

Metatranscriptomic investigation of single *Ixodes pacificus* ticks reveals diverse microbes, viruses, and novel mRNA-like endogenous viral elements

Calla Martyn,^{1,2} Beth M. Hayes,^{1,3} Domokos Lauko,¹ Edward Midthun,⁴ Gloria Castaneda,⁵ Angela Bosco-Lauth,⁴ Daniel J. Salkeld,⁶ Amy Kistler,⁵ Katherine S. Pollard,^{2,5,7} Seemay Chou^{1,5}

AUTHOR AFFILIATIONS See affiliation list on p. 16.

ABSTRACT Ticks are increasingly important vectors of human and agricultural diseases. While many studies have focused on tick-borne bacteria, far less is known about tick-associated viruses and their roles in public health or tick physiology. To address this, we investigated patterns of bacterial and viral communities across two field populations of western black-legged ticks (*Ixodes pacificus*). Through metatranscriptomic analysis of 100 individual ticks, we quantified taxon prevalence, abundance, and co-occurrence with other members of the tick microbiome. In addition to commonly found tick-associated microbes, we assembled 11 novel RNA virus genomes from *Rhabdoviridae*, *Chuviridae*, *Picornaviridae*, *Phenuiviridae*, *Reoviridae*, *Solemoviridae*, *Narnaviridae* and two highly divergent RNA virus genomes lacking sequence similarity to any known viral families. We experimentally verified the presence of these in *I. pacificus* ticks across several life stages. We also unexpectedly identified numerous virus-like transcripts that are likely encoded by tick genomic DNA, and which are distinct from known endogenous viral element-mediated immunity pathways in invertebrates. Taken together, our work reveals that *I. pacificus* ticks carry a greater diversity of viruses than previously appreciated, in some cases resulting in evolutionarily acquired virus-like transcripts. Our findings highlight how pervasive and intimate tick–virus interactions are, with major implications for both the fundamental biology and vectorial capacity of *I. pacificus* ticks.

IMPORTANCE Ticks are increasingly important vectors of disease, particularly in the United States where expanding tick ranges and intrusion into previously wild areas has resulted in increasing human exposure to ticks. Emerging human pathogens have been identified in ticks at an increasing rate, and yet little is known about the full community of microbes circulating in various tick species, a crucial first step to understanding how they interact with each and their tick host, as well as their ability to cause disease in humans. We investigated the bacterial and viral communities of the Western blacklegged tick in California and found 11 previously uncharacterized viruses circulating in this population.

KEYWORDS arthropod vectors, vector-borne disease, metagenomics, endosymbionts, nonhuman microbiome, innate immunity, bioinformatics, metatranscriptomics

Ticks are increasingly important disease vectors for humans and livestock, particularly in the United States, where they account for more cases of vector-borne diseases than mosquitoes. Approximately fifty thousand confirmed cases of tick-borne diseases are reported annually (1), which is likely an underestimate due to diagnostic challenges associated with Lyme disease and our poor understanding of rare tick-borne diseases or diseases of unknown etiology. Currently, the majority of field surveillance studies

Editor Jonathan L. Klassen, University of Connecticut, Storrs, Connecticut, USA

Address correspondence to Katherine S. Pollard, kpollard@gladstone.ucsf.edu, Seemay Chou, seemay.chou@ucsf.edu, or Amy Kistler, amy.kistler@czbiohub.org.

Seemay Chou is an employee of Arcadia Sciences.

See the funding table on p. 16.

Received 9 March 2024

Accepted 27 March 2024

Published 14 May 2024

Copyright © 2024 Martyn et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

of tick-associated microbes focus on *Borrelia burgdorferi*, the causative agent of Lyme disease, and a select number of other known human pathogens, such as *Rickettsia* (causative agents of Rocky Mountain spotted fever and other Rickettsioses), *Anaplasma phagocytophilum* (the etiologic agent of Anaplasmosis), and Powassan virus. Although the full diversity of microbes carried by ticks is much greater than those definitively linked to human disease (2–22), we know strikingly little about the ecology or disease implications of most tick-associated microbes.

Furthermore, we are only beginning to appreciate the broader role microbes play in tick biology. Like many other invertebrates, tick–microbe interactions go far beyond the transmission of human pathogens. Tick–microbe interactions can be antagonistic, neutral, or beneficial (23–26). Some microbes, such as the bacterial genus *Rickettsia*, play fundamental roles in tick physiology through stable and symbiotic interactions (26). It is not as well known whether ticks also form stable, symbiotic interactions with viruses and how such interactions may influence tick biology. Some viruses have been identified not only in live ticks but also in many laboratory-passaged tick cell lines (27). Microbes have also been implicated in shaping the evolution of ticks through the horizontal transfer of bacterial genes (28) and endogenization of viral sequences as an immune response (29–32).

The relationship between ticks and viruses is a subject of much interest. Ticks are able to carry viruses from a broad range of families, some of which have been identified as causative agents of disease in humans such as tick-borne encephalitis virus and Powassan virus of the *Flaviviridae* family of positive sense RNA viruses, and severe fever with thrombocytopenia syndrome virus (33) of the *Phenuiviridae* family of multi-segmented negative sense RNA viruses. While field microbiome studies have increased our catalog of tick-associated viruses (33), a number of experimental strategies and technical hurdles have limited the scope and depth of microbiome analyses in ticks. First, field studies often sequence pooled tick samples, preventing quantitative examination of microbial prevalence, co-occurrence, and per-sample relative abundance. These metrics could enable more sophisticated analyses of transmission dynamics and ecology. Furthermore, our ability to capture lower-abundance microbes is hampered by the dominance of tick host sequences in metatranscriptomic libraries. This limits our understanding to the most abundant microbes in ticks, which does not necessarily coincide with all microbes that have important roles in disease or tick physiology.

Moreover, our understanding of tick microbiota in North America is currently biased towards tick host species historically associated with human diseases, such as *Ixodes scapularis*, the primary vector for Lyme disease in the Eastern United States. However, in recent years, there has been an expansion of tick-borne disease cases on the West Coast of the United States that have been attributed to other, less-studied tick vector species. *Ixodes pacificus* is a major tick species extending from Northern Mexico to British Columbia (34). *I. pacificus* ticks are most abundant in California, where they cover 96% of all counties (35) and are responsible for the majority of human tick bites (36). They are vectors for a variety of well-characterized human pathogens such as *B. burgdorferi*, *Borrelia miyamotoi*, *Babesia odocoilei*, *Bartonella* spp., *A. phagocytophilum*, and *Ehrlichia* spp. (37). Despite this, *I. pacificus* is substantially understudied compared to the eastern black-legged tick *I. scapularis*.

To provide much-needed insight into tick-borne microbes in the western United States, we examined the microbiomes of *I. pacificus* ticks collected from two coastal habitats in California where humans are likely to encounter them (36, 38, 39). In order to capture lower-abundance microbes, we coupled an experimental microbial enrichment workflow with RNA sequencing to profile both bacteria and RNA viruses. Analysis of microbiomes at the level of individual ticks enabled us to also quantify patterns of microbial prevalence and abundance. We performed follow-up laboratory-controlled experiments examining microbial localization within tick compartments and across tick life stages, which provided additional insights into potential transmission dynamics and the symbiotic nature of tick-virus relationships.

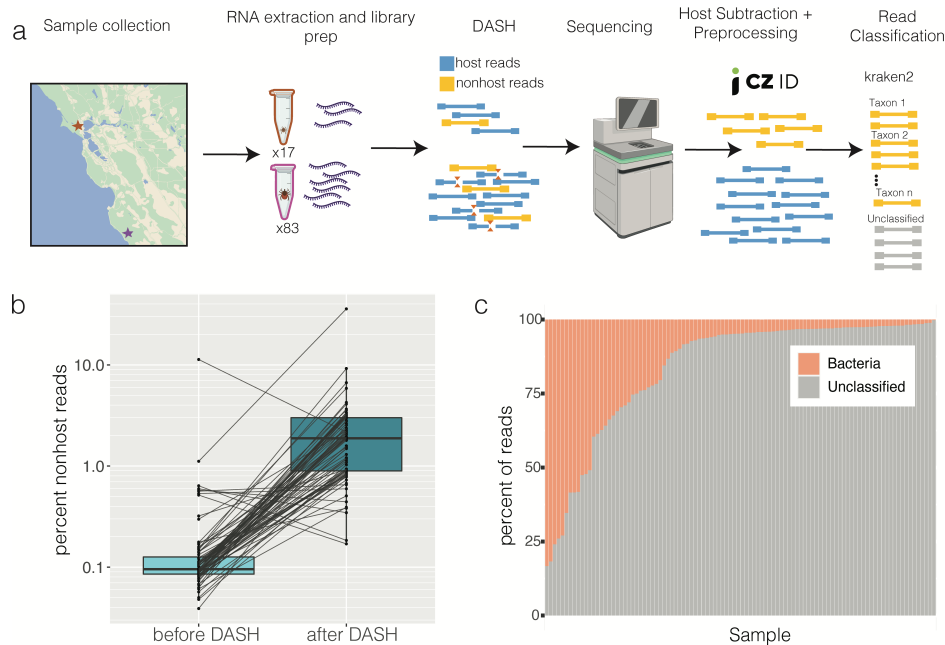


FIG 1 *Ixodes pacificus* metatranscriptomics experimental approach. (a) Eighty-three adults and 17 nymphs were collected from Garrapata State Park (purple) and China Camp State Park (red), respectively. RNA was extracted from whole bodies of individual *I. pacificus* ticks, sequencing libraries were prepared, and depletion of abundant sequences by hybridization (DASH) was performed to deplete abundant tick sequences. After sequencing, reads were quality filtered and tick-derived reads were removed using CZID. The remaining reads were classified by kraken2. Map was drawn based on an image from Google Maps. (b) The percent of non-host reads as classified by CZID for matched libraries before and after DASH. Lines connect individual libraries. (c) Percentage of non-host reads classified per sample by kraken2.

RESULTS

Establishment of an RNA-based approach to defining composition of field tick microbiomes

We set out to define the metatranscriptome of *I. pacificus* ticks collected from coastal California, focusing on two sites associated with human exposure (36–40). We examined the two most developmentally advanced life stages (nymphal and adult) that are more amenable to single-tick sequencing due to greater individual biomass. These sample sets included adult ticks from Garrapata State Park and nymphal ticks from China Camp State Park. Adults were collected in the winter (2018) and nymphs were collected in the spring (2019) so that we could investigate seasons and life stages enriched for human contact. In total, RNA libraries were sequenced for 100 individual ticks.

