



# Targeted Metagenomics for Clinical Detection and Discovery of Bacterial Tick-Borne Pathogens

 Luke Kingry,<sup>a</sup> Sarah Sheldon,<sup>a</sup> Stephanie Oatman,<sup>a</sup> Bobbi Pritt,<sup>b</sup> Melissa Anacker,<sup>c</sup> Jenna Bjork,<sup>c</sup> David Neitzel,<sup>c</sup> Anna Strain,<sup>c</sup> Jon Berry,<sup>b</sup> Lynne Sloan,<sup>b</sup> Laurel Respicio-Kingry,<sup>a</sup> Elizabeth Dietrich,<sup>a</sup> Karen Bloch,<sup>d</sup> Abelardo Moncayo,<sup>e</sup> Ganesh Srinivasamoorthy,<sup>f</sup> Bin Hu,<sup>f\*</sup> Alison Hinckley,<sup>a</sup> Paul Mead,<sup>a</sup> Kiersten Kugeler,<sup>a</sup> Jeannine Petersen<sup>a</sup>

<sup>a</sup>Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado, USA

<sup>b</sup>Division of Clinical Microbiology, Mayo Clinic, Rochester, Minnesota, USA

<sup>c</sup>Minnesota Department of Health, St. Paul, Minnesota, USA

<sup>d</sup>Vanderbilt University Medical Center, Nashville, Tennessee, USA

<sup>e</sup>Tennessee Department of Health, Nashville, Tennessee, USA

<sup>f</sup>Scientific Computing and Bioinformatics Support, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Luke Kingry, Sarah Sheldon, and Stephanie Oatman contributed equally; author order was determined by seniority.

**ABSTRACT** Tick-borne diseases, due to a diversity of bacterial pathogens, represent a significant and increasing public health threat throughout the Northern Hemisphere. A high-throughput 16S V1-V2 rRNA gene-based metagenomics assay was developed and evaluated using >13,000 residual samples from patients suspected of having tick-borne illness and >1,000 controls. Taxonomic predictions for tick-borne bacteria were exceptionally accurate, as independently validated by secondary testing. Overall, 881 specimens were positive for bacterial tick-borne agents. Twelve tick-borne bacterial species were detected, including two novel pathogens, representing a 100% increase in the number of tick-borne bacteria identified compared to what was possible by initial PCR testing. In three blood specimens, two tick-borne bacteria were simultaneously detected. Seven bacteria, not known to be tick transmitted, were also confirmed to be unique to samples from persons suspected of having tick-borne illness. These results indicate that 16S V1-V2 metagenomics can greatly simplify diagnosis and accelerate the discovery of bacterial tick-borne pathogens.

**KEYWORDS** anaplasmosis, Lyme disease, tick-borne bacteria, human granulocytic ehrlichiosis, metagenomics, vector-borne diseases

Tick-borne infections constitute a significant and growing public health threat, and millions of people seek medical care as a result of a tick bite each year (1, 2). Reported cases of tick-borne diseases have more than doubled in number since 2004 in the United States and continue to increase (2). Lyme disease is the most common tick-borne infection in the Northern Hemisphere, with an estimated 300,000 cases occurring annually in the United States alone (3–5). The economic impact of diagnostic testing for tick-borne diseases is considerable, with more than 2.4 million samples from U.S. patients tested annually for Lyme disease alone, at an estimated cost of \$492 million (3).

Infectious agents transmitted by ticks are etiologically and ecologically diverse. Bacteria, viruses, and protozoa are all causes of tick-borne disease, although bacteria are responsible for the vast majority of reported tick-transmitted infections (2). More than 13 bacterial species, encompassing multiple genera (*Borrelia*, *Anaplasma*, *Ehrlichia*, *Rickettsia*, and *Francisella*), have been shown to be the causative agents of multiple tick-borne diseases of public health importance, including Lyme disease, *Borrelia miyamotoi* disease, relapsing fever (RF), human granulocytic anaplasmosis, ehrlichiosis,

**Citation** Kingry L, Sheldon S, Oatman S, Pritt B, Anacker M, Bjork J, Neitzel D, Strain A, Berry J, Sloan L, Respicio-Kingry L, Dietrich E, Bloch K, Moncayo A, Srinivasamoorthy G, Hu B, Hinckley A, Mead P, Kugeler K, Petersen J. 2020. Targeted metagenomics for clinical detection and discovery of bacterial tick-borne pathogens. *J Clin Microbiol* 58:e00147-20. <https://doi.org/10.1128/JCM.00147-20>.

**Editor** John P. Dekker, National Institute of Allergy and Infectious Diseases

**Copyright** © 2020 American Society for Microbiology. All Rights Reserved.

Address correspondence to Luke Kingry, [vtx8@cdc.gov](mailto:vtx8@cdc.gov).

\* Present address: Bin Hu, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

Lynne Sloan met authorship criteria but was unreachable for final approval of the byline and article.

For a commentary on this article, see <https://doi.org/10.1128/JCM.01893-20>.

**Received** 23 January 2020

**Returned for modification** 17 February 2020

**Accepted** 21 July 2020

**Accepted manuscript posted online** 2 September 2020

**Published** 21 October 2020

spotted fever rickettsioses, and tularemia (1, 2). One-third of these bacterial agents were identified as causes of human illness only in the last 15 years (2, 6–9), with dozens more species of these same genera identified in ticks or animals. For most tick-transmitted agents, *in vitro* culture requires specialized growth conditions and can take several weeks or more and therefore is rarely performed as a diagnostic test (10, 11). Standard diagnostic testing is based largely on serology or PCR panels directed at known disease agents (10–13). As neither approach yields DNA sequence information, the detection of all bacterial tick-borne pathogens and genetic characterization of novel pathogens are not optimal.

The detection and discovery of tick-borne pathogens worldwide could directly benefit from the application of advanced molecular technologies such as next-generation sequencing. The amplification of one or more of the nine variable (V) regions of the 16S rRNA gene, which is present in all bacteria, coupled with deep sequencing and taxonomic assignment of the sequence reads, reveals an agnostic view of all bacterial taxa in a given sample. This type of diagnostic approach is particularly attractive for tick-borne bacteria given the polyvector and polymicrobial nature of bacterial tick-borne diseases. Here, we demonstrate the power of a single-assay high-throughput 16S rRNA V1-V2 sequence-based metagenomics approach to accurately yield taxonomic predictions for known and novel bacterial tick-borne pathogens in clinical samples from patients with suspected tick-borne illness.

## MATERIALS AND METHODS

**Sample overview. (i) Clinical specimens.** A convenience sample of residual clinical specimens originally submitted to Mayo Clinic Medical Laboratories from health care providers nationwide for PCR testing for tick-borne pathogens was retained following routine requested testing, frozen at  $-80^{\circ}\text{C}$ , and included in this analysis. Submission of a clinical specimen for testing for a tick-borne agent was considered to be a proxy for clinical suspicion of a tick-borne disease. Residual specimen types included blood ( $\text{K}_2$  EDTA), submitted for Lyme PCR (*Borrelia burgdorferi* and *Borrelia mayonii*) or tick-borne panel PCR (*Babesia microti*, *Ehrlichia chaffeensis*, *Ehrlichia muris* subsp. *eaucalarensis*, *Ehrlichia ewingii*, and *Anaplasma phagocytophilum*), and synovial fluid, cerebrospinal fluid (CSF), and fresh tissue, submitted for Lyme PCR. Residual blood specimens submitted to the Vanderbilt University Medical Center for *Ehrlichia* PCR testing from health care providers in Tennessee were also included. The study was approved by the Institutional Review Boards at Mayo Clinic (protocol identification number 14-001148), Vanderbilt University (protocol identification number 140738), and the CDC (protocol identification numbers 6635 and 6870).

**(ii) Controls.** To ascertain DNA amplification considered to be background, molecular-grade water and blood from healthy donors were obtained and screened via 16S V1-V2 metagenomics. Blood from healthy donors ( $\text{K}_2$  EDTA) ( $n = 500$ ) was obtained from Innovative Research (Novi, MI) ( $n = 100$ ), Biological Specialty Corporation (Colmar, PA) ( $n = 100$ ), BioIVT (Westbury, NY) ( $n = 120$ ), ZenBio (Research Triangle Park, NC) ( $n = 60$ ), BioChemed (Winchester, VA) ( $n = 60$ ), and Discovery Life Sciences (Los Osos, CA) ( $n = 60$ ). Healthy donor blood specimens originated from the Midwest (Innovative Research), Northeast (Biological Specialty Corporation), and Southeastern (BioIVT, ZenBio, BioChemed, and Discovery Life Sciences) United States. Molecular-grade water (three different lots) was obtained from Fisher (Pittsburgh, PA).

**DNA extraction.** DNA was extracted using the MagNA Pure 96 instrument (Roche, Indianapolis, IN) and the DNA and viral NA small-volume kit (Roche, Indianapolis, IN) with the associated DNA blood SV 3.1 extraction protocol. One hundred microliters was used for the input, and purified DNA was eluted in  $100\ \mu\text{l}$ . Each 96-well DNA plate included a negative and a positive control. The negative control consisted of pooled blood from healthy donors. The positive control consisted of *Escherichia coli* strain ATCC 8739 DNA spiked into pooled blood from healthy donors at a final concentration of  $127\ \text{fg}/\mu\text{l}$ .

