* sp. *nov.,* was first isolated from the midgut of the mosquito *Culex quinquefasciatus* [76] and later on described as having a beneficial role for their host, rescuing axenic larval mosquito development [20,70], while inducing a weak immune response. Previous work showed this bacterial species is well adapted to life in the gut of insect [77] and one can expect a high level of specificity from these mutualistic relationships.

Further, our data suggest the putative co-occurrence of several *Wolbachia* oligotypes in a single *Culex* specimen. Although systematic PCR errors in samples dominated by one or few bacterial taxa could influence the observed results, we retrieved the same oligotyping pattern for wild *Culex pipiens* individuals that have high to low *Wolbachia* densities, and that all together offer a *Wolbachia* relative abundance gradient. On the one hand, these data could putatively suggest possible co-infection patterns in independent samples and with no link to *Wolbachia* densities. On the other hand, the presence of varying *Wolbachia* 16S rRNA gene sequences within a single mosquito individual could not be confirmed with more than one hit when screening seven different *C. pipiens* ovary and egg-rafts metagenomes collected in Southern France and Northern Africa (Supplementary Fig. 21), avoiding any conclusive remark at this point. Nevertheless, the very low number of AC sequences retrieved in the metabarcoding datasets (ca. 2,500) suggests they might be too rare to be detected in the available Illumina metagenomes. In addition, the detection of some *Wolbachia* reads in few *Culex* Slab TC sample was not expected and might

be due to some DNA remnants rather than true bacterial cells or Horizontal Gene Transfer (HGT) events. Although HGT of functional genes between *Wolbachia* and its host has been reported [78], no HGT of 16S rRNA genes has been reported in *Wolbachia* to our knowledge. More generally, HGT of 16S rRNA genes seems to be a rare event occurring only at the intra-genus and intra-species levels [79].

The 100% BLAST matches with *Wolbachia* sequences for both AT and AC supervised oligotypes add more weight on the likelihood of having an "A" at the first high-variation position (260), while the second position (268) is possibly more variable. In other words, it reinforces the probability of a biological variation at position 268 while it diminishes the one at position 260. In line with this observation, we noted a homopolymer region before the first high entropy position "A", which could favor a possible sequencing or polymerase error while the second high-variation position variation remains independent of this region and might be biologically true. Although this hypothesis could not be verified here, these data would suggest there might be at least two *Wolbachia* variants (represented by supervised oligotypes AT and AC) while the remaining supervised oligotypes GT, AG and CC possibly represent artefactual variations due to sequencing errors or polymerase errors. These two potential *Wolbachia* variants AT and AC differed by one nucleotide, representing only 0.2% variation for a 428 bp fragment. These subtle nucleotide variations might have been obscured using OTU clustering at 97% or 99% in high-throughput sequencing and analysis, and therefore remained undetected till now. The presence of one nucleotide position variation at the 16S rRNA gene level could nevertheless suggest some larger differences at the genome level. These data are congruent with our previous analysis showing 168 to 612 Single Nucleotide Variants (SNV) for each of the four *Wolbachia* Metagenome-Assembled Genomes (MAGs) reconstructed from the ovaries metagenomes of *Culex* individuals [34]. Of note, these values might be underestimated since the different MAGs were not complete, showing between 93 and 127 contigs. In particular, MAGs were missing some of the highly-variable phage sequences that could not be assembled due to repetitive sequences leading to assembly breaks. These data could also indicate a horizontal transmission of rare *Wolbachia* variants.

These results differ from previous studies showing a unique *Wolbachia* 16S rDNA sequence and the lack of single nucleotide polymorphism in distinct *C. pipiens* populations from different geographic localities distances using Sanger sequencing [80], which imply a shallow sequencing effort (and therefore a small probability to detect low frequency variants with this sequencing method). However, other studies have been showing high levels of multiple *Wolbachia* infection in other insects like the ant *Formica exsecta* [81], with the hosts harboring up to five distinct *Wolbachia* strains. Authors revealed a maximum of 3% variations in the highly variable *wsp* gene, indicating highly divergent *Wolbachia* strains for this species. In addition, Arai and colleagues [82] also found a triple *Wolbachia* infection in tea pest *Homona magnanima*. The authors established uninfected and singly infected lines using antibiotics and revealed distinct CI intensities and/or mutualistic effects between the different *Wolbachia* lines [82]. Although the putative *Wolbachia* variants detected here differed by much less nucleotides, it remains unknown whether they may have distinct roles and induce distinct host-phenotype. Related to these issues, the presence of dominant and rare *Wolbachia* variants within the same individual would also suggest caution is needed in programs of *Wolbachia* transinfection. Overall, we believe the two supervised oligotypes AT and AC are worth to be noted, although the biological significance of these *Wolbachia* sequences is a question for further study.

