Análisis del metagenoma y viroma del linaje tropical de la garrapata *Rhipicephalus sanguineus* sensu lato (s.l.) en los departamentos de Santander y Casanare en Colombia

Luisa Fernanda Páez Triana

Universidad del Rosario
Facultad de Ciencias Naturales
Bogotá, Colombia
2023

Análisis del metagenoma y viroma del linaje tropical de la garrapata *Rhipicephalus sanguineus* sensu lato (s.l.) en los departamentos de Santander y Casanare en Colombia

Estudiante

Luisa Fernanda Páez Triana

Tesis presentada como requisito para obtener el título de:

Magíster en Ciencias Naturales

Director

Juan David Ramírez, Ph.D.

Profesor Asociado

Facultad de Ciencias Naturales, Universidad del Rosario

Co-director

Marina Muñoz Diaz, Ph.D.

Directora Técnica del Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR)

Facultad de Ciencias Naturales, Universidad del Rosario

Facultad de Ciencias Naturales
Maestría en Ciencias Naturales
Universidad del Rosario
Bogotá, Colombia
2023

Agradecimientos

No suelo ser una persona de abundantes palabras y grandes emociones. Sin embargo, este camino lleno de altibajos me ha puesto múltiples personas sin las que no habría llegado hasta aquí. Quisiera comenzar agradeciendo a mi familia y amigos que han compartido y apoyado este gran camino académico. Especialmente quisiera agradecer a mi padre *Fernando Páez*, no solo ha sido mi soporte a lo largo de estos últimos años; si no que me ha mostrado el tipo de persona que espero ser. Gracias a él tuve la oportunidad de cumplir mis sueños, y por el estoy aquí. Su apoyo en cada decisión, desde que emprendí este camino en la biología hasta ahora ha sido inigualable. Siempre ha estado para mí, escuchándome y dándome fuerzas para lograr completar este capítulo de mi vida de la mejor manera. También quisiera agradecer a mi madre *Norma Triana*, quien estuvo atenta y me impulso durante todos estos años. Quien, por su ejemplo, me ha enseñado a ser una mujer valiente, esforzada y resiliente. Ambos me enseñaron a ser responsable, a seguir con la cabeza en alto y no dejarme vencer por las dificultades, siempre serán mi inspiración.

Quisiera agradecer a mi director de tesis *Juan David Ramírez* por confiar en mi desde el primer día, y sobre todo por nutrir en mí una pasión por la microbiología y las garrapatas. Gracias por acompañarme y guiarme en esta investigación, por las enseñanzas tanto en la vida académica como en la vida personal y por la gran paciencia que me has tenido. Gracias por enseñarme a ver este mundo de una manera diferente y por mostrarme el tipo de científica que aspiro llegar a ser. Tu constante apoyo y tu forma de ser son una inspiración para mí. Gracias Juan por permitirme terminar este bonito proyecto, guiarme durante todos los altibajos y confiar en lo que puedo llegar a hacer en un futuro. Porque a pesar de que muchas veces yo no confió en mí, tú has creído en mis capacidades y en lo que puede llegar a ser, permitiéndome alcanzar esta primera pequeña meta y aspirar a mucho más en mi vida académica y personal.

También quisiera agradecer a mi codirectora *Marina Muñoz*, quien estuvo conmigo desde el día uno que comencé a investigar. Me has acompañado a lo largo de este viaje, desde mi primer día de laboratorio donde me ensañaste a usar una pipeta, mi primera entrega de tesis en el pregrado, ayudándome hasta media noche, hasta guiarme en esta tesis, aconsejándome hasta el último momento. Gracias por enseñarme, instruirme y acompañarme en todas las decisiones de estos últimos años. Me has escuchado todos mis problemas, desde el más pequeño hasta el más grande. Tu amor y forma de ser me enseñaron a amar la ciencia y a hacer las cosas cada vez mejor. Gracias por toda la confianza que has depositado en mí y por acompañarme e instruirme desde lo más mínimo hasta lo más grande en este camino académico. Espero algún día poder enseñar la ciencia de la manera que tú la enseñas, impartiendo amor por esto.

Finalmente quisiera agradecer a los miembros del Grupo de Investigación en Microbiología de la Universidad (GIMUR) por su apoyo y su contribución en el desarrollo de esta tesis. Sus aportes fueron de gran ayuda para mí. En especial quisiera agradecer a *Giovanny Herrera*, a *Laura Vega* y *Lissa Cruz* por su colaboración activa en el desarrollo de esta investigación. Ayudándome tanto en el desarrollo de cada artículo, como resolviéndome dudas por más bobas que fueran. También quisiera agradecer a mis

compañeros *Nicolas Luna* y *Vanessa Urrea* por acompañarme durante todo este proceso. Por escucharme, ayudarme y aconsejarme en cada paso. Por no dejarme rendir a pesar de todas las dificultades. Ambos han logado convertirse en una parte fundamental de mi diario vivir, y agradezco mucho a la vida por ponerlos en mi camino. Estuvieron para mí en cada momento de alegría, de tristeza y de frustración, y por eso estaré eternamente agradecida.

Por último, quisiera agradecer al comité de posgrados de la Universidad, y a los jurados, Dr. *Sebastián Muñoz* y Dra. *Jenny Chaparro* por leer, escuchar y aportar a esta tesis.

Resumen

Las garrapatas, reconocidas como vectores de gran relevancia a nivel mundial, se distinguen por su capacidad de transmitir una amplia gama de patógenos a diferentes especies de vertebrados. Entre estas garrapatas, se destaca *Rhipicephalus sanguineus* sensu lato (s.l.), conocida como la garrapata marrón de perro, que se considera de gran importancia tanto para la salud animal como humana. *R. sanguineus* s.l es una especie monotrópica, endófila y de tres hospederos, siendo notable por su amplia distribución global. Esta garrapata tiene la capacidad de transmitir diversos patógenos, ya sea como vectores mecánicos o biológicos, incluyendo bacterias, protozoos, hongos, nematodos y virus. Sin embargo, no son los únicos microorganismos significativos que conforman la microbiota de las garrapatas. En total, se pueden identificar tres comunidades ecológicas, tanto en el interior como en el exterior de *R. sanguineus* s.l. De esta manera, la microbiota de la garrapata está compuesta por patógenos transmitidos por ellas (causantes de enfermedades en humanos y animales), comensales y endosimbiontes. Estos últimos pueden aportar diversos beneficios a la garrapata, lo que incide en su aptitud biológica y, por lo tanto, en su abundancia.

Para identificar estos microorganismos, en particular aquellos patógenos transmitidos por garrapatas, se han empleado múltiples metodologías. Esto incluye métodos tradicionales como el análisis microscópico, el cultivo y la amplificación de genes, junto con la secuenciación Sanger. No obstante, estos métodos tienen limitaciones, como su falta de sensibilidad y especificidad para un solo patógeno. Recientemente, la metagenómica mediante el enfoque de shotgun metagenomics ha ampliado estas técnicas, lo que permite identificar múltiples patógenos al mismo tiempo y también detectar otras comunidades ecológicas relevantes, como los endosimbiontes. Esta técnica se ha utilizado ampliamente para caracterizar el viroma de garrapatas, incluyendo *R. sanguineus* s.l, utilizando su ARN. A pesar de las ventajas que ofrece el shotgun metagenomics, este no ha sido aplicado en *Rhipicephalus sanguineus* s.l en Colombia, a pesar de su importancia como vector de enfermedades para la salud humana y animal en todo el mundo. Esta garrapata provoca impactos directos e indirectos en varios hospederos, incluyendo a algunos de gran relevancia económica, como el ganado, donde se destaca la transmisión de una diversidad de patógenos.

Esta situación se agrava en Colombia, donde la mayoría de los patógenos identificados se han detectado mediante la técnica de reacción en cadena de la polimerasa (PCR), lo que limita la amplificación a un número reducido de patógenos y no permite conocer el amplio espectro de patógenos que *R. sanguineus* s.l puede albergar. Esto conlleva a una falta de comprensión de los efectos que esta garrapata puede estar generando en animales domésticos y seres humanos en el país, especialmente en lo que respecta a posibles enfermedades de relevancia médica y veterinaria que podrían estar transmitiéndose. Esto incluye enfermedades como la fiebre manchada de las Montañas Rocosas, causada por *R. rickettsia*, la anaplasmosis granulocítica humana, causada por *A. phagocytophilum*, la enfermedad de Lyme producida

por *B. burgdorferi*, y la babesiosis canina, principalmente causada por *B. vogeli* en Colombia. Debido a la falta general de conocimiento sobre esta garrapata en Colombia, así como su impacto en la transmisión de enfermedades y su diversidad de hospederos potenciales, surge la pregunta central: ¿Qué comunidades de microorganismos se pueden identificar en *R. sanguineus* s.l?. La realización de esta investigación nos permitió detectar organismos patógenos que circulan en el país a través de esta garrapata, contribuir al conocimiento general sobre los endosimbiontes y, sobre todo, avanzar en la comprensión del microbioma de *R. sanguineus* s.l.

Considerando lo mencionado anteriormente, el objetivo general de este estudio fue analizar el metagenoma del linaje tropical de la garrapata *Rhipicephalus sanguineus* sensu lato (s.l.) en las regiones de Boyacá y Casanare en Colombia. Este objetivo se desglosa en dos objetivos específicos:

1. Describir la composición de las comunidades microbianas de ADN presentes en *R. sanguineus* s.l. mediante un enfoque metagenómico.
2. Analizar la composición de los virus de ARN presentes en *R. sanguineus* s.l. a través de la evaluación del viroma utilizando secuenciación con la tecnología de Oxford Nanopore.

Cada uno de estos objetivos específicos corresponde a un capítulo independiente dentro de la tesis.

En el primer capítulo de la tesis, se llevó a cabo una extracción de ADN de las garrapatas, que posteriormente se sometió a secuenciación por Novoseq 6000 de Illumina, con una capacidad de 6 gigabytes por muestra. Para el análisis bioinformático, se emplearon dos aproximaciones: una asignación directa de las secuencias (reads) y la generación de ensamblajes metagenómicos (MAGs). A través de estas técnicas, se identificaron microorganismos, incluyendo endosimbiontes y patógenos. Entre los microorganismos identificados, los tres más abundantes en términos relativos fueron *Anaplasma phagocytophilum, Francisella tularensis* y *Theileria equi*. Además, se logró la identificación de endosimbiontes pertenecientes a los generos *Coxiella*, *Rickettsia* y *Wolbachia*. En particular, se logró ensamblar MAGs para la especie *Coxiella mudrowiae*, lo que permitió realizar análisis de genómica comparativa y funcional. Además, se exploraron las correlaciones entre los diversos microorganismos identificados, destacando que *C. mudrowiae* presentaba correlaciones negativas con otros endosimbiontes y patógenos.

En el segundo capítulo de la tesis, se realizó la extracción de ARN de las muestras, seguida de un proceso de eliminación de ARN de hospederos mediante el tratamiento RiboZero y un enriquecimiento viral mediante SMART9N. La secuenciación se llevó a cabo utilizando el dispositivo MinION de Oxford Nanopore. Se siguió un enfoque similar al del primer capítulo, asignando directamente las secuencias y generando MAGs para el conjunto de muestras, manteniendo restricciones geográficas y de sexos. Como resultado de esta investigación, se identificaron seis virus diferentes que conforman el viroma de esta

especie de garrapata. Estos virus incluyen una especie similar al virus Flavi asociado a *Rhipicephalus*, el virus Mogiana de la garrapata, un virus de la familia Iflaviridae, el virus Jingmen de la garrapata, Bole tick virus 4 y Mivirus sp. Para dos de estos virus, se logró un ensamblaje exitoso, lo que permitió llevar a cabo análisis filogenéticos y comparativos basados en genomas disponibles de diferentes regiones del mundo.

Abstract

Ticks, globally recognized as vectors of paramount significance, are characterized by their capacity to transmit an extensive array of pathogens to diverse vertebrate species. Among these ticks, *Rhipicephalus sanguineus* sensu lato (s.l.), known as the brown dog tick, holds a position of great importance in both animal and human health. *R. sanguineus* s.l. is a monophyletic, endophilic, three-host tick species, notable for its widespread global distribution. This tick species can act as a vector for various pathogens, serving as mechanical or biological transmitters of bacteria, protozoa, fungi, nematodes, and viruses. However, ticks do not operate in isolation; they harbor a complex microbiota comprising pathogens transmitted by them, commensals, and endosymbionts. These endosymbionts can offer a range of advantages to ticks, influencing their biological fitness and, consequently, their population dynamics.

To identify these microorganisms, particularly those transmitted by ticks, various methodologies have been employed. These encompass conventional techniques such as microscopic analysis, culture, and gene amplification by PCR, combined with Sanger sequencing. However, these methods have limitations, including their inability to provide sensitivity and specificity for individual pathogens. In recent times, metagenomics, specifically the shotgun metagenomics approach, has expanded the scope of these techniques. It enables the simultaneous identification of multiple pathogens and the detection of other ecologically significant communities, such as endosymbionts. Despite the potential advantages of shotgun metagenomics, it has not yet been applied to *Rhipicephalus sanguineus* s.l. in Colombia, despite its critical role as a disease vector for both human and animal health on a global scale. This tick species has direct and indirect impacts on various hosts, including economically significant ones such as livestock, where it transmits a diverse range of pathogens.

In Colombia, the situation is further exacerbated, as the majority of identified pathogens have been detected using the polymerase chain reaction (PCR) technique. PCR is limited in its capacity to amplify only a small number of pathogens, thereby failing to provide insight into the full spectrum of pathogens that *R. sanguineus* s.l. may carry. Consequently, this knowledge gap impedes our understanding of the potential effects this tick species may be having on domestic animals and humans in the country, particularly in relation to medical and veterinary relevant diseases. These diseases include Rocky Mountain spotted fever, caused by *R. rickettsia*, human granulocytic anaplasmosis, caused by *A. phagocytophilum*, Lyme disease caused by *B. burgdorferi*, and canine babesiosis, primarily caused by *B. vogeli*. Given the overall lack of knowledge about this tick species in Colombia, as well as its impact on disease transmission and its wide range of potential hosts, the central question arises: What communities of microorganisms can be identified in *R. sanguineus* s.l.? Conducting this research enabled us to detect circulating pathogenic organisms within the country via this tick and contribute to a better understanding of endosymbionts. Most importantly, it advanced our comprehension of the microbiome

of *R. sanguineus* s.l., laying the foundation for future strategies to control tick-borne diseases based on microbiome modifications.

In light of the above, the overarching objective of this study was to analyze the metagenome of the tropical lineage of the tick *Rhipicephalus sanguineus* sensu lato (s.l.) in the regions of Boyacá and Casanare in Colombia. This objective is further detailed through two specific aims:

1. Describe the composition of DNA microbial communities in *R. sanguineus* s.l. using a metagenomic approach.
2. Analyze the composition of RNA viruses in *R. sanguineus* s.l. through virome assessment using Oxford Nanopore sequencing technology.

Each of these specific objectives corresponds to a distinct chapter within the thesis.

In the first chapter of the thesis, DNA was extracted from ticks, which was subsequently subjected to Illumina's Novoseq 6000 sequencing platform, capable of handling 6 gigabytes per sample. For the bioinformatic analysis, two approaches were used: direct sequence assignment and metagenomic assembly (MAGs). Through these techniques, microorganisms were identified, including endosymbionts and pathogens. Among the microorganisms identified, the three most abundant in relative terms were *Anaplasma phagocytophilum, Francisella tularensis*, and *Theileria equi*. Additionally, endosymbionts belonging to the *Coxiella*, *Rickettsia*, and *Wolbachia* genera were identified. Notably, MAGs were successfully assembled for the species *Coxiella mudrowiae*, enabling comparative and functional genomics analysis. Furthermore, correlations between the various microorganisms identified were explored, with particular emphasis on the negative correlations observed between *C. mudrowiae* and other endosymbionts and pathogens.

In the second chapter of the thesis, RNA was extracted from the samples, followed by the elimination of host RNA using RiboZero treatment and viral enrichment via SMART9N. Sequencing was carried out using Oxford Nanopore's MinION device. A similar approach to the first chapter was followed, including direct sequence assignment and the generation of MAGs for the complete set of samples, all while maintaining geographical and gender-specific restrictions. As a result of this research, six different viruses constituting the virome of this tick species were identified. These viruses include a species closely related to the Rhipicephalus-associated Flavi virus, the tick Mogiana virus, a virus from the Iflaviridae family, the tick Jingmen virus, Bole tick virus 4, Mivirus sp. Successful assembly of two of these viruses facilitated phylogenetic and comparative analyses based on available genomes from various regions around the world.

Artículos

Capítulo I:

Paez-Triana, L., Herrera G., Vega L., García-Corredor D., Orlando M., Paniz-Mondolfi A., Muñoz M., Ramírez JD. Metagenomic Exploration of Endosymbionts and Pathogens in the Tropical Lineage of *Rhipicephalus sanguineus* sensu lato (s.l.) Ticks in Colombia. *Metabarcoding and Metagenomics.* **(ACEPTADO)**

Capítulo II:

Páez-Triana L., Luna N., Cruz-Saavedra L., Ramírez A., Medina J., Castañeda S., Gómez M., Garcia-Corredor DJ, Pulido Medellín M., Patiño LH., Muñoz M., Ramirez JD. Characterizing the Viral Diversity of *Rhipicephalus sanguinenus* s.l in Colombia: Insights into Tick- Borne Viral Ecology. *PLOS Neglected Tropical Diseases*. **(SOMETIDO)**

Capítulo I: Metagenomic Exploration of Endosymbionts and Pathogens in the Tropical Lineage of *Rhipicephalus sanguineus* sensu lato (s.l.) Ticks in Colombia

Metagenomic Exploration of Endosymbionts and Pathogens in the Tropical Lineage of *Rhipicephalus sanguineus* sensu lato (s.l.) Ticks in Colombia

Luisa Paez-Triana[1], Giovanny Herrera[1], Laura Vega[1], Diego García-Corredor[1,2], Martin Orlando Pulido Medellín[2], Alberto Paniz-Mondolfi[3], Marina Muñoz[1], Juan David Ramírez[1,3]*

[1] Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia

[2] Grupo de Investigación en Medicina Veterinaria y Zootecnia GIDIMEVETZ. Universidad Pedagógica y Tecnológica de Colombia.

[3] Molecular Microbiology Laboratory, Department of Pathology, Molecular and Cell-based Medicine, Icahn School of Medicine at Mount Sinai, New York, USA

*Correspondence: juand.ramirez@urosario.edu.co

ABSTRACT

Rhipicephalus sanguineus sensu lato (s.l.), commonly known as the domestic dog tick, is a globally distributed tick. Species that plays a significant role in human and animal health, as well as in economy due to its ability to infest livestock. The use of shotgun metagenomics has proven valuable in identifying tick-borne pathogens and key members of the tick microbiome, including endosymbionts. However, the application of shotgun metagenomics in *R. sanguineus* s.l. ticks in Latin America remains largely unexplored. Therefore, our objective aimed to explore and further analyze the metagenome of the tropical lineage of *R. sanguineus* s.l. ticks in Colombia. Through our analysis, we identified the three most prevalent pathogens harbored by these ticks, namely: *Anaplasma phagocytophilum*, *Francisella tularensis*, and *Theileria equi*. However, the most abundant microorganism detected was not a pathogen but the endosymbiont *Coxiella mudrowiae*. Interestingly, *Coxiella mudrowiae* exhibited significant negative correlations with several pathogens and other endosymbionts. Furthermore, we successfully constructed 27 medium-quality metagenome-assembled genomes (MAGs) for this microorganism, enabling us to conduct a pangenome analysis by comparing them with available genomes and identifying proteins of interest, such as those involved in vitamin B synthesis. This study represents the first implementation of shotgun metagenomics as a methodology to expand our understanding of pathogens and endosymbionts in the circulating tropical lineage of *R. sanguineus* s.l. ticks in Colombia. The findings of this research serve as a foundation for the development of prevention and mitigation strategies against pathogens transmitted by this tick species. Information gained from this study can contribute to the improvement of public health measures and veterinary practices aimed at controlling the impact of tick-borne diseases.

KEYWORDS

Metagenome, microbiome, shotgun-metagenomics, tick, *R. sanguineus* s.l.

