A Next-Generation Sequencing-Based Molecular Approach to Characterize a Tick Vector in Lyme Disease

Anil K. Madugundu,^{1-3,*} Babylakshmi Muthusamy,^{1,2,*} Sreelakshmi K. Sreenivasamurthy,^{1-4,*} Chandra Bhavani,² Jyoti Sharma,^{1,2} Bankatesh Kumar,⁴ Krishna R. Murthy,^{1,2,5} Raju Ravikumar,⁴ and Akhilesh Pandey^{3,6-8}

Abstract

Next-generation sequencing approaches have revolutionized genomic medicine and enabled rapid diagnosis for several diseases. These approaches are widely used for pathogen detection in several infectious diseases. Lyme disease is a tick-borne infectious disease, which affects multiple organs. The causative organism is a spirochete, *Borrelia burgdorferi*, which is transmitted by ticks. Lyme disease can be treated easily if detected early, but its diagnosis is often delayed or is incorrect leading to a chronic debilitating condition. Current confirmatory diagnostic tests for Lyme disease rely on detection of antigens derived from *B. burgdorferi*, which are prone to both false positives and false negatives. Instead of focusing only on the human host for the diagnosis of Lyme disease, one could also attempt to identify the vector (tick) and the causative organism carried by the tick. Since all ticks do not transmit Lyme disease, it can be informative to accurately identify the tick from the site of bite, which is often observed by the patient and discarded. However, identifying ticks based on morphology alone requires a trained operator and can still be incorrect. Thus, we decided to take a molecular approach by sequencing DNA and RNA from a tick collected from an individual bitten by the tick. Using next-generation sequencing, we confirmed the identity of the tick as a dog tick, *Dermacentor variabilis*, and did not identify any pathogenic bacterial sequences, including *Borrelia* species. Despite the limited availability of nucleotide sequences for many types of ticks, our approach correctly identified the tick species. This proof-of-principle study demonstrates the potential of next-generation sequencing in the diagnosis of tick-borne infections, which can also be extended to other zoonotic diseases.

Keywords: *Borrelia*, Lyme disease, tick-borne diseases, dog tick, molecular diagnosis

Introduction

TICKS ARE SECOND ONLY TO MOSQUITOES in the number of zoonotic diseases transmitted in the world (Parola and Raoult, 2001). Tick-borne diseases are common and caused by a variety of bacterial, viral, and protozoan pathogens (Edlow, 1999, 2008). Lyme disease was first clinically characterized in 1972 while studying a small group of children and adults with symptoms of arthritis in a city of Connecticut (Steere et al., 1977a, 1977b). Cases with symptoms

similar to Lyme disease were documented in Europe as early as 1882 and in archived vector ticks preserved in 1945 (Matuschka et al., 1995). Later reports confirmed Lyme disease in Asian countries, such as China, India, Russia, and Japan (Hao et al., 2011; Jairath et al., 2014; Steere, 2001; Takano et al., 2011).

Species of *Ixodes* are the principal vectors of Lyme disease in the United States (*Ixodes scapularis*), Europe (*Ixodes ricinus*), and Asia (*Ixodes persulcatus*) (Spielman, 1994). The incidence of tick-borne diseases has increased in the last two

¹Manipal Academy of Higher Education (MAHE), Manipal, Karnataka, India.

²Institute of Bioinformatics, International Technology Park, Bangalore, Karnataka, India.

³McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.

⁴National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India.

⁵Vittala International Institute of Ophthalmology, Bangalore, Karnataka, India.

⁶Departments of Biological Chemistry, ⁷Oncology, and ⁸Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland.

*These authors contributed equally to this work.

decades and is responsible for substantial morbidity in the United States (Wallace et al., 2016). According to the latest Centers for Disease Control and Prevention (CDC) report for the United States, 24 states in the Mid-Atlantic, Northeastern, South Atlantic, and North Central states of the USA accounted for 93 percent of all reported cases of Lyme disease (Adams et al., 2017). Changes in climate, such as an increase in average temperatures and humidity have been implicated in widening the geographical perimeters of tick distribution (Science, 2011).

Ticks are vectors transmitting different infectious agents and possess varying affinity to human and other hosts (Estrada-Pena and Jongejan, 1999). Some of the tick-borne diseases in humans include anaplasmosis, babesiosis, Colorado tick fever, ehrlichiosis, Lyme disease, Rocky Mountain spotted fever, tick-borne relapsing fever, and tularemia (Dantas-Torres et al., 2012; Preac-Mursic et al., 1986; Shaw et al., 2001; Spach et al., 1993; Walker, 1998; Zaidi and Singer, 2002). In these disorders, a wide array of neurological manifestations, including headache, meningitis, meningoencephalitis, and deafness can be observed and myocarditis, arthritis, splenomegaly, and hepatomegaly can be prominent during the chronic phases, which usually set in after the latent phase (Bogovic and Strle, 2015; Dworkin et al., 2008; McNair and Brown, 1991; Misić-Majerus et al., 2005; Parola and Raoult, 2001).

Lyme disease is caused by a spirochete, *Borrelia burgdorferi*, which is transmitted by a tick bite. Lyme disease occurs in temperate regions and is a common vector-borne disease in North America (Aucott et al., 2009; Brownstein et al., 2005). Ticks are well-known reservoirs of *Borrelia* and are usually infected during the nymph stage. Humans are accidental hosts for *Borrelia* during the feeding cycle of the tick to mature as an adult. The nymph stage of the tick transmits the bacteria after feeding for at least 24–48 h. The bite site reveals a rash commonly referred to as erythema migrans (Kirkland et al., 1997; Masters et al., 1998).