The majority of whole-tick RNA libraries are composed of tick ribosomal RNA, which reduces the power to detect lower-abundance sequences with homology to bacteria and viruses. To address this challenge, we enriched microbial sequences by experimentally depleting abundant tick sequences through depletion of abundant sequences by hybridization (DASH) (Fig. 1a) (41). For adult tick libraries, DASH-based depletion enriched non-host sequences nearly 10-fold (Fig. 1b). Sequencing libraries generated for smaller nymphal ticks were not of sufficient concentration to effectively perform DASH. After sequence quality filtering and removal of host sequences, the remaining non-host sequences were classified using the metagenomic classifier kraken2 (42) (Fig. 1a). The samples varied substantially in the proportion of non-host reads that could be classified, ranging from 0.2% to 83% (Fig. 1c). All classified reads by kraken2 were bacterial; no known viruses were identified. Because of this, unclassified sequences were analyzed with a custom pipeline that enabled the detection of divergent viruses (see below, Materials and Methods).

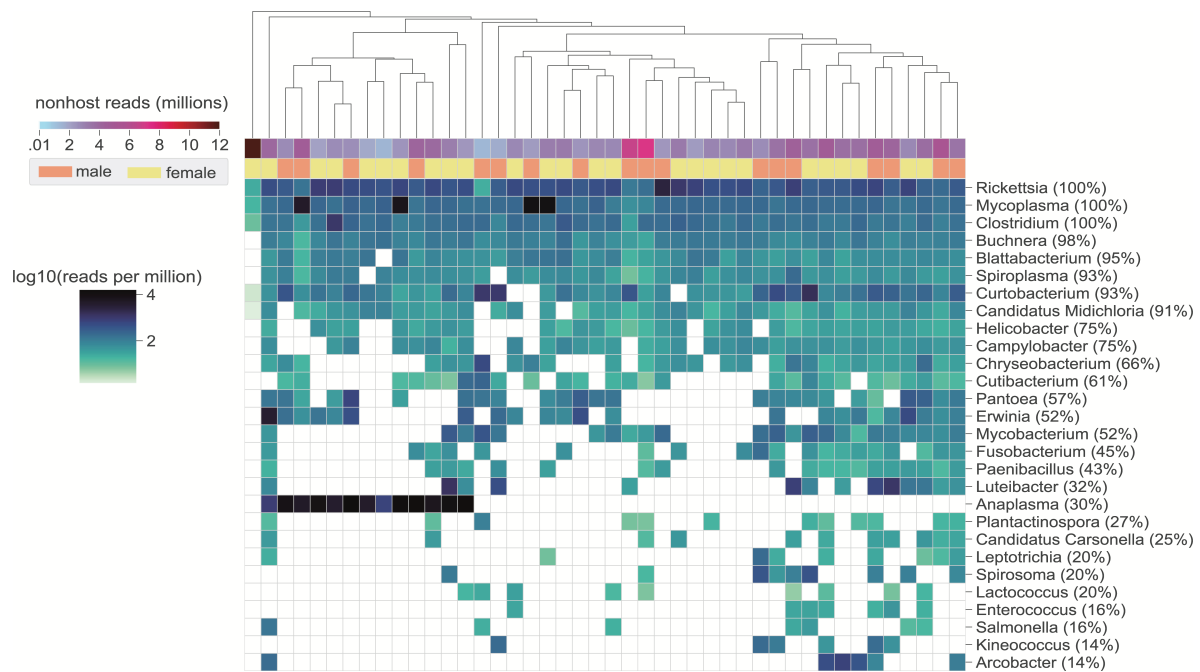


FIG 2 Bacterial genera detected in *Ixodes pacificus*. Heatmap showing reads per million (rpm) of bacterial genera as classified by kraken2. Plot is limited to samples with at least 1 million non-host reads and genera detected in at least five samples. Prevalence in the selected samples (rows) is shown next to the genus name, and genera are ordered by decreasing prevalence. Samples (columns) are hierarchically clustered using Euclidean distance.

To assess the general validity of our approach to characterizing the microbiota of field ticks, we first quantified the bacterial component of the tick metatranscriptomes. We identified 114 bacterial genera across the data set with a median of 11 genera per tick (Fig. S1a). Larger libraries generally had more classified genera, though there was substantial variability in the relationship between sequencing depth and detected richness (Fig. S1a and S2). We compared the taxonomic composition of our samples with previously reported tick-associated human pathogens, such as *Rickettsia*, *Anaplasma*, and *Coxiella*, which are most commonly linked to *I. pacificus* ticks (3, 6, 15, 37). *Borrelia*, *Borrelia*, *Ehrlichia*, and *Bartonella* are also human pathogens known to circulate in this species, although typically at lower frequencies. *Ehrlichia*, *Borrelia*, and *Borrelia* were identified at rates of 2–3% across the full data set but not in samples with a more stringent quality cutoff of at least 1 million non-host reads (Fig. 2; Fig. S2). We did not identify *Coxiella*, *Bartonella*, or *Francisella* in any ticks, indicating they are either absent in this population or present at levels too low to be detected.

Of the samples with at least 1 million non-host reads, several genera we identified were found at frequencies similar to previous reports (3, 6, 15, 37). We detected the endosymbiont *Rickettsia* in 100% and *Anaplasma* in approximately 30% of ticks (Fig. 2). We also observed several instances of coinfections between *Ehrlichia*, *Borrelia*, *Borrelia*, and *Anaplasma* species (Fig. S1b). In total, results from our bacterial analyses are largely consistent with previously reported analyses of bacterial constituents of tick microbiota, suggesting our RNA-based approach to characterizing tick-associated microbes can indeed be applied to field studies. These results gave us confidence that our sequencing and analysis pipeline could reliably report on microbiomes of field *I. pacificus* ticks at nymphal and adult life stages.

Optimized workflow enables detection of low-abundance bacteria and RNA viral transcripts

Our combined approach of host depletion and RNA-sequencing opened up several unique lines of inquiry. In addition to known tick-borne human pathogens, we identified

several bacterial genera previously undetected in *I. pacificus*. Although these genera had high prevalence across our samples, they were present at low relative abundances and likely escaped detection in previous studies in the absence of DASH-based microbial enrichment. *Mycoplasma*, a genus that has been linked to Lyme-like disease in patients with tick exposure (43, 44), was identified in all of the selected libraries (Fig. 2). *Blattabacterium*, *Buchnera*, *Spiroplasma*, and *Candidatus Midichloria* were all present in at least 90% of samples, and *Candidatus Carsonella* was identified in 25% of samples (Fig. 2). To our knowledge, none of these known endosymbionts have been commonly identified in *I. pacificus*, and this is the first report of *Blattabacterium*, *Buchnera*, and *Candidatus Carsonella* in any tick (45–47).

Sequencing individual ticks also provided sufficient resolution for co-occurrence analyses. We assessed whether the presence of one microbial genus increases the statistical likelihood that another microbial genus will be present in the same tick host (48). This revealed 14 pairs of bacterial genera detected together in a statistically significant number of samples (Fig. S3). Of note, there was a strong positive association between the endosymbiont *Candidatus Carsonella* and *Chryseobacterium*, a genus that has been shown to be pathogenic to soft ticks but tolerated by hard ticks (49). The remaining statistically significant co-occurrence relationships did not include any microbial genera known to be tick-associated, and we hypothesized that they may be environmental contaminants, such as soil bacteria (14), that were detectable despite our rinsing of the ticks and use of water controls. Therefore, outside of the *Candidatus–Chryseobacterium* case, we did not find clear evidence that any microbes associated with *I. pacificus* actively promote the colonization or growth of other microbes.

Our RNA-based sequencing approach also enabled the identification of both known and previously unidentified RNA viruses in *I. pacificus* ticks. Ticks are known to carry a diversity of viruses, and the majority of known transmissible arboviruses have RNA genomes (15, 50–54). We sought evidence of these and any novel viruses in our metatranscriptomes. To do so, we first developed a bioinformatics strategy because standard tools for microbiome analysis (e.g., kraken2 [42]) did not detect any known tick viruses in our data. This is a common phenomenon in RNA virus discovery, due to the diversity of RNA viruses not being well represented in reference databases. In keeping with other viral discovery efforts (55–57), we searched for sequences containing an RNA-dependent RNA polymerase (RdRp) domain using HMMER (58) (Fig. 3a).

Using this strategy, we detected a total of 13 distinct putative viral genomes in our *I. pacificus* field specimens and determined their prevalence across the data set as well as their relative abundance within each sample (Fig. 3b, Table S1). For each putative viral genome, we recovered was provisionally named based on the geographical features in the region in which the samples were collected. Underscoring the novelty of these putative viral genome sequences, the vast majority (11/13) showed less than 80% amino acid identity to their nearest relative in the NCBI non-redundant protein database (Fig. 3b). Despite this, phylogenetic analysis of their RdRp amino acid sequences using NCBI/ICTV methodology (Materials and Methods, Fig. S8) mapped most sequences to a viral order (*Bunyavirales* and *Monongavirales*) or a viral family (*Rhabdoviridae*, *Picornaviridae*, and *Narnaviridae*) that has been previously shown to have members infecting ticks. An exception corresponds to two partial putative viral genome sequences that we recovered, Kasler Point virus and Wildcat Canyon virus. These divergent sequences showed very low amino acid identity to their best RdRp matches and were provisionally assigned to the recently discovered ormycovirus clade, a group of unclassified RNA viruses that has not been phylogenetically placed within any existing order or family of viruses. The remaining two viral genome sequences we recovered, Notley's Landing virus and Rocky Ridge virus, showed >90% amino acid identity to previously described members of the *Solemoviridae* and *Chuviridae* families that were isolated from ticks (Table S1; Fig. S8). A detailed summary of the attributes of each of these viruses is provided in the Supplementary Note.

