Video Article

Tick Microbiome Characterization by Next-Generation 16S rRNA Amplicon Sequencing

Lisa Couper[1], Andrea Swei[1]

[1]Department of Biology, San Francisco State University

Correspondence to: Andrea Swei at aswei@sfsu.edu

URL: https://www.jove.com/video/58239
DOI: doi:10.3791/58239

Keywords: Biology, Issue 138, Tick, vector, microbiome, 16S rRNA, amplicon, next-generation sequencing, *Ixodes*

Date Published: 8/25/2018

Citation: Couper, L., Swei, A. Tick Microbiome Characterization by Next-Generation 16S rRNA Amplicon Sequencing. *J. Vis. Exp.* (138), e58239, doi:10.3791/58239 (2018).

Abstract

In recent decades, vector-borne diseases have re-emerged and expanded at alarming rates, causing considerable morbidity and mortality worldwide. Effective and widely available vaccines are lacking for a majority of these diseases, necessitating the development of novel disease mitigation strategies. To this end, a promising avenue of disease control involves targeting the vector microbiome, the community of microbes inhabiting the vector. The vector microbiome plays a pivotal role in pathogen dynamics, and manipulations of the microbiome have led to reduced vector abundance or pathogen transmission for a handful of vector-borne diseases. However, translating these findings into disease control applications requires a thorough understanding of vector microbial ecology, historically limited by insufficient technology in this field. The advent of next-generation sequencing approaches has enabled rapid, highly parallel sequencing of diverse microbial communities. Targeting the highly-conserved 16S rRNA gene has facilitated characterizations of microbes present within vectors under varying ecological and experimental conditions. This technique involves amplification of the 16S rRNA gene, sample barcoding via PCR, loading samples onto a flow cell for sequencing, and bioinformatics approaches to match sequence data with phylogenetic information. Species or genus-level identification for a high number of replicates can typically be achieved through this approach, thus circumventing challenges of low detection, resolution, and output from traditional culturing, microscopy, or histological staining techniques. Therefore, this method is well-suited for characterizing vector microbes under diverse conditions but cannot currently provide information on microbial function, location within the vector, or response to antibiotic treatment. Overall, 16S next-generation sequencing is a powerful technique for better understanding the identity and role of vector microbes in disease dynamics.

Video Link

The video component of this article can be found at https://www.jove.com/video/58239/

Introduction

The resurgence and spread of vector-borne diseases in recent decades pose a serious threat to global human and wildlife health. Effective vaccines are lacking for a majority of these diseases, and control efforts are hindered by the complex biological nature of vectors and vector-host interactions. Understanding the role of microbial interactions within a vector in pathogen transmission can allow for the development of novel strategies which circumvent these challenges. In particular, interactions between vector-associated microbial commensals, symbionts, and pathogens, referred to as the microbiome, may have important consequences for pathogen transmission. Overwhelming evidence now supports this assertion, with examples demonstrating a link between the vector microbiome and competence for diseases such as malaria, Zika virus, and Lyme disease[1,2,3]. However, translating these findings into strategies for disease control requires a far more detailed understanding of the structure, function, and origin of vector microbiomes. Identification and characterization of the vector microbial community under varying ecological and experimental conditions constitute an important path forward in this field.

A procedure for identifying the microbial residents of a pathogen vector is provided here by utilizing the Western black-legged tick, *Ixodes pacificus*, a vector species of the Lyme disease pathogen *Borrelia burgdorferi*. While ticks harbor more types of human pathogens than any other arthropod, relatively little is known about the biology and community ecology of tick microbiomes[4]. It is evident that ticks harbor a diverse array of viruses, bacteria, fungi, and protozoans which include commensals, endosymbionts, and transient microbial residents[5,4]. Prior work has demonstrated strong variations in *Ixodes* microbiomes associated with geography, species, sex, life stage, and blood meal source[6,7,8]. However, the mechanisms underlying this variation remain unknown and warrant more detailed investigations of the origin and assembly of these microbial communities. Ticks can acquire microbes through vertical transmission, contact with hosts, and uptake from the environment through the spiracles, mouth, and anal pore[9]. Understanding the factors shaping the initial formation and development of the tick microbiome, specifically the relative contribution of vertical and environmental transmission, is important for understanding the natural patterns and variations in tick microbiome diversity and how these communities interact during pathogen transmission, with possible applications to disease or vector control.

Powerful molecular techniques, such as next-generation sequencing, now exist for identifying microbial communities and can be employed to characterize vector microbiomes under diverse environmental or experimental conditions. Prior to the advent of these high-throughput sequencing approaches, the identification of microbes relied predominantly on microscopy and culture. While microscopy is a rapid and easy technique, morphological methods for identifying microbes are inherently subjective and coarse and limited by low sensitivity and detection[10]. Culture-based methods are broadly used for microbial identification and can be used to determine susceptibility of microbes to drug treatments[11]. However, this method also suffers from low sensitivity, as it has been estimated that fewer than 2% of environmental microbes can be cultured in a laboratory setting[12]. Histological staining approaches have also been employed to detect and localize specific microbes within vectors, enable investigations of various taxa distributions within the tick, and study hypotheses about microbial interactions. However, prior knowledge of microbial identity is required for selecting the appropriate stains, making this approach ill-suited for microbial characterization and identification. Furthermore, histological staining is a highly time-intensive, laborious process and does not scale well for large sample sizes. Traditional molecular approaches such as Sanger sequencing are similarly limited in their sensitivity and detection of diverse microbial communities.