**16S rRNA gene primer design and amplification.** Conserved regions flanking the V1-V2 region of the 16S rRNA gene were utilized for primer design based on a ClustalO DNA sequence alignment of tick-borne bacterial species, including *Borrelia* species, encompassing both the Lyme and RF groups; *Anaplasma*; *Ehrlichia*; *Rickettsia*; *Francisella*; and *E. coli* (see Fig. S2 in the supplemental material). Illumina Nextera XT index adaptor sequences (underlined) were included in the synthesis of forward primer 5'-TCGTCCGCAGCGTCAGATGTGTATAAGAGACAGCTGCTGCCCTCCCGTAGGAGT-3' and reverse primer 5'-GTCCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGTTTGATCTGGCTCAG-3' (IDT, Coralville, IA).

For 16S V1-V2 PCR amplification, the DNA template concentration was optimized for sensitivity (without inhibition) using DNA extracted from blood samples positive for either *B. burgdorferi* or *B. miyamotoi*. Template DNA volumes of 1 to  $5\ \mu\text{l}$  were tested, with the resulting amplicon being quantified using TapeStation high-sensitivity screen tape (Agilent, Santa Clara, CA). The final optimized  $50\text{-}\mu\text{l}$  PCR mixture consisted of  $25\ \mu\text{l}$  of premix ExTaq, DNA polymerase hot-start version (TaKaRa Bio, San Francisco, CA),  $100\ \text{ng}$  of each primer, and  $4\ \mu\text{l}$  of purified template DNA. Using the Veriti Dx 96-well thermal cycler,  $0.2\ \text{ml}$  (Thermo Fisher Scientific, Grand Island, NY), PCR cycling conditions were as follows: an initial denaturation step at  $95^{\circ}\text{C}$  for 4 min and 45 cycles of denaturation at  $94^{\circ}\text{C}$  for 30 s, annealing at  $56^{\circ}\text{C}$  for

30 s, and extension at 75°C for 30 s, followed by a final extension step at 72°C for 10 min. Each 96-well plate for PCR included the two DNA extraction controls described above, one negative PCR control of pooled healthy human donor blood DNA, and two PCR positive controls of *E. coli* DNA diluted in TE (Tris-EDTA) buffer (pH 8.0), at a final concentration of 127 fg/ $\mu$ l.

**Library preparation and multiplexing.** Dual Nextera XT indices (XT index kit v2, sets A to D; Illumina, San Diego, CA) were added via PCR to the V1-V2 amplicons to allow multiplexed sequencing of 384 samples. Each 50- $\mu$ l reaction mixture consisted of 25  $\mu$ l of premix ExTaq DNA polymerase hot-start version, 10  $\mu$ l of water (Fisher, Pittsburgh, PA), 5  $\mu$ l each of forward and reverse indices, and 5  $\mu$ l of the 16S V1-V2 amplicon. PCR conditions were 95°C for 3 min and 8 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, followed by a final extension step at 72°C for 5 min. The resulting uniquely indexed libraries were purified using 1.8 $\times$  AMPure XP (Beckman Coulter, Miami, FL) according to the manufacturer's protocol and eluted in 50  $\mu$ l of 10 mM Tris (pH 8.5) (Tris base; Sigma-Aldrich, St. Louis, MO).

Purified libraries for each sample were quantified using the FLx800T microplate fluorescence reader (BioTek, Winooski, VT) and the Quant-iT Qubit high-sensitivity double-stranded DNA (dsDNA) assay kit (Invitrogen, Carlsbad, CA). Each library was normalized to a final concentration of 4 nM by dilution in 10 mM Tris (pH 8.5). A *B. burgdorferi* positive clinical specimen with a threshold cycle ( $C_T$ ) value of 34 by pan-*Borrelia* PCR (14) was used as an internal sequencing control. The positive control was independently indexed, bead cleaned, and normalized to 4 nM like all other samples. The same indexed control was used throughout 16S metagenomics sequencing of all samples. A pool for each of the four 96-well plates was prepared and then combined into a single final library pool consisting of 384 uniquely indexed libraries. To remove DNA fragments smaller than 300 bp, the final pooled library was subjected to a 0.8 $\times$  bead cleanup by adding 40  $\mu$ l AMPure XP beads to a 50- $\mu$ l aliquot of the pooled library. The resulting final pooled library was quantified using the Qubit 3.0 fluorometer (Life Technologies, Carlsbad, CA) with the Quant-iT dsDNA high-sensitivity kit and diluted using 10 mM Tris buffer (pH 8.5) to obtain a final concentration of 4 nM.

**16S metagenomic sequencing.** Paired-end sequencing (2 by 250 bp) of the final pooled library (360 patient samples and 24 controls) was performed using an Illumina MiSeq and 500 cycle V2 reagent kit (Illumina). The pooled library was denatured by combining 5  $\mu$ l of the 4 nM library with 5  $\mu$ l of 0.2 N NaOH (Sigma-Aldrich, St. Louis, MO), briefly vortexed, and centrifuged at 280  $\times g$  for 1 min, followed by a 5-min incubation at room temperature. The denatured library was then diluted with 990  $\mu$ l chilled Illumina hybridization buffer (HT1) and put on ice, resulting in a 20 pM denatured library in 1 mM NaOH. The library was further diluted to an optimized 12.5 pM concentration by combining 375  $\mu$ l of the 20 pM library with 225  $\mu$ l chilled HT1, which was stored on ice until use. The 12.5 pM library was, unless otherwise indicated, combined with 12.5 pM PhiX (Illumina) at 60  $\mu$ l PhiX (10%) and 540  $\mu$ l of the denatured and diluted amplicon library. The combined library was heat denatured at 96°C for 2 min, followed by 5 min in an ice water bath immediately before loading 600  $\mu$ l into the MiSeq reagent cartridge.

**Bioinformatic analysis and taxonomic prediction.** Sequence reads were demultiplexed into 384 individual samples, followed by the removal of the adaptor and indices by internal MiSeq software prior to analysis using a custom bioinformatic workflow. Reads from each sample were subjected to adaptor and quality-based trimming (quality score Q30) using FaQCs 1.34 (15). Paired reads were merged using Flash 1.2.11 (16), with a mismatch density set to 0.25. The resulting quality-trimmed, merged reads were assigned taxonomic predictions with Kraken 0.10.5 (17) using the MiniKraken database (18 October 2017). A default kmer length of 31 bp was used. The final output for each sample was a taxon-based read abundance CSV file; all individual-sample CSV files were concatenated into a single file. Sequence read abundance and mean, median, minimum, and maximum read counts were determined using unmodified output files. For read abundance, for each sample with a tick-borne taxonomic prediction, the percentage of reads for the tick-borne agent relative to reads for all other bacteria taxa was calculated.

For analysis of taxonomic predictions in controls and clinical samples, only those with  $\geq 50$  sequence reads were included. This level of stringency was applied in order to avoid false positives in clinical samples due to error rates in Illumina sequencing and the high number of uniquely indexed libraries (384) included per sequencing run. For subtraction analysis, taxonomic predictions ( $\geq 50$  sequence reads) were manually collapsed to order, the highest taxon level with specificity for the most common tick-borne agents (i.e., *Rickettsiales* and *Spirochaetales*). Background taxonomic predictions were established from 500 individual blood specimens from healthy donors, molecular-grade water from three lots tested 805 times, and pooled blood from healthy donors tested 335 times. A breakdown of the taxonomic predictions from individual blood specimens from the 5 different companies is available in Table S5. Taxa were included in the background data set if they were present in at least two control specimens; singleton taxonomic predictions were not included. Background taxonomic predictions were manually subtracted from those taxonomic predictions in clinical specimens in order to determine unique taxa in clinical samples. After subtraction, the number of taxonomic predictions for *Rickettsiales* (54%) and *Spirochaetales* (10%) relative to all other taxonomic predictions in clinical samples was manually determined. Taxonomic predictions for clinical specimens were also analyzed at the genus level for *Anaplasma*, *Borrelia*, *Borrelia*, *Ehrlichia*, *Francisella*, and *Rickettsia*. The nomenclature for *B. burgdorferi sensu lato* genospecies has been in flux over the last several years. Both *Borrelia* and, more recently, *Borrelia* have been utilized for this species, although it has not been uniformly applied. For the purposes of this article and for clarity, regardless of the nomenclature used in databases, all *B. burgdorferi sensu lato* genospecies in this article are classified as *Borrelia* (18).