4.3. The importance of accounting for both bacterial relative and absolute abundance in fine-scale microbiota analysis

In general, this study confirms that the dominance of *Wolbachia* on most mosquito samples might have hampered estimating the actual diversity of symbionts associated with mosquitoes in previous studies. qPCR analysis targets the relative abundance of *Wolbachia* per host copy number and presumably better represents the actual *Wolbachia* titer. However, it might not be sensitive enough to distinguish and account for closely related and rare variants detected through high-throughput sequencing. As expected, analyzing set of mosquito samples with supposedly no to very low *Wolbachia* abundance (*e.g.*, Slab TC) and therefore not dominated by a single bacteria, allowed to shed light on clear patterns of closely related variants co-occurring in mosquito specimens, like for *Elizabethkingia* and *Erwinia*. Although the *Elizabethkingia* variants differed by a single nucleotide at the level of 16S rRNA gene amplicon-level, they might not be functionally redundant. The presence of closely related symbionts in multicellular hosts is now commonly reported, with several studies reporting some generalized niche partitioning between symbionts. Brochet and colleagues [83] showed niche partitioning allowed the co-existence of closely related bacteria in the gut of honey bees. Similarly, functional diversity was shown to enable the coexistence of multiple strains and niche partitioning in deep-sea *Bathymodiolus* mussels [84–85] or in the *Rimicaris exoculata* holobiont [86]. Overall, the mosquito microbiota can be acquired from several sources (transovarian through the germline or from the larvae during oviposition or acquisition from the aquatic environment). Although the microbiota supposedly evolves all along the life stage of mosquitoes (being at least partially eliminated from larvae to adults) [87–88], niche partitioning and functional diversity might allow symbiosis to sustain itself together with the resilience and evolution of its host.

5. Conclusion

Overall, this study allowed a high-resolution characterization of mosquito microbiota and suggests that co-infection pattern of symbionts could be common in mosquito species. Future studies on the maintenance, and importance of bacterial variation at the genomic-scale will allow understanding the possible phenotypic effect of multiple infection in mosquitoes as well as the evolutionary forces leading to this pattern. Shedding light on the functional implications of the less-abundant and neglected members of the mosquito microbiota (or the lack thereof) and their evolutionary history require additional strategies to reconstruct novel genomes from metagenomes, study microbial population genetics *in situ*, and pangenomic analyses to compare their gene pool.

6. Code availability

A reproducible bioinformatics workflow including scripts and supplementary datas used for this study is available at https://zenodo.org/badge/latestdoi/402375335.

7. Data availability

Raw sequencing data are available through the European Nucleotide Archive (PRJEB43079).

CRediT authorship contribution statement

Hans Schrieke: Investigation, Writing -original draft. **Loïs Maignien:** Investigation, Methodology, Writing - review & editing. **Florentin Constancias:** Investigation, Methodology. **Florian Trigodet:** Methodology. **Sarah Chakloute:** Methodology. **Ignace Rakotoarivony:** Methodology. **Albane Marie:** Methodology. **Gregory L'Ambert:** Methodology. **Patrick Makoundou:** Investigation, Methodology. **Nonito Pages:** Investigation, Methodology.

A. Murat Eren: Investigation, Methodology, Writing -review & editing. **Mylène Weill:** Investigation, Writing -review & editing. **Mathieu Sicard:** Investigation, Writing -review & editing. **Julie Reveillaud:** Investigation, Supervision, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Frederic Mahé, Hélène Vignes, and Renata Servan de Almeida for their insights and help with computational, sequencing and logistic matters, respectively. We are also thankful to Sandra Unal for insectories maintenance. This work was supported by the ERC RosaLind Starting Grant "948135" to JR, the INRAe Animal Health Department and Occitany Region to HS, the EU project MALIN and the Guadeloupe Regional Council under the European Research and Development Funds (ERDF) 2014- 2020 program (Grant 2018-FED-1084) to NP.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.12.019.