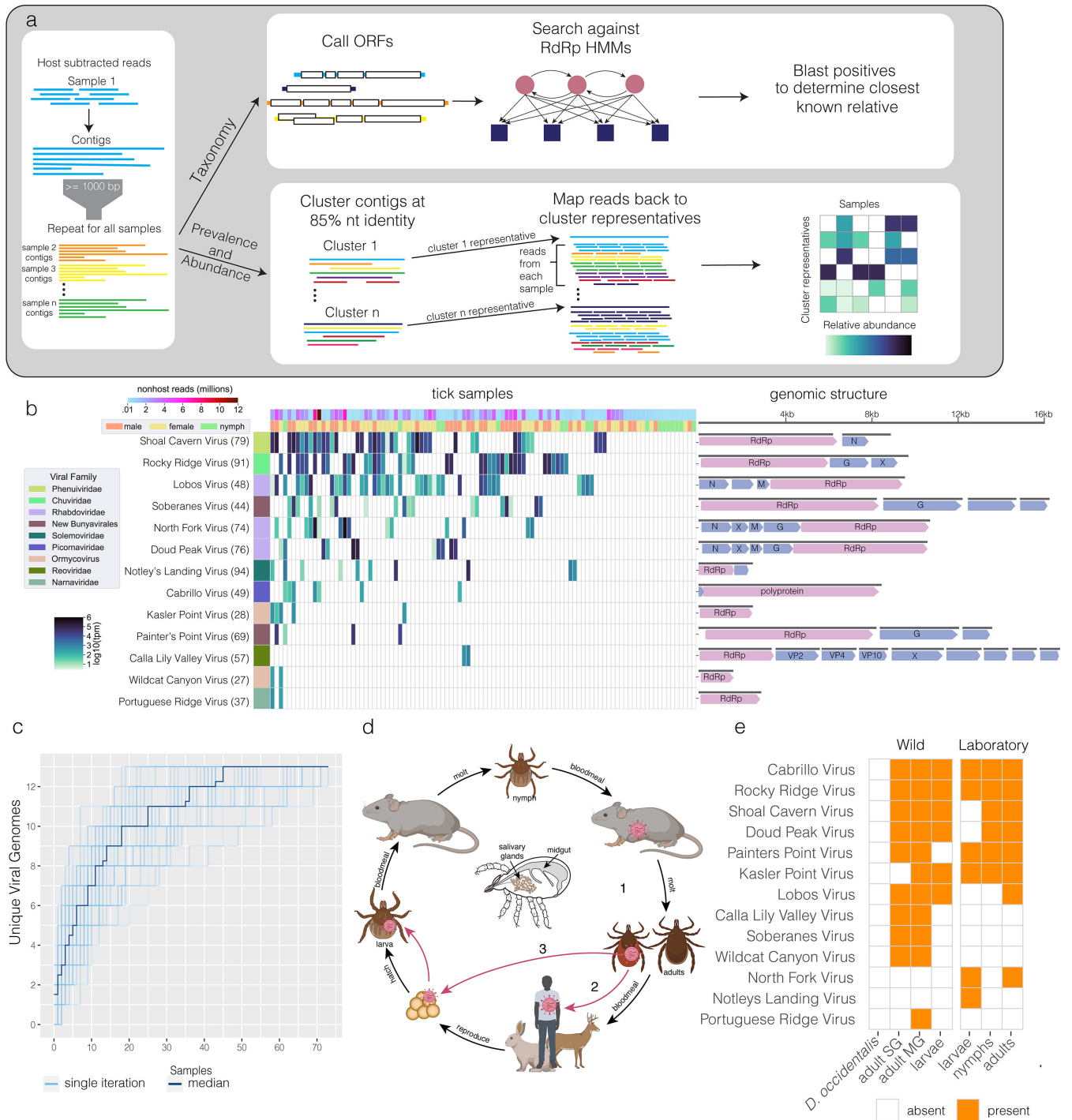


FIG 3 Discovery of novel viral sequences in *Ixodes pacificus*. (a) Analysis pipelines for identification of viruses: reads were assembled into contigs and open reading frames were predicted. Resulting protein sequences were scanned for the presence of RNA-dependent RNA polymerase (RdRp) using HMMER and classified by viral family using closest BLAST hits and taxonomically annotated following NCBI/ICTV methodology. Prevalence and abundance were determined by clustering contigs and mapping reads to the cluster representatives. (b) Heatmap of transcripts per million (tpm) of each identified viral RdRp-containing contig across the data set. Average amino acid identity of the aligned region to closest known sequence is displayed to the right of the virus name. Open reading frames are displayed to the right of the heatmap and annotated with identified proteins (RdRp, G = glycoprotein, N = nucleoprotein, M = matrix protein, VP = viral protein). Open reading frames with “X” contain homology to known viral proteins of unknown function, and those with no annotation had no identified homology to known proteins. (c) Rarefaction plot showing increase in number of viral genomes for each sample analyzed. Samples were randomly shuffled 50 times, and the median was plotted in dark blue. (d) Schematic of *I. pacificus* three-stage life cycle. Viruses can be transmitted horizontally between hosts and ticks (Continued on next page)

FIG 3 (Continued)

(1 and 2) or vertically from adult females to their offspring (3). (e) Table summarizing detection of each viral contig across tick samples. SG = salivary glands, MG = midgut.

RNA viruses are stable constituents of tick microbiota

Our RNA-based approach to tick microbiome characterization led to the discovery of highly divergent members of several viral families in *I. pacificus* field ticks. To follow up on these results, we next asked if the 13 distinct viral sequences we detected were likely to represent the full virome of the sampled ticks or only the most abundantly transcribed strains. To do so we performed rarefaction analysis, which shows the number of new taxa discovered as a function of the number of samples sequenced. Our rarefaction curve appears to be approaching an asymptote, and the estimated true number of viruses in this population using the Chao index is 14.2 (Fig. 3c). These results indicate that we have likely discovered the majority of viruses in this population with a median of 1.7 million non-host sequences after DASH host-subtraction. This is equivalent to an overall sequencing depth of 106 million reads given our median amount of host depletion. We also observed that relatively few samples are needed to saturate viral discovery for a given population at this sequencing depth.

We also performed co-occurrence analysis with our newly characterized tick viromes. Not only did we identify a broad diversity of viral genomes, but we also found evidence of co-occurrence within individual ticks. Ticks had a median of two viral species detected, with a maximum of 6 in one individual (Fig. S4a). We found a statistically significant positive relationship between Portuguese Ridge virus (*Narnaviridae*), Wildcat Canyon virus (*ormycovirus*), and Kasler Point virus (*ormycovirus*) (Fig. S3). Notably, all three of these viral genome assemblies contained only an RdRp; no additional segments or genes were identified. Since Narnaviruses are single-gene ribonucleoprotein complexes lacking structural proteins or capsids, it is possible that the two viruses of unknown origin replicate and transmit in a similar manner (see Supplemental Note: *Narnaviridae* and *Ormycovirus*). Our findings support the model that ticks can harbor multiple viral species per individual, suggesting that RNA viruses are consistent and stable members of the tick microbiome.

Life-stage and tissue tropism of viruses

Arthropods are known to tolerate viral infection more easily than vertebrates, often maintaining infections for life with no apparent ill-effects (59). Given that some of the viral species we identified are not only highly prevalent but also closely related to viruses previously discovered in tick cell lines (see Supplemental Note: *Rhabdoviridae*) (27), we next characterized several properties that could inform if and how these viruses are acquired and transmitted. Namely, we examined their distribution in ticks according to both life stages (larval, nymphal, and adult) and tissue localization (midgut and salivary glands). We performed multiplex PCR for the 13 viral sequences on cDNA generated from pools of salivary gland and midgut tissues dissected from an independent set of field-caught and laboratory-reared *I. pacificus* ticks of different life stages. Nine of the 13 viral sequences were detected in at least one of the dissected tissue pools across all life stages of the set of field or laboratory ticks (Fig. 3e).

Detection of virus in field larval ticks was particularly striking, since larval ticks have not yet consumed a bloodmeal, and ticks of any life stage are not known to transmit viruses directly within a population in the absence of a vertebrate host intermediate (60). Likewise, the detection of these viruses in laboratory-reared *I. pacificus* ticks indicated that the viruses can be maintained in the absence of natural bloodmeal hosts (Fig. 3; Fig. S4). These data suggest these viruses may be transmitted vertically within ticks, a phenomenon known to occur with several known tick viruses (60). Moreover, looking across life stages, we found that 8/9 viral sequences that we observed in the larvae were also detectable in either nymphs or adults, suggesting these viruses may be able

to persist or infect through multiple life stages. Our parallel analysis of the midgut and salivary tissues shows that the majority of the viral sequences (9/13) were also detectable in the salivary gland (Fig. 3e), the compartment of the tick generally associated with the potential for feeding-based (horizontal) transmission (Fig. 3d). While this does not definitively demonstrate that these viruses are transmissible by feeding, we hypothesize that this could be possible and worthy of further investigation.

Identification of novel mRNA-like virus-like transcripts

In addition to the 13 viral RdRps, we identified 21 sequences with homology to an RdRp but with an open reading frame (ORF) structure inconsistent with known RNA viral genomes. Specifically, these RNA sequences encoded clusters of small ORFs with RdRp homology and large gaps (hundreds of bases) between their predicted ORFs (Fig. 4a). Many also contained multiple overlapping ORFs and/or small ORFs in opposite orientations. These unusual sequences were highly prevalent (Fig. 4a) and were independently assembled from multiple different ticks. Our analysis of these sequences suggested that they originated from many of the same families as the viral genomes (Fig. 4b), but with distinct sequences that encoded ORFs that were smaller and more numerous than would be expected for that family.

We were intrigued by the observed irregular genomic organization of these 21 virus-like sequences; thus we next sought to better understand their possible origins and functions. We first conducted experiments to eliminate potential artifactual explanations for the irregular ORF structure. To test whether these sequences could be the result of a misassembly, we selected one of the longest (5,043 bp) and most highly prevalent sequences (vlt_111) for more in-depth evaluation. We performed PCR with tiled overlapping primers (Fig. S6a) and applied 5' and 3' RACE to test if we could recover the vlt_111 sequence from the cDNA of a Garrapata tick (Materials and Methods). Sanger sequencing of the resulting PCR products confirmed the accuracy of our vlt_111 sequence assembly and indicated that vlt_111 is expressed as a 3' poly-adenylated mRNA, ruling out the possibility that the non-canonical features of this RNA are due to misassembly of the original metatranscriptomic sequencing reads.

We also considered whether the irregular ORF structures we observed could be resolved with alternative codons, such as non-standard stop and start codons. We tested whether ORF prediction with any alternative genetic codes would result in an organization more consistent with that of a viral genome. Of the 25 known genetic codes tested, none substantially changed the ORF structure of any of the sequences (Fig. S6d), indicating that alternative genetic code alone cannot account for the observed genomic structure. Having eliminated possible artifactual explanations for how these sequences could originate from exogenous viral genomes, we termed these sequences of unknown origin and function "virus-like transcripts" (VLTs).