As the disease progresses, the spirochetes can disseminate through the hematogenous route to involve the nervous and cardiovascular systems, resulting in nerve palsies (Peltomaa et al., 1998) or carditis (Armstrong et al., 1992), respectively. Multiple secondary skin lesions (Mullegger, 2004), flu-like symptoms (Feder et al., 1993), and arthralgia (Berglund and Eitrem, 1993) are also noted during this stage. After a period of latent infection lasting several weeks, a chronic progressive phase ensues manifesting as recurrent attacks of arthritis, particularly affecting knee joints (Steere et al., 1977a).

Objective signs of the disease can be resolved with appropriate antibiotic intervention although some symptoms (fatigue, myalgia, and short-term memory) are stubborn, persisting for more than 6 months and are referred to as “post-Lyme disease syndrome” (Feder et al., 2007). Ocular manifestations have been reported to occur in different stages of Lyme disease. These include episcleritis, conjunctivitis, keratitis, diplopia, anterior uveitis, intermediate uveitis, retinal vasculitis, posterior uveitis, scleritis, neuroretinitis, and pupillary disorders such as mydriasis, Horner’s syndrome, Argyll Robertson pupil, and neuroretinitis (Amer et al., 2009; Babu and Murthy, 2010; Huppertz et al., 1999; Krist and Wenkel, 2002; Mikkila et al., 2000; Miyashiro et al., 1999; Uchio and Ohno, 1999).

Establishing an accurate diagnosis in a suspected case of Lyme disease is essential since an erroneous diagnosis can

lead to lack of appropriate treatment or inappropriate antibiotic administration. Presence of a rash typically seen in Lyme disease along with history of exposure in endemic areas and confirmatory tests can all help establish a diagnosis to initiate treatment. However, a classic bull’s-eye rash is seen in only 19% of the cases (Tibbles and Edlow, 2007) and an absence of rash is noted in at least 10% of the suspected cases of Lyme disease (Steere et al., 2003). Laboratory tests for Lyme disease are based on detecting antibodies to Lyme-specific antigens in the serum of the patients.

Recent CDC guidelines for serodiagnosis of Lyme disease recommend an initial enzyme immunoassay followed by western blots (both IgM and IgG) on positive and equivocal cases identified by the enzyme immunoassays. False negative results on serologic tests have been reported when performed in early stages of infection (Marques, 2015).

Previous studies report the association between the tick type and pathogenic bacteria. Several species of *Ixodes* ticks are strongly associated with the *Borrelia* infections compared with other tick types (Estrada-Pena et al., 2016; Lane et al., 1991). Thus, establishing the identity of the tick type and bacteria can resolve the possibility of infection by *Borrelia* and aid in the diagnosis of Lyme disease.

Although serological tests for detection of antibodies have been the mainstay of diagnosis (Lindblom et al., 2013), the diagnostic repertoire for tick-borne diseases has now expanded to include molecular methods for detection of ticks, which is based on duplex polymerase chain reaction (PCR) (Courtney et al., 2004). In this study, we applied next-generation sequencing methods to identify the type of tick as well as the various bacterial species present in the tick. Our analysis identified the tick as *Dermacentor variabilis* (American dog tick) and there were no pathogenic bacterial species present in the tick indicating that it is unlikely that the tick can cause Lyme disease in the individual with the tick bite.

Materials and Methods

DNA and RNA isolation

A tick still attached to a human subject was obtained in a deidentified fashion for molecular testing. The use of discarded deidentified specimens is not considered as human subjects research as per National Institutes of Health (NIH) guidelines and exempt from ethics review. Continued observation of the subject did not reveal any erythematous rash or any other symptoms consistent with Lyme disease excluding a diagnosis of Lyme disease. The collected tick was washed with phosphate-buffered saline three times to remove any contaminants. Genomic DNA and total RNA were isolated from the tick using the Qiagen AllPrep DNA/RNA Mini Kit from Qiagen (Cat No. 80204) as per the manufacturer’s instructions.

Briefly, the tick was homogenized and lysed using a TissueLyser II (Cat No. 85300) bead mill from Qiagen in RLT lysis buffer for 15 min at 30 Hz (1800 oscillations/min). The lysate was subsequently loaded on a DNA purification column and the filtrate was used for RNA isolation. The quality of genomic DNA and total RNA was checked using the Agilent 2100 BioAnalyzer and the quantity estimations were done using NanoDrop and Qubit Fluorometer 3.0. About 1 µg of genomic DNA and total RNA was used for library preparation. The DNA and total RNA isolated from the tick were

sequenced at the Genome Technology Center core facility at New York University School of Medicine.

Genomic DNA was fragmented using Covaris LE220 and the fragments were used for preparation of the library using KAPA general DNA-seq library prep protocol. Ribosomal RNA was depleted from the total RNA using Ribo-Zero Gold and the enriched RNA was used for library preparation as described in the manufacturer's protocol using TruSeq Stranded Total RNA Kit. The quality of library was assessed using the Agilent 2100 BioAnalyzer. The stranded TruSeq library and the genomic DNA library were each sequenced using the Illumina SBS Kit V4 for paired-end 100 bp reads on a single lane of Illumina HiSeq 2500 in the high-output mode.

Genomic DNA sequencing data analysis

Quality analysis and processing of the raw reads was carried out using FastQC (v0.11.5) and fqtrim (v0.9.4) (Perte, 2015) tools, respectively. All genomic DNA sequences available under the clade "Ixodida" (Taxonomy ID: 6935) were downloaded from National Center for Biotechnology Information (NCBI), which correspond to 757,132 nucleotide sequences. Quality processed reads were mapped against these sequences using Bowtie2 (v2.2.5) (Langmead and Salzberg, 2012). Default preset "—sensitive" parameters (-D 15 -R 2 -N 0 -L 22 -i S,1,1.15) in end-to-end mode were used for Bowtie2 alignments throughout this study.