INTRODUCTION

Ticks are blood-feeding arthropods that serve as obligate vectors, capable of transmitting a wide range of pathogens, surpassing other types of vectors in terms of diversity (Liu and Bonnet 2014). Among the hard-bodied ticks belonging to the family Ixodidae, several species have significant implications for public and animal health (Venzal et al. 2003). One such species is *Rhipicephalus sanguineus* sensu lato (s.l.), commonly known as the brown dog tick, which has a global distribution (Dantas-Torres 2010). This tick species displays endophilic and monotropic behavior, primarily infesting dogs, but it has also been found infesting wildlife, horses, and humans in rare instances (Dantas-Torres et al. 2013). *R. sanguineus* s.l. holds great importance in both human and animal health, exerting direct and indirect effects on various hosts, including economically significant livestock (Dantas-Torres et al. 2013). Within *R. sanguineus* s.l., five different lineages have been identified using mitochondrial molecular markers, namely *R. sanguineus* s.l. (tropical lineage), R. sp. I, R. sp. II (temperate lineage), R. sp. III, and R. sp. IV (Dantas-Torres et al. 2013). Among these lineages, *R. sanguineus* s.l. and R. sp. II are considered the primary lineages, with *R. sanguineus* s.l. being predominantly distributed in tropical regions, while R. sp. II is commonly found in temperate zones. However, it should be noted that certain countries exhibit a co-circulation of both lineages. Currently, only the circulation of the tropical lineage has been documented in Colombia (Paez-Triana et al. 2021).

It is widely recognized that ticks have the ability to harbor and transmit a diverse range of microorganisms, making them important vectors for diseases (Walker et al. 2000). While ticks can acquire numerous pathogens during their feeding, not all of them are capable of reproducing and being transmitted to new hosts. However, *Rhipicephalus sanguineus* s.l. is known to be a competent transmitter of a wide array of pathogens, including bacteria, protozoa, fungi, nematodes, and viruses (Jia et al. 2020). These tick-borne diseases (TBDs) contribute significantly to the overall burden of vector-borne diseases (VBDs), impacting human and animal health as well as national economies, especially in developing regions (Couper and Swei 2018, Ergunay et al. 2022). Therefore, there is a pressing need for comprehensive surveillance to identify pathogens and assess the potential risks to public health (Woolhouse and Gowtage-Sequeria 2005, Ergunay et al. 2022). However, studying the ecological community of ticks involves more than just identifying pathogens. The tick microbiota also includes commensal and endosymbiotic organisms (Pollet et al. 2020), understanding endosymbionts as those organisms that provide a benefit to the tick. These benefits include enhancing reproductive fitness, facilitating nutritional adaptation, influencing development and reproduction, providing defense against

environmental stressors, and modulating tick immunity (Pollet et al. 2020). Also, an impact of this community on pathogen transmission has been observed. For example, in *Demacentor andersoni*, negative correlations were found between parasites and endosymbionts *like Anaplasma marginale* and *R. belli* and positive between *Francisella* endosymbionts and *F. novicida* (Gall et al. 2016).The mechanisms involve the secretion of compounds that affect pathogen growth, either positively or negatively or the competition between the two organisms for nutrients, leading to reduced pathogen growth, or even the endosymbiont's impact on the tick's immune system results in increased pathogen growth (Gall et al. 2016). Many of these have been identified as endosymbionts based on their genomic characteristics (Gottlieb et al. 2015), the absence of virulence factors indicating their potential non-pathogenic nature, the presence of essential metabolic pathways for tick survival (Buysse and Duron 2021), and even through in-vivo experiments demonstrating the significant benefits of their presence in ticks (Ben-Yosef et al. 2020). Among these endosymbionts are the *Coxiella*-like endosymbionts (CLE), *Wolbachia* endosymbionts (WE), *Rickettsia* endosymbionts (RE), and *Francisella* endosymbionts (FE), among others (Hussain et al. 2022). Despite research on the potential pathogenicity of some of these organisms, it is currently assumed that they do not cause pathogenicity in humans and is limited to specific animals (Hussain et al. 2022). However, their significance in ticks as endosymbionts has indeed been widely recognized, rendering them a community of great importance for study.

Moreover, comprehensive understanding of the tick microbiota is crucial for unraveling the complex interactions within this community and their implications for tick biology and disease transmission. Traditional methods used to identify microorganisms in ticks have limitations in terms of sensitivity, as they only capture a fraction of the tick's microbiota (Wade 2002, Batool et al. 2021). These methods primarily focus on the most common pathogens associated with tick-borne diseases. They involve techniques such as culturing, microscopy, and PCR coupled with Sanger sequencing (Batool et al. 2021). In contrast, metagenomic techniques, particularly shotgun metagenomics, offer a more comprehensive understanding of the tick's microbial composition by identifying all genes and genomes present (Xia et al. 2015). This approach allows for the assessment of microbial diversity, abundance, and the analysis of complete or partial genomes (Xia et al. 2015). By utilizing metagenomics, we can strategically survey tick-borne pathogens, evaluate the associated risks, and gain insights into the mechanisms of disease transmission by these microorganisms. Additionally, this approach enables the identification of other ecologically significant communities, such as endosymbionts. Metagenomics goes beyond taxonomic classification and provides initial insights into the functional roles performed by these microorganisms (Bonnet and Pollet 2021), considering the diverse pathways of interaction between these communities and ticks. Previous studies have effectively employed this methodology. For example, a study conducted in Palestine focused on *Rhipicephalus* spp. and *Haemaphysalis* spp., successfully identified various pathogenic species and endosymbionts. They generated genomes for *Rickettsia massiliae*, Candidatus *Rickettsia barbariae*, and *Coxiella* endosymbionts (Ravi et al. 2019). Similarly, investigations involving *Amblyomma* species have revealed important pathogens such as *Coxiella burnetii* and *Francisella tularemia* (Ergunay et al. 2022). In the case *of R. sanguineus* s.l., several studies have explored its

microbiota using amplicon-based sequencing targeting the 16S rRNA gene. Pollet et al. identified Proteobacteria as the dominant phylum, with the genus *Rickettsia* and the family Coxiellaceae being the most abundant in these ticks (Pollet et al. 2020). Another study by René-Martellet et al. confirmed the prevalence of the genus *Rickettsia* and observed a high abundance of *Coxiella* and *Bacillus*, particularly in the tropical lineage, which exhibited the highest *Coxiella* abundance (René-Martellet et al. 2017). Additionally, Cassia-Luzzi et al. compared the microbiota composition among different lineages of *R. sanguineus* s.l. and between embryos and female adults, revealing a significant abundance of Proteobacteria in all lineages and Actinobacteria and Firmicutes in the tropical lineage (Luzzi et al. 2021).

The study of *R. sanguineus* s.l. ticks in South America, particularly in Colombia, is limited despite its significance in developing and tropical countries. A study by Cotes-Perdomo et al. (2020) conducted PCR analysis and identified *Anaplasma marginale*, *Babesia vogeli*, *Babesia bigemina*, and *Coxiella* spp. (*C. mudrowiae* and endosymbionts) in ticks collected from dogs in the Magdalena department. However, the lack of comprehensive research hinders our understanding of the potential impact of this tick species on domestic animals and humans in the country. It is crucial to consider the potential transmission of medically and veterinary important diseases by *R. sanguineus* s.l., including Rocky Mountain spotted fever, human granulocytic anaplasmosis, Lyme disease, and canine babesiosis. Despite the significant advantages of shotgun metagenomics in identifying a wide range of organisms, this technology has not yet been applied to *R. sanguineus* s.l. ticks in Latin America. However, this approach could provide valuable insights into the interactions and functions of the entire microbiome of these ticks, considering the diverse interaction pathways between these communities and the tick. It has been observed that pathogens can activate mechanisms and manipulate tick responses to facilitate their infection, while ticks can limit pathogen infection, and high bacterial infections can enhance tick survival. Based on the aforementioned information, our objective is to analyze the metagenome of the tropical lineage of *R. sanguineus* s.l. ticks in the Santander and Casanare departments in Colombia. By employing shotgun metagenomics, we aim to enhance our understanding of the microbial composition and potential pathogens associated with these ticks, thereby contributing to the development of effective prevention and control strategies for tick-borne diseases in the region.

METHODS

Sample collection and genetic material extraction for sequencing

A total of 38 adult ticks, collected from dogs, were obtained from the Casanare and Santander departments in Colombia. In the municipality of Yopal, Casanare (5°19’50”N 72°23’26”O), we collected nine females and eleven males. Similarly, in the municipality of Puente Nacional, Santander (5°52’38”N 73°40’43”O), we collected ten males and eight females (Fig. S1, Table S1). The ticks were manually removed from lesions that were clearly caused by ticks using tweezers and preserved in Zymo Shield solution (DNA/RNA Shield, Zymo Research, Cat.: R1100). Morphological identification of the ticks as *R. sanguineus* s.l. (tropical lineage) was done with the assistance of taxonomic keys. For DNA extraction, a

partial fraction of each tick was used, always the half of the tick divided longitudinally. The ticks were washed three times in PBS to remove impurities. Subsequently, the DNeasy Blood & Tissue Kit (Qiagen) was utilized following the manufacturer's instructions, except for extending the incubation time of the disruption buffer by 12 hours at 56°C. The concentration of each extracted DNA sample was evaluated using the Qubit 2.0 Fluorometer and the dsDNA High Sensitivity kit (Thermo Fisher) following the recommended protocol. The quality of the DNA was assessed by electrophoresis on 1.5% agarose gels. To perform the sequencing, the samples were sent to Novogene (Sacramento, CA, USA) and sequenced using the Illumina platform for shotgun metagenomics. The sequencing was carried out on the NovaSeq 6000 platform, employing paired-end 150 bp reads. Each sample generated approximately 6 Gbytes of storage for the raw data per sample. The number of reads is specified in Table S1.

Read-based bioinformatic analysis

A graphical summary for all the methods is in Fgiure S2. The quality assessment of the raw reads was conducted using FastQC v.0.11.9 (Andrews 2010), and MultiQC v.1.6 (Ewels et al. 2016) was employed for summarizing the results. Subsequently, Trimmomatic v.0.38 (Bolger et al. 2014) was utilized to filter the reads based on the following criteria: MINLEN:150, AVGQUAL:20, and TRAILING:20. The "Iluminaclip" parameter was applied to eliminate adapter sequences from the reads using the TruSeq3-PE sequence. On average, <0.00% of reads were removed in this filtering step. The filtered reads, which exhibited high quality, were then aligned to the available genome of *R. sanguineus* s.l (Jia et al. 2020), which can be found in the GenBank with the accession number GCA_013339695.1. This alignment was performed using Bowtie2 v.2.4.4 (Langmead and Salzberg 2012). For taxonomic assignment of the remaining reads, Centrifuge v.1.0.3-beta (Kim et al. 2016) was employed, utilizing the program's default database. The resulting outputs were transformed to Kraken-Report format using the Centrifuge-kreport function. The visual representation of the data was accomplished using Pavian (Breitwieser and Salzberg 2020) and the ggplot2 package in RStudio (Wickham 2008).

Identification of pathogens and endosymbionts

To identify pathogens and endosymbionts, a custom database (Number of genomes (N): 2153, Number of genomes remove (NR): 524) was created by compiling relevant information from the literature. Genomes of *Anaplasma* (N: 50, NR: 2), *Rickettsia* (N: 146, NR: 73), *Francisella* (N: 1075, NR: 161), *Ehrlichia* (N: 44, NR: 2), *Borreleila (Lyme Disease)* (N: 379, NR: 220), *Coxiella* (N: 158, NR: 0), *Babesia* (N: 5, NR: 2), and *Theileria* (N: 4, NR: 3) available in NCBI (RefSeq Annotated) were downloaded (Available genomes until July 2023). The database was then refined by excluding species that are not transmitted by ticks and those that are not known to circulate in the Americas. For the identification of endosymbionts, genomes of the *Wolbachia* (N: 292, NR: 62) were included in the genomes of endosymbionts belonging to the genus *Rickettsia* and *Coxiella*. Only the genera identified as endosymbionts of insects or arachnids were retained. The information of the genomes used are in Table S2.The created database was used with Centrifuge v.1.0.3-beta (Kim et al. 2016) to match against the reads. Centrifuge employed a

minimum length of partial hits (--min-hitlen) of 95 and a k of 1 (k classification parameter). The visual representation was made using the same specifications as mentioned earlier.

Analysis of correlation and abundance

To explore differences in read abundance between sexes and departments, Wilcoxon tests were performed. The significance level was set at 0.05, and the analysis was conducted using RStudio software. This statistical test helped identify any significant variations in read abundance, providing valuable insights into potential differences associated with sex and geographical location. Additionally, to examine the relationships within and between different ecological communities, such as endosymbionts and pathogens, the Spearman's nonparametric rank-order correlation with the Benjamini-Hochberg correction was employed. This statistical analysis was conducted using the "psych" package in RStudio. Significant correlations were identified when the p-value was less than 0.05. Strong correlations were defined as having a correlation coefficient (ρ) below -0.75 or above 0.75. The resulting correlations were then visualized using Cytoscape 3.9.0 (Paul Shannon et al. 1971, Gustavsen et al. 2019). Visualization was facilitated by utilizing R packages such as "igraph," "ggraph," and "Rcy3." These tools aided in creating a network representation of the correlations, allowing for a comprehensive understanding of the relationships among the various ecological communities.

Virulence factors and resistance markers

To identify virulence factors, the Basic Local Alignment Search Tool (BLAST) was utilized. The host-filtered reads, obtained after aligning the reads to the Virulence Factors Database (VFDB), were subjected to BLASTn analysis, considering a minimum of 95% identity and an at least an e-value of 7.04E-07 as a match. This approach allowed for the identification of genetic elements associated with the virulence of microorganisms (Altschul et al. 1990, Liu et al. 2019). In addition, to detect antibiotic resistance markers, the Resistance Gene Identifier (RGI) tool was employed. The clean reads obtained from the sequencing data were used as input for RGI. The CARD (Comprehensive Antibiotic Resistance Database) version 3.1.3, released on 5 July 2021, was employed as the reference database for identifying antibiotic resistance genes (Alcock et al. 2020).

Bioinformatic analysis for the construction of metagenomes (MAGs)

The clean reads were subjected to assembly using MetaSpades v3.15.3 (Bankevich et al. 2012, Nurk et al. 2017). Binning for each sample was performed using MetaBAT (Kang et al. 2019), CONCOCT (Alneberg et al. 2014), and MaxBin (Wu et al. 2016), with subsequent refinement using DAS Tool (Sieber et al. 2018), all employing default parameters. The quality of the assembled genomes was assessed using CheckM v1.1.3 (Parks et al. 2015, Bowers et al. 2017), following established quality parameters for metagenomic genomes (MAGs). Taxonomic assignment of the high-quality bins was carried out using the Genome Taxonomy Database (GTDB-Tk) v1.7.0 (Chaumeil et al. 2020), employing default parameters. Only bins with completeness greater than 50% and contamination less than 10% (medium-

quality MAGs) were obtained. All bins were taxonomically classified as Candidatus *Coxiella mudrowiae* and selected for further analysis. The obtained MAGs were annotated using the PROKKA program (Seemann 2014) using default parameters and visualized using PROKSEE (Stothard et al. 2018). All MAGs taxonomy assignments were double checked with pubMLST (Jolley et al. 2018).

Pangenome and comparative genomics of MAGs

A step of dereplication was perform for all the MAGs, using the program deRep (Olm et al. 2017), The parameter "genomeInfo" was included, containing the quality of the MAGs obtained from CheckM. Additionally, the "ignoreGenomeQuality" command was used due to the average quality of the MAGs. However, to perform the comparative genomics analysis and investigate the genetic diversity and variability of the species, we use the 27 obtained MAGs for *C. mudrowiae* with available ones and conducted a pangenome analysis. The genomes of Ca. *Coxiella mudrowiae* were downloaded from NCBI and combined with the herein recovered, Ca. *Coxiella mudrowiae* RSt (GCF_001077715.1) and Ca. *Coxiella mudrowiae* RSA-CAT (GCF_002804145.1). The Panaroo tool (Tonkin-Hill et al. 2020) was employed to analyze these genomes, considering a 95% identity threshold and their presence in at least 95% of the compared genomes. The parameters used included "--clean-mode strict -a core --aligner mafft --core_threshold 0.95". The output from Panaroo was utilized to visualize phylogenetic trees using iTol (Interactive Tree Of Life) (Letunic and Bork 2019), and Phandango (Hadfield et al. 2018) was used to create graphical representations of the genomes. However, because the analysis conducted using all obtained genomes seemed to produce biased results, we carried out an additional analysis. In this subsequent analysis, we excluded four specific MAGs while following the same protocol as previously described. To determine the relationships between the genomes obtained in this study and other Ca. *Coxiella mudrowiae* genomes, a phylogenetic network was constructed using the Neighbor Joining method in the SplitsTree5 program (Huson and Bryant 2006).

Furthermore, a comparative genomics analysis was conducted. Genomes identified as endosymbionts of ticks from the *Coxiella* genus and *Coxiella burnetii* RSA (GCF_000007765.2) were downloaded from NCBI. Amino acid trees of 40 proteins were constructed using the Uprot program in Taxxo (https://github.com/giraola/taxxo/wiki/Taxxo-wiki), which was executed within RStudio. These 40 proteins are universal phylogenetic markers that are single-copy and have been successfully used in various studies. ANIb (Average Nucleotide Identity using Blast) values were calculated between the complete genomes using the Taxxo tool. Genome pairs with an ANI score higher than 95.0% were considered to belong to the same species (Jain et al. 2018). Lastly, a distance analysis of SNPs was performed between the different genomes using snp-dist v. 2.4.1 and snp-sites v. 2.4.1 (Page et al. 2016), with default settings.

Functional analysis of MAGs

The annotation outputs from PROKKA were utilized to identify Clusters of Orthologous Groups (COG) with eggNOG-mapper v2 (Cantalapiedra et al. 2021), employing default settings. Additionally, the presence of crucial metabolic pathways was verified using KEGG KAS pathways (Kanehisa et al. 2016) with default setting for metagenomes. Specifically, genes associated with the biosynthesis of vitamin B and cofactors were selected and extracted from our MAGs and *Coxiella burnetii* RSA. The Ugene program (Okonechnikov et al. 2012) was employed to calculate the percentage of similarity between these genes. Furthermore, the corresponding genes from other *Coxiella* endosymbiont genomes were included and aligned using the MAFFT program, using default settings. Subsequently, phylogenetic trees were constructed using IQ-TREE (Minh et al. 2020) with an ultrafast bootstrap of 1000 repetitions, The best replacement model was chosen by jModelTest (Posada 2008). Finally, the trees were visualized using iTol (Minh et al. 2020).

Data availability statement

The raw sequences for each sample will be available under the project number PRJEB64029 in the The European Nucleotide Archive (ENA) repository. The BioSampleID is specified in Table S1. MAGs will be available in a GitHub repository (https://github.com/luisa-p-triana/MAGs-Metagenomic-Tick).

RESULTS

Read-based bioinformatic analysis

A total of 18.2 to 38.1 Mb of data per sample was obtained. However, the reads related to the host represented 88.5% of the reads (SD: ±6). After the filter, only a small percentage, of the sequences were successfully classified, with a mean of 19% (SD: ±5). This left an average of 81% of the sequences unclassified in each sample. Among the classified sequences (Tables S3, S4), bacteria accounted for the majority (3% to 24%), followed by archaea (0.07% to 0.09%), and viruses (0.02% to 0.04%). The family Coxiellaceae was found to be predominant in all samples (Fig. 1A), with a *Coxiella* endosymbiont representing a significant proportion ranging from 48% to 95% of the classified reads in bacteria. Apart from this species, the bacterial community in ticks was dominated by genera such as *Bacillus*, *Clostridium*, *Pseudomonas*, *Rickettsia*, *Streptomyces*, *Propionibacterium*, *Burkholderia*, *Fusobacterium*, *Lactobacillus*, and *Campylobacter*, in descending order of abundance. In terms of archaea (Fig. 1B), the dominant families were Methanobacteriaceae, Methanocaldococcaceae, Methanosarcinaceae, Thermoplasmataceae, and Sulfolobaceae. Each family was represented by a corresponding genus: *Methanothermobacter*, *Methanocaldococcus*, *Methanosarcina, Thermoplasma*, and *Sulfolobus*, respectively. As for viruses, they were distributed among the families Baculoviridae, Myoviridae, and Siphoviridae. However, the majority of reads were attributed specifically to the genera *Alphabaculovirus* and *Betabaculovirus*, indicating a clear dominance of the Baculoviridae family. Although other viral

families and genera were identified, the number of reads for each of them was minimal, making it challenging to accurately determine their taxonomic assignment.