VLTs likely serve a non-canonical function in *I. pacificus* ticks

Tick genomes are known to contain numerous endogenous viral elements (EVEs), which result from the horizontal integration of RNA viral sequences into tick genomes over the course of evolution (29). The observed high prevalence of VLTs led us to hypothesize that these mysterious sequences could have similar functions or histories. In the absence of a published genome assembly for *I. pacificus*, we could not check for corresponding sequences in a reference genome. Therefore, we developed PCR primers to screen for the VLTs in *I. pacificus* genomic DNA extracts isolated from the same field-collected ticks that were used for metatranscriptomic analysis (Fig. S6a). We confirmed the presence of vlt_111 in the field-collected and laboratory-reared tick DNA (Fig. S6b). To ensure this pattern was not specific to this particular VLT, we checked an additional five VLTs, all of which were present in lab-reared *I. pacificus* DNA (Fig. S6c), suggesting a genomic origin for these VLTs.

To confirm that the presence of DNA forms of transcripts is specific to the VLTs and not a general phenomenon, we additionally screened the genomic DNA for all of the



FIG 4 Virus-like transcripts detected across *Ixodes pacificus* ticks in our study. (a) Heatmap showing transcripts per million (tpm) of virus-like transcripts (VLTs) across the data set. VLTs are colored by viral family of the closest blast hit and marked with "*" if they were detected in DNA by PCR. Predicted open reading frames (ORFs) are shown to the right and colored by homology to viral RdRp (pink), tick proteins (turquoise; TRAF = TNF receptor-associated factor 6-like, TE = piggy Bac transposable element-derived protein 4-like, X = protein of unknown function), or no homology (blue). (b) Viral family assignment of exogenous viral sequences and VLTs. (c) Known functions of arthropod endogenous viral elements. (d) Heatmap displaying length of the longest perfectly matching sequences between each virus and each VLT. Rows and columns are colored by viral family.

presumed exogenous viral assemblies. Only one of the viral RNA genomes was detected in DNA. A faint band corresponding in size to Rocky Ridge virus was amplified from genomic DNA. This mivirus-like sequence could represent an intermediate between these two categories that is an exogenous RNA virus with a single or small number of recent genomic integrations into the *I. pacificus* genome. Alternatively, this could be caused by the production of DNA forms of the viral genome by endogenous retrotranscriptases (either prior to genomic integration or in the absence of integration) as in Salvati et al. (61). Given that its expression pattern mirrored that of the other exogenous viruses and its genome is both complete and contains the expected ORFs, we continued to classify this sequence as an exogenous virus and not a VLT.

To explore the possibility that our VLTs could have EVE-like functions, we next considered canonical pathways by which EVEs contribute to arthropod immunity (Fig. 4c). EVEs are most commonly known to function as non-coding RNAs. Much more rarely, they are expressed and translated as proteins that can act as dominant negative viral inhibitors or serve a new function (29, 30). As the fragmented ORF structure of our VLTs is inconsistent with the expression of full-length proteins, we focused on non-coding RNA functions. Typically, arthropod EVEs play antiviral roles by serving as a template for piwi

RNAs (piRNAs), 24–31 nucleotide (nt) RNAs that target an exogenous viral RNA genome for degradation by binding to a complementary sequence within it.

To test this model, we examined *in silico* whether our VLTs could give rise to small RNAs capable of binding the exogenous viral genomes in our data set through a matching sequence at least 24 nt long. Only one VLT contained a sequence of at least this length (vlt_307) matching one of our viral genome assemblies. Two others contained stretches longer than 20 nucleotides (Fig. 4d). The remaining 18 VLT sequences did not contain perfect matches longer than 17 nucleotides to any of the exogenous viruses. Of the three VLT-virus combinations with perfect matches of at least 20 nucleotides, none originated from the same tick sample (Fig. S6 f through h). In total, we did not uncover definitive evidence supporting the inhibition of viral replication through a canonical piRNA pathway. Our results point to either a non-canonical immunity mechanism or a different functional role entirely for the identified VLTs that could be explored in future studies.

Furthermore, we found evidence of this phenomenon beyond our own study. By examining the viral sequences deposited by Ni et al. (22) in a large metagenomic survey of the virome of ticks in China, we identified four sequences with an irregular ORF structure similar to that of our VLTs (Fig. S7a). Additionally, we used our VLTs to search for similar sequences in other tick genomes by BLAST. Most of the VLTs had either no homology to tick genomes or very short stretches of high homology (~100 bp). However, VLT 246, which has homology to segmented flaviviruses, also had significant homology to a Jingmen tick virus endogenous viral element in the *Ixodes ricinus* genome (Fig. S7b). These observations suggest that VLTs are features of tick genomes beyond our data set.

DISCUSSION

In this study, we show the power of combining experimental enrichment of microbial sequences with single-tick metatranscriptomics for the identification of bacterial and viral transcripts in field *I. pacificus* tick communities. The ability to deeply sequence the non-host fraction allowed us to identify several genera of bacteria in *I. pacificus* previously unidentified in any tick species. Further investigation is warranted into the consequences and mechanisms underlying these symbiotic tick-bacteria partnerships.

Our approach also enabled us to uncover many novel viral RNA transcripts which we further investigated in a series of laboratory experiments. We found that several of these viral sequences were not only highly prevalent but were also present in tick salivary glands. This has important implications for public health, as bloodmeal hosts (including humans) are very likely to be exposed to viruses present in the salivary glands during feeding. Currently, tick-borne disease surveillance in California is focused on Lyme disease and a small number of other bacterial pathogens, but our results indicate humans in the region may also be exposed to diverse viruses (1). While our approach cannot determine whether any of these viruses are human pathogens, screening of clinical samples for exposure to them is a critical first step to understanding their disease-causing potential.

In addition to providing more comprehensive and quantitative insights into the *I. pacificus* microbiome, one of the most exciting and unexpected themes that emerged from our work relates to how viruses fit into the broader framework of the tick microbiome and tick biology. We found evidence that several viruses persisted in ticks across multiple life stages, including juvenile naïve larvae, as well as across field-caught and laboratory-reared populations, suggesting they are stable constituents of the tick microbiome. Certain viruses in *I. pacificus* may be able to establish and maintain independent niches within their tick hosts. These findings lay the groundwork for future work aimed at understanding tick–virus dynamics and how such relationships fit into tick physiology.

Further underscoring the critical importance of tick–virus interactions for *I. pacificus* biology was our discovery that numerous VLTs may originate from EVEs in the tick genome. Closer experimental evaluation of our VLTs pointed to a non-canonical

mechanism that is distinct from known antiviral pathways. While we cannot eliminate the possibility that VLTs are spuriously expressed without serving an adaptive function, EVEs are an underexplored feature of tick genomes, and future studies could determine whether VLTs represent a new mechanistic class of EVEs with adaptive contributions to tick immunity and biology.

The integration of RNA viral genomes into the tick genome as EVEs also provides a unique historical footprint for viruses that may have infected that tick host in the near or distant past. Interestingly, the VLTs identified in *I. pacificus* appear to derive from several viral families from which no exogenous viral genomes were found in this study, including the recently discovered segmented flaviviruses that cause febrile illness (62–64). Future field studies that expand on our *I. pacificus* virome analyses will help determine whether VLTs stem from ancient tick–virus interactions or contemporaneous interactions that were not captured in this study due to low abundance, limited sample size, and our focus on two collection locations. Altogether, our results highlight the need for more studies such as this in order to capture the full range of tick-associated microbes that could represent critical components of tick physiology or poorly understood pathogenic threats to human health. Our work provides an improved experimental and computational framework with increased sensitivity for low-abundance bacterial and viral taxa present in this increasingly important class of arthropod disease vectors.

MATERIALS AND METHODS

Tick sources

Ticks for metagenomic sequencing were collected by dragging from Garrapata State Park, California, USA ($n = 83$ adults) in December 2018 and from China Camp State Park, California, USA ($n = 17$ nymphs) in May 2019. Ticks for follow-up VLT PCR experiments were from the original China Camp collection. Ticks for tissue-tropism experiments were collected from Garrapata State Park in March 2022 ($n = 50$). Adult ticks were collected separately by sex. Ticks were stored in conical tubes according to groups (adult females, adult males, nymphs, and larvae) during transportation to the lab. Once in the lab, adult and nymphal ticks were surface sterilized in 1% bleach and frozen individually. Larval ticks were frozen as a pool and were not surface sterilized due to difficulties presented by their small size. All frozen ticks were stored at -80°C .

Laboratory-reared *I. pacificus* were received from the tick lab at the Centers for Disease Control and Prevention tick lab (Atlanta, GA, USA) and provided through BEI Resources (a service funded by the National Institute of Allergy and Infectious Diseases and managed by ATCC). Ticks were maintained in glass jars with a relative humidity of 95% (saturated solution of potassium nitrate) in a sealed incubator at 22°C with a light cycle of 16 h/8 h (light/dark).

RNA extraction/library prep

Total RNA was extracted from the field-caught *I. pacificus* adult and nymph ticks in two separate batches. On ice, individual ticks were transferred to separate wells of a 96-well deepwell plate that was pre-loaded with a single 5 mm steel ball bearing (OMNI International, GA, USA) and 400 μL of 1 \times DNA/RNA shield (Zymo Research Corp., Irvine, CA, USA) in each well. The plates were sealed and subjected to bead bashing (3 \times 3 min, with 1 min rest on ice in between each round of bashing) on a TissueLyser II beadmill (Qiagen, Valencia, CA, USA), then clarified by centrifugation at 2,000 rpm at 4°C for 5 min in a refrigerated tabletop centrifuge (Beckman Coulter, Indianapolis, IN, USA) to remove large debris. About 350 μL of the supernatant was transferred to a fresh 96-well deepwell plate and re-centrifuged under the same conditions to further clarify the homogenate. And 90 μL of the resulting supernatant was used as input for total RNA extractions; 110 μL of supernatant was transferred to a separate plate and archived at -80°C for potential follow-up analyses.