To rule out the DNA contamination from possible host organisms, we aligned raw reads against the genome of human (*Homo sapiens*), dog (*Canis lupus familiaris*), and two deer species (*Capreolus capreolus* and *Odocoileus virginianus*). To identify the pathogenic bacteria from the whole genome sequencing (WGS) data, we aligned the processed reads against the representative bacterial/archaeal genome database available at NCBI using Bowtie2 with aforementioned parameters.

This database comprised of genome sequences from 4378 bacterial species, including various *Borrelia* species. Alignment was carried out using Bowtie2 with similar parameters as explained above. The alignment of mapped reads against tick genomic DNA sequences was carried out using Basic Local Alignment Search Tool (BLAST) to evaluate the sequence identity. All alignments with sequence identity <50% were filtered. The schematic of the workflow employed in the whole-genome sequencing analysis is depicted in Figure 1A.

Transcriptome sequencing data analysis

Quality analysis and processing of the raw reads was carried out using FastQC (v0.11.5) and fqtrim (v0.9.4) tools, respectively. A total of 161,514 mRNA sequences were downloaded from NCBI under the taxonomic order "Ixodida" (Taxonomy ID: 6935). The quality processed reads were mapped to these transcript sequences using Bowtie2 using the same parameters described above. Reads that uniquely aligned to *D. variabilis* were further extracted using "bam2fastx" utility tool of TopHat-v2.1.1 (Trapnell et al., 2009). To identify the presence of bacterial pathogen sequences in the RNA-seq data, the raw reads were aligned against representative bacterial/archaeal genome database using Bowtie2. The schematic of the workflow employed in the RNA-seq data analysis is depicted in Figure 1B.

Results

WGS data analysis to identify tick species

WGS on an Illumina HiSeq 2500 sequencer generated 224,784,714 paired-end reads of 100 bp read length. To remove the sequences corresponding to contaminants from possible host organisms, the processed raw reads were aligned against genome sequences of dog, cat, and deer. We observed that 0.25% of the raw reads aligned and those reads

FIG. 1. Workflow for analysis of next-generation sequencing data. Workflow employed for analysis of (A) whole-genome and (B) RNA-seq sequencing data. Genomic DNA and RNA sequences for *Ixodida* (taxonomy ID: txid6935) and bacterial genomes were downloaded from NCBI as indicated. NCBI, National Center for Biotechnology Information.

FIG. 2. Summary of alignment data from WGS. **(A)** Histogram of the number of WGS reads that mapped to various tick species. The size of *vertical bars* and *circles* represent the number of reads that map to the indicated tick species. Reads mapping to tick genomes were aligned against individual draft genomes of various ticks using BLAST to identify the exact tick species based on sequence identity. **(B)** Distribution of the number of mapped reads and sequence identity across tick species. The number of mapped reads against *Ixodes scapularis* was higher than against *Dermacentor variabilis* because of overrepresentation of *I. scapularis* sequences in NCBI. WGS, whole-genome sequencing.

were removed. Alignment of filtered reads against the available genome sequences of Ixodida taxonomical order, identified about 840,000 reads mapping to *Dermacentor* spp., 680,000 reads mapping to *Ixodes* spp. and 350,000 reads mapping to *Rhipicephalus* spp. (Fig. 2A).

Next-generation sequencing (NGS) aligners, such as Bowtie2 map short reads with high sequence identity and miss secondary alignments to other less similar reference sequences. To identify the exact tick species, we used BLAST tool to detect the sequence identity of mapped reads against various tick species. BLAST alignment of WGS reads against DNA sequences of various tick species identified about 1,774,000 reads mapping to species of *Dermacentor*, 450,000 to *Rhipicephalus*, 130,000 to *Amblyomma*, and 45,000 to *Ixodes*. Even though a complete genome sequence is not available, we observed many reads aligned to *D. variabilis* draft assembly with high sequence identity. Read alignment against all other tick species was limited (Fig. 2B).

These results reveal that the tick species is *D. variabilis*. As an example, the alignment between subolesin gene of *D. variabilis* and consensus sequence derived from WGS shows identical matches as shown in the Figure 3, whereas a much lower sequence identity was observed in the alignments against *Rhipicephalus*, *Amblyomma*, and *Ixodes*. These observations suggest that the tick species is indeed *D. variabilis*.

Transcriptome sequencing data confirms the identity of the tick species

The total RNA isolated from the tick was sequenced using HiSeq 2500, which resulted in generation of 262,215,549

paired-end reads of read length 100 bp. Unlike genomic sequence data, RNA sequences were available for a number of tick species in the NCBI databases. After quality processing of the raw reads, they were aligned against the available mRNA sequences of various tick species. We identified over 200 million reads (76%) that mapped to *D. variabilis*, followed by 72 million reads (28%) to *Rhipicephalus microplus* and 6 million reads (2%) to *Amblyomma* species (Fig. 4A). To determine the exact tick type, we carried out BLAST alignment of RNA-seq reads against mRNA sequences of various tick species.

We observed that the reads aligning to the mRNA sequences of *D. variabilis* were nearly identical but matches poorly to all other tick species (Fig. 4B). As an example, Figure 5 shows the alignment of a very highly conserved gene, beta-actin, against *D. variabilis* and related ticks. The lack of identity to other types of tick can be easily inferred based on the high number of mismatches in the cDNA sequence alignment. Overall, both DNA and RNA sequencing analyses provide definitive evidence that the tick species is the American dog tick, *D. variabilis*, and any overlap with other tick species is due to the fact that these species are closely related.