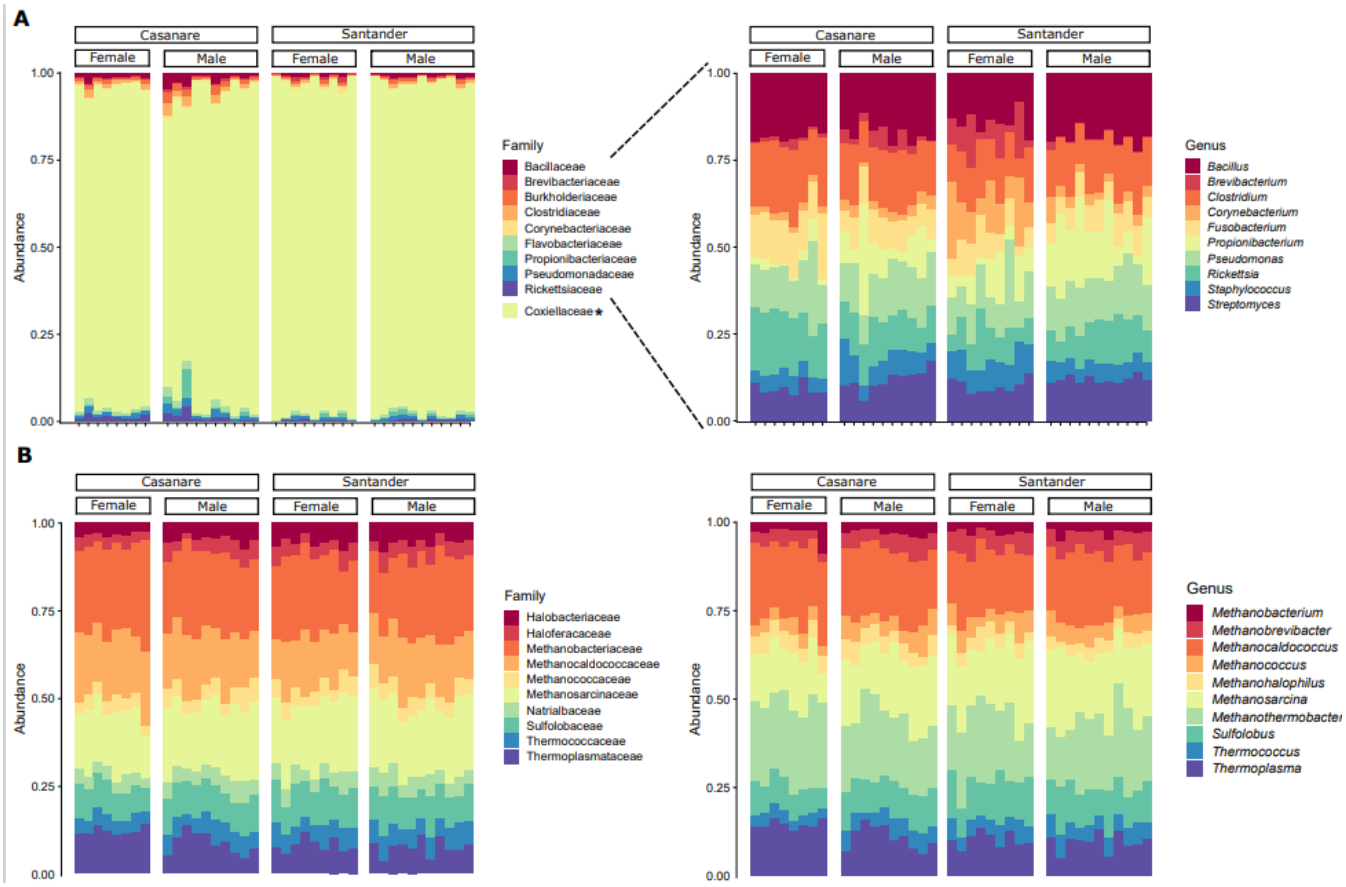

Figure 1. Relative abundance of the top 10 A. Bacteria and B. Archaea classified reads per tick, categorized by department and sexes. Families are indicated on the left side of the corresponding panel, while genera are presented on the right. Coxiellaceae *: For the bacterial genus (right panel), we excluded the reads from the Coxiellaceae family as they represent most of the relative abundance of the genus and were all assigned to Coxiella endosymbionts, which hindered the assessment of the abundance of other genera.

Although the overall composition remained consistent across all departments and sexes, significant differences were found in the most abundant genera (Fig S3). Within the bacteria (Fig. S3A) *Clostridium*, *Fusobacterium*, and *Rickettsia* showed greater abundance in Santander. These same genera exhibited higher abundance in females within Casanare. On the other hand, *Corynebacterium* and *Propionibacterium* demonstrated higher abundance in Casanare. Additionally, along with these, *Staphylococcus* and *Streptomyces* displayed higher abundance in males within Santander. Furthermore, *Propionibacterium* also showed higher abundance in Casanare. Regarding archaea (Fig. S3B), *Methanobrevibacter, Methanococcus, Methanosarcina,* and *Sulfolobus* exhibited higher abundance in

Casanare. Meanwhile, *Methanocaldococcus, Methanohalophilus,* and *Thermolasma* showed greater abundance in Santander. Only *Thermococcus* displayed sex-specific differences, with higher abundance in males.

Identification of pathogens and endosymbionts in the samples

Regarding pathogens (Table S6), the analysis revealed the dominance of three species in the samples (Fig. 2A). *Francisella tularensis* was the most prevalent in all ticks (MEAN: 23%, SD: 8), followed by *Anaplasma phagocytophilum* (MEAN: 7%, SD: 1), and to a lesser extent, *Theileria equi* (MEAN: 7%, SD: 3). Additional pathogens from the genus *Babesia, Ehrlichia Borrellia* and *Coxiella* were also identified on the top 10, albeit with lower abundance ranging from 2% to 0.02% (Fig. 2A). Other pathogens were detected in the samples, but their abundance was significantly lower (Fig. S4A), with percentages below 0.02% and a very limited number of reads per species. On the other hand, among the identified endosymbiotic species (Table S6), *Coxiella mudrowiae* accounted for a substantial proportion of the reads, with a mean of 47% (SD: 18) of the reads assigned with this database. With out *C. mudrowiae*, the top 10 endosymbionts identified were (Fig. 2B): WE of *Atemnus politus*, *Corcyra cephalonica*, *Culex molestus, Bemisia tabaci*, and *Nilaparvata lugens*, CLE of *Dermacentor marginatus, Rhipicephalus microplus*, *Ornithodoros amblus* and *Amblyomma americanum*, and RE of *Ixodes scapularis* (Fig. 2B). These endosymbionts exhibited abundance ranging from 0.02% to 3%. Other endosymbionts were identified but with even lower relative abundances (Fig. S4B), with percentages below 0.02%, and a minimal number of reads per species.

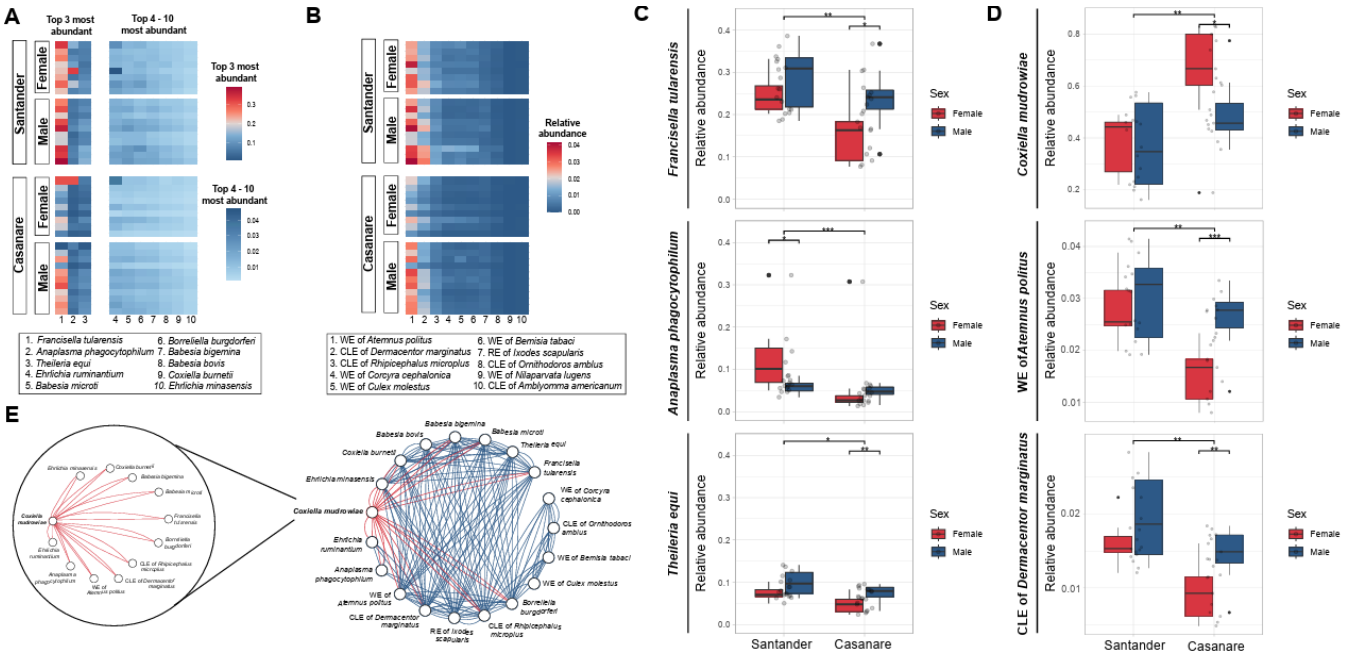

Figure 2. Pathogens and endosymbionts present in each tick. A. Top 10 pathogens. B. Top 10 Endosymbionts (excluding *C. mudrowiae*). C. Differences in relative abundance between departments and sexes for the top 3 pathogens. D. Differences in relative abundance between departments and sexes for the top 3 endosymbionts. E. Networks of positive (blue) and negative (red) correlations among the identified species. The significances are ranked as follows: The significances are ranked as follows: *: <0.05 **: <0.01 ***: <0.001. WE: *Wolbachia* endosymbiont. RE: *Rickettsia* endosymbiont. CLE: *Coxiella* like endosymbiont.

Analysis of correlation and abundance

Although no noticeable differences between sex and location were observed in the heatmaps (Figs. 2A, B), statistical analysis using Wilcox test (Table S7) revealed significant variations in the relative abundances of the top three pathogens with the highest number of reads (Fig. 2C). These differences were characterized by higher abundances in the Santander department, which was also consistent for the majority of the top 10 identified pathogens (Fig. S4C, Table S7). Moreover, significant differences in pathogen abundance were found between sexes within the Casanare department in *F. tularensis* and *T. equi* pathogens, where males consistently exhibited higher pathogen abundance (Fig. 2C). A similar pattern was observed for these two pathogens within the Santander department, although not statistically significant. This sex-based pattern was consistent across most of the species among the top 10 pathogens (Fig. S4C). However, *A. phagocytophilum* exhibits an inverse pattern in the Santander department. In this region, females displayed a significantly higher relative abundance (Fig. 2C). Regarding the endosymbionts, *Coxiella mudrowiae* and most other species showed significant differences between departments. Only *Coxiella mudrowiae* exhibited higher abundance in the Casanare department, while the rest of the endosymbionts were more abundant in Santander (Figs. 2D, S4D).

Furthermore, a significant difference was observed between sexes within the Casanare department (Fig. 2D). *Coxiella mudrowiae* displayed higher abundance in females, whereas the remaining endosymbionts exhibited greater abundance in males (Fig. 2D). Taking into account the differences in abundance, a correlation test was conducted among the top 10 species from each ecological community, revealing strong positive correlations between endosymbionts and pathogens, as well as within the same ecological communities (Fig. 2E). However, negative correlations were found between *Coxiella mudrowiae* and eleven different species (three endosymbionts and eight pathogens).

Virulence factors and resistance markers

The ticks showed a low abundance of reads (>52 reads per sample) related to virulence factors (Table S8). The majority of these reads were associated with the virulence factors Hemin binding protein b (hbpB), 6-phosphogluconate dehydrogenase (gnd), and adhesion and penetration protein (app). These factors are typically associated with *Bartonella quintana*, *Klebsiella pneumoniae* and *Neisseria meningitidis*, respectively. In contrast, a significantly higher number of reads were observed for the resistance markers. Five markers accounted for approximately 80% to 90% of the relative abundance of all the resistance markers. (Table S9). These markers were identified as APH(3'')-Ib, kdpE, aadA24, sul1, and acrS, listed in order of their abundance. Notably, the kdpE marker exhibited a significantly higher abundance (22%) in a male sample from Casanare compared to the remaining samples, where its abundance remained below 10%. These markers are associated with pathogens such as *Escherichia coli*, *Klebsiella pneumoniae*, *Salmonella enterica*.

Bioinformatic analysis for the construction of metagenomes (MAGs)

A total of 27 medium-quality MAGs belonging to the species Ca. *Coxiella mudrowiae* were successfully assembled (Fig. 3A). This assignment was validated using pubMLST (Jolley et al. 2018). The completeness of the assemblies ranged from 87.21% to 89.44%, with contamination ranging from 0.58% to 1.2% (Table S10). Our quality analysis revealed the absence of 20 genes that are typically used by checkM, along with the duplication of one of these genes (Table S1). We carried out a dereplication step. Upon observing the primary clustering dendrogram, despite identifying two separate clusters, both exhibited a MASH ANI identity percentage greater than 90%, thereby consolidating into a single cluster (Figure S5A). This unified cluster was then analyzed in the secondary clustering dendrogram, where the pair-wise distance between all organisms was minimal (Figure S5A). However, considering the distinction among the 4 genomes in the primary dendrogram, we made the decision to conduct independent comparative genomics and pangenome analysis for each of the 27 obtained MAGs, as well as for the available genomes. Pangenome analysis revealed that these four MAGs contained a significant number of genes in the accessory genome, which were absent in the other MAGs and available genomes (Fig. 3B, S5C). As a result, these four MAGs exhibited significantly different genome lengths compared to the others (Fig. S5, Table S10). This led to an imbalance in both gene percentages and quantities between the core genome and the accessory genome, posing challenges in effectively confining the structure of the

genome in our MAGS. As an illustration, one of these longer MAGs encompassed a total of 2907 unique genes in contrast to one normal length MAG and available genomes (Fig. 3B). In comparison, a representative of a Normal length MAGs demonstrated the presence of 252 genes exclusively found within them, distinct from the genomes available in the dataset (Fig. 3B). Notably, these extended-length MAGs predominantly featured hypothetical proteins, with no discernible plasmids identified among their genetic content. To accurately elucidate the composition of both the core genome and accessory genome, an additional pangenome analysis was executed (Fig. 3C), wherein these aforementioned extensive genomes were excluded from the analysis (this procedure was made with the same protocol detailed in the methods section). This procedure revealed a core genome comprising 41% (1089 genes) and an accessory genome accounting for 59% (1574 genes). To assess the coding potential of our MAGs compared to the available genomes of *C. mudrowiae* and *C. burnetii*, we examined the COG (Clusters of Orthologous Groups) sets (Table S11). We observed a similarity in the number of genes across all groups between the available genomes of *C. mudrowiae* and our MAGs, although there were slight reductions in certain groups. Additionally, a decrease in gene numbers compared to *C. burnetii* was observed in groups such as Transcription, Replication, recombination and repair, Cell cycle control, cell division, chromosome partitioning, among others.

The phylogenetic analysis based on the core genome (Fig. 3C) exhibited a clustering pattern where the MAGs from our study formed a distinct group, separated to some extent from the two available genomes. This separation was further supported by the SNP and genome distance analyses (Figs. 4A, B), as the MAGs from our study exhibited greater genetic distance from the available genomes compared to the genetic distance observed among the available genomes themselves. Considering these findings, a comprehensive phylogenetic tree was constructed using 40 proteins, incorporating the available genomes of *Coxiella mudrowiae*, the available *Coxiella* endosymbiont genomes, and our MAGs. This analysis, combined with ANIb calculations on these genomes, revealed a clear clustering pattern between the available genomes of *Coxiella mudrowiae* and our MAGs. Additionally, the ANIb values were over 95%, indicating no species differentiation (Jain et al. 2018) (Fig. 4C). This analysis provided compelling evidence for the distinction between *Coxiella mudrowiae* (represented by our MAGs and available genomes) and the other endosymbionts, as well as *Coxiella burnetii*.

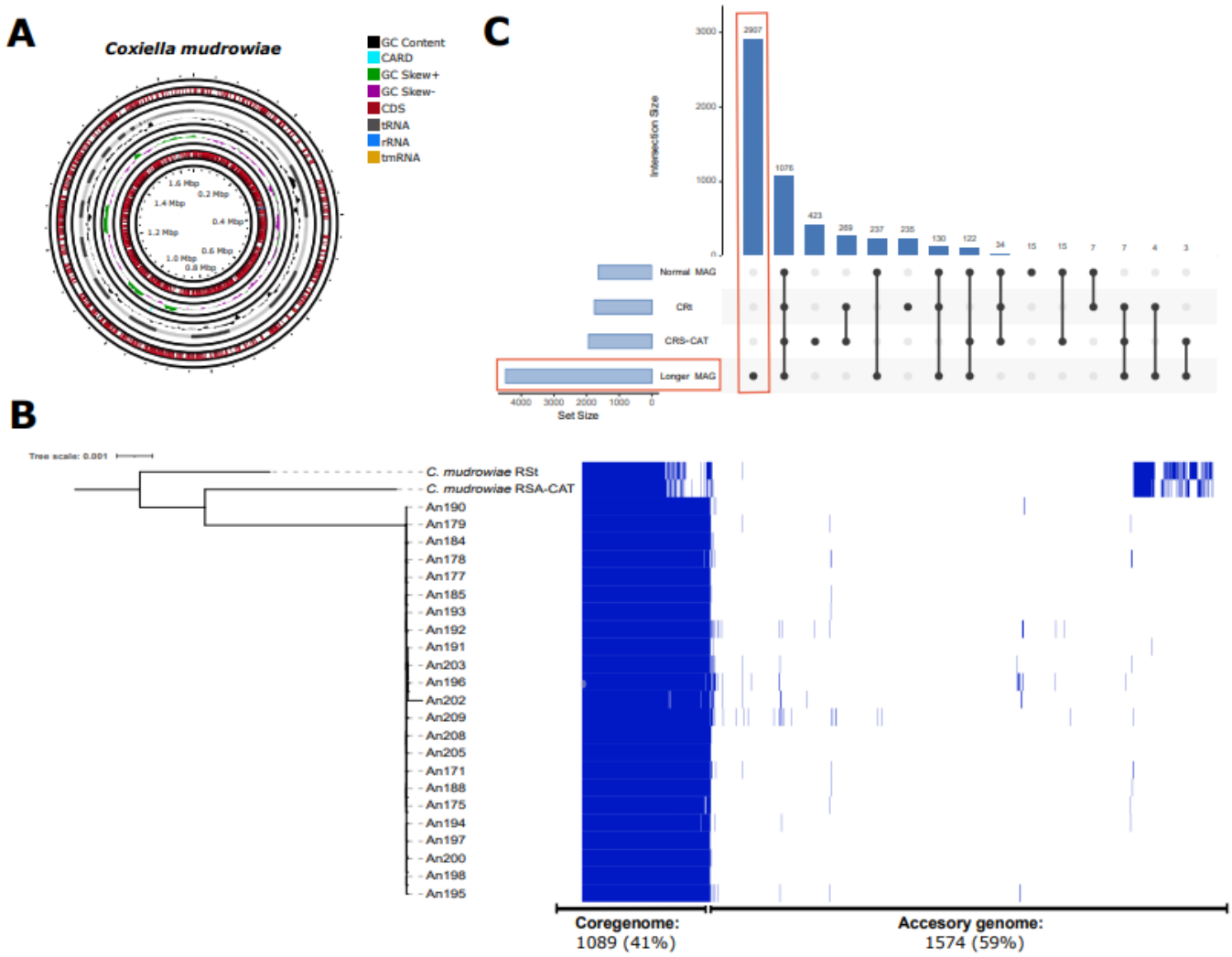

Figure 3. Assembled MAGs belonging to *C. mudrowiae* species. A. Example (Sample An171) of the assembled MAGs and its components. B. Shared genes between a representative of normal length MAG, a representative of longer length MAG, *C. mudrowiae* RSt, and *C. mudrowiae* RSA-CAT. This is a representation utilizing a reference MAG for each of the groups (Longer and Normal length). The red boxes highlight the Longer MAG, its contained gene count, and the substantial number of unique genes associated with it. C. Evolutionary insights between samples (without the Longer MAGs) based on the core genome (left panel) of the MAGs and available genomes, compared with a matrix where the core and accessory genes were either present or absent, which was graphically represented (right panel).