Next-generation sequencing allows for the rapid identification of microbes from a large number of samples. The presence of standard marker genes and reference databases further enables enhanced taxonomic resolution, often to the genus or species level. Small subunit ribosomal RNAs are frequently used to achieve this goal, with 16S rRNA being the most common due to the presence of conserved and variable regions within the gene, allowing for the creation of universal primers with unique amplicons for each bacterial species[13,14]. This report details a procedure for identifying taxa in the tick microbiome through 16S rRNA next-generation sequencing. In particular, this protocol emphasizes the steps involved in preparing samples for sequencing. More generalized details on the sequencing and bioinformatics steps are provided, as there are a variety of sequencing platforms and analysis programs currently available, each with extensive existing documentation. The overall feasibility of this next-generation sequencing approach is demonstrated by applying it to an investigation of microbial community assembly within a key disease vector.

Protocol

1. Tick Collection and Surface Sterilization

1. Collect ticks by dragging a 1 m^2 white cloth over a tick-associated habitat, removing ticks attached to host species, or rearing ticks in the lab[15,16]. Use fine forceps to manipulate ticks and store them at -80 °C.
2. Place ticks in the individual PCR tubes and remove surface contaminants by vortexing for 15 s successively with 500 µL of hydrogen peroxide (H_2O_2), 70% ethanol, and ddH_2O.
3. Place the ticks in a new PCR tube and allow them to air-dry.
4. In this tube, mechanically disrupt tick tissues by crushing the ticks with a mortar and pestle (used in this study), using small beads in a bead beater, or cutting the tick apart with a scalpel.

2. DNA Purification

1. Purify DNA from individual ticks, following the instructions provided in a commercially available DNA extraction kit. Elute for a final volume of 100 µL.
   NOTE: Refer to **Figure 1** for an overview of steps 2.2-2.8. Alternative extraction methods include phenol-chloroform extraction[17,18], ethanol precipitation[19], or extraction with a chelating material[20,21].
2. Confirm successful DNA purification using a fluorometer or spectrophotometer. For fluorometry, use 10 µL of DNA template in a 190 µL double-stranded DNA assay. The expected yield is 0.1-1.0 ng/µL[22]. For spectrophotometry, use 1 µL of DNA template in a nucleic acid quantification. The expected A260/280 ratio is approximately 1.8[23].
3. If not proceeding immediately to amplification, store the samples at -20 °C.

3. 16S rRNA Gene Amplification

1. Set up the amplicon PCR in a 27.5 µL reaction containing: 5 µL of each primer at 1 µM; 12.5 µL of commercially available PCR mix; and 5 µL of DNA extracted from individual ticks at a 5 ng/µL concentration.
   NOTE: Use the following primers for amplification of the hypervariable V3-V4 region of the 16S rRNA gene[13]:
   5'--TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG--3',
   5'--GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC--3'
2. Move tubes to a thermocycler programmed for: an initial denaturation at 95 °C for 3 min; 25 cycles of 1) 95 °C for 30 s, 2) 57 °C for 30 s, 3) 72 °C for 30 s; a final extension at 72 °C for 5 min; and a final hold at 10 °C.
3. Once the PCR is done, visualize the PCR product by loading 4-6 µL/sample on a 1.5% agarose gel. Look for a band at 460 bp to confirm amplification.
   NOTE: Amplicon PCR product may not be visible on the gel if the sample concentration is too low (< 10 ng). Primer dimer presence is common in low-concentration samples. If other non-target bands are present, adjust the annealing temperature or decrease the number of cycles.
4. If not proceeding immediately to purification, store the samples at 4 °C.

4. 16S Amplicon Purification

1. Using a new PCR tube for each sample, combine the PCR product with ddH_2O to obtain 60 µL total. For samples with low DNA concentrations, perform amplicon PCR in triplicate to reduce amplification bias and pool the samples to concentrate.
2. Bring paramagnetic beads to room temperature and vortex them well before use.
3. Add 48 µL of paramagnetic beads to 60 µL of the sample and incubate for 5 min. Place the tubes on a magnetic rack for 5 min until solution becomes clear. Remove the supernatant.

NOTE: This bead concentration targets the removal of primer dimers (< 60 bp). If needed, adjust the bead concentration to remove other non-specific banding.

4. With tubes on magnetic rack, add 500 µL of freshly-prepared 80% EtOH. Immediately after adding EtOH to all tubes, pipette out the liquid. Do not remove the beads. Repeat the EtOH addition and removal one more time.
5. Air-dry the samples to remove excess EtOH by leaving the tubes open on the magnetic rack for 5 min, or until small cracks are visible in the beads.
6. Add 20 µL of TE buffer and incubate the samples off the magnet, at room temperature, for 5-10 min.
7. Place the tubes back on the magnet. Once the beads and liquid are separated, transfer the supernatant to a fresh tube to obtain the cleaned PCR product.
8. (Optional) Visualize the product to confirm successful amplicon purification by loading 4-6 µL/sample on a 1.5% gel. The 460 bp band should be visible, and there should be no primer dimer.
9. If not proceeding immediately to index PCR, store the samples 4 °C.