**Validation of taxonomic predictions.** All samples positive for taxonomic predictions of *Anaplasma*, *Borrelia*, *Borrelia*, *Ehrlichia*, *Francisella*, and *Rickettsia* were subjected to secondary *in vitro* confirmation. If taxonomic predictions corresponding to two different tick-borne bacteria were identified, confirmatory assays for both predicted taxa were performed. For *Anaplasma* and *Ehrlichia*, a real-time PCR assay that uses hybridization probes and targets the *groEL* gene, followed by melting temperature analysis of PCR products, was utilized (19). For *Borrelia*, one to eight housekeeping genes, *uvrA*, *rplB*, *recG*, *pyrG*, *pepX*, *clpX*, *nifS*, and *clpA*, were amplified, sequenced, and analyzed (14). For *Francisella*, a real-time TaqMan PCR assay was employed for species identification (20, 21). For *Rickettsia*, a pan-*Rickettsia* TaqMan PCR assay (12) targeting the ribosomal protein L16P gene was utilized. For species that could not be identified by PCR, sequencing of the V1-V4 region of the 16S rRNA gene (~785 to 830 bp) was performed. Briefly, 16S amplicons were tagged and indexed according to the Nextera XT manufacturer's protocol and sequenced with the 300-cycle V2 reagent kit (Illumina). *De novo* consensus 16S sequences were generated using CLC Genomics Workbench (Qiagen), using one or more of the following parameters: word size of 20 to 30, default quality trimming of the paired reads, and 25 to 100% sampling of the paired reads. Consensus sequences were corroborated by mapping the paired reads to a reference sequence corresponding to the original taxonomic prediction (length and similarity fractions set to 0.9 and 0.95, respectively). Assemblies required >100× coverage and a Q score of >60 for each nucleotide site. If *de novo* and reference mapping assemblies yielded identical consensus sequences, primers were trimmed, and FASTA files were exported for further analysis. To maximize 16S sequence length for analyses, the V1-V4 sequence was stitched together with the V1-V2 consensus sequence derived from mapping and querying the NCBI nonredundant DNA database. Maximum likelihood trees were constructed using the general time-reversible model in MEGA (version 7.0) (22) with 1,000 bootstrap replicates. Pairwise genetic distances were calculated in MEGA (version 7.0) using the Kimura-2 model. Primer sequences and PCR and sequencing conditions are provided in Table S3. For specimens with taxonomic predictions that could not be confirmed by PCR or sequencing, DNA was reextracted from the original specimen and tested a second time by 16S metagenomics.

**BLASTn analysis.** The quality-trimmed, merged reads from each sample with a MiniKraken taxonomic prediction of *Anaplasma*, *Borrelia*, *Borrelia*, *Ehrlichia*, *Francisella*, *Rickettsia*, or other significant identifications were independently mapped (CLC Genomics) to reference sequences corresponding to their original taxonomic predictions. Default parameters were used for mapping, with the exception of the length and similarity fractions, set to 0.8 and 0.9, respectively. Consensus sequences derived from each mapping were subjected to BLASTn analysis to query the NCBI nonredundant DNA database (November 2018).

**Limit of detection of 16S metagenomics.** A limit of detection was determined by adding 10-fold dilutions of *B. burgdorferi* B31 DNA (diluted in pooled healthy donor blood DNA) to pooled healthy donor blood DNA to yield final concentrations of  $6 \times 10^6$  genomes/ml of blood to 60 genomes/ml of blood. To simulate coinfecting samples, DNA extracted from a human blood specimen strongly positive for *A. phagocytophilum* (*groEL* real-time PCR  $C_t$  of 26; read count of 42,027) was added to the 10-fold dilutions of *B. burgdorferi* B31 DNA. The volume of extracted human DNA was held consistent across all samples. For the calculation of genomic equivalents, a conversion value of 1.63 fg/genome was utilized.

**Data availability.** Raw sequence read data have been deposited for control specimens (molecular-grade water) in the Sequence Read Archive (SRA) under project accession number [PRJNA637310](https://www.ncbi.nlm.nih.gov/sra/PRJNA637310). Source codes for bioinformatic pipelines used in this study are available upon request. All unique 16S sequences have been uploaded to the GenBank database. The 16S V1-V4 sequence for the *Anaplasma* species identified here was deposited in the NCBI database with accession number [MG429812](https://www.ncbi.nlm.nih.gov/nuccore/MG429812). The novel *Rickettsia*-like and *N. risticii* 16S sequences were deposited in the NCBI database with accession numbers [MK580529](https://www.ncbi.nlm.nih.gov/nuccore/MK580529) and [MK580530](https://www.ncbi.nlm.nih.gov/nuccore/MK580530), respectively.

## RESULTS

**Sample overview.** A total of 13,038 residual clinical specimens (blood, cerebrospinal fluid [CSF], synovial fluid, and tissue) submitted by a health care provider for PCR testing for *Anaplasma phagocytophilum*, *Borrelia burgdorferi*, *Borrelia mayonii*, *Ehrlichia chaffeensis*, *Ehrlichia muris* subsp. *eaucalarensis*, *Ehrlichia ewingii*, or *Babesia microti* were available for 16S metagenomics testing. Specimens were originally submitted to either the Mayo Clinic ( $n = 12,494$ ) by providers throughout the United States or the Vanderbilt University Medical Center ( $n = 544$ ) by providers in Tennessee (see Fig. S1 in the supplemental material). The most common specimen type was blood (87.4%), followed by CSF (8.9%), synovial fluid (3.7%), and tissue (<0.1%).

**16S metagenomics.** The V1-V2 region of the 16S rRNA gene was chosen for the detection of tick-borne bacteria based on BLASTn analysis and multisequence alignments, which verified that nucleotide differences among 28 tick-borne bacterial species (encompassing known pathogens as well as tick-borne species not associated with human illness) within the genera *Borrelia*, *Anaplasma*, *Ehrlichia*, and *Francisella* were sufficient for correct identification (Fig. S2). Within the *Rickettsia* genus, interspecies sequence variation in this region is limited.

**TABLE 1** Comparisons of limit of detection for V1-V2 16S metagenomics and pan-*Borrelia* TaqMan PCR

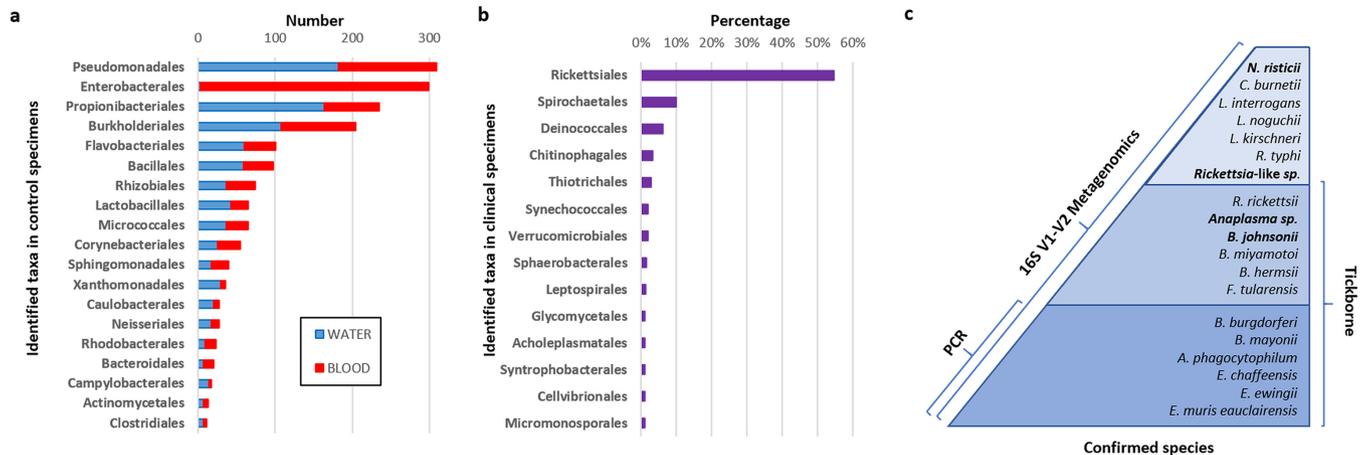
| <i>B. burgdorferi</i> DNA level (genomes/ml) | No. of positive samples/total no. of samples                                    |   |                            |   |  |
|--|---|---|----------------------------|---|--|
|  | 16S V1-V2 metagenomics  |   |                            | Pan- <i>Borrelia</i> real-time PCR  |  |
|  | <i>B. burgdorferi</i> DNA spiked into blood for <i>B. burgdorferi</i> detection | <i>B. burgdorferi</i> DNA spiked into an <i>Anaplasma</i> -positive blood sample for: |                            | <i>B. burgdorferi</i> DNA spiked into blood for <i>B. burgdorferi</i> detection | <i>B. burgdorferi</i> DNA spiked into an <i>Anaplasma</i> -positive blood sample for <i>B. burgdorferi</i> detection |
|  |   | <i>B. burgdorferi</i> detection   | <i>Anaplasma</i> detection |   |  |
| 6 × 10 <sup>6</sup>                          | 10/10   | 10/10   | 10/10                      | 10/10   | 10/10  |
| 6 × 10 <sup>5</sup>                          | 10/10   | 10/10   | 10/10                      | 10/10   | 10/10  |
| 6 × 10 <sup>4</sup>                          | 10/10   | 10/10   | 10/10                      | 10/10   | 10/10  |
| 6,000  | 10/10   | 7/10  | 10/10                      | 9/10  | 10/10  |
| 600  | 6/10  | 0/10  | 10/10                      | 6/10  | 8/10   |
| 60   | 2/10  | 0/10  | 10/10                      | 0/10  | 0/10   |

16S V1-V2 metagenomics yielded reproducible MiniKraken taxonomic predictions of *B. burgdorferi* down to 6 × 10<sup>3</sup> genomes/ml of blood (10/10 detected) in spiked blood samples (Table 1). This limit of detection was equivalent to that measured for pan-*Borrelia* real-time PCR (Table 1). When spiked into an *A. phagocytophilum*-positive blood sample to simulate coinfection, *B. burgdorferi* DNA remained consistently detectable by 16S V1-V2 rRNA metagenomics to 6 × 10<sup>4</sup> genomes/ml of blood (10/10), with intermittent detection at 6 × 10<sup>3</sup> genomes/ml of blood (7/10).