For both the adult tick and nymph tick homogenate preps, automated RNA extraction was performed in 96-well format (Bravo automated liquid handler, Agilent Technologies, Santa Clara, CA, USA) using a modified version of the Quick DNA/RNA pathogen magbead 96 extraction kit (Zymo Research Corp., Irvine, CA, USA) to automate total nucleic acid extraction and DNase treatment. RNA extracted from 90 μL of tick homogenates was eluted in a final volume of 25 μL of nuclease-free H_2O into 96-well PCR plates. An aliquot of 3 μL was used for quantitative and qualitative analysis of the total RNA for each sample via Qubit fluorometer assay (Thermo Fisher Scientific, Waltham, MA, USA) and Agilent Bioanalyzer Pico 6000 total Eukaryotic RNA electrophoresis (Agilent Technologies, Santa Clara, CA, USA). A separate 5 μL aliquot was used as input for RNAseq library prep, and $2 \times 7 \mu\text{L}$ aliquots were stamped into two separate daughter plates that were immediately frozen and archived at -80°C for potential follow-up studies.

The 96-well plates with 5 μL aliquots of adult tick and nymph tick total RNA preps were transferred to an automated liquid handler (Bravo automated liquid handler, Agilent Technologies, Santa Clara, CA, USA) for RNAseq library preparation in 96-well format. Briefly, the NEBNext Ultra II Non-directional RNAseq library preparation kit (New England Biolabs, Ipswich, MA, USA) was applied with the following modifications incorporated into the manufacturer's standard protocol: a 25 pg aliquot of External RNA Controls Consortium RNA spike-in mix ("ERCC," Thermo-Fisher, Waltham, MA, USA) was added to each sample prior to RNA fragmentation, and the input RNA mixture was fragmented for 8 min at 94°C prior to reverse transcription. A total of 12 cycles of PCR was performed using unique dual barcoded adapter primers for each sample at the amplification step. An initial equivolume pool was made up of 1 μL from each sample. Separate libraries were prepared for the adult and nymph ticks. SPRIselect (Beckman Coulter, Indianapolis, IN, USA) beads were used to size-select libraries with an average total length between 450 and 550 base pairs (bp). Library size distributions were verified by Agilent Bioanalyzer High Sensitivity DNA electrophoresis (Agilent Technologies, Santa Clara, CA, USA) and quantified by Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Paired-end 2×146 bp sequencing runs were separately performed on the adult and nymph equivolume pools with the Illumina MiSeq V2 Micro sequencing kit (Illumina, San Diego, CA, USA).

The yield of reads/ μL acquired from the small-scale MiSeq run of the equivolume pools of individual libraries were used to normalize volumes of the individual adult tick and nymph tick libraries to generate approximately equimolar pools of samples for subsequent large-scale metatranscriptomic sequence analysis. The equimolar library pools were then subjected to depletion of highly abundant sequences (DASH) (41, 65), using a previously described pool of tick gRNAs (66) complexed with in-house prep of purified recombinant Cas9 protein. Resulting DASH'd libraries were qualitatively and quantitatively analyzed by Agilent Bioanalyzer High Sensitivity DNA electrophoresis (Agilent Technologies, Santa Clara, CA, USA) and Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). While the DASH'd adult tick libraries provided sufficient material for large-scale metatranscriptomic sequencing, insufficient material remained in the tick nymph libraries that were DASH'd. Thus, for large-scale metatranscriptomic sequencing, the pool of DASH'd adult tick libraries was combined with a pool of un-DASH'd nymph tick libraries. This pooled prep was subjected to Paired-end 2×146 bp sequencing on the NextSeq2000 Illumina sequencing platform (Illumina, San Diego, CA, USA).

Host subtraction/pre-processing and contig assembly

Fastq reads from the run were demultiplexed based on their associated dual barcode index sequences, then further processed using the CZID analysis pipeline (67). Libraries underwent quality filtering and adaptor trimming. Host reads were then removed by mapping to closely related genomes. Host-subtracted reads from each library were then assembled into contigs using SPADEs (68) within the CZID pipeline (67) version 3.7.

Both these resulting contigs and the host-subtracted reads were used for downstream analysis.

Bacterial classification

Host-subtracted reads were classified using Kraken2 (version 2.1.1) (42). The full kraken2 database was used for classification. Only libraries with at least 1,000 classified reads were considered for analysis. Sequencing reads per taxon were converted to reads per million (rpm) using library size. To reduce false positives, the rpm value for each taxon was required to be at least 100 times the rpm in any of the control libraries (water and Hela cells) to be considered a positive. Additionally, at least 100 unique minimizers were required for each taxa. Taxa within each library not meeting these thresholds were excluded from the analysis. Fewer genera were detected in smaller libraries, including the nymphal ticks from China Camp SP (Fig. S1a), and samples clustered primarily by library size. We therefore focused subsequent analysis on libraries of at least one million non-host reads.

Virus and virus-like transcript identification

We focused our analysis on RNA viruses, as these tend to dominate arthropod viromes (33, 69, 70). Contigs were filtered to those of at least 1,500 bp in length. ORFs were predicted for these contigs using prodigal (71), and the resulting proteins were searched using HMMScan from HMMER3 (version 3.3.2) (58) against a collection of HMMER profiles of viral RdRps. The following RdRp HMMs were downloaded from the pfam database (72) on 4 March 2021; RdRP_1, RdRP_2, RdRP_3, RdRP_4, RdRP_5, Viral, RdRp_C, Mitovir_RNA_pol, Mononeg_RNA_pol, Birna_RdRp, and Bunya_RdRp. Additionally, custom HMMs were constructed from RdRp sequences for members of *Narnaviridae* and *Orthomyxoviridae* (sequences and combined HMM available in supplement).

All putative RdRp hits from HMMER were queried against the full NCBI nonredundant protein database (as of 24 January 2021) using diamond blastp version 0.9.24 to identify their closest hit (73). For putative RdRps with multiple hits, the alignment with the highest bitscore was reported. Putative RdRp sequences covering less than 30% of their closest blast hit were initially classified as putative virus-like transcripts, while those above that threshold were classified as exogenous viral sequences. The ORF structures of all contigs with RdRp sequences were manually inspected to confirm this classification. Those that showed an ORF structure similar to canonical viral genome sequences were classified as exogenous viral genome sequences; while those that contained significant gaps between ORFs or multiple reading frames with RdRp homology (where one was expected) were classified as VLTs.

To identify additional segments of multipartite viral genomes, we searched for contigs that strongly co-occurred with the contigs encoding an RdRp. Presence/absence of potential RdRp contig clusters in each library was based on evidence of quality- and host-filtered reads mapped to each contig. Presence was coded as 1 and absence as 0, and the Jaccard distance was calculated for all pairs of contigs. Any sequence with Jaccard distance <0.4 was classified as a putative additional genomic segment (Fig. S5). In our determination of whether a putative additional genomic segment represented a potential segment for its co-occurring RdRp contig cluster, we further considered (i) whether additional segments are expected for the viral family of the given RdRp contig cluster, and (ii) the sequence similarity (if any) of each putative additional genome segment to viral sequences in the NCBI databases. If additional segments were expected the contig was considered a segment originating from the same genome as the RdRp contig.

To infer the level of virus phylogeny to which each of the recovered exogenous viral sequences could be mapped, we downloaded the top 100 blast matches for each of the RdRp protein sequences that we recovered to create a representative set of related viral RdRp sequences for comparative sequence analysis. The downloaded sequences spanned the following viral taxa: *Bunyavirales*, *Mononegavirales*, *Reovirales*,

Picornaviridae, *Narnaviridae*, *Solemoviridae*, and *Chuviridae*. Within each of these taxa, we used CD-HIT (74) version 4.8.1 to filter the resulting initial sets of 100 RdRp sequences matches using a threshold of 85% nucleotide identity, then used MAFFT (75) version 7.475 to align each of the putative viral RdRp protein sequences with their corresponding taxonomic order of representative set of RdRp protein sequences. Maximum likelihood trees were estimated using iqtree 2.0.3 (76), using the model finder limited to all viral substitution models and with 1000 bootstraps. Resulting trees were visualized in iTOL version 5 (77) (Fig. S8).

Determination of prevalence and abundance

Assembled contigs were clustered using cd-hit-est (CD-HIT version 4.8.1) at a threshold of 85% nucleotide identity (74). Chuvirus genomes were rotated to a common start position using a custom python script to ensure accurate clustering. 85% nucleotide identity was chosen as a cutoff to minimize multi-mapping reads between closely related sequences. However, some clusters contained significant sequence diversity and could potentially be considered to contain multiple species. The representative sequences from this clustering were used for all downstream analyses.

Reads from each sequencing library were mapped back to the collection of cluster representatives using bowtie2 version 2.4.1 (78). Sequences aligning to each contig were counted using samtools idxstats version 1.9 (79). Aligned reads, contig length, and library size were used to calculate rpm and transcripts per million (tpm) values for each library. To consider a contig “present” in a given library, the rpm value was required to be greater than 10 times the value in any of the control libraries. This filter was designed to remove potential false positives caused by cross-contamination of high-titer species.

Rarefaction analysis

The viral genomes in each sample were determined by the presence of a contig of at least 1,000 base pairs (bp) that clustered with one of the 13 representative genomes identified. The presence of a contig rather than read mapping was used to simulate viral discovery in each sample, under the assumption that new viruses discovered may be too divergent to detect by read mapping. The samples were ordered by the number of new genomes seen (not seen in any of the previous samples). The number of new genomes was counted for the addition of each sample. This was repeated for 50 iterations and the median number of new samples at each step was determined. The Chao index was calculated using the R library fossil version 0.40 (80, 81).

Co-occurrence of taxa

To determine whether any pairings of taxa (either bacterial or viral) occur more or less frequently than expected given their prevalence, we utilized the recently developed metric α (48). The presence of each taxon was considered at the genus level for bacteria and at the species level for viruses. The distance α and associated P value were determined for all pairs of taxa using the CooccurrenceAffinity R package (version 1.0) (48). Pairs were filtered to those with a P value ≤ 0.005 . These relationships were visualized as a network with edges corresponding to α and nodes corresponding to taxa using the R package igraph (version 1.3.0) (82). Node size was scaled according to taxon prevalence in the data set.

Virus-like transcript mining

We used the bioinformatics methodology from this study to search for the presence of VLTs in the contigs from Ni et al. (22), a metagenomic data set of ticks from China. The open reading frame structure of all of the deposited contigs were visually examined using geneious for the same types of patterns as observed in our data set, including early stop codons, overlapping reading frames, multiple reading frames with RdRp homology.