5. Sample Barcoding and Purification

1. Assign unique primer combinations to each sample by selecting either forward or reverse primers, or both, from a commercially available library index kit.
   NOTE: Uniquely labeling samples allows for differentiation after sequencing. Kits typically provide enough primers to sequence 96-384 samples.
2. Attach the dual primers, or indices, to samples by performing PCR in a 25 µL reaction containing 2.5 µL of each primer (N7xx and S5xx), 12.5 µL of commercially available PCR master mix, 5 µL of ddH_2O, and 2.5 µL of cleaned amplicon product.
3. Move the tubes to a thermocycler programmed for: an initial denaturation at 95 °C for 3 min; 8-14 cycles of 1) 95 °C for 30 s, 2) 55 °C for 30 s, 3) 72 °C for 30 s; a final extension at 72 °C for 5 min; and a final hold at 10 °C.
4. Once the PCR is done, visualize the PCR product by loading 4-6 µL/sample on a 1.5% agarose gel. Look for a band at 550 bp to confirm amplification.
   NOTE: Use the visualization results to inform index PCR cycling conditions. Lower the cycle count to mitigate non-specific binding or increase the cycle count to obtain visible bands for each sample.
5. Repeat the clean-up procedure listed in step 4. To avoid dilution during clean-up, perform index PCR in duplicate and pool the product here.
6. If not proceeding immediately to library quantification and normalization, store the samples at 4 °C.

6. Library Quantification and Normalization

1. Estimate the concentration of each purified, barcoded product from step 5.5 using a fluorometer or spectrophotometer (see step 2.2). Dilute the samples in TE buffer to obtain concentrations of approximately 1 pM.
   NOTE: Library quantification is necessary to achieve the sample loading concentration recommended for the sequencing platform. Library quantification is achieved through qPCR, but estimating the sample concentration prior to qPCR saves reagents and time. The 1 pM concentration is recommended to maximize the accuracy of library quantification, as this is the mean concentration of the qPCR standards provided in the quantification kit[24].
2. Perform qPCR in a 10 µL reaction containing 6.0 µL of qPCR master mix (from library quantification kit), 2.0 µL of ddH_2O, and 2.0 µL of the sample or standard. Run each sample and standard in triplicate for greater accuracy and precision. Run each sample at three or more dilution levels (*e.g.*, 0.1 pm, 1 pm, and 10 pm for estimated starting concentrations) to ensure accurate quantification.
   NOTE: Detailed information about the provided primers and standards will vary based on the quantification kit used and are available in the products technical data sheet. Refer to the **Table of Materials** for the quantification kit used in this study.
3. Move the qPCR plate or tubes to a real-time PCR instrument programmed for: an initial denaturation of 95 °C for 5 min; 35 cycles of 1) 95 °C for 30 s and 2) 60 °C for 45 s; and a dissociation step of 1) 95 °C for 15 s, 2) 60 °C for 30 s, and 3) 95 °C for 15 s.
4. Calculate the average starting concentration for each sample, using the quantification values obtained from the qPCR results and the sample dilution levels used.
   NOTE: For example, an average concentration of 2 pM for a sample diluted 1:100,000 yields an original starting concentration of 200 nM. If none of the dilution levels for a given sample falls within the range of the standard curves, quantification results may not be accurate; in which case, adjust the dilution levels and re-perform qPCR.
5. Dilute each purified, barcoded sample to 4 nM in TE buffer, based on the average concentrations calculated in step 7.5.
   NOTE: For example, for an average sample concentration of 200 nM, dilute the sample 1:50 in TE buffer to achieve a 4 nM concentration. The precise sample concentration will vary based on the sequencing system and reagent kit used. Refer to the sequencing system user guide for the recommended concentration.
6. Create the combined library by adding equal volumes (typically 5-10 µL) of all individual libraries into a single tube.
   NOTE: As all samples should be at a 4 nM concentration, adding equal volumes achieves an equal concentration of all samples in the pooled library.
7. Repeat steps 6.2-6.4 on the combined library to confirm the 4 nM concentration.
8. Calculate the combined library concentration and dilute or re-constitute the final, pooled library using Tris buffer as necessary to achieve 4 nM.
9. If not proceeding immediately to sample loading, store the samples at -20 °C. Perform the sequencing run shortly after quantification to minimize loss or changes to DNA concentration during storage.

7. Library Denaturation and Dilution, and Sequencing Run (perform on the same day)

1. Denature the 4 nM combined library from step 6.9 with NaOH.

NOTE: These final library preparation steps vary for each sequencing system. Refer to the sequencing system user guide for more detailed and updated protocols. New users will likely need to be trained on the usage of the sequencing platform or may send their libraries to a core sequencing facility.

2. Dilute the denatured library to the desired loading concentration using the buffer provided in the sequencing reagent kit (**Table of Materials**).
NOTE: Optimal loading concentrations vary by sequencing system and typically range from 1-250 pM[25].
3. Denature and dilute the sequencing control.
NOTE: Adding a sequencing control corrects for sequencing issues arising from low diversity libraries. Denature the sequencing control using NaOH and dilute to the same concentration as the library.
4. Combine the library and sequencing control.
NOTE: The ratio of sequencing control to combined library will depend on the library diversity and sequencing system used, but it is typically 1:1[26].
5. Load the combined library and sequencing control mixture onto the sequencing flow cell.