Identification of microbes and other pathogens within the tick

Upon concluding the tick species as *D. variabilis*, we next proceeded to identify any infectious microbes that may be harbored by the tick. We rationalized that the presence of a bacterial species within the tick would be confirmed if we identify reads that map uniquely to the bacterial genome.

Alignment against *Dermacentor variabilis*

ATGCTTGTGGCACAATTAAGCGGACACACAGCATGGATCCCTCGACAGTCGGAGATTCGCCGGAAGAAGACGGAGATGTATGCTCTCTGAGTCTCCCGACCGGCTCTCAACAAGGGCGCACCAATCAATCCCTCGCGGTTT 150

ATGCTTGTGGCACAATTAAGCGGACACACAGCATGGATCCCTCGACAGTCGGAGATTCGCCGGAAGAAGACGGAGATGTATGCTCTCTGAGTCTCCCGACCGGCTCTCAACAAGGGCGCACCAATCAATCCCTCGCGGTTT 150

GGAGAGTGGCCACCGAAATTTGACTTCAGAGGAGATAGCAGCAACATTCGCGAGGAGATCGCAGCGGTG 219

GGAGAGTGGCCACCGAAATTTGACTTCAGAGGAGATAGCAGCAACATTCGCGAGGAGATCGCAGCGGTG 219

Sequence identity: 89%

Alignment against *Rhipicephalus sanguineus*

	Sequence identity: 89%
1	ATGCGTTGTGCGACATTAAACGGGACACAGCACTGGGATCCCTCGACAGTCGGAAGCGGAGATATATGCTCTGCCAGCGGCTCCCAACAGGGGGCGACCAATCAATCCCTCGCGGTTT 150
1	ATGCGTTGTGCGACATTAAACGGGACACAGCACTGGGATCCCTCGACAGTCGGAAGCGGAGATATATGCTCTGCCAGCGGCTCCCAACAGGGGGCGACCAATCAATCCCTCGCGGTTT 138
151	GGAGAGTGGCCACCGAATGTTGACTTCAGAGGAGATAGCAGCCCAACATTCGCGAGGAGATCGCAGCGGTG 219
299	GGAGAGTGGCCACCGAATGTTGACTTCAGAGGAGATAGCAGCCCAACATTCGCGAGGAGATCGCAGCGGCTA 207

Sequence identity: 89%

Alignment against <i>Amblyomma americanum</i>		Sequence identity: 89%
1	ATGCTTTGTGCGACATTAAAGCGGACACACGACTGGATCCCTCCTGCACAGTCCGAACGGCAGATCGCGAAAGACGGAGATGTATGCCTTCTCAGTCTCCCAACAGGGCGACCAATCAATCCCTCGCGTTT	150
1	ATGCTTTGCGCAACATTAAAGCGCAACGACGACTGGATCCCTGTCACAGTCCGAACGGGAAGATCAACAAAGCGGCGAGATGTATGCCTTTTTCAGTCAACCCAGCTCTCTCCCAAGGGCGACCAATCAATCCCTCGCGTTTC	150
151	GGAGAGTGGCCACCGAAAGTTGACTTTCAGAGGAGATAGCACCAACATTCGCGAGGAGATCGCAGCGCTG	219
21	GGCCACGTGGCCACCAAGCTGACTTTCAGAGGAGATAGCACCAACATTCGCTGAGGAAATCGCGGCGCTG	219

Sequence identity: 86%

Alignment against <i>Ixodes ricinus</i>		Sequence identity: 86%
1	ATGCGTTTGGCAGATTAAAGCGGACACACGACTGGGATCCCTGTCACAGTCCGAAACGGCAGATCGCCGAAAAGACGGAGATGTATGCTCTCTGTCACTCTCCCAACCGCG--	147
244	ATGCGTTTGGCAGATTAAAGCGGACACACGACTGGGATCCCTGTCACAGTCCGAAACGGCAGATCGCCGAAAAGACGGAGATGTATGCTCTCTCTGTCACTCTCCCAACCGCG--	150
448	TTTGGAGAGTCCGACAGAGATTGATTCAGAGGATAGCAGCAACTCCGCGAGGAGCGCGAGCGCTG	219
51	TTTCGCTGAACTGCCACCGAAATTAATTTACAGAGGAGATAGCAGCAACTTCGCGAGGAGAAATCGCAGCTGTG	222

FIG. 3. Conservation of subolesin gene across the tick species. Sequence alignment shows the conservation of subolesin gene identified from the sequenced tick against four related tick species along with sequence identity indicated in percent.

FIG. 4. Summary of alignment data from RNA-seq experiments. Sizes of *vertical bars* and *circles* represent the number of reads that map to the indicated tick species. Reads mapping to tick transcripts were aligned against individual transcript assemblies of various ticks using BLAST to confirm the identity of tick species based on sequence identity. **(A)** Histogram of the number of RNA-seq reads that mapped across tick species. **(B)** Distribution of number of reads and sequence identity across tick species.

Therefore, the genomic and transcriptomic data were aligned against the representative genomes of 4378 bacterial species, downloaded from NCBI. We observed that the large majority of sequences corresponded to only two bacterial genera—*Erythrobacter* and *Arsenophonus*, which are nonpathogenic. Species of *Arsenophonus* are earlier known to be transmitted by *Dermacentor* species (Borges et al., 2015; Dergousoff and Chilton, 2010; Grindle et al., 2003).