Sequencing statistics for runs encompassing all 13,038 clinical specimens, blood from healthy donors, molecular-grade water, and controls (reads >Q30, clusters passing filter, cluster density, and total read count) are provided in Table S1.

**MiniKraken taxonomic predictions.** A total of 92.3% (12,033/13,038) of clinical specimens, 63.2% (509/805) of molecular-grade water samples, and 65.8% (329/500) of blood samples from healthy donors yielded at least one 16S V1-V2 taxonomic prediction (with ≥50 sequence reads) for any bacterial taxa. Taxonomic predictions within the tick-borne genera *Anaplasma*, *Borrelia*, *Ehrlichia*, *Francisella*, and *Rickettsia* were specific to clinical specimens. A total of 878 clinical specimens yielded a single taxonomic prediction within one of the tick-borne genera *Anaplasma*, *Borrelia*, *Ehrlichia*, *Francisella*, and *Rickettsia*. Three specimens yielded simultaneous taxonomic predictions for two different tick-borne genera, *Anaplasma* and *Borrelia*, for a total of 881 specimens identified as being positive for tick-borne bacteria. Taxonomic predictions observed for blood from healthy donors (*n* = 500) and molecular-grade water samples (*n* = 805) demonstrated considerable overlap (35/41 orders; 85.4%) (Fig. 1a and Fig. S3). One exception was a taxonomic prediction of *Enterobacteriales*, which was common in blood from healthy donors (296/500; 59%) and rare in molecular-grade water (3/805; 0.4%). The removal of taxonomic predictions identified in individual healthy blood donors, molecular-grade water samples, and pooled healthy donor blood demonstrated that *Rickettsiales* and *Spirochaetales* were the two most prevalent bacterial taxa identified in the 13,038 clinical samples from patients with suspected tick-borne illness (Fig. 1b and Table S2).

**Increase in the number of detected tick-borne bacteria.** Specific 16S V1-V2 taxonomic predictions for 10 different tick-borne bacteria known to cause human illness included *A. phagocytophilum*, *B. burgdorferi*, *B. mayonii*, *B. miyamotoi*, *Borrelia hermsii*, *E. chaffeensis*, *E. muris*, *E. ewingii*, *Francisella tularensis*, and *Rickettsia* (Table 2). Additionally, taxonomic predictions corresponding to two species (*Anaplasma* sp. or *Borrelia* sp.) previously detected only in ticks or recently associated with human illness were discovered (14, 23, 24). Overall, these identifications doubled the number of detected tick-borne bacteria to 12 species compared to the number of bacterial species originally tested for by PCR (*n* = 6) (Fig. 1c). All clinical specimens positive for 1 of these 12 bacterial species were blood, except for *B. burgdorferi*-positive specimens, which included blood (*n* = 64), synovial fluid (*n* = 22), and CSF (*n* = 1). A single blood-contaminated synovial fluid specimen was positive for *A. phagocytophilum*.



**FIG 1** Identified taxa in controls and clinical specimens. (a) Distribution of order-level taxonomic predictions in healthy donor blood specimens and molecular-grade water arranged by descending prevalence. Data are shown only for those taxonomic predictions present in >10 individual specimens. A full list of taxonomic predictions is can be found in Fig. S3 in the supplemental material. (b) Proportion (percentage) of order-level taxonomic predictions in samples from patients suspected of having tick-borne illness relative to all other taxonomic predictions. The data shown are after the subtraction of background taxa identified in control specimens. Data are shown only for those taxonomic predictions represented in >1% of patient specimens. A full list of taxonomic predictions unique to clinical specimens is available in Table S2 in the supplemental material. (c) Identified bacteria (known and novel) in clinical samples from patients suspected of having tick-borne illness by 16S V1-V2 metagenomics compared to PCR. Dark blue, 6 tick-borne bacterial species initially tested for by targeted PCR; medium blue, 6 additional tick-borne bacterial species identified by 16S V1-V2 metagenomics; light blue, 7 other species identified by 16S V1-V2 metagenomics. The total number of bacterial species identified and independently confirmed was 19. Novel pathogens are indicated in boldface type ( $n = 4$ ).

The abundance (percentage) of 16S V1-V2 tick-borne bacterial sequence reads relative to all other bacterial taxa detected in each clinical specimen is shown in Fig. 2a; the corresponding read counts are shown in Table 2. The single sample positive for *F. tularensis* yielded a relative read abundance of 89.6% and a read count of 12,295. As expected, the average 16S read abundance and mean read count were highest for those bacterial species (*A. phagocytophilum*, *E. chaffeensis*, and *B. mayonii*) associated with higher levels of bacteremia ( $10^5$  to  $10^6$  bacteria/ml) (6, 25). Blood specimens positive for *B. burgdorferi* exhibited the lowest 16S read abundance, consistent with the low spirochetemia ( $10^2$  to  $10^3$  genome equivalents/ml) documented for *B. burgdorferi*-infected patients (11, 26).

Three blood specimens from patients suspected of having tick-borne illness contained simultaneous taxonomic predictions of *A. phagocytophilum* as well as a taxonomic prediction for *B. burgdorferi*, *B. mayonii*, or *B. miyamotoi* (Fig. 2c). The 16S V1-V2 relative read abundances were higher for *B. mayonii* and *B. miyamotoi* (67.9% and 74.0%, respectively) than for *A. phagocytophilum* (14.8% and 16.4%). For the sample with taxonomic predictions of *A. phagocytophilum*-*B. burgdorferi*, the relative read abundances were 86.2% for *A. phagocytophilum* and 1.7% for *B. burgdorferi*.

**Accuracy of taxonomic predictions for tick-borne bacterial species.** Of the 881 clinical specimens yielding a 16S V1-V2 MiniKraken taxonomic prediction of *Anaplasma*, *Borrelia*, *Ehrlichia*, *Francisella*, or *Rickettsia*, 877 (99.5%) were independently verified to be positive by secondary PCR and sequencing (Table 2 and Table S3). The four blood specimens (0.5%) that could not be confirmed yielded taxonomic predictions for *A. phagocytophilum* ( $n = 2$ ) or *E. chaffeensis* ( $n = 2$ ). DNA reextraction and testing by 16S V1-V2 metagenomics produced the same taxonomic predictions.

Comparison of 16S V1-V2 MiniKraken taxonomic predictions with identifications from independent PCR and sequencing methods demonstrated very high accuracies of taxonomic predictions for the 12 identified tick-borne bacterial species, with 100% agreement observed at the genus level (Table 2). For 9 known tick-borne bacteria with a complete genome available at the time when the MiniKraken database was constructed, 88.9% were correctly identified to the species level (Table 2). When consensus sequences from all positive specimens were queried by BLASTn against

**TABLE 2** Bacterial species detected in clinical samples by 16S V1-V2 metagenomics and independently confirmed by secondary assays<sup>d</sup>