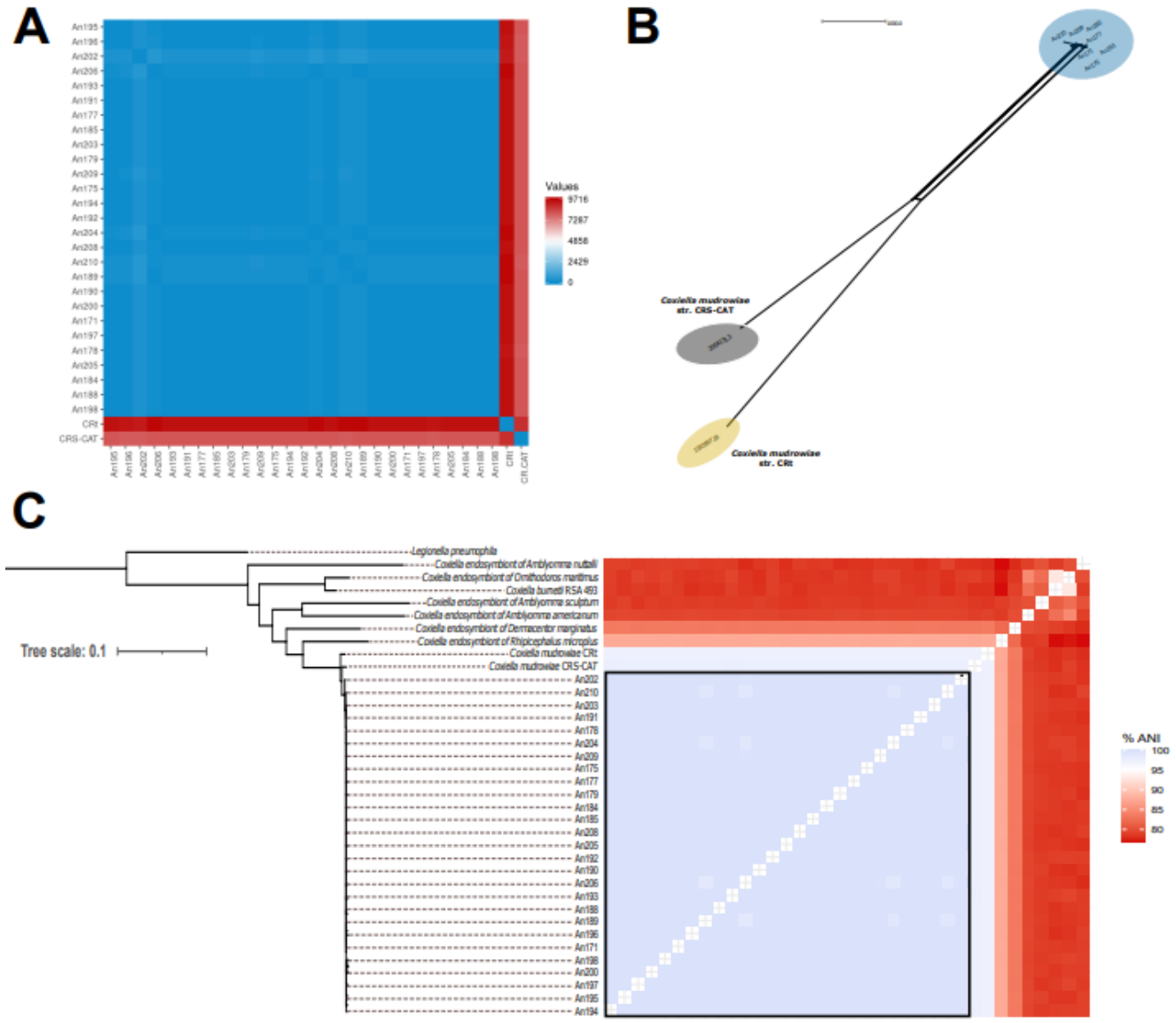

Figure 4. Comparative genomics and phylogenetic analysis of the obtained MAGs in relation to those available for *C. mudrowiae*. A. Genetic distance of SNPs between the available genomes and the obtained MAGs. B. Phylogenetic network representing the relationships among the MAGs and available genomes. C. Phylogenetic tree (bootstrap 1000 replicates, and jModelTest was used to select the best substitution method) based on 40 highly conserved proteins across various CLE, *Coxiella burnetii*, and MAGs (left panel). This phylogenetic tree is complemented by an ANIb analysis (right panel) for the entire genome conducted on CLE, *Coxiella burnetii*, and MAGs, with a threshold set at 95%.

Previous studies have suggested that the putative role of this endosymbiont involves synthesizing essential nutrients that are not readily available in the blood, considering the obligate hematophagy of this vector (Gottlieb et al. 2015, Brenner et al. 2021). Building upon this knowledge and the information

available on the genes responsible for the synthesis of vitamin B and cofactors in the existing genomes of *C. mudrowiae*, we investigated the presence of these genes in our MAGs. Our analysis revealed the presence of complete pathways for the synthesis of Biotin (B7), Riboflavin (B2), and the cofactors CoA and FAD in our MAGs (Fig. 4A). However, we did not identify the genes phoA, panE, and panD, which are involved in the biosynthesis pathways of Pyridoxine (B6) and Folic Acid (B9). These pathways have been previously reported in both *Coxiella* endosymbionts (CLE) and *Coxiella burnetii*. Nevertheless, upon comparing the proteins encoded by these genes, we observed very low similarity between our MAGs and *Coxiella burnetii* (Fig. 5A). This suggests significant non-synonymous mutations in these genes. Phylogenetic trees further support this observation, as they distinctly separate *Coxiella burnetii* from *Coxiella mudrowiae* and most previously reported endosymbionts (Figs. 5B, C).

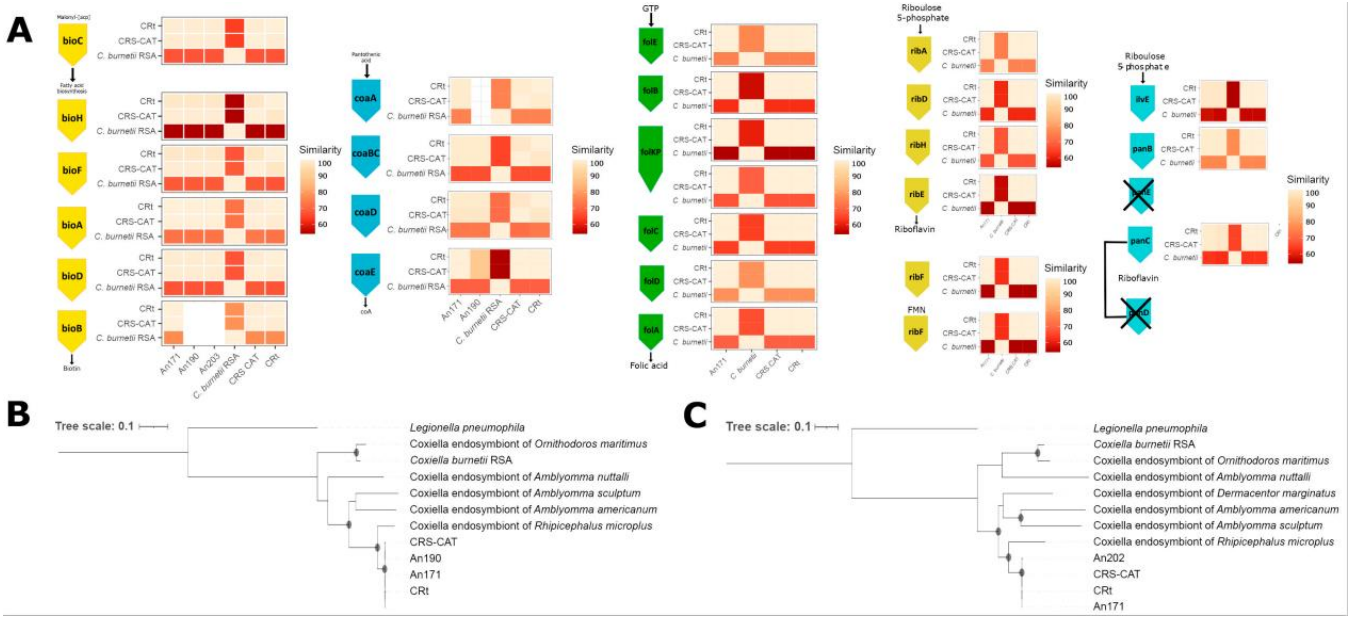

Figure 5. A. Genes involved in A. Vitamin B synthesis and B. cofactors identified in MAGs, *C. mudrowiae* RSt, *C. mudrowiae* RSA-CAT, and *C. burnetii*, along with the identity percentages of the proteins produced by them. B. Phylogenetic tree (Bootstrap 1000 replicates) of BioB protein of all avaible genomes for CLE, *C. burnetti* and MAGs. C. Phylogenetic tree (Bootstrap 1000 replicates) of coaE protein of all available genomes for CLE, *C. burnetti* and MAGs.

DISCUSSION

Ticks are recognized as the second most important vectors of diseases after mosquitoes (Chen et al., 2023). The Center for Disease Control and Prevention (CDC) reports that tick-borne diseases (TBDs) accounted for over 75% of the 650,000 reported vector-borne diseases in the United States between

2004 and 2016 (Rosenberg et al. 2018). The advent of advanced sequencing technologies, such as shotgun metagenomics, has revolutionized disease research by enabling rapid and accurate identification of emerging and re-emerging pathogens (Chen et al. 2023). This approach has also shed light on the diverse ecological communities that exist within ticks, providing a comprehensive understanding of tick microbiomes (Xia et al. 2015). Such comprehensive knowledge is crucial for the prevention and identification of both veterinary and human diseases, offering valuable insights into these disease-carrying vectors. Metagenomics has the potential to drive innovative strategies in vector control and the prevention of TBDs. By leveraging the information obtained through metagenomic studies, we can develop targeted interventions and enhance our ability to identify and combat TBDs effectively. This approach holds great promise for improving the health outcomes of both animals and humans by strengthening our prevention and identification efforts in the field of tick-borne diseases.

Here we employed shotgun metagenomics to characterize the microbiome of 38 *R. sanguineus* s.l. specimens. The predominant bacterial composition in our samples was found to be from the Coxiellaceae family (Fig. 1A), which is consistent with previous studies on the tropical lineage of *R. sanguineus* s.l. ticks worldwide (Portillo et al. 2019; Luzzi et al., 2021; René-Martellet et al., 2017). Notably, research by Luzzi et al. (2021) demonstrated a significant similarity in bacterial composition between embryos and adults, suggesting the potential for transovarial transmission of *Coxiella* spp. within this family. The identification of endosymbionts within the Coxiellaceae family (Fig. 1A) further emphasizes their importance in *R. sanguineus* s.l. ticks, as they have been consistently observed across different continents and studies. In addition to the Coxiellaceae family, our study identified other genera such as *Pseudomonas*, *Rickettsia*, *Propionibacterium*, and *Bacillus*, which have been previously associated with this tick species (Lalzar et al., 2014; René-Martellet et al., 2017). Furthermore, the presence of genera like *Clostridium*, *Streptomyces*, *Burkholderia*, *Fusobacterium*, *Lactobacillus*, and *Campylobacter*, although not previously reported in *R. sanguineus* s.l., has been documented in other tick species (Karim et al., 2017; Papa et al., 2020). It is likely that these bacteria may be acquired from the environment or host, such as the skin or cavities. It is important to note that certain bacteria, such as *Bacillus* and *Rickettsia* species, have been reported to exhibit a higher prevalence in one of the sexes (René-Martellet et al., 2017), our study found significant differences in various genera between sexes and departments. Including *Rickettsia*, where females displayed higher abundance, this observation aligns with the findings of René-Martellet et al. Despite these differences, the fundamental composition of the top 10 genera remains consistent across departments and sexes.

The presence of archaea and viruses in the samples was minimal in terms of abundance, but a diverse range of genera was identified within these groups (Fig. 1B). Archaea in ticks have received limited attention and research due to the scarcity of available data. The genus *Methanobacterium* has been previously detected in other tick species like *Ixodes granulatus* (Chen et al., 2023). As far as we know our study provides the first identification of archaea in this tick species using metagenomics. However, it is important to note that their role in ticks remains unclear and further investigation is necessary to

understand their potential interactions and contribution to the tick's bacterial and eukaryotic communities. With this finding it seems that the methane can play a role in the tick. However, a more comprehensive approach is required to explore the involvement of archaea in tick physiology and health and unravel their significance in tick interactions. In addition, this study identified the presence of the Baculoviridae family, which consists of arthropod viruses (Herniou & Jehle, 2007). Specifically, the genera *Alphabaculovirus* and *Betabaculovirus*, known as lepidopteran-specific nucleopolyhedroviruses (NPVs) (Herniou & Jehle, 2007), were found to be dominant among the DNA viruses detected in the samples. The presence of these viruses suggests their potential involvement in tick biology and their interaction with the tick's natural environment. Further research is needed to elucidate the specific roles and implications.

To identify the pathogens and endosymbionts present in the samples and explore their interactions, a database was created using available genomes. As a result, three major pathogens were detected in the samples: *A. phagocytophilum*, which causes human granulocytic anaplasmosis (HGA) and anaplasmosis in animals (Karshima et al. 2022); *F. tularensis*, the causative agent of tularemia (Telford and Goethert 2020); and *T. equi*, the causative agent of equine piroplasmosis (PE) (Wise et al. 2013). The vector competence of *R. sanguineus* s.l for these pathogens is currently unknown. However, previous studies have reported the detection of pathogen DNA in *R. sanguineus* s.l specimens worldwide (Ghafar and Amer 2012, Zhang et al. 2012, Lopes de Carvalho et al. 2016, Cabrera et al. 2022, Rocafort-FRocafort-Ferrer et al. 2022). Moreover, research conducted in Colombia has also identified these pathogens in both hosts and ticks, suggesting their likely circulation within the country (Máttar and Parra 2006, Molina-Guzmán et al. 2019, Santodomingo et al. 2019, Bonilla-Aldana et al. 2022, Cabrera et al. 2022). Considering the significance of these pathogens in terms of animal and human health, including the potential for humans to act as accidental hosts for these diseases (Cabrera et al. 2022), and the potential economic impact on a developing nation (Santodomingo et al. 2019), it is crucial to further investigate the vector competence of *R. sanguineus* s.l for these pathogens. However, it is important to consider in the case of *F. tularensis* that, as of now, genomes of FE have not been identified in RefSeq. Moreover, it has been observed that metagenomic studies can lead to confusion in distinguishing endosymbionts from their pathogenic species (Buysse and Duron 2021). Therefore, future research should be directed towards confirming the presence of *F. tularensis* or FE in these ticks. Even so, long-term surveillance studies focusing on these three tick-borne diseases should be implemented in Colombia. It is important to note that there is a hypothesis suggesting that the reported cases of TBDs in humans within the country may be underestimated due to underreporting (Máttar and Parra 2006). Additionally, considering the tendency of this tick species to feed on domestic animals such as dogs, cattle, and horses, there is a significant likelihood of human-animal contact in Colombia, highlighting the need for further research in this area.

On the other hand, a significant abundance of reads corresponded to various endosymbionts (Fig. 2B). It is well-known that ticks harbor a diverse microbiome primarily composed of non-pathogenic

organisms, and studying their ecology and evolutionary dynamics is crucial. In particular, a substantial number of reads aligned with *Coxiella* endosymbionts. However, some reads were also classified as *Wolbachia* endosymbionts (WE). Recent studies have reported the presence of *Wolbachia* in *R. sanguineus* s.l. ticks (Chao et al. 2021). Nevertheless, it is essential to identify and further investigate these endosymbionts due to the potential implications of *Wolbachia* in controlling vector-borne diseases in other arthropods (Chao et al. 2021). Therefore, future research on this tick species should expand, encompassing genome sequencing, characterization of their role as endosymbionts, and even exploring their potential experimental use in vector control. Regarding the identified *Coxiella* spp., a clear dominance *of Coxiella mudrowiae* (Fig. 2D) was observed, a candidate species previously identified and described as endosymbiont. *C. mudrowiae* has been detected in various geographical locations and different tick species (Ravi et al. 2019, Rahal et al. 2020). This candidate was recognized as an ongoing endosymbiont based on several genomic features that are characteristic of endosymbionts. These include a notable reduction in genome size when compared to *C. burnetii*, a pronounced rate of pseudogenization, and the pseudogenization of characteristic virulence genes such as the Dot/Icm Type IVB secretion system (T4BSS), a crucial system for the pathogenicity of *C. burnetii* (Gottlieb et al. 2015, Tsementzi et al. 2018, Brenner et al. 2021). However, imaging data or clinical evidence to verify whether this organism could potentially become a pathobiont were not found. Consequently, further development of this research area is imperative, given its widespread prevalence worldwide. Nevertheless, since all current evidence suggests its role as an endosymbiont and no pathogenic traits have been identified, it will be classified as an endosymbiont throughout the paper. Additionally, multiple *Coxiella*-like endosymbionts (CLE) were detected within the same tick individuals, indicating that multiple CLEs can establish themselves within the tick microbiota. This phenomenon has been previously discovered through the analysis of the 16S rRNA gene in tick species, including *R. sanguineus* s.l. (Machado-Ferreira et al. 2016). It is plausible that each endosymbiont provides a different function to the tick and coexists or that they compete for colonization. However, we observed a negative correlation between *C. mudrowiae* and the other CLE (Fig. 2E), suggesting that competition for colonization is more likely. Nonetheless, conducting further comprehensive research on both species is crucial to better understand their interactions, whether it be WE-WE, WE-CLE, or CLE-CLE, as well as to comprehend the possible coevolution occurring between different endosymbionts and ticks, and even among the endosymbionts themselves.

It is well-established that endosymbionts engage in ongoing interactions with the transmitted pathogens, which can either facilitate, restrict, or impede their transmission (Bonnet et al. 2017). Examining the correlations, *Coxiella mudrowiae* showed negative associations with multiple pathogens and endosymbionts. This finding was further supported by observing regions with a higher abundance of this CLE, where pathogen read abundance was reduced. Similarly, negative correlations were identified with other endosymbionts. In the Casanare department, where *Coxiella mudrowiae* was more prevalent, the remaining endosymbionts showed lower abundance, and vice versa (Fig. 2C, D). Previous studies have highlighted the importance of investigating these interactions. For example, in *Demacentor*

andersoni ticks, a negative correlation was found between pathogens such as *Anaplasma marginale* and *Rickettsia belli* and endosymbionts, while a positive correlation was observed between *Francisella* endosymbionts and *Francisella novicida* (Gall et al., 2016). Therefore, further exploration of these negative interactions involving *Coxiella mudrowiae*, which is known to be abundant in this and other tick species, is crucial. This investigation should encompass not only next-generation sequencing techniques but also biological studies aimed at validating and enhancing our understanding of these interactions.

Similarly, considering the prominence of this endosymbiont in our samples, we assembled multiple MAGs and compared them to the genomes constructed thus far. Currently, two genomes have been characterized, one derived from *R. sanguineus* s.l and another from *R. turanicus* (Gottlieb et al. 2015, Tsementzi et al. 2018), both originating from Israel. These genomes highlight the functional role and significance of *Coxiella* endosymbionts in ticks, as they are involved in vitamin synthesis, cofactor production, and other metabolic processes (Tsementzi et al. 2018). In our comparison of the MAGs to these reference genomes, we observed the absence of several genes (Fig. 3B, C), which could be attributed to their incomplete nature. However, analysis of the clusters of orthologous groups (COGs) revealed that the quantities were highly similar to those in the available genomes, albeit with a reduction in certain functional groups compared to *C. burnetii*. This reduction has been previously documented and is associated with pseudogenization and gene loss resulting from the obligate endosymbiotic nature of this bacterium (Gottlieb et al. 2015).