8. Amplicon Sequence Analysis

1. Assess the overall success of the run by examining the run metrics on the cloud computing environment corresponding to the sequencing system used.
NOTE: The run metrics, such as number of sequencing reads, will vary based on sequencing platform and reagent kit used. Target values for common sequencing systems are listed in **Table 1**.
2. Download the raw sequencing data and desired bioinformatics open source software, Quantitative Insights into Microbial Ecology (QIIME)[27] or Mothur[28].
NOTE: Both QIIME and Mothur are open-source and free to download. Detailed instructions on using these programs can be found online and are sufficiently detailed for first-time users. The steps below provide a broad overview of a bioinformatics pipeline conducted in QIIME. Familiarity with python is not necessary but will facilitate the implementation of the following scripts
3. Create and validate the mapping file using the "validate_mapping_file.py" script in QIIME.
NOTE: The mapping file contains all the metadata necessary to perform data analysis, including sample ID, amplicon primer sequences, and sample description. The validation step checks that all necessary data have been entered in the proper format.
4. Demultiplex and filter sequences using the "split_libraries.py" script (**Figure 1**).
NOTE: This script assigns barcoded reads to samples based on the index primer combinations and sample IDs input in the mapping file. It also performs several quality filtering steps to control for sequencing error based on user-defined cut-offs for minimum quality score, sequence lengths, and end-trimming.
5. Assign operational taxonomic units (OTUs) to sequences using the "pick_open_references_otus.py" script.
NOTE: In this step, sequences are clustered against a reference sequence collection based on a threshold of identity (typically 97%). Reads which do not match the reference sequence collection are clustered against one another. De novo and closed-reference OTU picking options are also available, but open-reference picking is recommended by QIIME developers.
6. Normalize the OTU table using the "alpha_rarefaction.py" or "normalize_table.py" script.
NOTE: This step corrects for variation in column sums, or total sequence reads per sample, that result from modern sequencing technologies. Normalization can be performed using traditional rarefaction, or through alternative methods such as cumulative sum scaling.
7. Perform several alpha-diversity, beta-diversity, and taxonomic composition diversity analyses at once using the "core_diversity_analysis.py" script.
NOTE: Alternatively, these analyses can be run separately using the individual scripts (*e.g., "*alpha_diversity.py").

Representative Results

A total of 42 ticks from three separate egg clutches and two environmental exposure periods, 0 and 2 weeks in soil, were processed for microbiome sequencing. Each treatment group, considered to be a single clutch and exposure time, contained 6-8 replicate tick samples. These processed tick extracts were loaded onto a next-generation sequencer and yielded 12,885,713 paired-end reads passing filter. Included in this run were 3 negative controls from the extraction step, yielding a total of 211,214 reads (included in previous count). Further run quality metrics along with optimal values for each metric are detailed in **Table 2**. Rarefaction curves, which relate sequencing effort to number of OTUs per sample, indicated that a sequencing depth, or number of unique sequence reads per sample, of 2,129 reads would be sufficient to adequately capture the diversity of the microbial community (**Figure 2**). Rarefaction levels will vary based on sample type and must be determined individually for each sequencing run. After rarefying to this depth, 1,714 OTUs were identified across all samples with an average of 93.3 ± 4.3 OTUs per sample. To avoid downstream analysis issues arising from sparse matrices and to remove potential contaminants, all OTUs not found in at least one sample at ≥ 1% abundance were pooled into a rare general category. Further, the decontam package in R was utilized to identify OTUs over-represented in negative controls relative to real samples, and these OTUs were removed from downstream analysis.

Community ecology analyses were performed using the QIIME diversity analyses workflow to demonstrate the types of alpha diversity, beta diversity, and taxonomy composition diversity output generated using a simple bioinformatics pipeline. For example, weighted and unweighted Unifrac principal coordinates analyses were produced using the "core_diversity_analyses.py" script, and they revealed spatial clustering of larval ticks based on clutch identity (**Figure 3**). Boxplots of OTU counts at varying environmental exposure times were also generated through this script, demonstrating differences in microbiome species richness over time (**Figure 4**). These alpha and beta diversity analyses are performed based on the user-defined categories listed in the mapping file, but figures and statistics on general taxonomic information are also generated for all samples (**Figure 5**).

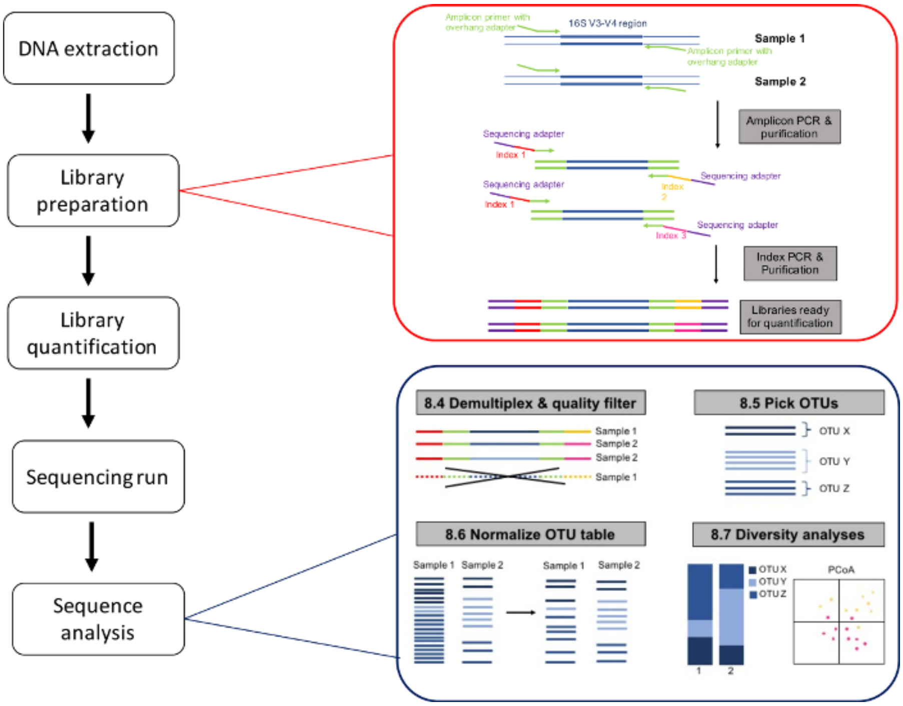

Figure 1: Microbiome sequencing workflow. Major steps of the microbiome sequencing workflow are displayed with call-outs for the library preparation and sequencing analysis steps. Please click here to view a larger version of this figure.