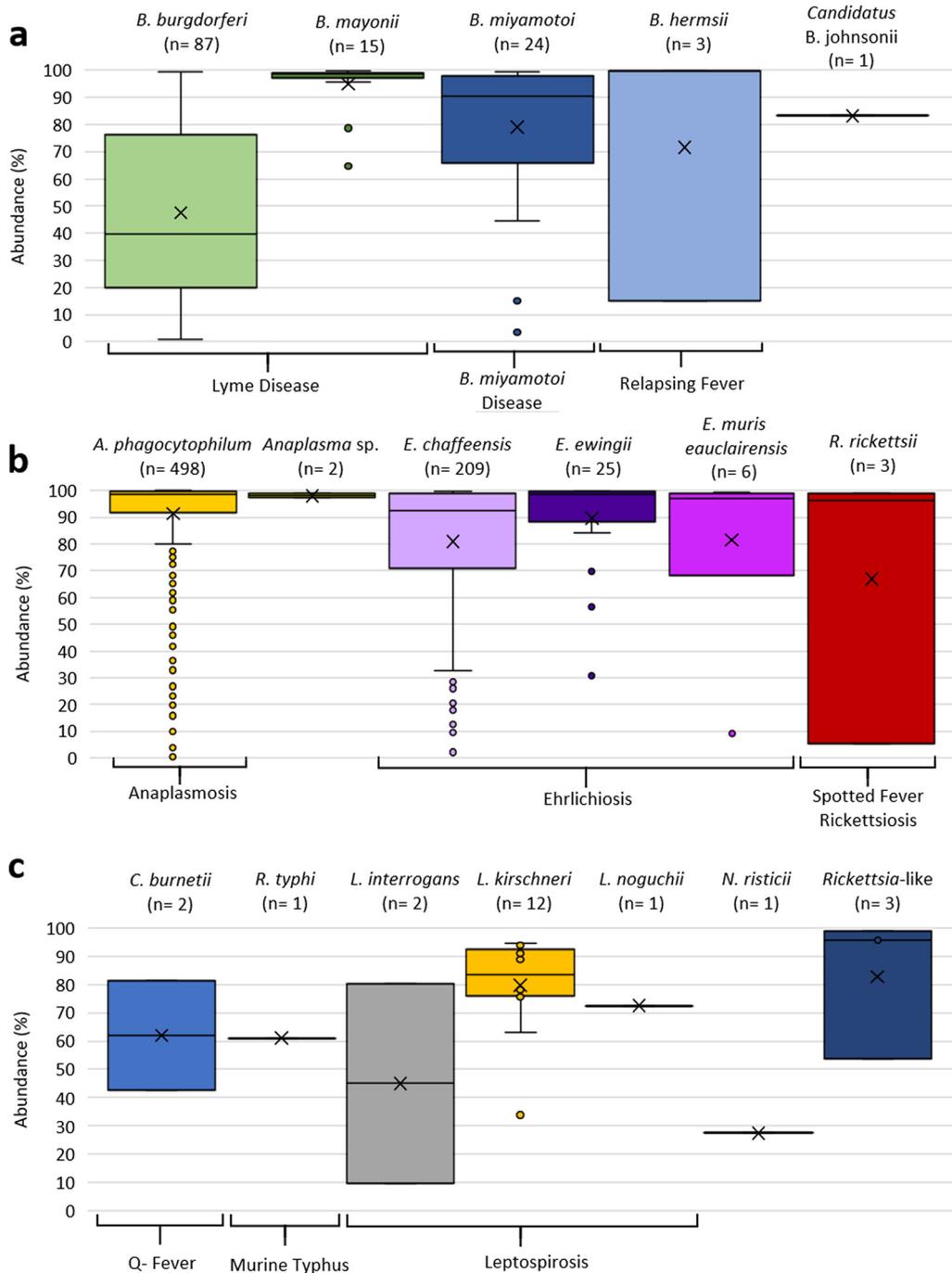
| Confirmed identification <sup>a</sup><br>(no. of isolates) | V1-V2 sufficient<br>for species<br>identification | MiniKraken<br>taxonomic<br>prediction | BLASTn taxonomic<br>prediction (% identity)  | Sequence in<br>MiniKraken<br>database | Tick<br>borne  | Previously<br>associated<br>with human<br>illness | Mean<br>read<br>count | Median<br>read<br>count | Min<br>read<br>count | Max<br>read<br>count |
|--|---|---------------------------------------|--|---------------------------------------|----------------|---|-----------------------|-------------------------|----------------------|----------------------|
| <i>Anaplasma phagocytophilum</i> (498)                     | Yes   | <i>A. phagocytophilum</i>             | <i>A. phagocytophilum</i> (100)  | Yes                                   | Yes            | Yes   | 31,864.3              | 29,649.5                | 105                  | 153,246              |
| <i>Borrelia burgdorferi</i> (87)                           | Yes   | <i>B. burgdorferi</i>                 | <i>B. burgdorferi</i> (≥99.7)  | Yes                                   | Yes            | Yes   | 5,868.2 <sup>c</sup>  | 3,773.0                 | 73                   | 46,420               |
| <i>Borrelia hermsii</i> (3)                                | Yes   | <i>B. hermsii</i>                     | <i>B. hermsii</i> (100)  | Yes                                   | Yes            | Yes   | 33,128.3              | 48,119.0                | 1,729                | 49,537               |
| <i>Borrelia miyamotoi</i> (24)                             | Yes   | <i>B. miyamotoi</i>                   | <i>B. miyamotoi</i> (≥99.7)  | Yes                                   | Yes            | Yes   | 12,041.9              | 12,552.0                | 277                  | 28,158               |
| <i>Borrelia mayonii</i> (15)                               | Yes   | <i>B. mayonii</i>                     | <i>B. mayonii</i> (100)  | Yes                                   | Yes            | Yes   | 26,614.5              | 20,836.0                | 7,223                | 52,264               |
| <i>Ehrlichia chaffeensis</i> (209)                         | Yes   | <i>E. chaffeensis</i>                 | <i>E. chaffeensis</i> (100)  | Yes                                   | Yes            | Yes   | 25,603.2              | 19,689.0                | 629                  | 118,209              |
| <i>Ehrlichia muris</i> subsp.<br><i>euclairensis</i> (6)   | Yes   | <i>E. muris</i>                       | <i>E. muris</i> (100)  | Yes                                   | Yes            | Yes   | 23,461.0              | 20,820.0                | 925                  | 58,206               |
| <i>Francisella tularensis</i> (1)                          | Yes   | <i>F. tularensis</i>                  | <i>F. tularensis</i> (100)   | Yes                                   | Yes            | Yes   | NA                    | NA                      | NA                   | 12,295               |
| <i>Rickettsia rickettsii</i> (3)                           | No  | <i>Rickettsia</i>                     | <i>R. rickettsii</i> / <i>R. peacockii</i> /<br><i>R. philipii</i> / <i>R. slovaca</i> (100) | Yes                                   | Yes            | Yes   | 19,474.7              | 27,178.0                | 1,216                | 30,030               |
| <i>Ehrlichia ewingii</i> (25)                              | Yes   | <i>E. chaffeensis</i>                 | <i>E. ewingii</i> (100)  | No                                    | Yes            | Yes   | 37,312.0              | 35,024.0                | 4,827                | 118,405              |
| <i>Anaplasma</i> sp. (2)                                   | Yes   | <i>Anaplasma centrale</i>             | <i>Anaplasma</i> sp. (98.2)  | No                                    | Yes            | No  | 24,911.0              | 24,911.0                | 16,295               | 33,527               |
| " <i>Candidatus</i> <i>Borrelia johnsonii</i> " (1)        | Yes   | <i>Borrelia</i>                       | <i>Borrelia</i> sp. strain EFL (99.7)  | No                                    | Yes            | No <sup>b</sup>                                   | NA                    | NA                      | NA                   | 14,461               |
| <i>Coxiella burnetii</i> (2)                               | Yes   | <i>C. burnetii</i>                    | <i>C. burnetii</i> (100)   | Yes                                   | No             | Yes   | 2,987.5               | 2,987.5                 | 2,974                | 3,001                |
| <i>Rickettsia typhi</i> (1)                                | Yes   | <i>R. typhi</i>                       | <i>R. typhi</i> (100)  | Yes                                   | No; flea borne | Yes   | NA                    | NA                      | NA                   | 1,024                |
| <i>Neorickettsia risticii</i> (1)                          | Yes   | <i>N. risticii</i>                    | <i>N. risticii</i> (99.7)  | Yes                                   | No             | No  | NA                    | NA                      | NA                   | 8,994                |
| <i>Leptospira interrogans</i> (2)                          | Yes   | <i>L. interrogans</i>                 | <i>L. interrogans</i> (100)  | Yes                                   | No             | Yes   | 11,776.0              | 11,776                  | 21,666               | 1,886                |
| <i>Leptospira kirschneri</i> (12)                          | Yes   | <i>L. interrogans</i>                 | <i>L. kirschneri</i> (100)   | No                                    | No             | Yes   | 14,682.3              | 13,200.5                | 5,380                | 27,940               |
| <i>Leptospira noguchii</i> (1)                             | Yes   | <i>Leptospira</i>                     | <i>L. noguchii</i> (100)   | No                                    | No             | Yes   | NA                    | NA                      | NA                   | 11,587               |
| <i>Rickettsia</i> -like (3)                                | Yes   | <i>Rickettsia</i>                     | <i>Rickettsia raoultii</i> (97.9)  | No                                    | Unknown        | No  | 22,113.3              | 30,499.0                | 3,484                | 32,357               |

<sup>a</sup>All species designations were independently validated by confirmatory assays.

<sup>b</sup>Same specimen reported in reference 14.

<sup>c</sup>The mean read counts for *B. burgdorferi*-positive blood (*n* = 64), synovial fluid (*n* = 22), and CSF (*n* = 1) were 5,523.9, 6,919.9, and 4,763.0, respectively.

<sup>d</sup>NA, not applicable.