Furthermore, the MAGs formed a distinct monophyletic group separate from the assembled genomes. However, when analyzing average nucleotide identity (ANI) percentages and constructing a phylogenetic tree based on 40 informative proteins, the MAGs clustered with known *Coxiella* species and indicated a low degree of diversity within the country. This aligns with previous research that has attributed the limited intrapopulational diversity of *C. mudrowiae* to the restricted number of hosts, as represented by the two *Coxiella* strains derived from different tick species (Gottlieb et al. 2015). These findings are consistent with our observations and the anticipated limited diversity within the country (Paez-Triana et al. 2021). However, the substantial single-nucleotide polymorphism (SNP) differences between our MAGs and the available genomes could be attributed to the considerable geographic distance separating the reported origin countries of these genomes. Therefore, we hypothesize that a distinct strain of *C. mudrowiae* is circulating within Colombia. Such an undertaking would enable comprehensive genomic and phylogenetic analyses, thereby enhancing our understanding of the coevolutionary mechanisms between *C. mudrowiae* and its tick host.

Finally, we conducted an assessment of specific genes that have been previously reported in this endosymbiont (Fig. 4), particularly those involved in vitamin B synthesis (Tsementzi et al. 2018, Brenner et al. 2021). We observed the presence of these genes, consistent with findings from the other two genomes of *C. mudrowiae*. However, we noted the absence of panD and panE, which have been documented in other *Coxiella*-like endosymbionts, such as the CLE of *Rhipicephalus sanguineus* s.l l

(Brenner et al. 2021). Nevertheless, it has been postulated that these missing genes may be substituted by other genes present in the genome of *C. mudrowiae* (Tsementzi et al. 2018). Additionally, previous studies have indicated that these genes are also present in *C. burnetii*, suggesting their essential role in the pathways required by both organisms and their interaction with their respective hosts (human and tick) (Brenner et al. 2021). However, upon analyzing the protein sequences, we discovered substantial variations that resulted in low similarity between the proteins and even divergence in the protein trees (Fig. 4). Despite previous investigations in proteomics highlighting the significance of these genes and proteins in the endosymbiotic association between *C. mudrowiae* and its tick host (Cibichakravarthy et al., 2022), it is crucial to assess how these dissimilarities in the proteins might impact their structure and functionality compared to what *C. burnetii* can produce. Understanding the metabolic aspects of this endosymbiont could potentially contribute to future biotechnological strategies for controlling tick populations and managing tick-borne diseases both within the country and globally. Previous studies have indicated that the absence of *Coxiella*-like endosymbionts in ticks results in delayed post-feeding development, prolonged feeding periods, and reduced fertility and fecundity. Larvae lacking CLE endosymbionts have been observed to exhibit unsuccessful feeding (Ben-Yosef et al. 2020). Overall, further investigations focusing on the genetic and functional characteristics of these genes and proteins are necessary to fully comprehend their roles in the endosymbiotic relationship between *C. mudrowiae* and ticks. Such understanding can inform the development of innovative approaches for tick control and the management of tick-borne diseases.

Despite certain limitations, such as competition for genomic resources between ticks and their pathogens, which can limit the number of assigned reads, and challenges associated with taxonomic assignment due to the scarcity of research on vector microbiomes, particularly in ticks, understanding the microbiome of these vectors is of utmost importance. It not only allows us to identify the pathogens that the tropical lineage of *R. sanguineus* s.l. could transmit or carry to a wide range of animals, including humans due to their close contact with domestic dogs, but it is also essential for the control of diseases transmitted by these ticks. This is particularly crucial considering the increasing emergence of resistance markers in pathogens worldwide. Understanding the dynamics of these microorganisms and their interaction with ticks is essential to effectively address the challenges related to veterinary and animal health. Additionally, it can provide valuable information for implementing more effective prevention and control strategies, thereby minimizing the impact of diseases transmitted by *R. sanguineus* s.l. on public health and the livestock industry. To gain a comprehensive understanding, it is imperative that future investigations focus on exploring the prevalence and circulation of the different pathogens identified in this study and implement active surveillance measures within the country.

Furthermore, conducting in-depth studies on the endosymbiont *C. mudrowiae* is crucial. This involves not only comprehending its behavior within the country through complete genome analysis but also assessing potential strategies that could pave the way for future vector control approaches utilizing this endosymbiont. Moreover, to further enhance our understanding of pathogen transmission in *R.*

sanguineus s.l., it is imperative to expand this study to encompass the diverse lineages of this tick species worldwide. This expansion would allow for the identification of distinct pathogens that may be transmitted by different lineages. Additionally, considering the knowledge generated regarding endosymbionts in these ticks, it becomes important to explore their prevalence and potential interactions within each lineage. By doing so, we can develop comprehensive surveillance strategies that encompass the entire *R. sanguineus* s.l. species complex, thereby enabling a more thorough assessment of pathogen diversity and transmission dynamics. Such an approach would provide valuable insights for the implementation of effective surveillance and control measures targeting this complex species.

CONCLUSION

Our study has yielded a comprehensive understanding of the pathogens and endosymbionts present in *R. sanguineus* s.l ticks in Colombia. We have successfully identified three significant pathogens, namely *A. phagocytophilum*, *F. tularensis*, and *T. equi*, which have implications for both animal and human health. Implications ranging from the occurrence of mild symptoms to the death of both humans and animals. This information is invaluable for guiding tick-borne disease surveillance efforts in the country. Furthermore, we have made notable discoveries regarding the presence of highly relevant endosymbionts, including *Coxiella mudrowiae*, among others, within this tick species. One intriguing finding is the negative correlations observed between the presence of pathogens and endosymbionts. This suggests complex interactions within the tick microbiome that warrant further investigation. In addition, our study represents a milestone as we have successfully assembled multiple MAGs for *C. mudrowiae*, making them the first MAGs of this endosymbiont reported in Colombia. These MAGs have provided valuable insights into the phylogenetic relationships and functional characteristics of *C. mudrowiae* in the country. Overall, our findings contribute significantly to the understanding of pathogen and endosymbiont dynamics in *R. sanguineus* s.l ticks in Colombia. This knowledge serves as a crucial resource for guiding tick-borne disease surveillance efforts and has implications for public health and animal welfare.

FUNDING

This work was funded by Dirección de Investigación e Innovación from Universidad del Rosario

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGMENTS

We thank the personnel who helped in the collection of the specimens in the field.

REFERENCES

Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AL V., Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar G V., McArthur AG (2020) CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Research 48: D517–D525. https://doi.org/10.1093/nar/gkz935

Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. Nature Methods 2014 11:11 11: 1144–1146. https://doi.org/10.1038/nmeth.3103

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455–477. https://doi.org/10.1089/cmb.2012.0021

Batool M, Blazier JC, Rogovska Y V., Wang J, Liu S, Nebogatkin I V., Rogovskyy AS (2021) Metagenomic analysis of individually analyzed ticks from Eastern Europe demonstrates regional and sex-dependent differences in the microbiota of Ixodes ricinus. Ticks and Tick-borne Diseases 12: 101768. https://doi.org/10.1016/j.ttbdis.2021.101768

Ben-Yosef M, Rot A, Mahagna M, Kapri E, Behar A, Gottlieb Y (2020) Coxiella-Like Endosymbiont of Rhipicephalus sanguineus Is Required for Physiological Processes During Ontogeny. Frontiers in Microbiology 11: 1–16. https://doi.org/10.3389/fmicb.2020.00493

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonilla-Aldana DK, Gutiérrez-Grajales EJ, Martínez-Arboleda JP, Reina-Mora MA, Trejos-Mendoza AE, Pérez-Vargas S, Valencia-Mejía L, Marín-Arboleda LF, Osorio-Navia D, Chacón-Peña M, González-Colonia LV, Cardona-Ospina JA, Jiménez-Posada EV, Diaz A, Salazar JC, Sierra M, Muñoz-Lara F, Zambrano LI, Ramírez-Vallejo E, Álvarez JC, Jaramillo-Delgado IL, Pecho-Silva S, Paniz-Mondolfi A, Faccini-Martínez ÁA, Rodríguez-Morales AJ (2022) Seroprevalence canine survey for selected

vector-borne pathogens and its relationship with poverty in metropolitan Pereira, Colombia, 2020. Parasite Epidemiology and Control 17: 1–11. https://doi.org/10.1016/j.parepi.2022.e00249

Bonnet SI, Pollet T (2021) Update on the intricate tango between tick microbiomes and tick-borne pathogens. Parasite Immunology 43. https://doi.org/10.1111/PIM.12813

Bonnet SI, Binetruy F, Hernández-Jarguín AM, Duron O (2017) The tick microbiome: Why non-pathogenic microorganisms matter in tick biology and pathogen transmission. Frontiers in Cellular and Infection Microbiology 7: 1–14. https://doi.org/10.3389/fcimb.2017.00236

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology 35: 725–731. https://doi.org/10.1038/nbt.3893

Breitwieser FP, Salzberg SL (2020) Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics 36: 1303–1304. https://doi.org/10.1093/bioinformatics/btz715

Brenner AE, Muñoz-Leal S, Sachan M, Labruna MB, Raghavan R (2021) Coxiella burnetii and Related Tick Endosymbionts Evolved from Pathogenic Ancestors. Genome Biology and Evolution 13: 1–15. https://doi.org/10.1093/gbe/evab108

Buysse M, Duron O (2021) Evidence that microbes identified as tick-borne pathogens are nutritional endosymbionts. Cell 184: 2259–2260. https://doi.org/10.1016/j.cell.2021.03.053

Cabrera R, Mendoza W, López-Mosquera L, Cano MA, Ortiz N, Campo V, Keynan Y, López L, Rueda ZV, Gutiérrez LA (2022) Tick-Borne-Agents Detection in Patients with Acute Febrile Syndrome and Ticks from Magdalena Medio, Colombia. Pathogens 11. https://doi.org/10.3390/pathogens11101090

Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38: 5825–5829. https://doi.org/10.1093/molbev/msab293

Chao LL, Castillo CT, Shih CM (2021) Molecular detection and genetic identification of Wolbachia endosymbiont in Rhipicephalus sanguineus (Acari: Ixodidae) ticks of Taiwan. Experimental and

Applied Acarology 83: 115–130. https://doi.org/10.1007/s10493-020-00574-3

Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH (2020) GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. Bioinformatics 36: 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Chen B, Lu X, Wang X (2023) High-throughput sequencing to detect the bacterial microbiome in Ixodes granulatus in China. Research Square: 1–15. https://doi.org/https://doi.org/10.21203/rs.3.rs-2433236/v1 License:

Couper L, Swei A (2018) Tick microbiome characterization by next-generation 16S rRNA amplicon sequencing. Journal of Visualized Experiments 2018: 1–10. https://doi.org/10.3791/58239

Dantas-Torres F (2010) Biology and ecology of the brown dog tick, Rhipicephalus sanguineus. Parasites & Vectors 3.

Dantas-Torres F, Latrofa MS, Annoscia G, Giannelli A, Parisi A, Otranto D (2013) Morphological and genetic diversity of Rhipicephalus sanguineus sensu lato from the New and Old Worlds. Parasites and Vectors 6. https://doi.org/10.1186/1756-3305-6-213

Ergunay K, Mutinda M, Bourke B, Justi SA, Caicedo-Quiroga L, Kamau J, Mutura S, Akunda IK, Cook E, Gakuya F, Omondi P, Murray S, Zimmerman D, Linton YM (2022) Metagenomic Investigation of Ticks From Kenyan Wildlife Reveals Diverse Microbial Pathogens and New Country Pathogen Records. Frontiers in Microbiology 13: 1–16. https://doi.org/10.3389/fmicb.2022.932224

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Gall CA, Reif KE, Scoles GA, Mason KL, Mousel M, Noh SM, Brayton KA (2016) The bacterial microbiome of Dermacentor andersoni ticks influences pathogen susceptibility. Nature Publishing Group 10: 1846–1855. https://doi.org/10.1038/ismej.2015.266

Ghafar MW, Amer SA (2012) Prevalence and first molecular characterization of Anaplasma phagocytophilum, the agent of human granulocytic anaplasmosis, in Rhipicephalus sanguineus ticks attached to dogs from Egypt. Journal of Advanced Research 3: 189–194. https://doi.org/10.1016/j.jare.2011.08.002

Gottlieb Y, Lalzar I, Klasson L (2015) Distinctive genome reduction rates revealed by genomic analyses of two Coxiella-like endosymbionts in ticks. Genome Biology and Evolution 7: 1779–1796. https://doi.org/10.1093/gbe/evv108

Gustavsen JA, Pai S, Isserlin R, Demchak B, Pico AR (2019) Rcy3: Network biology using cytoscape from

within r [version 1; peer review: 2 approved]. F1000Research 8: 1–21. https://doi.org/10.12688/f1000research.20887.1

Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR (2018) Phandango: An interactive viewer for bacterial population genomics. Bioinformatics 34: 292–293. https://doi.org/10.1093/bioinformatics/btx610

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254–267. https://doi.org/10.1093/molbev/msj030

Hussain S, Perveen N, Hussain A, Song B, Aziz MU, Zeb J, Li J, George D, Cabezas-Cruz A, Sparagano O (2022) The Symbiotic Continuum Within Ticks: Opportunities for Disease Control. Frontiers in Microbiology 13: 1–17. https://doi.org/10.3389/fmicb.2022.854803

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications 9. https://doi.org/10.1038/s41467-018-07641-9

Jia N, Wang J, Shi W, Du L, Sun Y, Zhan W, Jiang JF, Wang Q, Zhang B, Ji P, Bell-Sakyi L, Cui XM, Yuan TT, Jiang BG, Yang WF, Lam TTY, Chang QC, Ding SJ, Wang XJ, Zhu JG, Ruan XD, Zhao L, Wei J Te, Ye RZ, Que TC, Du CH, Zhou YH, Cheng JX, Dai PF, Guo W Bin, Han XH, Huang EJ, Li LF, Wei W, Gao YC, Liu JZ, Shao HZ, Wang X, Wang CC, Yang TC, Huo QB, Li W, Chen HY, Chen SE, Zhou LG, Ni XB, Tian JH, Sheng Y, Liu T, Pan YS, Xia LY, Li J, Zhao F, Cao WC (2020) Large-Scale Comparative Analyses of Tick Genomes Elucidate Their Genetic Diversity and Vector Capacities. Cell 182: 1328-1340.e13. https://doi.org/10.1016/j.cell.2020.07.023

Jolley KA, Bray JE, Maiden MCJ (2018) Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Research 3. https://doi.org/10.12688/WELLCOMEOPENRES.14826.1

Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. Journal of Molecular Biology 428: 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z (2019) MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 2019: 1–13. https://doi.org/10.7717/peerj.7359

Karshima SN, Ahmed MI, Kogi CA, Iliya PS (2022) Anaplasma phagocytophilum infection rates in questing and host-attached ticks: a global systematic review and meta-analysis. Acta Tropica 228: 106299. https://doi.org/10.1016/j.actatropica.2021.106299

Kim D, Song L, Breitwieser FP, Salzberg SL (2016) Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Research 26: 1721–1729. https://doi.org/10.1101/gr.210641.116

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359. https://doi.org/10.1038/nmeth.1923

Letunic I, Bork P (2019) Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Research 47. https://doi.org/10.1093/nar/gkz239

Liu B, Zheng D, Jin Q, Chen L, Yang J (2019) VFDB 2019: A comparative pathogenomic platform with an interactive web interface. Nucleic Acids Research 47: D687–D692. https://doi.org/10.1093/nar/gky1080

Liu XY, Bonnet SI (2014) Hard Tick Factors Implicated in Pathogen Transmission. PLoS Neglected Tropical Diseases 8: 5. https://doi.org/10.1371/journal.pntd.0002566

Lopes de Carvalho I, Toledo A, Carvalho CL, Barandika JF, Respicio-Kingry LB, Garcia-Amil C, García-Pérez AL, Olmeda AS, Zé-Zé L, Petersen JM, Anda P, Núncio MS, Escudero R (2016) Francisella species in ticks and animals, Iberian Peninsula. Ticks and Tick-borne Diseases 7: 159–165. https://doi.org/10.1016/j.ttbdis.2015.10.009

Luzzi M de C, de Carvalho LAL, Pinheiro DG, Lima-Duarte L, Camargo JV, Kishi LT, Fernandes CC, Machado RZ, Soares JF, André MR, Barros-Battesti DM (2021) Analysis on the prokaryotic microbiome in females and embryonic cell cultures of rhipicephalus sanguineus tropical and temperate lineages from two specific localities in brazil. Revista Brasileira de Parasitologia Veterinaria 30: 1–15. https://doi.org/10.1590/S1984-29612021066

Machado-Ferreira E, Vizzoni VF, Balsemão-Pires E, Moerbeck L, Gazeta GS, Piesman J, Voloch CM, Soares CAG (2016) Coxiella symbionts are widespread into hard ticks. Parasitology Research 115: 4691–4699. https://doi.org/10.1007/s00436-016-5230-z

Máttar S, Parra M (2006) Detection of Antibodies To Anaplasma, Bartonella and Coxiella in Rural Inhabitants of the Caribbean Area of Colombia. Rev.MVZ Córdoba 11: 781–789.

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. Teeling E (Ed.). Molecular Biology and Evolution 37: 1530–1534. https://doi.org/10.1093/molbev/msaa015

Molina-Guzmán LP, Ríos-Tobón S, Cardona-Lopera X, Lopera JA, Ríos-Osorio LA, Gutiérrez-Builes LA (2019) Occupational history of exposure to zoonotic agents in people dedicated to livestock in san

pedro de los milagros, antioquia, Colombia. Revista Facultad de Medicina 67: 399–405. https://doi.org/10.15446/revfacmed.v67n4.72585

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) MetaSPAdes: A new versatile metagenomic assembler. Genome Research 27: 824–834. https://doi.org/10.1101/gr.213959.116

Okonechnikov K, Golosova O, Fursov M (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28: 1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Olm MR, Brown CT, Brooks B, Banfield JF (2017) DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME Journal 11: 2864–2868. https://doi.org/10.1038/ismej.2017.126

Paez-Triana L, Muñoz M, Herrera G, Moreno-Perez D, Taffur-Gomez G, Montenegro D, Patarroyo MA, Paniz-Mondolfi A, Ramirez JD (2021) Genetic diversity and population structure of Rhipicephalus sanguineus s.l across different regions from Colombia. Parasites & Vectors 14: 1–11. https://doi.org/10.1186/S13071-021-04898-W

Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR (2016) SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial genomics 2: e000056. https://doi.org/10.1099/mgen.0.000056

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research: 1043–1055. https://doi.org/10.1101/gr.186072.114.Freely

Paul Shannon 1, Andrew Markiel 1, Owen Ozier, 2 Nitin S. Baliga, 1 Jonathan T. Wang, 2 Daniel Ramage 2, Nada Amin 2, Benno Schwikowski, 1, 5 and Trey Ideker2, 3, 4 5, 山本隆久, 豊田直平, 深瀬吉邦, 大森敏行 (1971) Cytoscape: A Software Environment for Integrated Models. Genome Research 13: 426. https://doi.org/10.1101/gr.1239303.metabolite

Pollet T, Sprong H, Lejal E, Krawczyk AI, Moutailler S, Cosson JF, Vayssier-Taussat M, Estrada-Peña A (2020) The scale affects our view on the identification and distribution of microbial communities in ticks. Parasites and Vectors 13: 1–13. https://doi.org/10.1186/s13071-020-3908-7

Portillo A, Palomar A, Toro M (2019) Exploring the bacteriome in anthropophilic ticks : To investigate the vectors for diagnosis. : 1–16.