To determine whether the VLTs observed in our data set are present in other tick genomes, we compared the 25 representative VLT nucleotide sequences to tick genomes using BLAST. BLAST was performed using the NCBI web server on 19 December 2023, employing the blastn algorithm with word size of 11, *e*-value of 0.05, and limiting results to taxid 6939 (*Ixodidae*).

Virus-like transcript PCR

To verify that VLTs were present in tick genomes and expressed in field-caught and laboratory-reared ticks, we extracted tick RNA and genomic DNA and confirmed the presence of VLTs with PCR. Adult male *I. pacificus* ticks from China Camp ($n = 3$) were pooled and homogenized by beating for two increments of 30 s at 4,000 bpm in a bench homogenizer (Bead Bug, Benchmark Scientific) with 1.4 mm zirconium oxide ceramic beads (Fisher Scientific) in ice-cold TRIzol reagent (Thermo Fisher Scientific). RNA extraction was performed using a Zymo Research Direct-zol RNA Microprep kit (Zymo Research), and RNA was converted to cDNA using Primescript RT reagent kit (Takara Bio) in 10 μ L reactions using random hexamer primers and following manufacturer protocols. Genomic DNA was extracted from adult male *I. pacificus* ticks ($n = 3$, separate individuals from RNA), which were pooled, flash-frozen, and ground to powder. A DNeasy Blood & Tissue Kit (Qiagen) was used to extract genomic DNA following the manufacturer's protocols.

3' RACE was conducted using the 3' RACE system for rapid amplification of cDNA ends kit (Thermo Fisher Scientific) according to the manufacturer's protocols. Nested PCR using VLT-specific primers was used to amplify regions of VLTs from tick cDNA and genomic DNA in two steps using 10 μ M VLT-specific primers and the UAP primer provided in the kit. Nested PCR primer sequences were designed by hand, following the protocol from the kit for the 3' RACE system for rapid amplification of cDNA ends. We verified the absence of hairpin loops with IDT oligoanalyzer, and then we used NCBI BLAST to verify that there were no non-specific targets. The resulting primer sequences can be found in Table S2. We loaded 5 μ L of product onto 0.7% agarose gels and ran them at 160 V for 2 h before imaging. PCRs were purified using a QIAquick PCR purification kit (Qiagen) and unidirectionally Sanger sequenced (Genewiz) with the nested PCR primer.

Virus multiplex PCR

Viral genome sequences were analyzed with Snappgene (Dotmatics) and PrimerPlex software (Premier Biosoft) to design four sets of multiplex primers that amplify 100–550 bp regions. Platinum SuperFi II Green PCR Master Mix (Thermo Scientific) was used for all PCR reactions. Mixed cDNA from the original sequenced field-collected ticks was used as a positive control. No-template (water only) reactions were also included as negative controls. Primer pair sequences are listed in Table S2. PCR reactions were analyzed by electrophoresis using a 2% agarose gel containing GelRed (Biotium) and visualized using an Azure c400 imager (Azure Biosystems). At least one band corresponding to each virus was cut out of the agarose gel, purified using the QIAquick Gel extraction kit (Qiagen), and Sanger sequenced (GeneWiz) to confirm correct PCR specificity.

Tick dissections/extractions

For tissue-tropism determination, wild-collected ticks ($n = 20$) from Garrapata State Park were dissected using a micro scalpel cleaned with 70% isopropanol and a sterile needle. Tissues were pooled from multiple ticks rather than analyzed individually as it was not possible to obtain sufficient RNA from the midgut or salivary glands of a single tick. Males and females were processed separately and each pool of tissues contained material from 3 to 10 individuals.

Ticks were pooled. The scalpel was cleaned with 70% isopropanol and the needle was replaced between pools. The tick cuticle was excised and the midgut and salivary

glands were removed using tweezers cleaned with 70% isopropanol (cleaned between each pool). Tissues were pooled and rinsed in droplets of phosphate-buffered saline, then transferred by pipette into 300 μ L of TRIzol. PCR results from males and females, as well as biological replicates are collapsed for simplicity in Fig. 3d, but sex-specific and replicate-specific results can be found in Fig. S4d.

Whole adult ticks were added to 300 μ L of TRIzol in pools of four to five individuals, grouped by sex. Nymphal ticks were added to TRIzol in a pool of 3, and larval ticks were flash-frozen and added to TRIzol in pools of 10–15. All ticks and tick tissues were homogenized by bead-beating with ceramic beads in increments of 30 s. Samples were placed on ice between cycles and cycles continued until tissue was visually homogenized.

RNA extraction was performed using Directzol RNA Extraction kits, with on-column DNase1 treatment. RNA was converted to single-stranded cDNA using Quantabio cDNA mastermix in 10 μ L reactions.

ACKNOWLEDGMENTS

We would like to thank Anne Sapiro, Gytis Dudas, Kishen Patel, and Raul Andino for their helpful conversations and feedback on this project. We are grateful to Greg Huber, Olga Botvinnik, and Norma Neff at Chan Zuckerberg Biohub San Francisco for their strategic and technical input. We also thank members of both the Chou lab and the lab of Andrea Swei for assistance collecting ticks in the field. We greatly appreciate all members of the Chou and Pollard labs for their input and feedback on experiments, analyses, and manuscript preparation.

AUTHOR AFFILIATIONS

¹Department of Biochemistry & Biophysics, University of California–San Francisco, San Francisco, California, USA

²Gladstone Institute of Data Science & Biotechnology, San Francisco, California, USA

³One Health Institute, Colorado State University–Fort Collins, Fort Collins, Colorado, USA

⁴Department of Biomedical Sciences, Colorado State University–Fort Collins, Fort Collins, Colorado, USA

⁵Chan Zuckerberg Biohub, San Francisco, San Francisco, California, USA

⁶Department of Biology, Colorado State University–Fort Collins, Fort Collins, Colorado, USA

⁷Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, California, USA

AUTHOR ORCIDs

Calla Martyn  <http://orcid.org/0000-0002-1886-5198>

Amy Kistler  <http://orcid.org/0000-0003-1112-719X>

Katherine S. Pollard  <http://orcid.org/0000-0002-9870-6196>

Seemay Chou  <http://orcid.org/0000-0002-7271-303X>

FUNDING

Funder	Grant(s)	Author(s)
Chan Zuckerberg Biohub		Calla Martyn
		Domokos Lauko
		Gloria Castaneda
		Amy Kistler
		Katherine S. Pollard
Gladstone Institutes (J. David Gladstone Institutes)		Calla Martyn

Funder	Grant(s)	Author(s)
		Katherine S. Pollard
HHS National Institutes of Health (NIH)	R01 AI32851	Beth M. Hayes
Pew Charitable Trusts (PCT)		Beth M. Hayes

AUTHOR CONTRIBUTIONS

Calla Martyn, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Beth M. Hayes, Validation, Writing – original draft, Writing – review and editing | Domokos Lauko, Validation, Writing – original draft | Edward Midthun, Investigation, Validation | Gloria Castaneda, Investigation, Methodology | Angela Bosco-Lauth, Funding acquisition, Investigation, Supervision | Daniel J. Salkeld, Investigation, Methodology | Amy Kistler, Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing | Katherine S. Pollard, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing | Seemay Chou, Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Sequencing data are available at the NCBI Sequence Read Archives under project [PRJNA870442](#). Intermediate files, including assembled contigs and read coverage can be found on figshare at https://figshare.com/projects/_pacifics_mNGS/144081. Viral assemblies have been deposited at NCBI GenBank under accession numbers [PP415825](#), [PP415826](#), [PP415827](#), [PP415828](#), [PP415829](#), [PP415830](#), [PP415831](#), [PP415832](#), [PP415833](#), [PP415834](#), [PP415835](#), [PP415836](#), [PP415837](#), [PP415838](#), [PP415839](#), [PP415840](#), [PP415841](#), [PP415842](#), [PP415843](#), [PP415844](#), [PP415845](#), [PP415846](#), [PP415847](#), [PP415848](#), and [PP415849](#). In addition, they can be accessed at FigShare using DOI [10.6084/m9.figshare.20497227](https://doi.org/10.6084/m9.figshare.20497227). *I. pacificus* genome sequence (released February 2018) is available at [s3.us-west-2.amazonaws.com: https://czid-public-references.s3.us-west-2.amazonaws.com/host_filter/ticks/2018-02-15-utc-1518652800-unixtime__2018-02-15-utc-1518652800-unixtime/tick_genomes.fa](https://s3.us-west-2.amazonaws.com/czid-public-references.s3.us-west-2.amazonaws.com/host_filter/ticks/2018-02-15-utc-1518652800-unixtime__2018-02-15-utc-1518652800-unixtime/tick_genomes.fa) and as a supplemental data file. Analysis notebooks and scripts can be found on github at https://github.com/callamartyn/ipac_virus. CZID pipeline results, including intermediate host subtracted reads, assemblies, and other intermediate files can be found at <https://czid.org/pub/uuFkacq3hT> (Garrapata adults), <https://czid.org/pub/EEJfJPNhYP> (Garrapata adults, resequencing of low-coverage samples), and <https://czid.org/pub/ss3AnxpDbU> (China Camp nymphs).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental Figures (mSystems00321-24-S0001.pdf). Figures S1-S8.

Supplemental Note (mSystems00321-24-S0002.docx). Descriptions of viral genome assemblies.

Table S1 (mSystems00321-24-S0003.docx). Summary of the viruses identified.

Table S2 (mSystems00321-24-S0004.docx). All PCR primer sequences used.

Open Peer Review

PEER REVIEW HISTORY (review-history.pdf). An accounting of the reviewer comments and feedback.