References

[1] Rahman MT, Sobur MA, Islam MS, Ievy S, Hossain MJ, et al. Zoonotic diseases: etiology, impact, and control. Microorganisms 2020;8(9).
[2] Ahmed S, Dávila JD, Allen A, Haklay MM, Tacoli C, et al. Does urbanization make emergence of zoonosis more likely? Evidence, myths and gaps. Environment and Urbanization. 2019;31(2):443–60.
[3] Cunningham AA, Daszak P, Wood JLN. One Health, emerging infectious diseases and wildlife: two decades of progress? Philos Trans R Soc B: Biol Sci 2017;372(1725):20160167.
[4] Franklinos LHV, Jones KE, Redding DW, Abubakar I. The effect of global change on mosquito-borne disease. Lancet Infect Dis 2019;19(9):e302–12.
[5] Hemingway J, Hawkes NJ, McCarroll L, Ranson H. The molecular basis of insecticide resistance in mosquitoes. Insect Biochem Mol Biol 2004;34 (7):653–65.
[6] Dusfour I, Vontas J, David J-P, Weetman D, Fonseca DM, et al. Management of insecticide resistance in the major *Aedes* vectors of arboviruses: advances and challenges. PLoS NeglTrop Dis 2019;13(10):e0007615.
[7] Huang YJ, Higgs S, Vanlandingham DL. Biological control strategies for mosquito vectors of arboviruses. Insects 2017;8(1):21.
[8] Benelli G, Jeffries C, Walker T. Biological control of mosquito vectors: past, present, and future. Insects 2016;7(4):52.
[9] Achee NL, Grieco JP, Vatandoost H, Seixas G, Pinto J, et al. Alternative strategies for mosquito-borne arbovirus control. PLoS NeglTrop Dis 2019;13(1): e0006822.
[10] Cuthbert RN, Callaghan A, Sentis A, Dalal A, Dick JTA. Additive multiple predator effects can reduce mosquito populations. Ecol Entomol 2020;45 (2):243–50.
[11] Flores HA, O'Neill SL. Controlling vector-borne diseases by releasing modified mosquitoes. Nat Rev Microbiol 2018;16(8):508–18.
[12] Bonneau M, Caputo B, Ligier A, Caparros R, Unal S, et al. Variation in *Wolbachia cidB* gene, but not *cidA*, is associated with cytoplasmic incompatibility *mod* phenotype diversity in *Culex pipiens*. Mol Ecol 2019;28(21):4725–36.
[13] Wang S, Ghosh AK, Bongio N, Stebbings KA, Lampe DJ, et al. Fighting malaria with engineered symbiotic bacteria from vector mosquitoes. PNAS 2012;109 (31):12734–9.
[14] Capone A, Ricci I, Damiani C, Mosca M, Rossi P, et al. Interactions between *Asaia*, *Plasmodium* and *Anopheles*: new insights into mosquito symbiosis and implications in Malaria symbiotic control. Parasites Vectors 2013;6(1):182.
[15] Wu P, Sun P, Nie K, Zhu Y, Shi M, et al. A gut commensal bacterium promotes mosquito permissiveness to arboviruses. Cell Host Microbe 2019;25 (1):101–112.e5.
[16] Bai L, Wang L, Vega-Rodríguez J, Wang G, Wang S. A gut symbiotic bacterium *Serratia marcescens* renders mosquito resistance to *Plasmodium* infection through activation of mosquito immune responses. Front Microbiol 2019;10:1580.
[17] Correa MA, Matusovsky B, Brackney DE, Steven B. Generation of axenic *Aedes aegypti* demonstrate live bacteria are not required for mosquito development. Nat Commun 2018;9(1):4464.
[18] Romoli O, Schönbeck JC, Hapfelmeier S, Gendrin M. Production of germ-free mosquitoes via transient colonisation allows stage-specific investigation of host-microbiota interactions. Nat Commun 2021;12(1):942.
[19] Chouaia B, Rossi P, Epis S, Mosca M, Ricci I, et al. Delayed larval development in *Anopheles* mosquitoes deprived of *Asaia* bacterial symbionts. BMC Microbiol 2012;12(1):S2.
[20] Coon KL, Vogel KJ, Brown MR, Strand MR. Mosquitoes rely on their gut microbiota for development. Mol Ecol 2014;23(11):2727–39.
[21] Gaio A de O, Gusmão DS, Santos AV, Berbert-Molina MA, Pimenta PFP et al. Contribution of midgut bacteria to blood digestion and egg production in *Aedes aegypti* (diptera: culicidae) (L.). Parasites & Vectors. 2011;4(1):105.