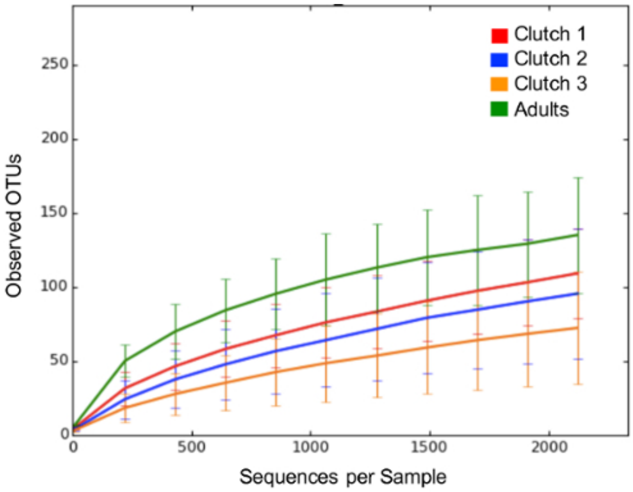

Figure 2: Rarefaction curves for sequence count normalization. Rarefaction curves, shown for each cohort of ticks, relate observed OTU counts to sampling effort. Error bars are present due to replicate samples present within each clutch, and they denote one standard deviation. A sequencing depth should be selected at or beyond the point where the curve becomes stable to adequately capture the full diversity of OTUs. Here, a sequencing depth of 2,000 reads appears appropriate. Please click here to view a larger version of this figure.

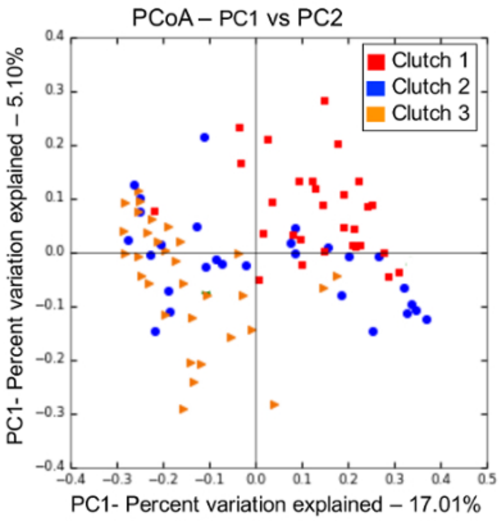

Figure 3: Principal coordinate analysis by clutch. Unweighted Unifrac principal coordinate analysis (PCoA) shows variation in microbiome composition between larval ticks from different clutches, and from adults. Each data point denotes an individual tick. This figure is automatically generated through the QIIME "core_diversity_analyses.py" script. Please click here to view a larger version of this figure.

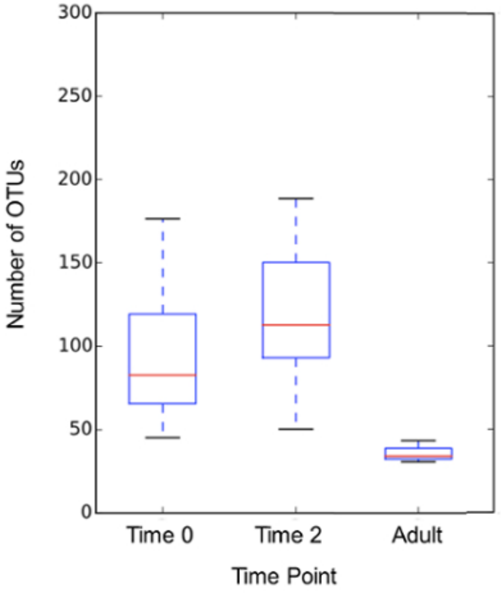

Figure 4: Alpha diversity boxplots by exposure time. Boxplots depicting OTU counts for environmental exposure groups show variation in microbial diversity over time. This figure is automatically generated through the QIIME "core_diversity_analyses.py" script. Please click here to view a larger version of this figure.

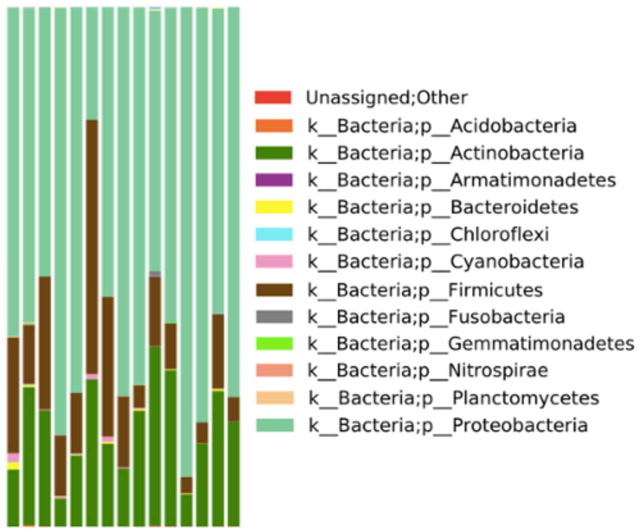

Figure 5: Phylum-level microbial identification for clutch one. Microbiome composition for individual tick samples from clutch one is shown at the phylum level. This figure, as well as summary information at lower taxonomic levels, is automatically generated through the QIIME "core_diversity_analyses.py" script. Please click here to view a larger version of this figure.