We neither identified any pathogenic bacteria that are known to be transmitted by dog ticks—*Borrelia*, *Rickettsiella*, *Anaplasma*, and *Ehrlichia* nor by other types of ticks (Ammerman et al., 2004; Gall et al., 2016; Holden et al., 2003; Rynkiewicz et al., 2015). Thus, we conclude the absence of any pathogenic bacteria being transmitted by the tick and, in particular, *B. burgdorferi*.

Discussion

An early detection of tick-borne diseases is suboptimal because of nonspecific and common symptoms until a visible rash is noticed after few days (Steere et al., 1977a). Current diagnosis of tick-borne infections, such as immunoassays and PCR are less specific in detecting the type of bacteria and associated disease to guide effective treatment (Johnson et al., 1997; Rijpkema et al., 1997; Sanchez et al., 2016). Serological testing is the mainstay for diagnosis of Lyme disease and the available tests include, enzyme-linked immunosorbent assay (ELISA), immunofluorescence assays, and western blot.

However, the interpretation of these immunological tests has caveats. Since these tests detect antibodies to *Borrelia* antigens, patients may test negative until seroconversion. In the United States, only 20–50% cases were positive for IgM antibodies in the acute phase of the disease (Leefflang et al., 2016;

Steere et al., 2008). During convalescence, 70–80% cases were positive for the IgM antibodies, and after 4–8 weeks of infection, nearly 100% cases were positive for IgG antibodies. Nevertheless, there is clear evidence that early initiation of treatment helps avoid late complications of Lyme disease.

The CDC recommends a two-tier testing to define Lyme borreliosis cases. The first step is an ELISA or immunofluorescence assay. Any positive or equivocal tests are tested by a western blot confirmatory test. A major limitation of serological testing is that the IgG antibodies remain in the patients' blood for years after a borreliosis infection (Kalish et al., 2001; Steere et al., 2016). Thus, when the subject suffers another tick bite, it is impossible to differentiate a second acute infection from the previous infection and thus leads to a dilemma in starting therapy. This problem is exaggerated in highly endemic areas, where a large population may have already been exposed to subclinical borreliosis infection.

Given these challenges in the diagnosis of Lyme disease and the need to initiate early treatment, our study shows that if the offending tick is available with the subject for molecular characterization, next-generation sequencing of the tick can highlight not only the species of the tick but also indicate if it is harboring *Borrelia* or any other infectious pathogens. A clinical decision to treat the patient with antibiotics can be fast tracked when such information is available, especially in Lyme endemic regions.

The *I. scapularis* was the first genome to be sequenced among the subphylum Chelicerata (Gulia-Nuss et al., 2016; Pagel Van Zee et al., 2007), followed by *I. ricinus* very recently (Cramaro et al., 2015). In spite of limited availability of whole-genome sequences for most tick species in public repositories, we were able to identify the tick. However, the availability of RNA sequences from ticks enabled us to unambiguously

Beta-actin transcript cDNA-cDNA comparison

Alignment against *Dermacentor variabilis*

Alignment against <i>Dermacentor variabilis</i>		Sequence identity: 99%
30301	GACATGGAGAAAGATCTGGCACACACTTTCTACACGAGCTGCGTGTGGCCCGCCGAGGAGCACCCCGTTCTGCTGACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	450
30301	GACATGGAGAAAGATCTGGCACACACTTTCTACACGAGCTGCGTGTGGCCCGCCGAGGAGCACCCCGTTCTGCTGACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	450
451	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGGCTCTCGACTTG	600
451	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGGCTCTCGACTTG	600

Alignment against *Rhipicephalus appendiculatus*

Alignment against <i>Rhipicephalus appendiculatus</i>		Sequence identity: 97%
3301	GACATGGAGAAGATCTGGCACACACTTTCTACACGAGCTGCGTGTGGCCCGCCGAGGAGCACCCCGTTCTGCTGACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	450
244	GACATGGAGAAGATCTGGCACACACTTTCTACACGAGCTGCGTGTGGCCCGCCGAGGAGCACCCCGTTCTGCTGACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	393
451	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGCTCTCGACTTG	600
394	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGCTCTCGACTTG	543

Alignment against *Ixodes ricinus*

Alignment against <i>Ixodes ricinus</i>		Sequence identity: 93%
301	GACATGGAGAAAGATCTGGACACACACTTTCTACACGAGCTGCGTGTGGCCCGCCGAGGAGCACCCCGTTCTGCTGACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	450
244	GACATGGAGAAAGATCTGGACACACACTTTCTACACGAGCTGCGTGTGGCCCGCCGAGGAGCACCCCGTTCTGCTGACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	393
451	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGCTCTCGACTTG	600
394	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGCTCTCGACTTG	543

Alignment against *Boophilus microplus*

Alignment against <i>Boophilus microplus</i>		Sequence identity: 91%
301	GACATGGAGAAAGATCTGGACACACACTTTCTACACGAGCTGCGTGTGGCCCGCCGAGGAGCACCCCGTTCTGCTGACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	450
334	GACATGGAGAAAGATCTGGACACACACTTTCTACACGAGCTGCGTGTGGCTCCCGAGGAGCACCCCGTTCTGCTCACCGAGGCCCGCCCTGAACCCCAAGGCCAACCGTGAGAGATGACCCAGATATGTTTGAGACCTTCAACACGCC	483
451	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGCTCTCGACTTG	600
484	GCCATGTACGTTGCCATCCAGGCCCGTGTGCTGTACGCTTCCGGTTCGTACACCGGTAATTGTGCTCGACTCCGGCGACGGTGCTCTCCCACTGTGCGCATCTACGAAAGTTACGCCCTTCCCCACGCCATCCTCGCTCTCGACTTG	634

FIG. 5. Conservation of beta-actin transcript across tick species. Sequence alignment shows conservation of beta-actin transcript from the sequenced tick against four related tick species along with sequence identity indicated in percent.

identify the tick as *D. variabilis*. Earlier studies of bacteria hosted by dog ticks have identified several infectious (Ammerman et al., 2004; Clay et al., 2008) and noninfectious bacteria (Dergousoff and Chilton, 2010; Grindle et al., 2003).