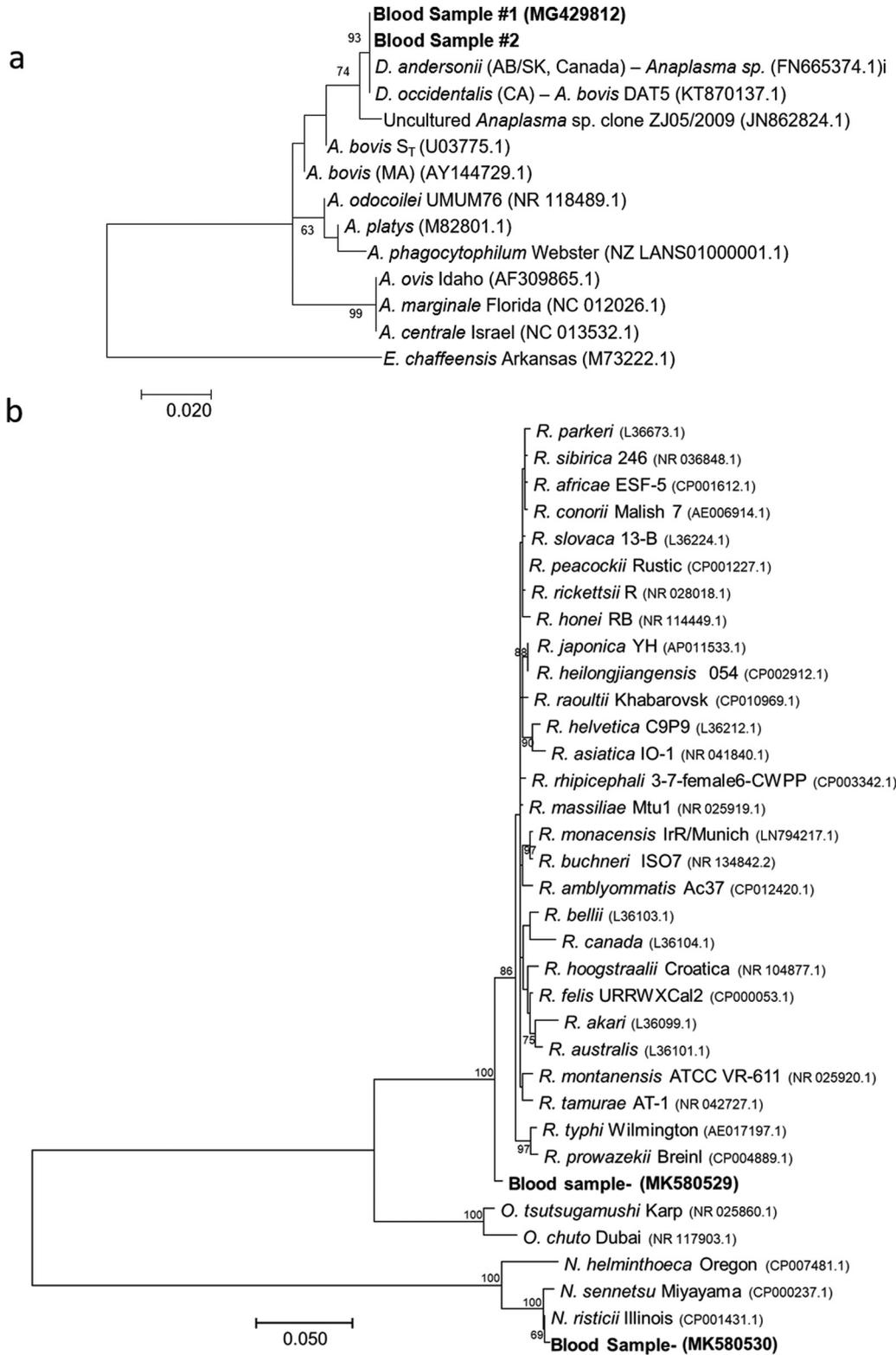
**FIG 2** Box-and-whisker plots showing the distribution of V1-V2 16S read abundances (percentages) for bacterial taxa identified in clinical specimens. The distribution of read abundances for a given taxon relative to all other bacterial taxa detected in the same specimen is shown. The number of clinical samples positive for each of the taxa is indicated at the top of each plot. If known, the disease associated with the bacterial taxon is indicated at the bottom of the y axis. The middle line of the box-and-whisker plot indicates the median, and “X” indicates the mean. The dots outside boxes indicate outliers. Read abundances are shown for *Spirochaetales* (a), *Rickettsiales* (b), and other validated taxa not known to be tick transmitted (c).

the NCBI nonredundant DNA database, species-level taxonomic identifications were achieved for 9/10 (90%) known tick-borne pathogens (Table 2). The *Anaplasma* species identified in two blood specimens demonstrated only 98.2% sequence identity to all *Anaplasma* V1-V2 16S sequences in the NCBI database, providing evidence that it may be a novel pathogen. Sequencing and analysis of a larger

region of the 16S rRNA gene encompassing the V1-V4 region (837 nucleotides) confirmed that the *Anaplasma* species in the two blood specimens shared 100% identity to each other and only 98.7% identity to an uncultured *Anaplasma* sp. (GenBank accession number [JN862824.1](#)), which was the highest identity to any sequence in the NCBI nonredundant database (Table S4). Notably, 100% nucleotide identity was shared across a smaller fragment (357 bp) of the 16S rRNA gene with two *Anaplasma* sequences identified in human-biting ticks: *Dermacentor andersoni* collected in Alberta and Saskatchewan, Canada, and *Dermacentor occidentalis* collected in northern California (23, 24). A phylogenetic comparison of the *Anaplasma* species identified in the two patient blood specimens and *Dermacentor* ticks to other *Anaplasma* species is shown in Fig. 3a. The 16S V1-V4 sequence for the *Anaplasma* species identified here was deposited in the NCBI database with accession number [MG429812](#). The blood specimen positive for "*Candidatus* *Borrelia johnsonii*" was previously reported and confirmed by sequencing and phylogenetic analysis of the 16S rRNA, *flaB*, and *glpQ* genes (14).

**Other taxonomic identifications unique to clinical samples.** Taxonomic predictions for other bacteria in the *Rickettsia* genus unique to samples from patients suspected of having tick-borne illness were also validated by secondary methods (Table 2 and Fig. 1c). Five blood specimens yielded taxonomic predictions within the *Rickettsiales* that were not known tick-borne agents. These included one blood specimen with a MiniKraken and BLASTn taxonomic prediction of *Neorickettsia risticii* (99.7% identity), one blood specimen with a MiniKraken and BLASTn taxonomic prediction of *Rickettsia typhi* (100% identity), and three blood specimens with MiniKraken taxonomic predictions of *Rickettsia* and only 97.9% identity to all other *Rickettsia* species in the NCBI database. The 16S V1-V2 sequence amplified from these three blood specimens displayed 100% nucleotide identity to each other. Sequencing of the 16S V1-V4 region confirmed the accuracy of the MiniKraken and BLASTn predictions at the genus level for all five specimens and at the species level for the *N. risticii*- and *R. typhi*-positive specimens. The V1-V4 16S sequence for the *N. risticii*-positive specimen showed 100% identity (797 bp) to *N. risticii* 812 (GenBank accession number [KX001784.1](#)) and grouped most closely with *N. risticii* in a phylogenetic analysis of *Neorickettsia*, *Orientia*, and *Rickettsia* species (Fig. 3b). The 16S V1-V4 sequence amplified from one of the three blood specimens with taxonomic prediction of *Rickettsia* indicated only 98.6% identity to any other *Rickettsia* species in the NCBI database; phylogenetic analysis revealed that the identified agent falls outside the genus *Rickettsia* and is therefore denoted a *Rickettsia*-like species here (Fig. 3b). The novel *Rickettsia*-like and *N. risticii* 16S sequences were deposited in the NCBI database with accession numbers [MK580529](#) and [MK580530](#), respectively.

Taxonomic predictions for the causative agents of two other zoonotic diseases, Q fever and leptospirosis, were also confirmed by secondary testing (Table 2). The 16S V1-V2 read abundances for these samples are shown in Fig. 2c. Two blood specimens yielded MiniKraken and BLASTn taxonomic predictions of *Coxiella burnetii* (100% identity to each other) and 100% identity to the *C. burnetii* genome reported under GenBank accession number [CP001020.1](#) across the 16S V1-V4 region. Notably, a taxonomic prediction of *Leptospirales* was identified in 15 blood specimens (Table 2). Primary MiniKraken taxonomic predictions of *Leptospira interrogans* (the only *Leptospira* species present in the database) or *Leptospira* were returned for 14 and 1 specimen, respectively. NCBI taxonomic predictions for these specimens included 12 *L. kirschneri*, 2 *L. interrogans*, and 1 *L. noguchii* specimen. Sequencing of the V1-V4 region of 16S confirmed the BLASTn taxonomic predictions for 14 specimens, with 16S sequences displaying 100% identity to *L. kirschneri* (NCBI accession number [NR\\_043051.1](#)) ( $n = 12$ ), 100% identity to *L. interrogans* (NCBI accession number [AY996800](#)) ( $n = 1$ ), and 99% identity to *L. noguchii* (NCBI accession number [NR\\_115926.1](#)) ( $n = 1$ ). One specimen with both MiniKraken and BLASTn taxonomic predictions of *L. interrogans* could not be confirmed by secondary methods.



**FIG 3** Maximum likelihood trees for bacterial species not previously associated with human illness. (a) Phylogenetic analysis of a 357-bp fragment of the 16S rRNA gene amplified from two patient blood specimens compared to *Anaplasma* species and *E. chaffeensis*. The bar corresponds to 0.02 substitutions per site. (b) Phylogenetic analysis of an 834-bp fragment of the 16S rRNA gene amplified from patient blood specimens positive for *N. risticii* and the *Rickettsia*-like agent compared to other *Rickettsia*, *Orientia*, and *Neorickettsia* species. The bar corresponds to 0.05 substitutions per site. Bootstrap values of >60 are shown.

## DISCUSSION

The complexity of tick species and bacterial agents causing tick-borne infections as well as the substantial and increasing disease burden necessitate the development and use of innovative laboratory-based methods for simplified detection. Here, we demonstrate the utility of a single-assay high-throughput 16S V1-V2 metagenomics approach to provide a comprehensive understanding of tick-borne bacteria in specimens from patients clinically suspected of having tick-borne illness. This method displayed a limit of detection comparable to that of real-time PCR and an accuracy of 100% for genus-level taxonomic predictions of *Anaplasma*, *Borrelia*, *Ehrlichia*, *Francisella*, and *Rickettsia*. 16S V1-V2 taxonomic predictions at the species level for tick-borne agents (with the exception of *Rickettsia*) also proved to be precise if a reference genome was present in the MiniKraken database or if a complete V1-V2 sequence was publicly available in the NCBI nonredundant database. This level of accuracy was anticipated based on assessments of species-specific nucleotide differences in the V1-V2 regions of known tick-borne pathogens and highlights the general applicability of utilizing V1-V2 for the detection of tick-borne bacteria, irrespective of the sequencing platform, and the value of *in silico* analyses to delineate suitable variable regions of 16S for the intended use.