[22] Carissimo G, Pondeville E, McFarlane M, Dietrich I, Mitri C, et al. Antiviral immunity of *Anopheles gambiae* is highly compartmentalized, with distinct roles for RNA interference and gut microbiota. Proc Natl Acad Sci U S A 2015;112(2):E176–85.
[23] Saraiva RG, Kang S, Simões ML, Angleró-Rodríguez YI, Dimopoulos G. Mosquito gut antiparasitic and antiviral immunity. Dev Comp Immunol 2016;64:53–64.
[24] Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A 2006;103(32):12115–20.
[25] Lynch MDJ, Neufeld JD. Ecology and exploration of the rare biosphere. Nat Rev Microbiol 2015;13(4):217–29.
[26] Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, et al. Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J 2017;11 (4):853–62.
[27] Eren AM, Morrison HG, Lescault PJ, Reveillaud J. Vineis JH and al. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J 2015;9(4):968–79.
[28] Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, et al. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol 2013;4(12):1111–9.
[29] Fisher JC, Eren AM, Green HC, Shanks OC, Morrison HG, et al. Comparison of sewage and animal fecal microbiomes by using oligotyping reveals potential human fecal indicators in multiple taxonomic groups. Appl Environ Microbiol 2015;81(20):7023–33.
[30] Reveillaud J, Maignien L, Eren AM, Huber JA, Apprill A, et al. Host-specificity among abundant and rare taxa in the sponge microbiome. ISME J 2014;8 (6):1198–209.
[31] Utter DR, Mark Welch JL, Borisy GG. Individuality, stability, and variability of the plaque microbiome. Front Microbiol 2016;7.
[32] Coon KL, Brown MR, Strand MR. Mosquitoes host communities of bacteria that are essential for development but vary greatly between local habitats. Mol Ecol 2016;25(22):5806–26.
[33] Duron O, Bernard C, Unal S, Berthomieu A, Berticat C, et al. Tracking factors modulating cytoplasmic incompatibilities in the mosquito *Culex pipiens*. Mol Ecol 2006;15(10):3061–71.
[34] Reveillaud J, Bordenstein SR, Cruaud C, Shaiber A, Esen ÖC, et al. The *Wolbachia* mobilome in *Culex pipiens* includes a putative plasmid. Nat Commun 2019;10 (1):1051.
[35] Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 2013;79(17):5112–20.
[36] Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using Illumina paired-end technology. PLoS ONE 2013;8(6):e66643.
[37] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses, BMC Biology. 2014; 12:87.
[38] Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 2018;6(1):226.
[39] Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol 2007;73(16):5261–7.
[40] Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, et al. Dada2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13(7):581–3.
[41] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl Acids Res. 2013;41(Database issue):D590-D596.
[42] McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 2013;8(4):e61217.
[43] R Core Team. R: The R Project for statistical computing. R version 3.6.3. Published February 29, 2020. Accessed June 7, 2021. https://www.r-project.org/.
[44] McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, et al. Methods for normalizing microbiome data: an ecological perspective. Methods Ecol Evol 2019;10(3):389–400.
[45] Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P et al. vegan: community ecology package. R package version 2.5-6. Published 2019. Accessed June 7, 2021. https://CRAN.R-project.org/package=vegan.
[46] Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biometrical J 2008;50(3):346–63.
[47] Alekseyenko AV. Multivariate Welch *t* -test on distances. Bioinformatics 2016;32(23):3552–8.