Posada D (2008) jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256. https://doi.org/10.1093/molbev/msn083

Rahal M, Medkour H, Diarra AZ, Bitam I, Parola P, Mediannikov O (2020) Molecular identification and

evaluation of Coxiella-like endosymbionts genetic diversity carried by cattle ticks in Algeria. Ticks and Tick-borne Diseases 11: 101493. https://doi.org/10.1016/j.ttbdis.2020.101493

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 Neglected Tropical Diseases 13: 1–19. https://doi.org/10.1371/journal.pntd.0006805

René-Martellet M, Minard G, Massot R, Van Tran V, Valiente Moro C, Chabanne L, Mavingui P (2017) Bacterial microbiota associated with Rhipicephalus sanguineus (s.l.) ticks from France, Senegal and Arizona. Parasites and Vectors 10: 1–10. https://doi.org/10.1186/s13071-017-2352-9

Rocafort-FRocafort-Ferrer G, Leblond A, Joulié A, René-Martellet M, Sandoz A, Poux V, Pradier S, Barry S, Vial L, Legrand L (2022) Molecular assessment of Theileria equi and Babesia caballi prevalence in horses and ticks on horses in southeastern France. Parasitology Research 121: 999–1008. https://doi.org/10.1007/s00436-022-07441-7

Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, Hooks H, Partridge SK, Visser SN, Beard CB, Petersen LR (2018) Vital Signs : Trends in Reported Vectorborne Disease Cases — United States and Territories, 2004–2016 . MMWR. Morbidity and Mortality Weekly Report 67: 496–501. https://doi.org/10.15585/mmwr.mm6717e1

Santodomingo A, Sierra-Orozco K, Cotes-Perdomo A, Castro LR (2019) Molecular detection of Rickettsia spp., Anaplasma platys and Theileria equi in ticks collected from horses in Tayrona National Park, Colombia. Experimental and Applied Acarology 77: 411–423. https://doi.org/10.1007/s10493-019-00354-8

Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30: 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology 3: 836–843. https://doi.org/10.1038/s41564-018-0171-1

Stothard P, Grant JR, Van Domselaar G (2018) Visualizing and comparing circular genomes using the CGView family of tools. Briefings in Bioinformatics 20: 1576–1582. https://doi.org/10.1093/bib/bbx081

Telford SR, Goethert HK (2020) Ecology of francisella tularensis. Annual Review of Entomology 65: 351–372. https://doi.org/10.1146/annurev-ento-011019-025134

Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J (2020) Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biology 21: 1–21. https://doi.org/10.1186/s13059-020-02090-4

Tsementzi D, Castro Gordillo J, Mahagna M, Gottlieb Y, Konstantinidis KT (2018) Comparison of closely related, uncultivated Coxiella tick endosymbiont population genomes reveals clues about the mechanisms of symbiosis. Environmental Microbiology 20: 1751–1764. https://doi.org/10.1111/1462-2920.14104

Venzal JM, Castro O, Cabrera PA, de Souza CG, Guglielmone AA (2003) Las garrapatas de Uruguay: especies, hospedadores, distribución e importancia sanitaria. Veterinaria (Montevideo) 30: 17–28.

Wade W (2002) Unculturable bacteria - The uncharacterized organisms that cause oral infections. Journal of the Royal Society of Medicine 95: 81–83. https://doi.org/10.1258/jrsm.95.2.81

Walker J, Keinrans J, Horak I (2000) The Genus Rhipicephalus (Acari, Ixodidae): A Guide to the Brown Ticks of the World. Cambridge University Press.

Wickham H (2008) Applied Spatial Data Analysis with R Elegant Graphics for Data Analysis: ggplot2. 21–54 pp.

Wise LN, Kappmeyer LS, Mealey RH, Knowles DP (2013) Review of equine piroplasmosis. Journal of Veterinary Internal Medicine 27: 1334–1346. https://doi.org/10.1111/jvim.12168

Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. Emerging Infectious Diseases 11: 1842–1847. https://doi.org/10.3201/eid1112.050997

Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using. Microbiome 2: 4904–4909. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/2049-2618-2-26.

Wu YW, Simmons BA, Singer SW (2016) MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32: 605–607. https://doi.org/10.1093/bioinformatics/btv638

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: 1–16. https://doi.org/10.1371/journal.pone.0121609

Zhang L, Liu H, Xu B, Lu Q, Li L, Chang L, Zhang X, Fan D, Li G, Jin Y, Cui F, Shi Y, Li W, Xu J, Yu XJ (2012) Anaplasma phagocytophilum infection in domestic animals in ten provinces/cities of China. American Journal of Tropical Medicine and Hygiene 87: 185–189.

https://doi.org/10.4269/ajtmh.2012.12-0005

Supplementary figures and tables available in: Tesis-Figuras

Capítulo II: Characterizing the Viral Diversity of *Rhipicephalus sanguineus* s.l in Colombia: Insights into Tick-Borne Viral Ecology

Characterizing the Viral Diversity of *Rhipicephalus sanguineus* sensu lato (s.l) in Colombia: Insights into Tick-Borne Viral Ecology

Luisa Páez-Triana[1], Nicolás Luna[1], Lissa Cruz-Saavedra[1], Angie L Ramírez[1], Julián Esteban Medina[1], Sergio Castañeda[1], Marcela Gómez[1,2], Diego J. Garcia-Corredor[1,3], Martin Orlando Pulido Medellín[3], Luz H. Patiño[1], Marina Muñoz[1], Juan David Ramirez[1, 4]*

[1]Centro de Investigaciones en Microbiología y Biotecnología – UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá – Colombia.

[2]Grupo de Investigación en Ciencias Básicas (NÚCLEO), Facultad de Ciencias e Ingeniería, Universidad de Boyacá, Tunja, Colombia.

[3]Grupo de Investigación en Medicina Veterinaria y Zootecnia, Facultad de Ciencias Agropecuarias, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia.

[4]Molecular Microbiology Laboratory, Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States.

* Corresponding author

Email: juand.ramirez@urosario.edu.co, juan.ramirezgonzalez@mssm.edu

Abstract

Rhipicephalus sanguineus sensu lato (s.l), also known as the brown dog tick, is a significant vector for various pathogens worldwide. One major concern relates to the tick-borne viruses, many of which are RNA viruses. These viruses have attracted growing attention due to their high transmission rates, the expanding distribution of ticks, and the emergence of associated diseases in human populations. Our main goal was to analyze the RNA virus composition in *R. sanguineus* s.l ticks collected from the Santander and Casanare regions in eastern Colombia. To achieve this, we used Oxford Nanopore sequencing technology following a viral enrichment process. We worked with 39 adult ticks collected from dogs and conducted two distinct analyses. One analysis focused on sequencing reads and used Centrifuge for direct assignment, while the other involved assembly using the Genome Detective Tool program. Our investigation revealed the presence of six distinct viruses in the virome of this tick species: *Rhipicephalus*-associated flavi-like virus, Mogiana tick, Iflaviridae sp., Jingmen tick virus, Bole tick virus 4, and Mivirus sp. Notably, we successfully assembled the final two species, allowing us to perform phylogenetic and comparative analyses using genomes from various regions worldwide. It is important to mention that we did not observe significant differences based on gender or geographical location. This observation enabled us to identify potentially pathogenic viruses, some of which might be associated with febrile cases, as well as endosymbionts. The latter group presents promising research opportunities for developing biotechnological tools that could aid future tick vector control strategies.

Additionally, our research provided valuable insights into the primary viruses present in *R. sanguineus* s.l. ticks in two previously unexplored regions within the country.

Author summary

Ticks are recognized as the second most significant disease vectors worldwide, playing a crucial role in transmitting a wide range of pathogens to their hosts. Among these pathogens, viruses have gained increasing attention due to their potential to cause emerging diseases with significant impacts on public health for both humans and animals, as well as the economy. It is vital to conduct research to monitor and understand the virome (the collection of viruses) in these vectors, especially in the various species that affect us. With this perspective in mind, our goal was to assess the primary viruses present in *Rhipicephalus sanguineus* s.l., commonly known as the brown dog tick, using Oxford Nanopore sequencing technologies. This sequencing platform offers long read lengths in less time and at a lower cost. Our comprehensive evaluation goes beyond previously identified organisms, including those with proposed roles as both pathogens and endosymbionts. In doing so, we are broadening the scope of virus research in this tick species to cover two previously unexplored regions within our country.

Introduction

Ticks are blood-feeding arthropods and are regarded as the second most significant disease vectors in public health, trailing behind mosquitoes [1]. Their remarkable ability to transmit a wide array of pathogens [2,3] to various hosts, including mammals, birds, reptiles, and amphibians [4], combined with their potential for future geographical expansion [5] due to climate chance and deforestation and feeding habits, makes them an enduring and substantial public health threat [3,6]. Ticks are broadly categorized into two prominent groups: hard ticks (Ixodidae) and soft ticks (Argasidae) [7]. Among hard ticks, *Rhipicephalus sanguineus* sensu lato (s.l), commonly known as the brown dog tick, holds a prominent position [8]. This tick species boasts one of the most extensive global distributions and is often considered cosmopolitan [8].

R. sanguineus s.l. has gained notoriety for serving as a host for various pathogens [9], establishing itself as a vector for numerous prevalent diseases worldwide. Although its primary host is the domestic dog, documented instances exist of this tick species feeding on a wide range of animals, including cattle, horses, and even humans [10–12]. Consequently, it plays a pivotal role in the realms of animal and human health, as well as the economy [10]. In the context of Colombia, it is noteworthy that Páez-Triana et al. 2021*,* have identified the tropical lineage of *R. sanguineus* s.l. as the only circulating lineage within the country [13]. The other four lineages (R. sp. I, and R. sp. II - template lineage, R. sp. III, R. sp. IV) have not been described in Colombia, despite the co-circulation of lineages in tropical countries like Brazil. Lineage differentiation holds great significance, especially in light of studies that have revealed variations in vector capacity, as exemplified by *Ehrlichia canis* [14].

Ticks transmit a wide range of pathogens, including protozoa, fungi, nematodes, and viruses [9]. Among these pathogens, tick-borne viruses (TBVs) have garnered significant attention. TBVs span across two orders, nine families, and at least 12 genera [15]. Many of these viruses have been identified as causative agents of severe illnesses, some leading to fatalities and triggering multiple epidemics [15,16]. For instance, by 2017, a staggering 160 arboviruses were associated with ticks, with 25% of them linked to diseases affecting humans or animals [17]. Prominent TBVs include pathogens like the Crimean-Congo hemorrhagic fever virus (CCHFV), considered one of the most significant TBVs to date [18], the tick-borne encephalitis virus (TBEV), responsible for some of the most dangerous human infections in Europe and Africa [19], the African swine fever virus (ASFV), causing substantial economic losses due to fatal diseases in pigs [20], and the Nairobi sheep disease virus (NSDV), with the potential to inflict significant losses on economies dependent on small ruminants [21], among others [6]. RNA viruses constitute the majority of TBVs [17], making them a pivotal category to consider when discussing these vectors. In recent times, TBVs have become a primary focus of infectious disease research. This heightened attention is not only due to the geographic expansion of tick vectors driven by climate change [5], but also the unique ecological dynamics associated with these vectors. Vertical and horizontal transmission, coupled with the ability to replicate within various vertebrate hosts, provide these viruses with the capacity to overcome diverse barriers when encountering different hosts [22]. Consequently, TBVs present a formidable challenge in the prevention and control of diseases, given their substantial transmission capabilities [22].

With this in main, in recent years, there has been a noticeable surge in research articles focusing on the viral composition of ticks. These studies, employing advanced next-generation sequencing (NGS) techniques, have greatly contributed to the discovery of novel viruses [6,23,24]. These newly identified viruses have a significant impact on both human and animal health and can also influence interactions within the broader tick microbiome. For instance, entirely new viral families, such as *Chuviridae*, have been found across various tick species (including *Rhipicephalus* spp., *Dermacentor* spp., *Haemaphysalis* spp., *Hyalomma* spp., among others) and geographical regions, emphasizing their significant relevance within the arthropod viral community [6,25–27]. However, due to their recent discovery, ongoing research is necessary to determine their pathogenicity and interactions. Furthermore, investigations into *Rhipicephalus* spp. have revealed a predominance of animal viruses, particularly *Nairovirus,* in some samples, while others have shown a prevalence of phages followed by invertebrate viruses [24]. Within the *Nairoviruses*, belonging to the *Bunyaviridae* family, a novel virus named Nayun tick nairovirus has been identified. Additionally, newly identified viruses have been categorized under the *Anelloviridae* and *Rhabdoviridae* families [24]. Interestingly, sequences closely related to plant viruses, such as *Caulimoviridae*, *Nanoviridae*, *Geminiviridae*, *Virgaviridae*, and *Sobemovirus*, have also been detected [24]. Specifically concerning *R. sanguineus* s.l., multiple viral families have been characterized, including *Phenuiviridae*, *Rhabdoviridae*, *Flaviviridae*, *Chuviridae*, *Peribunyaviridae*, *Totiviridae,* among others [4]. Remarkably, variations in viral composition based on geographic location and differences in alpha diversity related to sex have been documented [4].

In Colombia, a prior study conducted in the Antioquia department delved into tick viromes, with a specific focus on specimens of *R. sanguineus* s.l [28]. This investigation successfully identified viruses falling within the Tymovirales, Amarillovirales, and Mononegavirales orders. Likewise, other studies have explored the viral composition in various tick species, including *R. microplus* [29]. However, there remains a crucial need for comprehensive studies assessing the viral composition of *R. sanguineus* s.l. in different regions of the country. The urgency for such studies is amplified by the limited scope of virome research in developing countries, which increases the likelihood of underreporting diseases transmitted by ticks in Colombia and South America [30]. Hence, it becomes imperative to establish an efficient approach for evaluating viral composition—one that optimizes both time and resources while enhancing our capacity for assessing prevalence and conducting tick-transmitted virus surveillance. Oxford Nanopore sequencing emerges as a practical and efficient alternative for this purpose. This technology's capability to deliver rapid results with longer reads [31,32] provides valuable insights into the primary viral constituents of organisms under investigation. As a result, it enables more consistent surveillance, aiding in the prevention of tick-transmitted diseases and the mitigation of potential future epidemics. By harnessing this technology, our aim is to describe the composition of RNA viruses in *R. sanguineus* s.l. from the Santander and Casanare departments in eastern Colombia through virome evaluation using Oxford Nanopore sequencing.

Methods

Ethical statement

This study was approved by the Ethics committe of the Universidad de Ciencias Aplicadas y Ambientales (U.D.C.A) under act number 001/2018. The authors declare that dog handling and tick collection were conducted according to international principles of animal research ethics. Animals were enrolled in the study after obtaining the owner´s written consent to participate in the study.

Sample collection and genetic material extraction for sequencing

A total of 39 adult ticks were collected from dogs in the Casanare and Santander departments (government and territorial divisions of the country) in eastern Colombia. In Yopal, Casanare (coordinates: 5°19’50”N 72°23’26”W), we gathered nine females and eleven male ticks. Similarly, in Puente Nacional, Santander (coordinates: 5°52’38”N 73°40’43”W), we collected nine females and ten males (S1 Table, S1 Fig). These ticks were carefully removed from lesions attributed to tick infestations using tweezers and preserved in Zymo Shield solution (DNA/RNA Shield, Zymo Research, Cat.: R1100). The morphological identification of ticks as *R. sanguineus* s.l. (tropical lineage) was carried out with taxonomic keys. For RNA extraction, a partial section of each tick was used, consistently taking half of the tick by splitting it longitudinally. The ticks underwent a thorough washing procedure in PBS to eliminate any impurities. RNA extraction was performed using the Quick-RNA Viral kit adding a

disruption step of 1 min with 30 Hz and a dilution step on 30 µl. Subsequently, RNA was quantified using the NanoDrop™ Spectrophotometer, and its integrity was assessed through a 1.5% agarose gel.

RiboZero treatment, viral enrichment, and sequencing

We employ the Ribo-Zero Plus rRNA Depletion kit (Illumina, REF: 20036696, https://www.illumina.com/products/by-type/accessory-products/ribo-zero-plus-rrna-depletion.html) on the RNA samples, following the manufacturer's instructions, to eliminate host related ARN. Subsequently, viral enrichment was carried out using the Rapid-SMART9n methodology, as per Claro et al., 2021 [33], with a minor modification. This enrichment methodology has demonstrated its sensitivity, compatibility with low input, and suitability for long-read sequencing, leading to significant improvements in terms of cost, time, and laboratory complexity [33]. In brief, cDNA synthesis involved 5 µl of RNA, 0.5 µl of RLB-RT9N primer (TTTTTCGTGCGCCGCTTCAACNNNNNNNNN, 2 µM), and 0.5 µl of dNTPs (10 mM) (New England BioLabs, USA), followed by incubation at 65°C for 5 minutes. A mix containing 2 µl of SuperScript IV First-strand Buffer, 0.5 µL of 0.1 M DTT, 0.5 µl RNase OUT, 0.5 µl RLB TSO (GCTAATCATTGCTTTTTCGTGCGCCGCTTCAACATrGrGrG, 2 µM), and 0.5 µL SuperScript IV (Invitrogen, Carlsbad, CA, USA) was then added to the annealed RNA. The mixture underwent an incubation at 42°C for 90 minutes followed by 10 minutes at 70°C to yield the cDNA. Next, the cDNA product was amplified (modified step: separation of ligation barcodes) using 6.25 µl of LongAmp Taq 2X master mix (New England BioLabs, USA), 4.875 µl of NFW, 0.125 µl of RLB primer (TTTTTCGTGCGCCGCTTCA, 20 µM), and 1.25 µl of cDNA, employing the following conditions: 98°C for 45 seconds, 30 cycles of 98°C for 15 seconds, 62°C for 15 seconds, and 65°C for 5 minutes, with a final step at 65°C for 10 minutes. During both the cDNA construction and post-viral enrichment steps, all samples were quantified to ensure the proper execution of the process. This quantification was performed using the Qubit dsDNA High Sensitivity Assay (Life Technologies, USA) on the Qubit 3.0 instrument (Life Technologies, USA), following the manufacturer's instructions.

To sequence the samples, we employed the NEBNext® Ultra™ II End Repair/dA-Tailing Module commercial kit for the end preparation. Subsequently, barcode ligation was carried out using the NEBNext® Ultra™ II Ligation Module kit. Finally, adapters were ligated using the NEBNext® Quick Ligation Module kit in conjunction with the ligation kit from Oxford Nanopore Technologies (SQK-LSK109) with flowcell FLO-MIN106.

Read-based bioinformatic analyses

Basecalling was performed using Guppy v3.1.5 (ONT), with minimum quality score of 7, and low-quality reads were subsequently removed with Porechop v0.2.3_seqan2.1.1. The remaining reads were aligned using the minimap2 software to the available genome of *R. sanguineus* s.l [9], with accession number GCA_013339695.1. Following the removal of host-related reads, an additional filter step was carried out using the SortMeRNA software [34] to eliminate bacterial and archaeal ribosomal reads. To assign reads

to specific viruses, a database was created using sequences from NCBI Virus with tick as the specified host (Number of sequences: 2593, July 2023, Table S2). This database was employed in conjunction with Centrifuge v.1.0.3-beta [35] to match the reads. Centrifuge utilized a minimum length for partial hits (--min-hitlen) of 95 and a *k* classification parameter of 1. The resulting outputs were converted to Kraken-Report format using the Centrifuge-kreport function. Visual representation of the data was achieved using Pavian [36] and the ggplot2 package in RStudio [37]. To ensure that no other viruses of public health importance were missed, unclassified reads were extracted and assigned using a more broad second database. This second database was constructed using all available genome and fragment sequences in NCBI Virus (Number of sequences: 13686, August 2022, Table S3). The Centrifuge program and visual representation followed the same specifications as previously described [35–37]. To investigate potential variations in read abundance associated with sex and geographical location, we employed Wilcoxon tests with a significance level set at 0.05. These statistical tests, conducted using RStudio software, allowed us to identify significant differences in read abundance, offering valuable insights into sex and location-based variations.