16S V1-V2 metagenomic testing of very large numbers of controls (>1,000) alongside >13,000 clinical samples provided convincing evidence that taxonomic predictions for *Anaplasma*, *Borrelia*, *Ehrlichia*, *Francisella*, *Rickettsia*, *Neorickettsia*, *Coxiella*, and *Leptospira* were unique to clinical specimens and that *Rickettsiales* and *Spirochaetales* represent the most commonly identified taxa in specimens from patients suspected of having tick-borne illness. The efficacy of incorporating controls into 16S metagenomic studies, as pertains to ubiquitous DNA contamination arising from commensal, water, or soil bacteria, has been well documented, particularly as pertains to misidentifying suspected causes of human infection (27). A list of bacteria commonly detected in reagents via contamination was previously published, with the majority also identified in our controls (27). The remaining identified taxa unique to clinical samples (see Table S2 in the supplemental material) were primarily environmental bacteria and included previously identified background contaminants (e.g., *Deinococcales* and *Chitinophagales*), suggesting that many of these taxonomic predictions also result from reagent contamination. Proof that 16S V1-V2 taxonomic predictions for tick-borne bacterial agents are unique to clinical specimens and exceptionally accurate establishes the groundwork necessary for the future diagnostic use of V1-V2 for the identification of true infections due to tick-borne bacteria.

Twelve different tick-borne bacterial species were uncovered by 16S V1-V2 metagenomic testing of residual clinical specimens from patients throughout the United States suspected of having one of four different tick-borne diseases (Lyme disease, anaplasmosis, ehrlichiosis, or babesiosis), highlighting the potential complexity involved in diagnosis. Ten known tick-borne pathogens that are the causative agents of seven different diseases, Lyme disease (*B. burgdorferi* and *B. mayonii*), anaplasmosis (*A. phagocytophilum*), ehrlichiosis (*E. chaffeensis*, *E. muris* subsp. *eaucalirensis*, and *E. ewingii*), *Borrelia miyamotoi* disease (*B. miyamotoi*), relapsing fever (*B. hermsii*), spotted fever rickettsiosis (*Rickettsia rickettsii*), and tularemia (*F. tularensis*), were detected. Additionally, two tick-borne bacterial species not previously associated with human illness ("*Candidatus Borrelia johnsonii*" and *Anaplasma* sp.) were discovered (14). As the latter two bacteria were documented previously in the human-biting ticks *Dermacentor* (*D. occidentalis* and *D. andersoni*) and *Carios kelleyi*, they likely represent novel or previously unrecognized tick-transmitted agents of human illness (23, 24, 28). Beyond the detection of a single tick-borne pathogen, for three blood specimens, 16S V1-V2 testing accurately yielded simultaneous taxonomic predictions for *A. phagocytophilum* and one of the following: *B. burgdorferi*, *B. mayonii*, or *B. miyamotoi*. Importantly, all three agents are known to be transmitted by the same tick, *Ixodes scapularis* (29).

The extensive view of bacterial taxa afforded by V1-V2 16S metagenomics also provided insight into other zoonotic bacterial infections that may be clinically confused with tick-borne diseases. Five bacterial pathogens unique to clinical specimens from patients suspected of having tick-borne illness included the causative agents of leptospirosis, Q fever (*C. burnetii*), and murine typhus (*R. typhi*), all of which can present with generalized symptoms that overlap those of several tick-borne infections. Notably, 15 blood samples were positive for one of three different *Leptospira* species, with *L. kirschneri* representing the most commonly identified species. *Leptospira* and *C. burnetii* are not known to be transmitted by ticks to humans (30, 31). *R. typhi* is a flea-transmitted vector-borne pathogen (32). Two other bacteria unique to clinical specimens included a novel *Rickettsia*-like species and *N. risticii*, neither of which has been associated previously with human infection. The *Rickettsia*-like species found in three blood specimens represented a unique sequence in the NCBI database, with 16S phylogenetic analysis placing it between the *Rickettsia* and *Orientia* genera. These specimens originated from two different patients and were submitted by separate providers. Whether these infections were acquired by a tick bite is unknown given that members of the *Rickettsiales* can be transmitted by a variety of arthropod vectors (32). *N. risticii* is the causative agent of equine neorickettsiosis, which horses acquire via the ingestion of infected aquatic insects containing encysted *N. risticii*-infected trematodes (33, 34).

For the diagnosis of tick-borne diseases, health care providers currently must order a specific diagnostic test or panel of specific-pathogen-based tests based on clinical suspicion. Due to the polyvector and polymicrobial nature of bacterial tick-borne diseases, this paradigm of a single or a panel of diagnostic assays is far from an ideal diagnostic approach for suspected bacterial tick-borne infection. To this end, our results indicate that V1-V2 16S metagenomics shows promise as a single-assay, comprehensive diagnostic test for both known and novel tick-borne pathogens where the number of bacterial genomic copies in blood is above or within the range for direct detection. The potential opportunity for physicians to order a single test could help simplify the diagnosis of bacterial tick-borne diseases and directly benefit patient care. Here, a 100% increase in the number of different tick-borne bacterial pathogens identified by 16S V1-V2 metagenomics was demonstrated compared to initial targeted PCR assays. Moving forward, the possibility exists for the development of similar assays targeting 18S rRNA or pan-arboviral sequences for the detection of tick-borne protozoa and viruses, respectively.

Limitations of this diagnostic approach exist given that direct detection of bacterial tick-borne pathogens by V1-V2 16S metagenomics relies on several factors, including the level of bacteremia resulting from the infecting pathogen, the timing of specimen sampling in relation to illness onset, the specimen type tested, and whether specimen sampling was performed pre- or posttreatment. This is particularly relevant for some tick-borne bacteria, most notably, *B. burgdorferi*, where the number of bacteria/genomic copies in blood is low, very often below the limit of PCR detection (11, 26). In addition, taxonomic predictions are only as accurate and up-to-date as the database utilized. As metagenomic studies continue to identify bacterial taxa associated with human-biting ticks and their zoonotic hosts, the addition of complete 16S sequences, encompassing the V1-V2 region, to public databases is imperative for ensuring accurate taxonomic predictions.

In summary, this large-scale study demonstrates that a single 16S V1-V2 rRNA gene metagenomic-based assay can be used to (i) accurately detect bacterial tick-borne pathogens in clinical samples, (ii) uncover novel bacterial agents likely transmitted to humans by a tick bite as well as other novel bacterial species causing human illness, (iii) identify and quantify the frequency of tick-transmitted bacterial coinfections, and (iv) define bacterial pathogens that may cause symptoms that are clinically confused with those of tick-borne diseases. Ultimately, this agnostic methodology may help to simplify diagnostic testing as well as fill surveillance gaps and

keep public health officials and clinicians on the leading edge of the established and emerging threat posed by tick-borne infections.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 0.7 MB.

## ACKNOWLEDGMENTS

We greatly acknowledge Amy Schwartz for assistance with IRB continuations. We also thank John Young and Ryan Pappert for sample receipt and accessioning; Parker Abbot, Zoe Geisen, and Sam Melchior for support with sample shipping and aliquoting at MDH; Gil Kersh, Alex Hoffmaster, Paige Armstrong, and Renee Galloway for critical review of the manuscript; and Jonathan Schmitz for study design, testing, and specimen handling at VUMC.

The views and opinions expressed here are those of the authors alone and do not represent the official position of the Centers for Disease Control and Prevention.

This work was supported by the CDC's Office of Advanced Molecular Detection (project identifier AMD 90) and the CDC's TickNet Emerging Infections Program. S.O. was supported in part by an appointment to the Research Participation Program at the Centers for Disease Control and Prevention administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the CDC.