Metric	Definition	Our results (MiSeq V3)	Optimal for MiSeq V3	Optimal for HiSeq Rapid Mode
Reads PF	Number of sequencing reads passing the chastity filter. A read passes filter if no more than 1 base call has a chastity value below 0.6 in the first 25 cycles	12,885,713	14-16 million	600 million
Error Rate	Percentage of base pairs called incorrectly during a cycle, based on reads aligned to PhiX control	3.28 ±0.10	*Depends on values for other metrics	*Depends on values for other metrics
% ≥Q30	Percentage of bases with a Q score ≥ 30, indicating a base call accuracy of 99.9%	68.94	> 70%	> 80%
Cluster PF (%)	Percentage of clusters passing the chastity filter.	85.61 ±0.81	90%	90%
Density	Density of clusters on the flowcell - a key metric for data quality and output	552 ± 35 cluster/mm^2	1200-1400 cluster/ mm^2	850-1000 cluster/ mm^2

Table 1: Key performance parameters for NGS output. User data and target values reported based on the sequencing platform and reagent kit used in this study (**Table of Materials**) and a 25% sequencing control addition.

Discussion

Next-generation sequencing of 16S rRNA has become a standard approach for microbial identification and enabled the study of how vector microbiomes affect pathogen transmission. The protocol outlined here details the use of this method to investigate microbial community assembly in *I. pacificus*, a vector species for Lyme disease; however, it can easily be applied to study other tick species or arthropod vector species.

Indeed, 16S rRNA sequencing for microbiome analysis has been used broadly to study the microbiomes of vectors including mosquitoes, psyllids, and tsetse flies[29,30,31]. Other methods available for microbial identification within vectors include microscopy, culturing, and histological staining. These methods may be more appropriate than sequencing if the goal is to identify and describe a novel microbe (microscopy), evaluate the effect of antimicrobial drugs (culture), or localize specific and known microbes within the vector (histological staining). However, these methods suffer from low specificity, detection, and scalability, and thus are less appropriate for identifying the full community of microbes within a vector or characterizing the vector microbiome under varying ecological and experimental conditions. Conversely, high-throughput sequencing of the 16S rRNA gene enables identification of low-abundance and non-culturable bacteria, provides high resolution and detection given the comprehensive reference databases, and can provide high replication depending on coverage needs.

While 16S rRNA sequencing is now widely used for microbial identification, this technique is not without limitations. Principally, microbial contamination can obscure interpretation of the sequencing results and confound biological meaning[32]. Furthermore, given the use of universal bacterial primers and sensitivity to low starting concentrations that are inherent to 16S rRNA sequencing, microbial contamination is common[32]. Sources of contamination include PCR reagents, DNA extraction kits, laboratory surfaces, and the skin and clothing of researchers[33,34,35]. The effects of microbial contaminants can be minimized by working in a sterile lab environment, using negative controls and technical replicates, and keeping a record of all kits and reagents used[36].

In addition to microbial contamination, low-quality sequencing output can greatly hinder the usability of microbiome sequencing data. Data quality can be assessed by a number of run metrics including the number of reads passing filter, the percentage of reads above a Q-score (a measure of the predicted probability of error in base-calling) of 30, and cluster density. Values for these metrics will vary based on the number of samples run, the sequencing system, the reagent cartridge, and the percent sequencing control used, but optimal values based on run conditions are available online. In particular, cluster density, the number of library clusters on a given plane of the flow cell prior to sequencing, is a key parameter for optimizing data quality and yield. Both over and under-clustering can reduce data output and result in samples being excluded from analysis due to insufficient coverage. Poor cluster density often reflects inaccurate library quantification; thus, care should be taken to perform proper DNA clean-up, individual sample quantification, and whole library quantification.

Assuming high data quality and yield are achieved, divergent approaches in data analysis used to overcome statistical challenges of large datasets pose another limitation of this approach. For example, multiple methods exist to address variation in read counts between samples. Rarefaction, which involves sub-sampling reads to achieve an equal sequencing depth across all samples, is frequently used but has been subject to critique recently for wasting large amounts of data and the subjective selection of minimum sequencing depth[37,38,39]. Cumulative sum scaling (CSS), which keeps all sequence reads but weighs them based on the cumulative sum of counts within a given percentile, has been developed as an alternative technique. However, CSS has not been widely adopted due to its relative novelty and the confirmed utility of rarefaction for normalization prior to presence/absence analyses.

Standard data analysis procedures are also lacking for handling sequence reads in negative controls and distinguishing these from low abundance reads. As mentioned, sequencing negative controls generated from the DNA extraction step is recommended to help differentiate true vector microbiome residents from microbial contaminants during analysis. Yet, a standard and statistically rigorous approach for identifying and filtering suspected contaminants is lacking. A common approach is to remove OTUs present in the negative controls from all samples. However, this method may be overly conservative since many of these microbes likely originate from real samples rather than kit reagents[40]. Another common technique involves grouping microbes present at < 1% into a "rare" category under the assumption that microbial contaminants are rare in real samples[8,41]. However, this method may remove true vector microbes that are present at low abundances, particularly when the microbiome is dominated by an endosymbiont as seen in *I. pacificus* and *I. holocyclus*[7,42]. In these cases, detecting rare microbes would require deeper sequencing, developing primers that inhibit the amplification of the endosymbiont during the amplicon PCR[42], or computationally removing the endosymbiont during sequence analysis[7]. Selecting among the various methods to address rare and contaminant OTUs creates the opportunity for subjectivity and bias in microbiome data analysis, which may limit the ability to compare results across studies.