REFERENCES

- CDC. Tickborne disease surveillance data summary. CDC Centers for Disease Control and Prevention <https://www.cdc.gov/ticks/data-summary/index.html> (2021).
- Tokarz R, Tagliaferro T, Sameroff S, Cucura DM, Oleynik A, Che X, Jain K, Lipkin WI. 2019. Microbiome analysis of *Ixodes scapularis* ticks from New York and Connecticut. *Ticks Tick Borne Dis* 10:894–900. <https://doi.org/10.1016/j.ttbdis.2019.04.011>
- Ross BD, Hayes B, Radey MC, Lee X, Josek T, Bjork J, Neitzel D, Paskewitz S, Chou S, Mougous JD. 2018. *Ixodes scapularis* does not harbor a stable midgut microbiome. *ISME J* 12:2596–2607. <https://doi.org/10.1038/s41396-018-0161-6>
- Avelo T, Norberg A, Tschirren B. 2019. Bacterial microbiota composition of *Ixodes ricinus* ticks: the role of environmental variation, tick characteristics and microbial interactions. *PeerJ* 7:e8217. <https://doi.org/10.7717/peerj.8217>
- Chandra S, Harvey E, Emery D, Holmes EC, Šlapeta J. 2021. Unbiased characterization of the microbiome and virome of questing ticks. *Front Microbiol* 12:627327. <https://doi.org/10.3389/fmicb.2021.627327>
- Socarras KM, Earl JP, Krol JE, Bhat A, Pablonia M, Harrison MH, Lang SP, Sen B, Ahmed A, Hester M, Mell JC, Vandegrift K, Ehrlich GD. 2021. Species-level profiling of *Ixodes pacificus* bacterial microbiomes reveals high variability across short spatial scales at different taxonomic resolutions. *Genet Test Mol Biomarkers* 25:551–562. <https://doi.org/10.1089/gtmb.2021.0088>
- Adegoke A, Kumar D, Bobo C, Rashid MI, Durrani AZ, Sajid MS, Karim S. 2020. Tick-borne pathogens shape the native Microbiome within tick vectors. *Microorganisms* 8:1299. <https://doi.org/10.3390/microorganisms8091299>
- Cross S, Kapuscinski M, Perino J, Maertens B, Weger-Lucarelli J, Ebel G, Stenglein M. 2018. Co-infection patterns in individual *Ixodes scapularis* ticks reveal associations between viral. *Viruses* 10:388. <https://doi.org/10.3390/v10070388>
- Bonnet SI, Binetruy F, Hernández-Jarguín AM, Duron O. 2017. The tick microbiome: why non-pathogenic microorganisms matter in tick biology and pathogen transmission. *Front Cell Infect Microbiol* 7. <https://doi.org/10.3389/fcimb.2017.00236>
- Chicana B, Couper LI, Kwan JY, Tahiraj E, Swee A. 2019. Comparative microbiome profiles of sympatric tick species from the far-Western United States. *Insects* 10:353. <https://doi.org/10.3390/insects10100353>
- Díaz-Sánchez S, Hernández-Jarguín A, Torina A, de Mera IGF, Blanda V, Caracappa S, Gortazar C, de la Fuente J. 2019. Characterization of the bacterial microbiota in wild-caught *Ixodes ventralis*. *Ticks Tick Borne Dis* 10:336–343. <https://doi.org/10.1016/j.ttbdis.2018.11.014>
- Estrada-Peña A, Cabezas-Cruz A, Pollet T, Vayssier-Tausat M, Cosson J-F. 2018. High throughput sequencing and network analysis disentangle the microbial communities of ticks and hosts within and between ecosystems. *Front Cell Infect Microbiol* 8:236. <https://doi.org/10.3389/fcimb.2018.00236>
- Lee S, Kim JY, Yi M-H, Lee I-Y, Fyomagwa R, Yong T-S. 2019. Comparative microbiomes of ticks collected from a black rhino and its surrounding environment. *Int J Parasitol Parasites Wildl* 9:239–243. <https://doi.org/10.1016/j.ijppaw.2019.05.008>
- Maldonado-Ruiz LP, Neupane S, Park Y, Zurek L. 2021. The bacterial community of the lone Star tick (*Amblyomma americanum*). *Parasit Vectors* 14:49. <https://doi.org/10.1186/s13071-020-04550-z>
- Bouquet J, Melgar M, Swee A, Delwart E, Lane RS, Chiu CY. 2017. Metagenomic-based surveillance of pacific coast tick *Dermacentor occidentalis* identifies two novel bunyaviruses and an emerging human rickettsial pathogen. *Sci Rep* 7:12234. <https://doi.org/10.1038/s41598-017-12047-6>
- Couper LI, Kwan JY, Ma J, Swee A. 2019. Drivers and patterns of microbial community assembly in a Lyme disease vector. *Ecol Evol* 9:7768–7779. <https://doi.org/10.1002/ece3.5361>
- Ravi A, Ereqat S, Al-Jawabreh A, Abdeen Z, Abu Shamma O, Hall H, Pallen MJ, Nasereddin A. 2019. Metagenomic profiling of ticks: identification of novel rickettsial genomes and detection of tick-borne canine parvovirus. *PLoS Negl Trop Dis* 13:e0006805. <https://doi.org/10.1371/journal.pntd.0006805>
- Tufts DM, Sameroff S, Tagliaferro T, Jain K, Oleynik A, VanAcker MC, Diuk-Wasser MA, Lipkin WI, Tokarz R. 2020. A metagenomic examination of the pathobiome of the invasive tick species, *Haemaphysalis longicornis*, collected from a New York City borough, USA. *Ticks Tick Borne Dis* 11:101516. <https://doi.org/10.1016/j.ttbdis.2020.101516>
- Kurilshikov A, Livanova NN, Fomenko NV, Tupikin AE, Rar VA, Kabilov MR, Livanov SG, Tikunova NV. 2015. Comparative metagenomic profiling of symbiotic bacterial communities associated with *Ixodes persulcatus*, *Ixodes pavlovskii* and *Dermacentor reticulatus* ticks. *PLoS One* 10:e0131413. <https://doi.org/10.1371/journal.pone.0131413>
- Xia H, Hu C, Zhang D, Tang S, Zhang Z, Kou Z, Fan Z, Bente D, Zeng C, Li T. 2015. Metagenomic profile of the viral communities in *Rhipicephalus* spp. ticks from Yunnan, China. *PLoS One* 10:e0121609. <https://doi.org/10.1371/journal.pone.0121609>
- Zakham F, Albalawi AE, Alanazi AD, Truong Nguyen P, Alouffi AS, Alaoui A, Sironen T, Smura T, Vapalahti O. 2021. Viral RNA metagenomics of hyalomma ticks collected from dromedary camels in Makkah province, Saudi Arabia. *Viruses* 13:1396. <https://doi.org/10.3390/v13071396>
- Ni X-B, Cui X-M, Liu J-Y, Ye R-Z, Wu Y-Q, Jiang J-F, Sun Y, Wang Q, Shum MH-H, Chang Q-C, et al. 2023. Metavirome of 31 tick species provides a compendium of 1,801 RNA virus genomes. *Nat Microbiol* 8:162–173. <https://doi.org/10.1038/s41564-022-01275-w>
- Duron O, Morel O, Noël V, Buysse M, Binetruy F, Lancelot R, Loire E, Ménard C, Bouchez O, Vavre F, Vial L. 2018. Tick-bacteria mutualism depends on B vitamin synthesis pathways. *Curr Biol* 28:1896–1902. <https://doi.org/10.1016/j.cub.2018.04.038>
- Guizzo MG, Parizi LF, Nunes RD, Schama R, Albano RM, Tironi L, Oldiges DP, Vieira RP, Oliveira WHC, Leite M de S, Gonzales SA, Farber M, Martins O, Vaz I da S, Oliveira PL. 2017. A coxiella mutualist symbiont is essential to the development of *Rhipicephalus microplus*. *Sci Rep* 7:17554. <https://doi.org/10.1038/s41598-017-17309-x>
- Hofer U. 2021. Tick symbiont promotes blood feeding. *Nat Rev Microbiol* 19:744. <https://doi.org/10.1038/s41579-021-00645-1>
- Hunter DJ, Torkelson JL, Bodnar J, Mortazavi B, Laurent T, Deason J, Thephavongsa K, Zhong J. 2015. The rickettsia endosymbiont of *Ixodes pacificus* contains all the genes of *de novo* folate biosynthesis. *PLoS One* 10:e0144552. <https://doi.org/10.1371/journal.pone.0144552>
- Alberdi MP, Dalby MJ, Rodriguez-Andres J, Fazakerley JK, Kohl A, Bell-Sakyi L. 2012. Detection and identification of putative bacterial endosymbionts and endogenous viruses in tick cell lines. *Ticks Tick-borne Dis* 3:137–146. <https://doi.org/10.1016/j.ttbdis.2012.05.002>
- Chou S, Daugherty MD, Peterson SB, Biboy J, Yang Y, Jutras BL, Fritz-Laylin LK, Ferrin MA, Harding BN, Jacobs-Wagner C, Yang XF, Vollmer W, Malik HS, Mougous JD. 2015. Transferred interbacterial antagonism genes augment eukaryotic innate immune function. *Nature* 518:98–101. <https://doi.org/10.1038/nature13965>
- Ter Horst AM, Nigg JC, Dekker FM, Falk BW. 2019. Endogenous viral elements are widespread in arthropod genomes and commonly give rise to PIWI-interacting RNAs. *J Virol* 93:e02124-18. <https://doi.org/10.1128/JVI.02124-18>
- Palatini U, Miesen P, Carballar-Lejarazu R, Ometto L, Rizzo E, Tu Z, van Rij RP, Bonizzoni M. 2017. Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. *BMC Genomics* 18:512. <https://doi.org/10.1186/s12864-017-3903-3>
- Russo AG, Kelly AG, Enosi Tuipulotu D, Tanaka MM, White PA. 2019. Novel insights into endogenous RNA viral elements in *Ixodes scapularis* and other arbovirus vector genomes. *Virus Evol* 5:vez010. <https://doi.org/10.1093/ve/vez010>
- Bell-Sakyi L, Attoui H. 2013. Endogenous tick viruses and modulation of tick-borne pathogen growth. *Front Cell Infect Microbiol* 3:25. <https://doi.org/10.3389/fcimb.2013.00025>
- Vandegrift KJ, Kapoor A. 2019. The ecology of new constituents of the tick virome and their relevance to public health. *Viruses* 11:529. <https://doi.org/10.3390/v11060529>
- Hahn MB, Jarnevich CS, Monaghan AJ, Eisen RJ. 2016. Modeling the geographic distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the contiguous United States. *J Med Entomol* 53:1176–1191. <https://doi.org/10.1093/jme/tjw076>