Our analysis showed the presence of *Arsenophonus*, which is known to be noninfectious and previously reported in dog ticks (Grindle et al., 2003). We did not identify DNA derived from any pathogenic bacteria known to cause either Lyme disease or any other tick-borne diseases. In the context of Lyme disease, testing DNA and RNA from ticks using next-generation sequencing technology could assist clinicians help make a diagnosis for administering appropriate antibiotic therapy promptly as necessary.

Conclusions

In this study, we carried out a systematic bioinformatics analysis of whole-genome and mRNA sequences obtained by next-generation sequencing to identify the type of tick as a member of *D. variabilis* species (American dog tick). Although whole-genome sequencing data could not reliably confirm the tick species owing to limited availability of genomic sequences representing various tick species, the availability of a mRNA sequences from a wide variety of tick species allowed us to confidently identify the tick to the species level. The approach demonstrated in this study also identified nonpathogenic bacterial species hosted by the tick. This is a proof-of-concept study to help diagnose Lyme disease through identification of the tick and associated pathogenic bacteria contained in the tick.

Instead of using specific gene panels for identifying particular sequences of the tick or the *Borrelia* species, our “metagenomic” analysis through next-generation sequencing allowed an unbiased opportunity to identify both. As the costs of NGS-based sequencing drop further, we believe that our approach will be an attractive alternative to immunological testing or at least complement serological testing when the tick is available and when the results are ambiguous. This could enable physicians to make informed and more accurate decisions regarding the treatment of cases with suspected Lyme, and possibly other, zoonotic diseases.

Availability of Supporting Data

The data discussed in this publication have been deposited in NCBI's Sequence Read Archive (SRA), which is accessible through project accession number “SRP101465.”

Acknowledgment

The authors thank the Department of Biotechnology (DBT), Government of India for research support to the Institute of Bioinformatics.

Author Disclosure Statement

The authors declare that no competing financial interests exist.

References

Adams DA, Thomas KR, Jajosky RA, et al. (2017). Summary of notifiable infectious diseases and conditions—United States, 2015. *MMWR Morb Mortal Wkly Rep* 64, 1–143.

- Amer R, Brannan S, and Forrester JV. (2009). Inflammatory choroidal neovascular membrane in presumed ocular Lyme borreliosis. *Acta Ophthalmol* 87, 346–348.
- Ammerman NC, Swanson KI, Anderson JM, et al. (2004). Spotted-fever group Rickettsia in *Dermacentor variabilis*, Maryland. *Emerg Infect Dis* 10, 1478–1481.
- Armstrong AL, Barthold SW, Persing DH, and Beck DS. (1992). Carditis in Lyme disease susceptible and resistant strains of laboratory mice infected with *Borrelia burgdorferi*. *Am J Trop Med Hyg* 47, 249–258.
- Aucott J, Morrison C, Munoz B, et al. (2009). Diagnostic challenges of early Lyme disease: Lessons from a community case series. *BMC Infect Dis* 9, 79.
- Babu K, and Murthy PR. (2010). Neuroretinitis as a manifestation of Lyme disease in South India: A case report. *Ocul Immunol Inflamm* 18, 97–98.
- Berglund J, and Eitrem R. (1993). Tick-borne borreliosis in the archipelago of southern Sweden. *Scand J Infect Dis* 25, 67–72.
- Bogovic P, and Strle F. (2015). Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management. *World J Clin Cases* 3, 430–441.
- Borges LM, de Oliveira Filho JG, Ferreira LL, et al. (2015). Identification of non-host semiochemicals for the brown dog tick, *Rhipicephalus sanguineus* sensu lato (Acari: Ixodidae), from tick-resistant beagles, *Canis lupus familiaris*. *Ticks Tick Borne Dis* 6, 676–682.
- Brownstein JS, Holford TR, and Fish D. (2005). Effect of climate change on Lyme disease risk in North America. *Ecohealth* 2, 38–46.
- Clay K, Klyachko O, Grindle N, et al. (2008). Microbial communities and interactions in the lone star tick, *Amblyomma americanum*. *Mol Ecol* 17, 4371–4381.
- Courtney JW, Kostelnik LM, Zeidner NS, and Massung RF. (2004). Multiplex real-time PCR for detection of *Anaplasma phagocytophilum* and *Borrelia burgdorferi*. *J Clin Microbiol* 42, 3164–3168.
- Cramaro WJ, Revets D, Hunewald OE, et al. (2015). Integration of *Ixodes ricinus* genome sequencing with transcriptome and proteome annotation of the naive midgut. *BMC Genomics* 16, 871.
- Dantas-Torres F, Chomel BB, Otranto D. (2012). Ticks and tick-borne diseases: A One Health perspective. *Trends Parasitol* 28, 437–446.
- Dergousoff SJ, and Chilton NB. (2010). Detection of a new *Arsenophonus*-type bacterium in Canadian populations of the Rocky Mountain wood tick, *Dermacentor andersoni*. *Exp Appl Acarol* 52, 85–91.
- Dworkin MS, Schwan TG, Anderson DE, and Borchardt SM. (2008). Tick-borne relapsing fever. *Infect Dis Clin North Am* 22, 449–468, viii.
- Edlow JA. (1999). Lyme disease and related tick-borne illnesses. *Ann Emerg Med* 33, 680–693.
- Edlow JA. (2008). Preface: Tick-borne diseases, part II. *Infect Dis Clin North Am* 22, xiii–xv.
- Estrada-Pena A, and Jongejans F. (1999). Ticks feeding on humans: A review of records on human-biting Ixodoidea with special reference to pathogen transmission. *Exp Appl Acarol* 23, 685–715.
- Estrada-Pena A, Sprong H, Cabezas-Cruz A, et al. (2016). Nested coevolutionary networks shape the ecological relationships of ticks, hosts, and the Lyme disease bacteria of the *Borrelia burgdorferi* (s.l.) complex. *Parasit Vectors* 9, 517.