[48] Kandlikar GS, Gold ZJ, Cowen MC, Meyer RC, Freise AC et al. ranacapa: An R package and Shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations. *F1000Res*. 2018;7:1734.
[49] Berticat C, Rousset F, Raymond M, Berthomieu A, Weill M. High *Wolbachia* density in insecticide-resistant mosquitoes. Proceedings Biological sciences / The Royal Society. 2002;269:1413–6.
[50] Weill M, Berticat C, Raymond M, Chevillon C. Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. Anal Biochem 2000;285(2):267–70.
[51] Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 2021;6(1):3–6.
[52] Bonneau M, Atyame C, Beji M, Justy F, Cohen-Gonsaud et al. *Culex pipiens* crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*. Nat Commun. 2018;9(1):319.
[53] Ross PA, Callahan AG, Yang Q, Jasper M, Arif MAK, et al. An elusive endosymbiont: Does *Wolbachia* occur naturally in *Aedes aegypti*? Ecol Evol 2020;10(3):1581–91.
[54] Sicard M, Bonneau M, Weill M. *Wolbachia* prevalence, diversity, and ability to induce cytoplasmic incompatibility in mosquitoes. Curr Opin Insect Sci 2019;34:12–20.
[55] Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res*. 2015;43(Database issue):D593-D598.
[56] Muturi EJ, Kim C-H, Bara J, Bach EM, Siddappaji MH. *Culex pipiens* and *Culex restuans* mosquitoes harbor distinct microbiota dominated by few bacterial taxa. Parasites Vectors 2016;9(1):18.
[57] Novakova E, Woodhams DC, Rodríguez-Ruano SM, Brucker RM, Leff JW, et al. Mosquito microbiome dynamics, a background for prevalence and seasonality of West Nile Virus. Front Microbiol 2017;8.
[58] Muturi EJ, Ramirez JL, Rooney AP, Kim CH. Comparative analysis of gut microbiota of mosquito communities in central Illinois. PLoS NeglTrop Dis 2017;11(2):e0005377.
[59] Hegde S, Khanipov K, Albayrak L, Golovko G, Pimenova M, et al. Microbiome interaction networks and community structure from laboratory-reared and field-collected *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus* mosquito vectors. Front Microbiol 2018;9.
[60] Thongsripong P, Chandler JA, Green AB, Kittayapong P, Wilcox BA, et al. Mosquito vector-associated microbiota: metabarcoding bacteria and eukaryotic symbionts across habitat types in Thailand endemic for dengue and other arthropod-borne diseases. Ecol Evol 2017;8(2):1352–68.
[61] Scolari F, Casiraghi M, Bonizzoni M. *Aedes* spp. and their microbiota: a review. Front Microbiol 2019;10.
[62] Gusmão DS, Santos AV, Marini DC, Bacci M, Berbert-Molina MA, et al. Culture-dependent and culture-independent characterization of microorganisms associated with *Aedes aegypti* (Diptera: Culicidae) (L.) and dynamics of bacterial colonization in the midgut. Acta Trop 2010;115(3):275–81.
[63] Mancini MV, Damiani C, Accoti A, Tallarita E, Cappelli A, et al. Estimating bacteria diversity in different organs of nine species of mosquito by next generation sequencing. BMC Microbiol 2018;18(1):126.
[64] Ramirez JL, Souza-Neto J, Cosme RT, Rovira J, Ortiz A, et al. Reciprocal tripartite interactions between the *Aedes aegypti* midgut microbiota, innate immune system and Dengue virus influences vector competence. PLoS NeglTrop Dis 2012;6(3):e1561.
[65] David MR, Santos LMB, Vicente ACP, Maciel-de-Freitas R. Effects of environment, dietary regime and ageing on the dengue vector microbiota: evidence of a core microbiota throughout *Aedes aegypti* lifespan. Mem Inst Oswaldo Cruz 2016;111(9):577–87.
[66] Muturi EJ, Lagos-Kutz D, Dunlap C, et al. Mosquito microbiota cluster by host sampling location. Parasites Vectors 2018;11(1):468.
[67] Boissière A, Tchioffo MT, Bachar D, Abate L, Marie A, et al. Midgut microbiota of the Malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. PLoS Pathog 2012;8(5):e1002742.
[68] Williamson BD, Hughes JP, Willis AD. A multiview model for relative and absolute microbial abundances. Biometrics 2021;1:14.