Alpha and Beta diversity analyses

We conducted diversity analyses to assess variations between departments and sexes, considering results from both the Tick viruses sequences database and NCBI Virus sequences database. The same procedures were applied in both cases. For alpha diversity evaluation, we calculated Shannon-Wiener (species diversity) and Simpson (species dominance) indices using phyloseq package version 1.40.0[38] and microbiome package of R version 1.18.0 [39]. These calculations were based on Centrifuge results, which were imported into R for analysis. To test for significant differences between the diversity of sexes and departments, the Wilcoxon test was again applied, maintaining a significance level of 0.05. To explore variations in diversity between these categories further, we turned to beta (β) diversity analysis. This involved constructing a Bray-Curtis dissimilarity matrix and performing principal coordinate analysis (PCoA) using relative abundance data. These analyses were executed using the vegan and phyloseq packages within RStudio. The results were then visually represented using the ggplot2 package. Moreover, we conducted similarity analysis and applied a ANOSIM test with 1,000 permutations to assess changes in viral communities. These statistical analyses were carried out using the vegan package version 2.6-2.

Bioinformatic analysis for the reconstruction of metagenomic assembled genomes (MAGs)

Considering the results obtained from the diversity analysis and differential abundance of the species based on reads, we proceeded to concatenate the filtered reads by departments and sexes. This resulted in concatenated FASTQ files that were uploaded to the online software Genome Detective Tool [40]. Utilizing the default virus analysis option with Oxford Nanopore reads, we aimed to assign and assemble the reads to construct metagenomic genomes using long reads. High-quality assemblies with satisfactory identity (at least 70% of identity) and completeness (at least 80% coverture) were selected to verify the

taxonomic assignment. An online BLASTn [41] analysis was conducted with the entire genome to check the correct assignment. To visualize the MAGs, we employed the Proksee program [42], which utilizes the BLAST formatter tool to display genetic identity on a gene-by-gene basis with previously identified genomes from the online BLASTn search. For evaluating phylogenetic relationships, we obtained genomes from the same genus as our MAGs from the NCBI database. We performed alignments using the MAFFT program [43] with default parameters and visualized them using Ugene [44]. A specific segment of the RNA-dependent RNA polymerase (Rdpd) or polyprotein was selected, depending on the genus. This was made to retain lengthy informative sections with minimal Ns in the obtained MAGs. These gene segments were utilized to construct phylogenetic trees employing the IQtree tool [45] following default parameters, with the best substitution model using JModelTest and generated 1000 Bootstrap replicates. Finally, to validate the identity among all genomes, including the MAGs, we conducted a FastANI analysis [46] following the default program settings.

Results

Read-based bioinformatic analyses

For each of the 39 samples, an average of 176,271 reads was obtained (SD: ± 68,077), with an average length per sample of 438.49 base pairs (SD: ± 63.7) and an average quality score of 9.1 (SD: ± 0.08). Nevertheless, reads associated with the host comprised 71% of the total reads (SD: ±20). Subsequently, the filtering of reads corresponding to bacterial ribosomes was performed, accounting for an average of 53% (SD: ± 7) of the reads filtered by the host.

Subsequently, a primary assignment of reads post filtration of host reads was conducted with a specific focus on viruses found in ticks (Fig 1A). It was observed that most of the composition was represented by the viral family *Flaviviridae.* It was followed, to a much lesser extent, by the families *Chuviridae, Iflaviridae, Phenuiviridae, Nairoviridae,* and *Totiviridae* in order of relative abundance. However, upon analyzing the composition at the species level, it was noted that more than 50% of the reads assigned to this family could not be attributed to a specific species. Within this family, only *Rhipicephalus*-associated flavi-like virus, and Mogiana tick virus (Fig 1B) were successfully assigned with a substantial number of reads. The latter two species exhibited the highest abundance at this taxonomic level, followed by species such as Bole tick virus 4, Iflaviridae sp., Jingmen tick virus, and Mivirus sp. Despite the absence of clear distinctions in the heatmaps for either sex or departments (Fig 1B), some significant differences were observed when conducting statistical analyses (Fig 1C). This analysis revealed that differences were observed only in the case of Iflaviridae sp. among different departments, whereas Mogiana tick virus and Jingmen tick virus exhibited variations between sexes within the Santander department (Table S4). Furthermore, when examining viral families, no significant differences were detected through statistical analysis. Finally, the reads that were not previously assigned were analyzed using the general virus database (NCBI virus database) (Fig 1D), revealing the presence of the following viral genera: *Betapartitivirus, Goravirus, Ilarvirus, Okubovirus, and Pomovirus*.

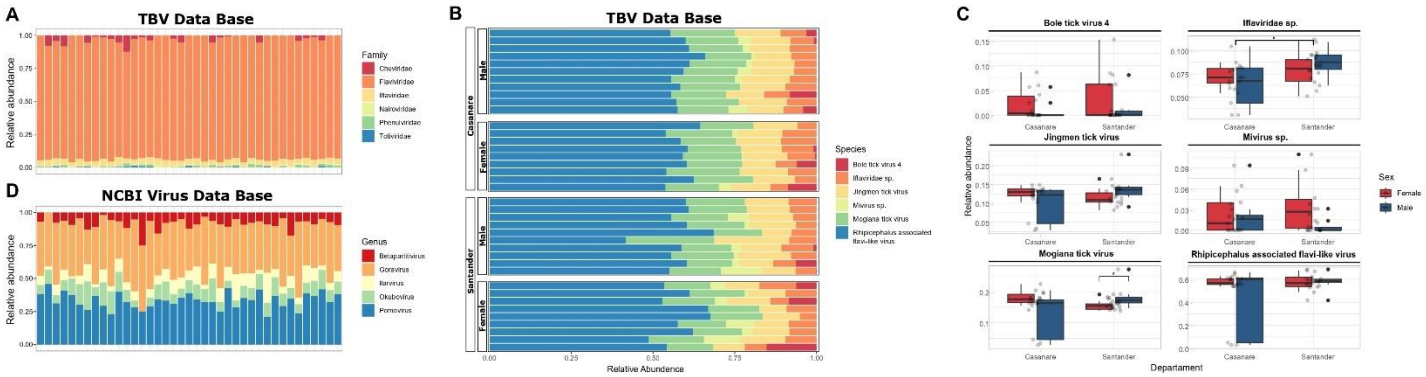

Fig 1. Taxonomic assignment of reads for viruses. A. Viral families identified with the database of sequences of viruses found in ticks. B. Top 5 species identified with the database of virus sequences found in ticks. C. Differences in relative abundance of the previously described viral species. D. Type of viruses identified with the NCBI generic virus database. Significant *p-value*: * <0.05.

Alpha and Beta diversity analyses

Regarding the assessment of diversity variations between departments and sexes, we employed two diversity indices: the Shannon-Wiener and Simpson indices. We applied these indices at both the species level using the tick database (Fig 2) and at the genus level with the virus database from NCBI (Fig S2). Our observations revealed that the diversity, as indicated by the Shannon index, for the viruses obtained from the tick database at the species level (Fig 2A) did not exhibit notably high values within either category. Additionally, there was no significant dominance of any single species, despite the relatively high abundance of the *Rhipicephalus*-associated flavi-like virus (Fig 1B). Furthermore, no significant differences were detected in both diversity indices, whether concerning sexes or departments. Moving on to beta diversity analysis, we observed that there was no distinct grouping that separated the departments (Fig 2B). Instead, the samples exhibited overlapping patterns in a seemingly random distribution in the principal components analysis (R: 0.052, *p-value:* 0.0589). However, when analyzing the sexes, a subtle separation between categories was discernible (Fig 2C). Nevertheless, this separation did not reach statistical significance (R: 0.00868, *p-value*: 0.3136). This finding aligns with the limited significant changes found in the relative abundances of the species (Fig 1C). This pattern remained consistent when using the NCBI virus database (S2 Fig) (Department: R: 0.036, *p-value*: 0.1278; Sex: R: -0.024, p-value: 0.7532), indicating an absence of significant diversity differences and a random distribution of samples for both departments and sexes across all described viruses.

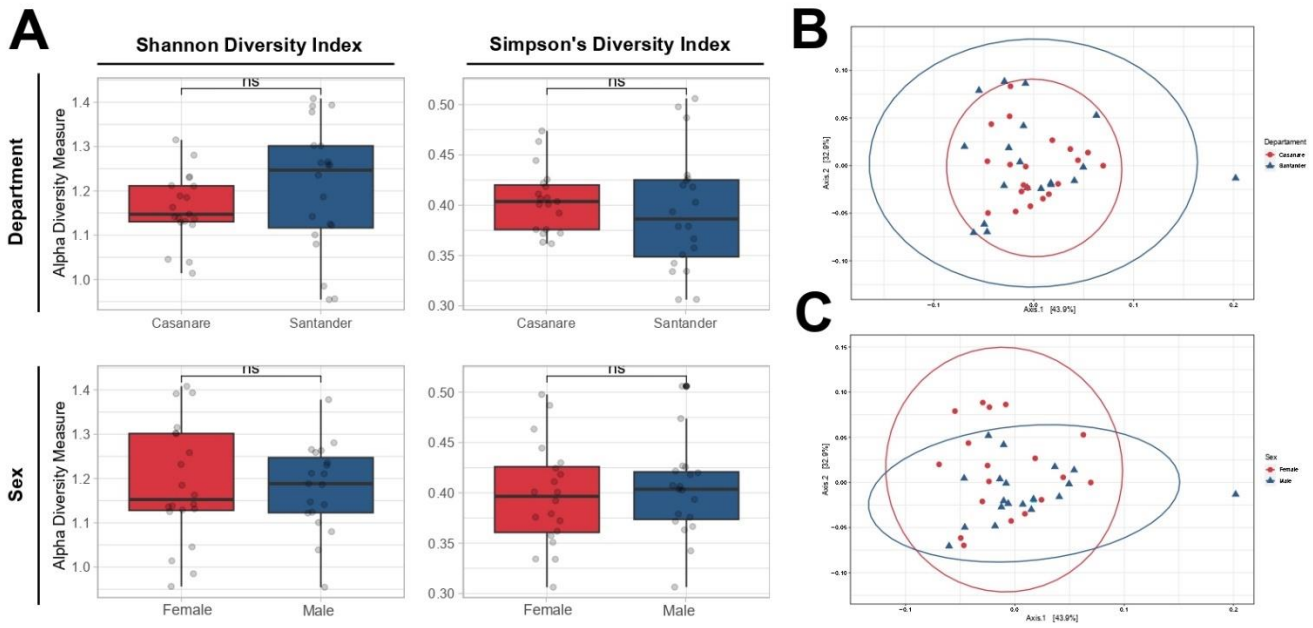

Fig 2. Analysis of diversity in the species identified through the tick virus sequence database. A. Alpha Diversity for sexes and departments. B. Beta Diversity for sexes. C. Beta Diversity for departments. Significant *p-value:* * <0.05.

Bioinformatic analysis for the reconstruction of metagenomic assembled genomes (MAGs)

Considering the previous results regarding viral species diversity and differential abundance, the decision was made to concatenate the reads obtained by department, separating them by sex. Utilizing the Genome Detective Tool, various viral contigs were identified (Table 1). Notably, the species Bole tick virus 4 and Mivirus sp. exhibited the highest coverage and identity percentages, as reported by the Genome Detective Tool. In addition to Bole tick virus 4 and Mivirus sp., contigs were identified using this methodology for other species as well (Table 1), which were shared across all four concatenated datasets analyzed. It is worth noting that Scheffersomyces segobiensis virus L was only identified in the Casanare-Female and Santander-Male samples. Furthermore, despite having low identity and coverage, the presence of tick viruses of importance, such as Pacific coast tick phlebovirus (segment L) and American dog tick phlebovirus (segment L), was observed in all concatenated datasets.

Table 1. Results for Genome Detective Tool mapping of concatenated reads

Department	Sex	Species	Number of Reads	Nt Identity	Genome Coverage
Santander	Female	Bole tick virus 4	139005	69.40%	86.70%
		Mivirus sp.	356	96.30%	99.30%
		Uukuvirus tachengense (segment L)	77	65.10%	40.00%

		Operophtera brumata reovirus (segment 1)	66	59.30%	5.50%
		Uukuvirus lihanense (segment L)	37	61.70%	24.50%
	Male	Mivirus sp.	355	94.70%	94.10%
		Bole tick virus 4	93	75.80%	53.20%
		Pacific coast tick phlebovirus (segment L)	478	57.30%	22.00%
		Scheffersomyces segobiensis virus L	10	56.60%	8.60%
		American dog tick phlebovirus (segment L)	9	65.30%	14.00%
		Uukuvirus tachengense (segment L)	7	64.20%	17.00%
		Uukuvirus lihanense (segment L)	7	68.40%	14.10%
Casanare	Female	Bole tick virus 4	704	78.20%	85.70%
		Mivirus sp.	61	88.90%	69.00%
		Pacific coast tick phlebovirus (segment L)	12	67.40%	19.50%
		Scheffersomyces segobiensis virus L	6927	62.60%	8.00%
		Uukuvirus tachengense (segment L)	4	63.40%	9.60%
	Male	Bole tick virus 4	327	75.70%	89.20%
		Mivirus sp.	324	98.50%	49.70%
		Uukuvirus tachengense (segment L)	26	64.90%	33.50%
		Uukuvirus lihanense (segment L)	21	65.30%	21.50%
		Pacific coast tick phlebovirus (segment L)	16	66.40%	22.00%

Analysis for *Chuviridae* family

To validate the identity of the MAG obtained for the species Mivirus sp., a BLASTn analysis was conducted online for the complete genome, revealing a 96.01% identity with an e-value of 0 to the species Mivirus sp. isolate TTP-Pool-7 (Accession number: NC_076416.1). Additionally, a BLASTn analysis was performed using the same sequence broken down by gene, showing identity percentages ranging from 90.0% to 100% across its genome (Fig 3A).

Subsequently, a phylogenetic tree was constructed for a segment of the Rpdp gene from Chuviridae family sequences, to which the species Mivirus sp. belongs (Fig 3B). A window size of 4,169 base pairs was used for this analysis. It was observed that our MAG closely clustered with the available sequences of Mivirus sp. and exhibited clear distinctions from other species within the same viral family. Within the Mivirus group, the Changping mivirus (MN095545.1) was also grouped. To assess the similarity of this sequence to our MAG and the rest of the available sequences for Mivirus sp., a FastANI analysis was conducted for the entire genome (Fig 3C). This analysis revealed that the Mivirus sp. group shares an identity of over 95% among themselves, with values higher than those found in the rest of the species

within the Chuviridae family. Nonetheless, the Changping virus sequence remains grouped with the Mivirus species, showing high identity with our MAG and even higher identity with the available sequences of Mivirus sp.

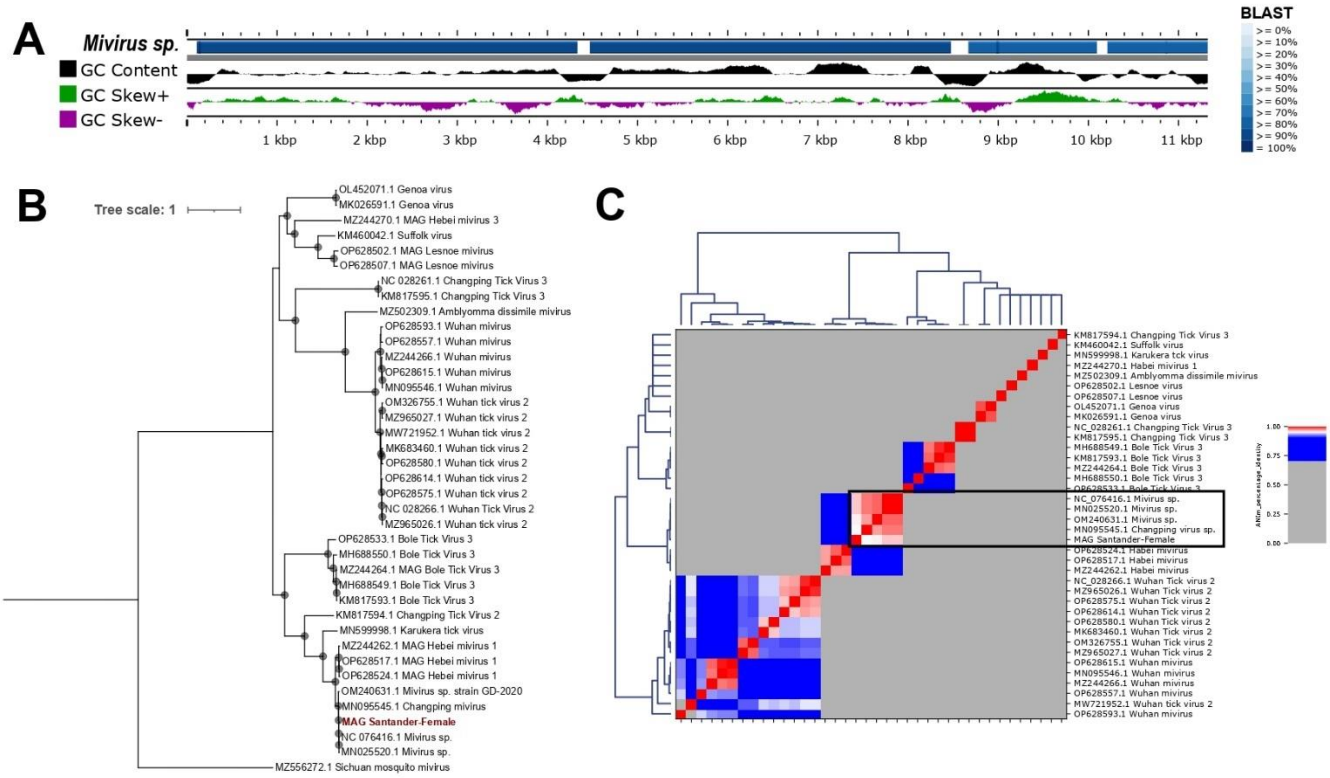

Fig 3. Analysis for the MAGs obtained for Mivirus sp. A. Visualization of the genome with identity percentages by BLAST with available genome. B. Phylogenetic tree of the partial Rpdp gene, showing values greater than 85% for Bootstrap. Tree rooted in the species: Sichuan mosquito mivirus (MZ556272.1). C. Percentage of identity by ANI of the sequences used. Black box shows the grouping of the MAG obtained with the available sequences.

Analysis for *Pestivirus*-like group

To validate the identity of the MAG obtained for the species Bole tick virus 4, a BLASTn analysis was conducted online for the complete genome. However, in this case, two different species exhibited similar values of identity, coverage, and e-value, with one of them not aligning with the results reported by the Genome Detective Tool. The first match was Trinbago virus isolate TTP-Pool-4 (Accession number: MN025505.1) with a percentage of identity of 91.59%, coverage of 87%, and an e-value of 0. The second match corresponded to Bole tick virus 4 isolate TIGMIC_1 (Accession number: ON811701.1) with an identity percentage of 88.35%, coverage of 87%, and an e-value of 0. Taking this into account, a gene-wise BLASTn analysis was performed (Fig 4A), revealing very similar identity values for both species compared to the MAG obtained in this article. Where, Trinbago virus displayed slightly higher and

consistent identity values across the entire genome in comparison to Bole tick virus 4, ranging from 78.2% to 90.3%; in comparation with the Bole Tick Virus 76.3% to 88.1%.

To further this analysis, a phylogenetic tree of the Pesti-like viruses group was constructed with an observation window of 6507 base pairs. In this tree, it was observed that the MAG obtained in the sample closely clustered with the Trinbago species. Both are categorized within the broader group of Bole Tick Virus 4 (Fig 4B), that are phylogenetic apart from other species in this group. Finally, a FastANI analysis was performed (Fig 4C), revealing that for this species, there is no distinct clustering for our MAG with identities greater than 95%. However, low identities were also observed for any of the other available sequences for Bole Tick Virus 4. All genomes, including our MAG, displayed values exceeding 70% identity but lower than 95% among them. All findings demonstrate a broad clustering among the available species identified as Bole tick virus 4, the genome available for Trinbago virus, and our MAG. However, nucleotide identities via FastANI were observed to be low among them, as illustrated in Fig 3C with a large black rectangle.