## REFERENCES

1. Paules CI, Marston HD, Bloom ME, Fauci AS. 2018. Tickborne diseases—confronting a growing threat. *N Engl J Med* 379:701–703. <https://doi.org/10.1056/NEJMp1807870>.
2. Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, Hooks H, Partridge SK, Visser SN, Beard CB, Petersen LR. 2018. Vital signs: trends in reported vectorborne disease cases—United States and territories, 2004–2016. *MMWR Morb Mortal Wkly Rep* 67:496–501. <https://doi.org/10.15585/mmwr.mm6717e1>.
3. Hinckley AF, Connally NP, Meek JI, Johnson BJ, Kemperman MM, Feldman KA, White JL, Mead PS. 2014. Lyme disease testing by large commercial laboratories in the United States. *Clin Infect Dis* 59:676–681. <https://doi.org/10.1093/cid/ciu397>.
4. Nelson CA, Saha S, Kugeler KJ, Delorey MJ, Shankar MB, Hinckley AF, Mead PS. 2015. Incidence of clinician-diagnosed Lyme disease, United States, 2005–2010. *Emerg Infect Dis* 21:1625–1631. <https://doi.org/10.3201/eid2109.150417>.
5. Stanek G, Fingerle V, Hunfeld KP, Jaulhac B, Kaiser R, Krause A, Kristoferitsch W, O'Connell S, Ornstein K, Strle F, Gray J. 2011. Lyme borreliosis: clinical case definitions for diagnosis and management in Europe. *Clin Microbiol Infect* 17:69–79. <https://doi.org/10.1111/j.1469-0691.2010.03175.x>.
6. Pritt BS, Mead PS, Johnson DKH, Neitzel DF, Respicio-Kingry LB, Davis JP, Schiffman E, Sloan LM, Schriefer ME, Replogle AJ, Paskewitz SM, Ray JA, Bjork J, Steward CR, Deedon A, Lee X, Kingry LC, Miller TK, Feist MA, Theel ES, Patel R, Irish CL, Petersen JM. 2016. Identification of a novel pathogenic *Borrelia* species causing Lyme borreliosis with unusually high spirochaetemia: a descriptive study. *Lancet Infect Dis* 16:556–564. [https://doi.org/10.1016/S1473-3099\(15\)00464-8](https://doi.org/10.1016/S1473-3099(15)00464-8).
7. Pritt BS, Sloan LM, Johnson DK, Munderloh UG, Paskewitz SM, McElroy KM, McFadden JD, Binnicker MJ, Neitzel DF, Liu G, Nicholson WL, Nelson CM, Franson JJ, Martin SA, Cunningham SA, Steward CR, Bogumill K, Bjorgaard ME, Davis JP, McQuiston JH, Warshauer DM, Wilhelm MP, Patel R, Trivedi VA, Eremeeva ME. 2011. Emergence of a new pathogenic *Ehrlichia* species, Wisconsin and Minnesota, 2009. *N Engl J Med* 365:422–429. <https://doi.org/10.1056/NEJMoa1010493>.
8. Chowdri HR, Gugliotta JL, Berardi VP, Goethert HK, Molloy PJ, Sterling SL, Telford SR. 2013. *Borrelia miyamotoi* infection presenting as human granulocytic anaplasmosis: a case report. *Ann Intern Med* 159:21–27. <https://doi.org/10.7326/0003-4819-159-1-201307020-00005>.
9. Shapiro MR, Fritz CL, Tait K, Paddock CD, Nicholson WL, Abramowicz KF, Karpathy SE, Dasch GA, Sumner JW, Adem PV, Scott JJ, Padgett KA, Zaki SR, Eremeeva ME. 2010. *Rickettsia* 364D: a newly recognized cause of eschar-associated illness in California. *Clin Infect Dis* 50:541–548. <https://doi.org/10.1086/649926>.
10. La Scola B, Raoult D. 1997. Laboratory diagnosis of rickettsioses: current approaches to diagnosis of old and new rickettsial diseases. *J Clin Microbiol* 35:2715–2727. <https://doi.org/10.1128/JCM.35.11.2715-2727.1997>.
11. Aguero-Rosenfeld ME, Wang G, Schwartz I, Wormser GP. 2005. Diagnosis of Lyme borreliosis. *Clin Microbiol Rev* 18:484–509. <https://doi.org/10.1128/CMR.18.3.484-509.2005>.
12. Kato CY, Chung IH, Robinson LK, Austin AL, Dasch GA, Massung RF. 2013. Assessment of real-time PCR assay for detection of *Rickettsia* spp. and *Rickettsia rickettsii* in banked clinical samples. *J Clin Microbiol* 51:314–317. <https://doi.org/10.1128/JCM.01723-12>.
13. Lee-Lewandrowski E, Chen Z, Branda J, Baron J, Kaufman HW. 2020. Laboratory blood-based testing for non-Lyme disease tick-borne infections at a national reference laboratory. *Am J Clin Pathol* 153:139–145. <https://doi.org/10.1093/ajcp/aqz139>.
14. Kingry LC, Anacker M, Pritt B, Bjork J, Respicio-Kingry L, Liu G, Sheldon S, Boxrud D, Strain A, Oatman S, Berry J, Sloan L, Mead P, Neitzel D, Kugeler KJ, Petersen JM. 2018. Surveillance for and discovery of *Borrelia* species in US patients suspected of tickborne illness. *Clin Infect Dis* 66:1864–1871. <https://doi.org/10.1093/cid/cix1107>.
15. Lo C-C, Chain PSG. 2014. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics* 15:366. <https://doi.org/10.1186/s12859-014-0366-2>.
16. Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
17. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
18. Margos G, Castillo-Ramirez S, Cutler S, Dessau RB, Eikeland R, Estrada-Pena A, Gofton A, Grana-Miraglia L, Hunfeld KP, Krause A, Lienhard R, Lindgren PE, Oskam C, Rudolf I, Schwartz I, Sing A, Stevenson B, Wormser GP, Fingerle V. 2020. Rejection of the name *Borrelia* and all proposed species comb. nov. placed therein. *Int J Syst Evol Microbiol* 70:3577–3581. <https://doi.org/10.1099/ijsem.0.004149>.
19. Johnson DK, Schiffman EK, Davis JP, Neitzel DF, Sloan LM, Nicholson WL, Fritsche TR, Steward CR, Ray JA, Miller TK, Feist MA, Uphoff TS, Franson

- JJ, Livermore AL, Deedon AK, Theel ES, Pritt BS. 2015. Human infection with Ehrlichia muris-like pathogen, United States, 2007-2013. *Emerg Infect Dis* 21:1794–1799. <https://doi.org/10.3201/eid2110.150143>.
20. Kugeler KJ, Pappert R, Zhou Y, Petersen JM. 2006. Real-time PCR for Francisella tularensis types A and B. *Emerg Infect Dis* 12:1799–1801. <https://doi.org/10.3201/eid1211.060629>.
21. Versage JL, Severin DD, Chu MC, Petersen JM. 2003. Development of a multitarget real-time TaqMan PCR assay for enhanced detection of Francisella tularensis in complex specimens. *J Clin Microbiol* 41:5492–5499. <https://doi.org/10.1128/JCM.41.12.5492-5499.2003>.
22. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
23. Chilton NB, Dergousoff SJ, Lysyk TJ. 2018. Prevalence of Anaplasma bovis in Canadian populations of the Rocky Mountain wood tick, Dermacentor andersoni. *Ticks Tick Borne Dis* 9:1528–1531. <https://doi.org/10.1016/j.ttbdis.2018.07.003>.
24. Lane RS, Mun J, Peribanez MA, Fedorova N. 2010. Differences in prevalence of Borrelia burgdorferi and Anaplasma spp. infection among host-seeking Dermacentor occidentalis, Ixodes pacificus, and Ornithodoros coriaceus ticks in northwestern California. *Ticks Tick Borne Dis* 1:159–167. <https://doi.org/10.1016/j.ttbdis.2010.09.004>.
25. Reller ME, Dumler JS. 2018. Development and clinical validation of a multiplex real-time quantitative PCR assay for human infection by Anaplasma phagocytophilum and Ehrlichia chaffeensis. *Trop Med Infect Dis* 3:14. <https://doi.org/10.3390/tropicalmed3010014>.
26. Liveris D, Schwartz I, McKenna D, Nowakowski J, Nadelman R, Demarco J, Iyer R, Bittker S, Cooper D, Holmgren D, Wormser GP. 2012. Comparison of five diagnostic modalities for direct detection of Borrelia burgdorferi in patients with early Lyme disease. *Diagn Microbiol Infect Dis* 73:243–245. <https://doi.org/10.1016/j.diagmicrobio.2012.03.026>.
27. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:87. <https://doi.org/10.1186/s12915-014-0087-z>.
28. Gill JS, Ullmann AJ, Loftis AD, Schwan TG, Raffel SJ, Schrupf ME, Piesman J. 2008. Novel relapsing fever spirochete in bat tick. *Emerg Infect Dis* 14:522–523. <https://doi.org/10.3201/eid1403.070766>.
29. Eisen RJ, Eisen L. 2018. The blacklegged tick, Ixodes scapularis: an increasing public health concern. *Trends Parasitol* 34:295–309. <https://doi.org/10.1016/j.pt.2017.12.006>.
30. Duron O, Sidi-Boumedine K, Rousset E, Moutailler S, Jourdain E. 2015. The importance of ticks in Q fever transmission: what has (and has not) been demonstrated? *Trends Parasitol* 31:536–552. <https://doi.org/10.1016/j.pt.2015.06.014>.
31. Johnson RC. 1996. Chapter 35: Leptospira. In Baron S (ed), *Medical microbiology*, 4th ed. University of Texas Medical Branch at Galveston, Galveston, TX.
32. Azad AF, Beard CB. 1998. Rickettsial pathogens and their arthropod vectors. *Emerg Infect Dis* 4:179–186. <https://doi.org/10.3201/eid0402.980205>.
33. Mott J, Muramatsu Y, Seaton E, Martin C, Reed S, Rikihisa Y. 2002. Molecular analysis of Neorickettsia risticii in adult aquatic insects in Pennsylvania, in horses infected by ingestion of insects, and isolated in cell culture. *J Clin Microbiol* 40:690–693. <https://doi.org/10.1128/jcm.40.2.690-693.2002>.
34. Baird JD, Arroyo LG. 2013. Historical aspects of Potomac horse fever in Ontario (1924-2010). *Can Vet J* 54:565–572.