[69] Kämpfer P, Matthews H, Glaeser SP, Martin K, Lodders N, et al. *Elizabethkingia anophelis* sp. nov., isolated from the midgut of the mosquito *Anopheles gambiae*. Int J Syst Evol Microbiol 2011;61(11):2670–5.
[70] Chen S, Bagdasarian M, Walker ED. *Elizabethkingia anophelis*: Molecular manipulation and interactions with mosquito hosts. Appl Environ Microbiol 2015;81(6):2233–43.
[71] Ngwa CJ, Glöckner V, Abdelmohsen UR, Scheuermayer M, Fischer M, et al. 16S rRNA gene-based identification of *Elizabethkingia meningoseptica* (Flavobacteriales: Flavobacteriaceae) as a dominant midgut bacterium of the Asian malaria vector *Anopheles stephensi* (Dipteria: Culicidae) with antimicrobial activities. J Med Entomol 2013;50(2):404–14.
[72] Bahia AC, Dong Y, Blumberg BJ, Mlambo G, Tripathi A, et al. Exploring *Anopheles* gut bacteria for *Plasmodium* blocking activity. Environ Microbiol 2014;16(9):2980–94.
[73] Perrin A, Larsonneur E, Nicholson AC, Edwards DJ, Gundlach KM, et al. Evolutionary dynamics and genomic features of the *Elizabethkingia anophelis* 2015 to 2016 Wisconsin outbreak strain. Nat Commun 2017;8(1):15483.
[74] Basset A, Khush RS, Braun A, Gardan L, Boccard F, et al. The phytopathogenic bacteria *Erwinia carotovora* infects *Drosophila* and activates an immune response. Proc Natl Acad Sci U S A. 2000;97(7):3376–81.
[75] Chouaia B, Rossi P, Montagna M, Ricci I, Crotti E, et al. Molecular evidence for multiple infections as revealed by typing of *Asaia* bacterial symbionts of four mosquito species. Appl Environ Microbiol 2010;76:7.
[76] Kämpfer P, Chandel K, Prasad GBKS, Shouche YS, Veer V. *Chryseobacterium culicis* sp. nov., isolated from the midgut of the mosquito *Culex quinquefasciatus*. Int J Syst Evol Microbiol 2010;60(10):2387–91.
[77] Lee J, Hwang S, Cho S. Immune tolerance to an intestine-adapted bacteria, *Chryseobacterium* sp., injected into the hemocoel of *Protaetia brevitarsis seulensis*. Sci Rep. 2016;6(1):31722.
[78] Cordaux R, Gilbert C. Evolutionary significance of *Wolbachia*-to-animal horizontal gene transfer: female sex determination and the f element in the isopod *Armadillidium vulgare*. Genes 2017;8(7).
[79] Tian R-M, Cai L, Zhang W-P, Cao H-L, Qian P-Y. Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA Gene. Genome Biol Evol 2015;7(8):2310–20.
[80] O'Neill SL, Giordano R, Colbert AM, Karr TL, Robertson HM. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc Natl Acad Sci U S A 1992;89(7):2699–702.
[81] Reuter M, Keller L. High Levels of multiple *Wolbachia* infection and recombination in the ant *Formica exsecta*. Mol Biol Evol 2003;20(5):748–53.
[82] Arai H, Hirano T, Akizuki N, Abe A, Nakai M, et al. Multiple infection and reproductive manipulations of *Wolbachia* in *Homona magnanima* (Lepidoptera: Tortricidae). Microb Ecol 2019;77(1):257–66.
[83] Brochet S, Quinn A, Mars RA, Neuschwander N, Sauer U et al. Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. eLife. 2021;10:e68583.
[84] Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, et al. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J 2016;10(4):990–1001.
[85] Ansorge R, Romano S, Sayavedra L, Porras MÁG, Kupczok A, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol 2019;4(12):2487–97.
[86] Cambon-Bonavita M-A, Aubé J, Cueff-Gauchard V, Reveillaud J. Niche partitioning in the *Rimicaris Exoculata* holobiont: the case of the first symbiotic *Zetaproteobacteria*. Microbiome 2021;9(87).
[87] Dickson LB, Jiolle D, Minard G, Moltini-Conclois I, Volant S, et al. Carryover effects of larval exposure to different environmental bacteria drive adult trait variation in a mosquito vector. Sci Adv 2017;3(8):e1700585.
[88] Heu K, Gendrin M. Le microbiote de moustique et son influence sur la transmission vectorielle. Biologie Aujourd'hui 2018;212(3–4):119–36.