- Feder HM, Johnson BJB, O'Connell S, et al. (2007). A critical appraisal of "chronic Lyme disease". *N Engl J Med* 357, 1422–1430.
- Feder HM Jr., Gerber MA, Krause PJ, Ryan R, and Shapiro ED. (1993). Early Lyme disease: A flu-like illness without erythema migrans. *Pediatrics* 91, 456–459.
- Gall CA, Reif KE, Scoles GA, et al. (2016). The bacterial microbiome of *Dermacentor andersoni* ticks influences pathogen susceptibility. *ISME J* 10, 1846–1855.
- Grindle N, Tyner JJ, Clay K, and Fuqua C. (2003). Identification of *Arsenophonus*-type bacteria from the dog tick *Dermacentor variabilis*. *J Invertebr Pathol* 83, 264–266.
- Gulia-Nuss M, Nuss AB, Meyer JM, et al. (2016). Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat Commun* 7, 10507.
- Hao Q, Hou X, Geng Z, and Wan K. (2011). Distribution of *Borrelia burgdorferi* sensu lato in China. *J Clin Microbiol* 49, 647–650.
- Holden K, Boothby JT, Anand S, and Massung RF. (2003). Detection of *Borrelia burgdorferi*, *Ehrlichia chaffeensis*, and *Anaplasma phagocytophilum* in ticks (Acari: Ixodidae) from a coastal region of California. *J Med Entomol* 40, 534–539.
- Huppertz HI, Munchmeier D, and Lieb W. (1999). Ocular manifestations in children and adolescents with Lyme arthritis. *Br J Ophthalmol* 83, 1149–1152.
- Jairath V, Sehrawat M, Jindal N, Jain VK, and Aggarwal P. (2014). Lyme disease in Haryana, India. *Indian J Dermatol Venereol Leprol* 80, 320–323.
- Johnson AJ, Karabatsos N, and Lanciotti RS. (1997). Detection of Colorado tick fever virus by using reverse transcriptase PCR and application of the technique in laboratory diagnosis. *J Clin Microbiol* 35, 1203–1208.
- Kalish RA, McHugh G, Granquist J, et al. (2001). Persistence of immunoglobulin M or immunoglobulin G antibody responses to *Borrelia burgdorferi* 10–20 years after active Lyme disease. *Clin Infect Dis* 33, 780–785.
- Kirkland KB, Klimko TB, Meriwether RA, et al. (1997). Erythema migrans-like rash illness at a camp in North Carolina: A new tick-borne disease? *Arch Intern Med* 157, 2635–2641.
- Krist D, and Wenkel H. (2002). Posterior scleritis associated with *Borrelia burgdorferi* (Lyme disease) infection. *Ophthalmology* 109, 143–145.
- Lane RS, Piesman J, and Burgdorfer W. (1991). Lyme borreliosis: Relation of its causative agent to its vectors and hosts in North America and Europe. *Annu Rev Entomol* 36, 587–609.
- Langmead B, and Salzberg SL. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359.
- Leeflang MM, Ang CW, Berkhout J, et al. (2016). The diagnostic accuracy of serological tests for Lyme borreliosis in Europe: A systematic review and meta-analysis. *BMC Infect Dis* 16, 140.
- Lindblom A, Wallménus K, Nordberg M, et al. (2013). Seroreactivity for spotted fever rickettsiae and co-infections with other tick-borne agents among habitants in central and southern Sweden. *Eur J Clin Microbiol Infect Dis* 32, 317–323.
- Marques AR. (2015). Laboratory diagnosis of Lyme disease: Advances and challenges. *Infect Dis Clin North Am* 29, 295–307.
- Masters E, Granter S, Duray P, and Cordes P. (1998). Physician-diagnosed erythema migrans and erythema migrans-like rashes following Lone Star tick bites. *Arch Dermatol* 134, 955–960.
- Matuschka FR, Ohlenbusch A, Eiffert H, Richter D, and Spielman A. (1995). Antiquity of the Lyme-disease spirochaete in Europe. *Lancet* 346, 1367.
- McNair AN, and Brown JL. (1991). Tick-borne encephalitis complicated by monoplegia and sensorineural deafness. *J Infect* 22, 81–86.
- Mikkila HO, Seppala IJ, Viljanen MK, Peltomaa MP, and Karma A. (2000). The expanding clinical spectrum of ocular Lyme borreliosis. *Ophthalmology* 107, 581–587.
- Misić-Majerus L, Bujić N, Madarić V, and Avsić-Zupanc T. (2005). [Hepatitis caused by tick-borne meningoencephalitis virus (TBEV)—A rare clinical manifestation outside the central nervous system involvement]. [Article in Croatian]. *Acta Medica Croatica* 59, 347–352.
- Miyashiro MJ, Yee RW, Patel G, and Ruiz RS. (1999). Lyme disease associated with unilateral interstitial keratitis. *Cornea* 18, 115–116.
- Mullegger RR. (2004). Dermatological manifestations of Lyme borreliosis. *Eur J Dermatol* 14, 296–309.
- Pagel Van Zee J, Geraci NS, Guerrero FD, et al. (2007). Tick genomics: The *Ixodes* genome project and beyond. *Int J Parasitol* 37, 1297–1305.
- Parola P, and Raoult D. (2001). Ticks and tickborne bacterial diseases in humans: An emerging infectious threat. *Clin Infect Dis* 32, 897–928.
- Peltomaa M, Saxen H, Seppala I, Viljanen M, and Pyykko I. (1998). Paediatric facial paralysis caused by Lyme borreliosis: A prospective and retrospective analysis. *Scand J Infect Dis* 30, 269–275.
- Pertea G. (2015). fqtrim: v0.9.4 release. <https://ccb.jhu.edu/software/fqtrim/>
- Preac-Mursic V, Wilske B, and Schierz G. (1986). European *Borrelia burgdorferi* isolated from humans and ticks culture conditions and antibiotic susceptibility. *Zentralbl Bakteriol Mikrobiol Hyg A* 263, 112–118.
- Rijpkema SG, Tazelaar DJ, Molkenboer MJ, et al. (1997). Detection of *Borrelia afzelii*, *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* and group VS116 by PCR in skin biopsies of patients with erythema migrans and acrodermatitis chronica atrophicans. *Clin Microbiol Infect* 3, 109–116.
- Rynkiewicz EC, Hemmerich C, Rusch DB, Fuqua C, and Clay K. (2015). Concordance of bacterial communities of two tick species and blood of their shared rodent host. *Mol Ecol* 24, 2566–2579.
- Sanchez E, Vannier E, Wormser GP, and Hu LT. (2016). Diagnosis, treatment, and prevention of Lyme disease, human granulocytic *Anaplasmosis*, and *Babesiosis*: A Review. *JAMA* 315, 1767–1777.
- Science IoMUCoLDaOT-BDTSot. (2011). *Critical Needs and Gaps in Understanding Prevention, Amelioration, and Resolution of Lyme and Other Tick-Borne Diseases: The Short-Term and Long-Term Outcomes: Workshop Report*. Washington, DC: National Academies Press.
- Shaw SE, Day MJ, Birtles RJ, and Breitschwerdt EB. (2001). Tick-borne infectious diseases of dogs. *Trends Parasitol* 17, 74–80.
- Spach DH, Liles WC, Campbell GL, et al. (1993). Tick-borne diseases in the United States. *N Engl J Med* 329, 936–947.
- Spielman A. (1994). The emergence of Lyme disease and human babesiosis in a changing environment. *Ann N Y Acad Sci* 740, 146–156.
- Steere AC. (2001). Lyme disease. *N Engl J Med* 345, 115–125.
- Steere AC, Dhar A, Hernandez J, et al. (2003). Systemic symptoms without erythema migrans as the presenting picture of early Lyme disease. *Am J Med* 114, 58–62.