The choice of reference database, necessary for assigning taxonomy and phylogenetic information to sequence reads, presents another opportunity for divergence and subjectivity. While the task of relating sequence reads from a variable gene region to an identified parent genome is inherently challenging, a reliable reference database is crucial for accurate phylogenetic assignment. Multiple databases have emerged to meet this challenge, such as SILVA[43], Greengenes[44], RDP[45], and NCBI[46]. SILVA is the largest of the 16S-based taxonomies, but SILVA, as well as RDP, only provides taxonomic information down to the genus level. Greengenes provides species-level information and is included in metagenomic analyses packages like QIIME, but it has not been updated in over four years. NCBI, while not a primary source for taxonomic information, provides daily updated classifications from user-submitted sequences, but it is uncurated. While selection among the databases typically depends on the users' needs regarding resolution, coverage, and currency, it has a significant impact on phylogenetic assignment and downstream analysis[47,48]. Consensus among investigators regarding which reference database to use is thus imperative for avoiding additional bias in analyses.

Standard approaches to these data processing challenges will likely emerge as 16S rRNA sequencing becomes an increasingly common technique. Continued reductions in sequencing costs and time will further popularize this method. The increased usage of this technology will enable deeper investigations into the composition and ecology of vector microbiomes under diverse conditions. As knowledge of vector microbiome biology is still in its infancy, these types of descriptive studies are a critical first step preceding attempts to leverage the microbiome as a means of vector control. However, to truly understand the role of the microbiome in pathogen transmission, microbial identification must be coupled with knowledge of the functional role of these microbes. RNA sequencing and transcriptomic approaches, which involve mapping and quantifying gene expression, enable inferences into the functional role of microbes but require deep sequencing and fully assembled genomes of target species. Circumventing these challenges, computational tools to predict functional composition from 16S rRNA sequencing data have recently been developed but are not widely adopted yet[49]. The development of such tools, as well as ecological theory, linking microbial identity and the functional role within vectors will increase the utility of 16S rRNA sequencing data in vector microbiome studies.

Disclosures

The authors have nothing to disclose.

Acknowledgements

This work was supported by National Science Foundation grants to A.S. (DEB #1427772, 1745411, 1750037).

References

1. Dong, Y., Manfredini, F., Dimopoulos, G. Implication of the mosquito midgut microbiota in the defense against malaria parasites. *Public Library of Science Pathogens*. **5** (5), (2009).
2. Aliota, M.T., Peinado, S.A., Velez, I.D., Osorio, J.E. The wMel strain of Wolbachia reduces transmission of Zika virus by Aedes aegypti. *Scientific Reports*. **6** (July), 1-7 (2016).
3. Narasimhan, S., *et al.* Gut microbiota of the tick vector Ixodes scapularis modulate colonization of the Lyme disease spirochete. *Cell Host and Microbe*. **15** (1), 58-71 (2014).