35. Eisen RJ, Eisen L, Beard CB. 2016. County-scale distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the continental United States. *J Med Entomol* 53:349–386. <https://doi.org/10.1093/jme/tjv237>
36. Salkeld DJ, Porter WT, Loh SM, Nieto NC. 2019. Time of year and outdoor recreation affect human exposure to ticks in California, United States. *Ticks Tick Borne Dis* 10:1113–1117. <https://doi.org/10.1016/j.ttbdis.2019.06.004>
37. McVicar M, Rivera I, Reyes JB, Gulia-Nuss M. 2022. Ecology of *Ixodes pacificus* ticks and associated pathogens in the Western United States. *Pathogens* 11:89. <https://doi.org/10.3390/pathogens11010089>
38. Porter WT, Barrand ZA, Wachara J, DaVall K, Mihaljevic JR, Pearson T, Salkeld DJ, Nieto NC. 2021. Predicting the current and future distribution of the Western black-legged tick, *Ixodes pacificus*, across the Western US using citizen science collections. *PLoS One* 16:e0244754. <https://doi.org/10.1371/journal.pone.0244754>
39. Dennis DT, Nekomoto TS, Victor JC, Paul WS, Piesman J. 1998. Forum: reported distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the United States. *J Med Entomol* 35:629–638. <https://doi.org/10.1093/jmedent/35.5.629>
40. Xu G, Pearson P, Dykstra E, Andrews ES, Rich SM. 2019. Human-biting *Ixodes* ticks and pathogen prevalence from California, Oregon, and Washington. *Vector Borne Zoonotic Dis* 19:106–114. <https://doi.org/10.1089/vbz.2018.2323>
41. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H, DeRisi JL. 2016. Depletion of abundant sequences by hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol* 17:41. <https://doi.org/10.1186/s13059-016-0904-5>
42. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
43. Eskow E, Adelson ME, Rao R-V, Mordechai E. 2003. Evidence for disseminated mycoplasma fermentans in New Jersey residents with antecedent tick attachment and subsequent musculoskeletal symptoms. *J Clin Rheumatol* 9:77–87. <https://doi.org/10.1097/01.RHU.0000062510.04724.07>
44. Straub RK, Powers CM. 2021. Chronic fatigue syndrome: a case report highlighting diagnosing and treatment challenges and the possibility of Jarisch–Herxheimer reactions if high infectious loads are present. *Healthcare (Basel)* 9:1537. <https://doi.org/10.3390/healthcare9111537>
45. Sasserá D, Beninati T, Bandi C, Bouman EAP, Sacchi L, Fabbri M, Lo N. 2006. *Candidatus* *Midichloria mitochondrii*, an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol* 56:2535–2540. <https://doi.org/10.1099/ijs.0.64386-0>
46. Ogata S, Mohamed WMA, Kusakisako K, Thu MJ, Qiu Y, Moustafa MAM, Matsuno K, Katakura K, Nonaka N, Nakao R. 2021. Spiroplasma infection among Ixodid ticks exhibits species dependence and suggests a vertical pattern of transmission. *Microorganisms* 9:333. <https://doi.org/10.3390/microorganisms9020333>
47. Beliavskaia A, Höning V, Erhart J, Vyhlidalova T, Palus M, Cerny J, Kozlova I, Ruzek D, Palomar AM, Bell-Sakyi L. 2021. Spiroplasma isolated from third-generation laboratory colony *Ixodes persulcatus* ticks. *Front. Vet. Sci* 8. <https://doi.org/10.3389/fvets.2021.659786>
48. Mainali KP, Slud E, Singer MC, Fagan WF. 2022. A better index for analysis of co-occurrence and similarity. *Sci Adv* 8:eabj9204. <https://doi.org/10.1126/sciadv.abj9204>
49. Buresová V, Franta Z, Kopáček P. 2006. A comparison of chryseobacterium indologenes pathogenicity to the soft tick ornithodoros moubata and hard tick *Ixodes ricinus*. *J Invertebr Pathol* 93:96–104. <https://doi.org/10.1016/j.jip.2006.05.006>
50. Blomström A-L, Luz HR, Öhlund P, Lukenge M, Brandão PE, Labruna MB, Berg M. 2020. Novel viruses found in antricola ticks collected in bat caves in the Western Amazonia of Brazil. *Viruses* 12:48. <https://doi.org/10.3390/v12010048>
51. Brinkmann A, Dinçer E, Polat C, Hekimoğlu O, Hacıoğlu S, Földes K, Özkul A, Öktem İMA, Nitsche A, Ergünay K. 2018. A metagenomic survey identifies tamdy orthonairovirus as well as divergent phlebo-, rhabdo-, chu- and favi-like viruses in anatolia, Turkey. *Ticks Tick Borne Dis* 9:1173–1183. <https://doi.org/10.1016/j.ttbdis.2018.04.017>
52. Cholleti H, Hayer J, Mulandane FC, Falk K, Fafetine J, Berg M, Blomström A-L. 2018. Viral metagenomics reveals the presence of highly divergent quaranjavirus in rhipicephalus ticks from Mozambique. *Infect Ecol Epidemiol* 8:1478585. <https://doi.org/10.1080/2008686.2018.1478585>
53. Harvey E, Rose K, Eden J-S, Lo N, Abeyesuriya T, Shi M, Doggett SL, Holmes EC. 2019. Extensive diversity of RNA viruses in Australian ticks. *J Virol* 93:e01358-18. <https://doi.org/10.1128/JVI.01358-18>
54. Kobayashi D, Murota K, Itokawa K, Ejiri H, Amoa-Bosompem M, Faizah AN, Watanabe M, Maekawa Y, Hayashi T, Noda S, Yamauchi T, Komagata O, Sawabe K, Isawa H. 2020. RNA virome analysis of questing ticks from Hokuriku district, Japan, and the evolutionary dynamics of tick-borne phleboviruses. *Ticks Tick Borne Dis* 11:101364. <https://doi.org/10.1016/j.ttbdis.2019.101364>
55. Batson J, Dudas G, Haas-Stapleton E, Kistler AL, Li LM, Logan P, Ratnasiri K, Retallack H. 2021. Single mosquito metatranscriptomics identifies vectors, emerging pathogens and reservoirs in one assay. *Elife* 10:e68353. <https://doi.org/10.7554/eLife.68353>
56. Reyes A, Alves JMP, Durham AM, Gruber A. 2017. Use of profile hidden Markov models in viral discovery: current insights. *Adv Genom Genet* 7:29–45. <https://doi.org/10.2147/AGG.S136574>
57. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. 2014. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PLOS ONE* 9:e105067. <https://doi.org/10.1371/journal.pone.0105067>
58. Eddy SR. 2011. Accelerated profile HMM searches. *PLOS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
59. Oliveira JH, Bahia AC, Vale PF. 2020. How are arbovirus vectors able to tolerate infection? *Dev Comp Immunol* 103:103514. <https://doi.org/10.1016/j.dci.2019.103514>
60. Maqbool M, Sajid MS, Saqib M, Anjum FR, Tayyab MH, Rizwan HM, Rashid MI, Rashid I, Iqbal A, Siddique RM, Shamim A, Hassan MA, Atif FA, Razzaq A, Zeeshan M, Hussain K, Nisar RHA, Tanveer A, Younas S, Kamran K, Rahman SU. 2022. Potential mechanisms of transmission of tick-borne viruses at the virus-tick interface. *Front Microbiol* 13:846884. <https://doi.org/10.3389/fmicb.2022.846884>
61. Salvati MV, Salaris C, Monteil V, Del Vecchio C, Palù G, Parolin C, Calistri A, Bell-Sakyi L, Mirazimi A, Salata C. 2021. Virus-derived DNA forms mediate the persistent infection of tick cells by hazaara virus and crimean-congo hemorrhagic fever virus. *J Virol* 95:e0163821. <https://doi.org/10.1128/JVI.01638-21>
62. Kholodilov IS, Litov AG, Klimentov AS, Belova OA, Polienko AE, Nikitin NA, Shchetinin AM, Ivannikova AY, Bell-Sakyi L, Yakovlev AS, Bugmyrin SV, Bespyatova LA, Gmyl LV, Luchinina SV, Gmyl AP, Gushchin VA, Karganova GG. 2020. Isolation and characterisation of alongshan virus in Russia. *Viruses* 12:362. <https://doi.org/10.3390/v12040362>
63. Vandegrift KJ, Kumar A, Sharma H, Murthy S, Kramer LD, Ostfeld R, Hudson PJ, Kapoor A. 2020. Presence of segmented flavivirus infections in North America. *Emerg Infect Dis* 26:1810–1817. <https://doi.org/10.3201/eid2608.190986>
64. Kumar A, et al. 2020 Early release - virome analysis for detection of segmented flaviviruses in wild rodents. *Emerg Infect Dis J - CDC*. <https://doi.org/10.3201/eid2608.190986>
65. Lyden A, Crawford E, Quan J, Caldera S, Dwynerman D. 2019. DASH protocol. Available from: <https://www.protocols.io/view/dash-protocol-6rjhd4n>
66. Ring K, Couper LI, Sapiro AL, Yarza F, Yang XF, Clay K, Mateusiak C, Chou S, Swei A. 2022. Host blood meal identity modifies vector gene expression and competency. *Mol Ecol* 31:2698–2711. <https://doi.org/10.1111/mec.16413>
67. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, Han J, Holmes OB, Juan Y-F, King R, et al. 2020. IDseq-an open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* 9:giaa111. <https://doi.org/10.1093/gigascience/giaa111>
68. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
69. Wu H, Pang R, Cheng T, Xue L, Zeng H, Lei T, Chen M, Wu S, Ding Y, Zhang J, Shi M, Wu Q. 2020. Abundant and diverse RNA viruses in insects

- revealed by RNA-Seq analysis ecological and evolutionary implications. *mSystems* 5:e00039-20. <https://doi.org/10.1128/mSystems.00039-20>
70. Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, Chen L-J, Qin X-C, Xu J, Holmes EC, Zhang Y-Z. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* 4:e05378. <https://doi.org/10.7554/eLife.05378>
 71. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
 72. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res* 49:D412–D419. <https://doi.org/10.1093/nar/gkaa913>
 73. Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18:366–368. <https://doi.org/10.1038/s41592-021-01101-x>
 74. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
 75. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
 76. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>
 77. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>
 78. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
 79. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and Samtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
 80. Chao A. 1984. Nonparametric estimation of the number of classes in a population. *Scand J Stat* 11:265–270. <http://www.jstor.org/stable/4615964>.
 81. MatthewJV. 2011. fossil: Palaeoecological and palaeogeographical analysis tools
 82. Csardi G, Nepusz T. 2006. The Igraph software package for complex network research. *InterJ Complex Syst* 1695. <https://zenodo.org/>.