- Steere AC, Malawista SE, Hardin JA, et al. (1977a). Erythema chronicum migrans and Lyme arthritis. The enlarging clinical spectrum. *Ann Intern Med* 86, 685–698.
- Steere AC, Malawista SE, Snyderman DR, et al. (1977b). Lyme arthritis: An epidemic of oligoarticular arthritis in children and adults in three connecticut communities. *Arthritis Rheum* 20, 7–17.
- Steere AC, McHugh G, Damle N, and Sikand VK. (2008). Prospective study of serologic tests for lyme disease. *Clin Infect Dis* 47, 188–195.
- Steere AC, Strle F, Wormser GP, et al. (2016). Lyme borreliosis. *Nat Rev Dis Primers* 2, 16090.
- Takano A, Nakao M, Masuzawa T, et al. (2011). Multilocus sequence typing implicates rodents as the main reservoir host of human-pathogenic *Borrelia garinii* in Japan. *J Clin Microbiol* 49, 2035–2039.
- Tibbles CD, and Edlow JA. (2007). Does this patient have erythema migrans? *JAMA* 297, 2617–2627.
- Trapnell C, Pachter L, and Salzberg SL. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Uchio E, and Ohno S. (1999). Ocular manifestations of systemic infections. *Curr Opin Ophthalmol* 10, 452–457.
- Walker DH. (1998). Tick-transmitted infectious diseases in the United States. *Annu Rev Public Health* 19, 237–269.
- Wallace JW, Nicholson WL, Perniciaro JL, et al. (2016). Incident tick-borne infections in a cohort of North Carolina outdoor workers. *Vector Borne and Zoonotic Dis* 16, 302–308.
- Zaidi SA, and Singer C. (2002). Gastrointestinal and hepatic manifestations of tickborne diseases in the United States. *Clin Infect Dis* 34, 1206–1212.

Address correspondence to:

Akhilesh Pandey, MD, PhD

McKusick-Nathans Institute of Genetic Medicine

Johns Hopkins University School of Medicine

733 N. Broadway

Baltimore, MD 21205

E-mail: pandey@jhmi.edu

Abbreviations Used

BLAST = Basic Local Alignment Search Tool
 CDC = Centers for Disease Control and Prevention
 ELISA = enzyme-linked immunosorbent assay
 NCBI = National Center for Biotechnology
 Information
 NGS = Next-generation sequencing
 NIH = National Institutes of Health
 PCR = polymerase chain reaction
 WGS = whole-genome sequencing