4. Clay, K., Fuqua, C. The Tick Microbiome: Diversity, Distribution and Influence of the Internal Microbial Community for a Blood-Feeding Disease Vector. *Critical Needs and Gaps in Understanding Prevention, Amelioration, and Resolution of Lyme and Other Tick-Borne Diseases: The Short-Term and Long-Term Outcomes.* Washington, D.C., October 11-12, 1-22 (2010).
5. Noda, H., Munderloh, U.G., Kurtti, T.J. Endosymbionts of Ticks and Their Relationship to Wolbachia spp . and Tick-Borne Pathogens of Humans and Animals. *Applied and Environmental Microbiology.* **63** (10), 3926-3932 (1997).
6. van Treuren, W., *et al.* Variation in the microbiota of Ixodes ticks with regard to geography, species, and sex. *Applied and Environmental Microbiology*. **81** (18), 6200-6209 (2015).
7. Swei, A., Kwan, J.Y. Tick microbiome and pathogen acquisition altered by host blood meal. *The ISME Journal: Multidisciplinary Journal of Microbial Ecology*. **11** (3), 813-816 (2017).
8. Kwan, J.Y., Griggs, R., Chicana, B., Miller, C., Swei, A. Vertical *vs.* horizontal transmission of the microbiome in a key disease vector, Ixodes pacificus. *Molecular Ecology.* **26** (23), 6578-6589 (2017).
9. Narasimhan, S., Fikrig, E. Tick microbiome: The force within. *Trends in Parasitology*. **31** (7), 315-323 (2015).
10. Houpikian, P., Raoult, D. Traditional and molecular techniques for the study of emerging bacterial diseases: One laboratory's perspective. *Emerging Infectious Diseases*. **8** (2), 122-131 (2002).
11. Kotsilkov, K., Popova, C., Boyanova, L., Setchanova, L., Mitov, I. Comparison of culture method and real-time PCR for detection of putative periodontopathogenic bacteria in deep periodontal pockets. *Biotechnology and Biotechnological Equipment*. **29** (5), 996-1002 (2015).
12. Wade, W. Unculturable bacteria - The uncharacterized organisms that cause oral infections. *Journal of the Royal Society of Medicine*. **95** (2), 81-83 (2002).
13. Klindworth, A., *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*. **41** (1), 1-11 (2013).
14. Janda, J.M., Abbott, S.L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*. **45** (9), 2761-2764 (2007).
15. Falco, R.C., Fish, D. A comparison of methods for sampling the deer tick, Ixodes dammini, in a Lyme disease endemic area. *Experimental & Applied Acarology*. **14** (2), 165-173 (1992).
16. Patrick, C.D., Hair J.A. Laboratory rearing procedures and equipment for multi-host ticks (Acarina: Ixodidae). *Journal of Medical Entomology.* **12** (3), 389-390 (1975).
17. Köchl, S., Niederstätter, H., Parson, W. DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. *Methods in Molecular Biology.* **297**, 13-30 (2005).
18. Wallace, D.M. Large- and small- scale phenol extractions, in: Berger S.L., Kimmel R. (Eds.), *Guide to molecular cloning techniques.* Academic Press, Orlando, **152,** 33-41 (1987).
19. Zeugin, J.A., Hartley, J.L. "Ethanol Precipitation of DNA." *Focus.* **7** (4), 1-2 (1985).
20. Walsh, P.S, Metzger D.A., Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*. **10**, 506-518 (1991).
21. Gariepy, T.D., Lindsay, R., Ogden, N., Gregory, T.R. Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. *Molecular Ecology Resources*. **12**, 646-652 (2012).
22. Ammazzalorso, A.D., Zolnik, C.P., Daniels, T.J., Kolokotronis, S.O. To beat or not to beat a tick: comparison of DNA extraction methods for ticks (*Ixodes scapularis*). *PeerJ*. **3**, 1-14 (2015).
23. Desjardins, P., Conklin, D. NanoDrop microvolume quantitation of nucleic acids. *Journal of Visualized Experiments.* (2010).
24. TaKara Bio. *Library Quantification Kit: User Manual.* Mountain View, CA. (2018).
25. Genohub. *Cluster density optimization on Illumina sequencing instruments.* (2018).
26. Fadrosh D.W., *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome.* **2** (6), (2014).
27. Caporaso, G.J., *et al*. QIIME allows analysis of high-throughput community sequencing data, *Nature Methods.* **7**, 335-336 (2010).
28. Schloss, P.D., *et al.*Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology.* **75**, 7537-7541 (2009).
29. Duguma, D., *et al.* Developmental succession of the microbiome of Culex mosquitoes Ecological and evolutionary microbiology. *BMC Microbiology*. **15** (1), 1-13 (2015).
30. Fagen, J.R. Characterization of the Relative Abundance of the Citrus Pathogen Ca. Liberibacter Asiaticus in the Microbiome of Its Insect Vector, Diaphorina citri, using High Throughput 16S rRNA Sequencing. *The Open Microbiology Journal*. **6** (1), 29-33 (2012).
31. Geiger, A., *et al.* First isolation of Enterobacter, Enterococcus, and Acinetobacter spp. as inhabitants of the tsetse fly (Glossina palpalis palpalis) midgut. *Infection, Genetics and Evolution*. **9** (6), 1364-1370 (2009).
32. Salter, S.J., *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*. **12** (1), 1-12 (2014).
33. Rand, K.H., Houck, H. Taq polymerase contains bacterial DNA of unknown origin. *Molecular and Cellular Probes.* **4**, 445-450 (1990).
34. Grahn, N., Olofsson, M., Ellnebo-Svedlund K., Monstein H.J., Jonasson, J. Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiology Letters*. **219**, 87-91 (2003).
35. Mohammadi T., Reesink H.W., Vandenbroucke-Grauls C.M., Savelkoul P.H. Removal of contaminating DNA from commercial nucleic acid extraction kit reagents. *Journal of Microbiological Methods*. **61**, 285-288 (2005).
36. Weiss, S., Amir, A., Hyde, E.R., Metcalf, J.L., Song, S.J., Knight, R. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology.* **15**, 564 (2014).
37. Paulson, J.N., Stine, O.C., Bravo, H.C., Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*. **10** (12), 1200-1202 (2013).
38. Weiss, S., *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*. **5** (1), 1-18 (2017).
39. McMurdie, P.J., Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *Public Library of Science Computational Biology*. **10** (4), (2014).
40. Edmonds, K., Williams, L. The role of the negative control in microbiome analyses. *The Federation of American Societies for Experimental Biology Journal. **23*** (Suppl 1), (2017).

41. Gall, C.A*., et al*. The bacterial microbiome of Dermacentor andersoni ticks influences pathogen susceptibility. *The ISME Journal: Multidisciplinary Journal of Microbial Ecology*. **10**, 1846-1855 (2016).
42. Gofton, A.W., *et al*. Inhibition of the endosymbiont "Candidatus Midichloria mitochondrii" during 16S rRNA gene profiling reveals potential pathogens in Ixodes ticks from Australia. *Parasites & Vectors*. 1-11 (2015).
43. Pruesse, E., *et al*. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*. **35,** 7188-7196 (2007).
44. DeSantis, T.Z., e*t al*. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*. **72,** 5069-5072 (2006).
45. Cole, J.R., *et al.* The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*. **37**, D141-D145 (2009).
46. Benson D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. GenBank*. Nucleic Acids Research*. **34**, D16-D20 (2006).
47. Schloss, P.D. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *Public Library of Science Computational Biology*. **6** (7), 19 (2010).
48. Balvočiute, M., Huson, D.H. SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics*. **18** (Suppl 2), 1-8 (2017).
49. Langille, M.G., *et al*. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*. **31** (9), 814-821 (2013).