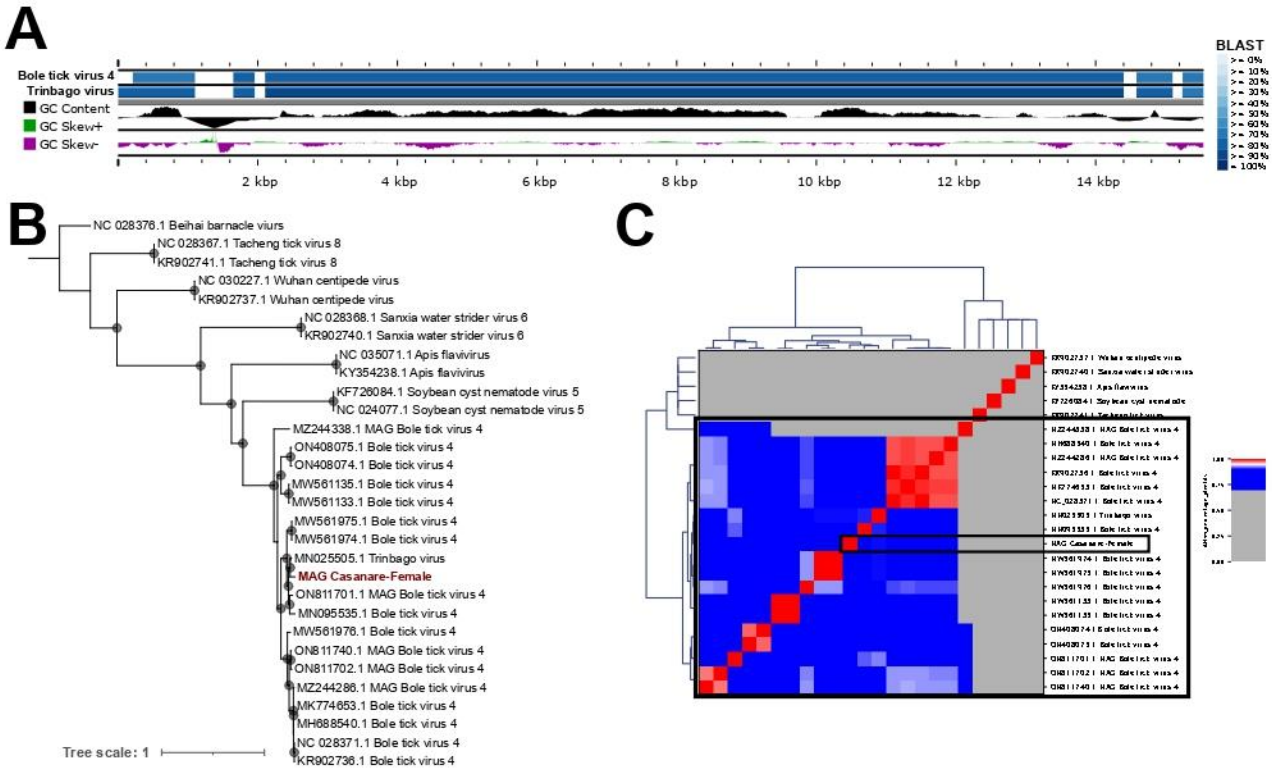

Fig 4. Analysis for the MAGs obtained for Bole Tick Virus species. A. Visualization of the genome with identity percentages by BLAST with the available genome of Trinbago Virus and Bole Tick Virus 4. B. Phylogenetic tree of the partial pliprotein gene, showing values greater than 85% for Bootstrap. Tree rooted in the species: Beihai barnacle viurs (NC 028376.1). C. Percentage of identity by ANI of the sequences used. Black box shows the grouping of the MAG obtained with the available sequences**.**

Discussion

Ticks are well-known vectors with the capacity to transmit a wide range of pathogens to various hosts [47]. The diseases they transmit contribute significantly to the overall burden of vector-borne diseases (VBD), which have substantial economic and public health implications, particularly in developing regions [48,49]. Among these pathogens, various viruses play a crucial role in both human and animal health, affecting domestic and wildlife populations [25]. The global concern regarding tick-transmitted viruses is driven by several complex factors. These include the rising prevalence and geographic expansion of ticks, largely attributed to climate change [4,50], the potential for ticks to act as vectors for emerging and re-emerging viruses [4], and the suspected high efficiency of these vectors in transmitting diseases [22]. This makes tick-borne viruses a formidable group of pathogens, making prevention and control during outbreaks a challenging endeavor. Various techniques have been employed to evaluate tick-borne viruses, including metagenomics using Illumina sequencing or PCR. In our study, we utilized Oxford Nanopore sequencing technology to offer a comprehensive overview of the predominant viruses found in 39 ticks of the *R. sanguineus* s.l. species. This technology provides several advantages, including longer read lengths, quicker analysis times, cost-effectiveness, and portability [31,32]. Consequently, it enables us to contribute valuable information that advances our understanding of tick-borne viruses in Colombia.

In our initial analysis, we performed read assignments, uncovering the presence of six distinct viral families previously documented in this tick species (Fig 1A, B), as well as in other arthropods, encompassing both hematophagous and non-hematophagous varieties. Notably, we identified a significant prevalence of the *Flaviviridae* family (Fig 1A). However, it is essential to highlight that a substantial portion of the reads could not be attributed to specific species, a phenomenon previously observed in virome studies of *R. sanguineus* s.l [6]. While research on tick viromes has gained momentum in recent years, investigations remain limited, particularly in developing nations where the spectrum of circulating diseases and pathogens may differ. Given the well-established role of arthropod-borne viruses (arboviruses) within the *Flaviviridae* family, responsible for various diseases in vertebrates and humans, it becomes imperative to continue exploring new viruses. Further research should focus on elucidating their potential pathogenicity, interactions within the broader microbial community, and their implications for both the tick and its host organisms. Importantly, it is crucial to consider that in other vectors like *Aedes* spp. and *Culex* spp., research has demonstrated that certain insect-specific viruses (ISVs) within the *Flaviviridae* family have the ability to influence vector competence [51]. For instance, the cell fusing agent virus (CFAV) notably inhibits the replication of the Dengue virus [52], while the Nhumirim virus (NHUV) has shown significant reductions in the transmission of Dengue and Zika viruses [53]. Evaluating these interactions becomes paramount, particularly when dealing with tick-borne viruses that have severe public health consequences, such as CCHFV and TBEV.

Regarding the *Flaviviridae* family, we successfully identified four distinct species (Fig 1B), including Jingmen tick virus, Mogiana tick virus, and *Rhipicephalus*-associated flavi-like virus, all belonging to the

Jingmenvirus group [54]. These recently discovered and unclassified viruses have been associated with various insects and ticks, including cases of febrile illness in humans [6,24,55–57]. They have been found in a range of tick species, such as *Rhipicephalus microplus*, *Haemaphysalis longicornis*, *Haemaphysalis campanulata*, *Haemaphysalis flava*, *Rhipicephalus sanguineus*, *Ixodide sinensis*, *and Ixodes granulatus* [54,58]. Both Jingmen tick virus and *Rhipicephalus*-associated flavi-like virus have previously been documented in *R. sanguineus* s.l. within the country [28]. However, this is not the case with Mogiana tick virus, which was initially discovered in the salivary glands of *R. microplus* in 2014 [59]. Subsequently, it has been reported in both cattle and *R. microplus* in several countries, including Brazil, Trinidad and Tobago, China, and Guinea [24,25,56,60]. Nevertheless, it is important to note that all three species exhibit high similarities and closeness in phylogenetic trees, and due to the sequence used here, the genetic differentiation of these three viruses could be a limitation. Nonetheless, it remains crucial for future research to delve further into the presence and differentiation of these three species in *R. sanguineus* s.l.. This is particularly important due to the association with potentially severe illnesses caused by Jingmen tick virus in humans and the significant genetic similarities these viruses may exhibit. This importance is further highlighted when considering that Mogiana tick virus has already been identified in cattle in a neighboring country, Brazil. It is worth noting that cattle have also been reported as hosts for *R. sanguineus* s.l. ticks in Colombia.

Finally, we also identified species belonging to the *Iflaviridae* family (Fig 1A, B). This family has become increasingly common in various metagenomic analyses of ticks, spanning different species of *Ixodida* and *Rhipicephalus* [16,61,62]. *Iflaviridae* has also been detected in other arthropod species, some of which can induce pathogenicity or even mortality [62]. Nonetheless, there is still a need for further investigation into the biological aspects of *Iflaviridae,* particularly its implications for tick fitness and development [62]. There is a need for research to focus on evaluating possibilities of biotechnological tools for controlling ticks and other insects of significant public health importance and expanding our understanding of this field, which is currently quite limited.

Additionally, we conducted an evaluation of the differences in relative abundance between different departments and sexes (Fig 1C). Previous studies conducted in other countries have demonstrated variations in viral composition across geographical locations for *R. sanguineus* s.l ticks and other species [4,22]. However, in our study, the overall composition remained consistent, and no variations were observed between different departments. This aligns with previous hypotheses, considering the limited diversity of this tick species within our country. Additionally, the compositions identified in this investigation closely resembled the results obtained in another department from the same country [28]. Similarly, we did not identify variations in the relative abundance of species between different sexes. While variations in abundance between sexes are well-documented in microbial communities, particularly among bacteria like *Rickettsia* [63], which are of public health significance, the variations in viral composition between sexes have received limited attention. Recent research suggested potential alpha diversity differences between sexes of *Rhipicephalus* spp. [4], but the specific nature and reasons

for these variations remain uncharted territory. In our study, we did not uncover differences in alpha diversity across our samples, nor did we detect distinctive clustering in beta diversity, either by departments or by sex (Fig 2). However, this analysis should be considered for future studies in ticks, including *R. sanguineus* s.l., to thoroughly investigate variations between sexes. It is crucial to consider that behavioral and biological differences between male and female ticks could hold significant implications for TBVs.

For the assembly analysis, we made the decision to concatenate samples by departments, while maintaining the separation of sexes within each geographic location. This approach allowed us to achieve a higher number of reads per sample (Table 1), consequently yielding MAGs with enhanced completeness and identity. Using this methodology, we successfully assembled genomes for two species that were previously identified through reads (Figs 3, 4). The first species is Mivirus sp. (Fig 3A), within the *Chuviridae* family. Mivirus sp. has been documented in a wide range of hosts, including both soft-bodied and hard-bodied ticks, with various *Rhipicephalus* species among them [16,26,64,65]. This genus, classified in the order Jingchucirales, has recently gained acceptance by the International Committee on Taxonomy of Viruses (ICTV) [25]. Mivirus sp. belongs to the category of negative-sense RNA viruses with genomic variability, this genus features monopartite linear, monopartite circular, and multipartite circular genomes [25]. In the case of our *Rpdp* sequence, it was grouped with available genomes known for their circular genome configuration (Fig 3B). These genomes have been reported in various countries, including Trinidad and Tobago and Thailand [25,64,65].

Studies have indicated that this genus of viruses tends to exhibit clustering patterns based on tick genera [64], hinting at a potential co-speciation relationship with specific vectors and could support theories about the potential pathogenesis of this species. Currently, the theories for this viral group lean more towards an endosymbiotic relationship [25–27]. Some authors have suggested that the low sequence conservation within their genomes and their high prevalence could indicate a possible endosymbiotic relationship [25–27]. Additionally, seroprevalence studies conducted on the hosts of these ticks have shown the absence of horizontal transmission [26,27]. Therefore, future research should focus on understanding the type of endosymbiotic relationship this virus has with ticks and whether it provides any benefits to them, as observed in some endosymbiotic bacteria and other vector viruses. Consequently, future studies should not only investigate the whole genome but also conduct *in vivo* experiments to assess its pathogenicity and the interaction with other TBV as done in other vectors to explore the potential of these endosymbiont viruses as biological control agents against vector-borne diseases. Investigations have shown promise in various vectors like *Aedes* spp. and *Culex* spp., where virus co-infections, such as those involving CFAV and the Phasi Charoen-like virus (PCLV), have demonstrated significant reductions in the transmission of Dengue, Zika, and La Crosse encephalitis virus (LACV) by up to 90% [66]. Additionally, even in the case of viruses known to be transmitted by ticks, such as the West Nile virus (WNV), it has been established that other viruses like NHUV in *C. quinquefasciatus* can lead to reductions in its transmission [67]. Therefore, evaluating these endosymbiont viruses for

their potential in diagnostic therapies and as new vaccine platforms is of paramount importance, as it has been demonstrated in other vector species.

To validate the genomes exhibiting the closest similarity to our MAG, we conducted FastANI analysis (Fig 3C). This analysis corroborated the findings from the genetic tree, showing significant similarity to the Champing virus isolated from Thailand [25], as well as other Miviruses from the monophyletic group. This group, previously referred to as "Champing viruses" [27], represent a distinct cluster from other sequences of Mivirus. Nevertheless, due to the challenges in demarcating viruses, future research should concentrate on delineating the species identified in various metagenomic investigations, including the present one.

Finally, MAGs were successfully reconstructed for the Bole Tick Virus 4 species (Fig 4A), in the *Flaviviridae* family, this virus belongs to the category of "large genome flaviviruses" (LGFs), a group of viruses characterized by genomes ranging from 16 to 26 kb, and they share regions of homologous sequences with flaviviruses [54]. Bole Tick Virus 4 had been previously reported in ticks of the species *Hyalomma asiaticum* in China, and in Colombia, it was previously documented in the department of Antioquia [28]. Phylogenetic analyses have revealed that it forms a sister clade with the genus Pestivirus, and as a result, it has also been referred to as Pesti-like viruses [25]. Pestiviruses exclusively infect vertebrates and can be pathogenic for livestock [54]. In contrast to Mivirus sp., Bole Tick Virus 4 has been detected in a wide range of tick species, including *Rhipicephalus* spp., *Amblyoma* spp., and *Hyalomma* spp., across various regions of the world [28,49,68]. This suggests the possibility of horizontal transmission of this virus [25]. Currently, research on its potential pathogenicity and its association with infections in vertebrates is limited. It is worth noting that while the assembly tool initially identified these MAGs as Bole Tick Virus 4, subsequent BLAST and phylogenetic analyses grouped it with Trinbago virus, a virus previously reported on the islands of Trinidad and Tobago (Fig 4B). However, multiple studies have demonstrated the phylogenetic grouping of this species with other sequences of Bole Tick Virus 4 from various regions worldwide [16,49], which supports the findings of this research. Nevertheless, the identity indicated by Fast ANI (Fig 4C) and the protein similarity reported in other studies remain relatively low. Therefore, further research is essential to fully comprehend the taxonomy, biology, and implications of this group of viruses in the country. This is particularly important considering that this is only the second study in the country reporting these viruses circulating in *R. sanguineus* s.l. [28], there is a possibility of their presence in other tick species as well due to previous reports in other countries [25], the close relationship of the group (FLGs) to pathogenic viruses to livestock [54] and the report of *Rhipicephalus sanguineus* feeding on cows in the country [11].

Despite certain limitations, such as sequence competition between ticks and viruses that restricted species assignment and the number of reads obtained per sample due to the chosen sequencing technology, we managed to outline the overall landscape of viruses present in these ticks. This study extends our knowledge to new regions in the country. However, In the future, the study of microbial

communities in tropical areas must be strengthened to favor future viral assignments in this country and others. We also employed a novel sequencing technology for virus identification, which, despite its limitations, offers a new approach that can enhance virus surveillance in tick-transmitted diseases in our country. Nonetheless, it is crucial to acknowledge certain limitations associated with the sequencing technique employed, such as error rates and sequencing depth. Future enhancements in this surveillance approach could lead to more accurate results. One avenue of improvement could involve adopting new techniques to improve the error rate of this sequencer. Additionally, reducing the number of samples per cell used in the process may increase the quantity of reads obtained. Furthermore, in the future, it is imperative to refine the available bioinformatics techniques for evaluating the virome through long reads sequencing. At present, most tools are primarily designed for analyzing short reads with adequate depth, and a shift towards accommodating long reads would be highly beneficial. To improve the comprehensibility and fluency of our study, it is essential to note that we focused exclusively on two departments within a country characterized by vast territory and significant diversity. Furthermore, it is important to acknowledge that our study had limitations in terms of sample size. Therefore, to continue conducting comprehensive surveillance of the viruses transmitted by this tick, it is imperative to expand both the sampling efforts and the exploration of new territories in the future.

On the other hand, based on these results, several considerations must be made for future virus descriptions in this tick species. Firstly, while virome studies in *R. sanguineus* s.l. have increased worldwide, most do not distinguish between different lineages. This is important because variations in composition between lineages have been observed in bacteria [63], suggesting the potential for different disease vectors. Even though only the tropical lineage is found in Colombia, this is a crucial factor for countries with multiple lineage cycles (e.g. Brazil). Moreover, future research should consider sex-based differentiation when assessing the virome, as variations in virus diversity and bacterial composition have been observed between the sexes. Additionally, this research should not be limited to RNA viruses, as was the case in this study. To gain a comprehensive understanding of the virome, including DNA viruses, is essential, as they can significantly impact and interact with the broader microbiome. Such a comprehensive understanding will play a critical role in disease prevention, safeguarding both human and animal health, and monitoring tick-transmitted viruses across the entire country.

As conclusion, by employing next-generation long-read sequencing, we conducted a comprehensive analysis of the primary RNA viruses found in *R. sanguineus* s.l. Through this investigation, we identified several noteworthy viruses with implications for public health. These include *Rhipicephalus*-associated flavi-like virus and Bole tick virus 4. We also detected viruses of interest due to their endosymbiotic relationships with ticks, such as *Mivirus* sp., and viruses that could potentially influence tick survival and development, such as Iflaviviridae sp. Furthermore, we successfully assembled genomes with high identity and coverage, enabling us to explore the phylogenetic relationships between these viruses and their counterparts reported worldwide. In the future, it is imperative to continue studying these viruses, elucidating their relationships and potential pathogenicity. Ongoing monitoring of these viruses is crucial

to prevent diseases that could impact both human and animal health, as well as the country's economy, particularly concerning animals like cattle and horses.

Acknowledgments

We thank the personnel who helped in the collection of the specimens in the field.

Data availability statement

The raw data of this study was deposited as an ENA Bioproject under the accession number PRJEB67590.

Financial Disclosure Statement

This work was funded by Dirección de Investigación e Innovación from Universidad del Rosario.

Author Contributions

Conceptualization: Luisa Páez-Triana, Marina Muñoz, Juan David Ramirez

Formal analysis: Luisa Páez-Triana, Nicolás Luna, Marina Muñoz, Juan David Ramirez

Funding acquisition: Marina Muñoz, Juan David Ramirez

Methodology: Luisa Páez-Triana, Nicolás Luna, Lissa Cruz-Saavedra, Diego J. Garcia-Corredor, Martin Orlando Pulido Medellín, Luz Helena Patiño, Marina Muñoz, Juan David Ramirez

Supervision: Marina Muñoz, Juan David Ramirez

Writing – original draft: Luisa Páez-Triana, Juan David Ramirez

References

1. Chen B, Lu X yan, Wang X. High-throughput sequencing to detect the bacterial microbiome in Ixodes granulatus in China. Res Sq. 2023;1–15.
2. Liu XY, Bonnet SI. Hard Tick Factors Implicated in Pathogen Transmission. PLoS Negl Trop Dis. 2014;8(1):5.
3. Jongejan F, Uilenberg G. The global importance of ticks. Parasitology. 2004;129(SUPPL.):S3–14.
4. Ni XB, Cui XM, Liu JY, Ye RZ, Wu YQ, Jiang JF, et al. Metavirome of 31 tick species provides a compendium of 1,801 RNA virus genomes. Nat Microbiol. 2023;8(1):162–73.
5. Alkishe A, Raghavan RK, Peterson AT. Likely geographic distributional shifts among medically important tick species and tick-associated diseases under climate change in north america: A review. Insects. 2021;12(3):1–28.

6. Bratuleanu BE, Temmam S, Chrétien D, Regnault B, Pérot P, Bouchier C, et al. The virome of Rhipicephalus, Dermacentor and Haemaphysalis ticks from Eastern Romania includes novel viruses with potential relevance for public health. Transbound Emerg Dis. 2021;69(3):1387–403.

7. Nuttall PA, Labuda M. Dynamics of infection in tick vectors and at the tick-host interface. Adv Virus Res. 2003;60:233–72.

8. Dantas-Torres F. Biology and ecology of the brown dog tick, Rhipicephalus sanguineus. Parasit Vectors. 2010;3(26).

9. Jia N, Wang J, Shi W, Du L, Sun Y, Zhan W, et al. Large-Scale Comparative Analyses of Tick Genomes Elucidate Their Genetic Diversity and Vector Capacities. Cell. 2020;182(5):1328-1340.e13.

10. Gray J, Dantas-Torres F, Estrada-Peña A, Levin M. Systematics and ecology of the brown dog tick, Rhipicephalus sanguineus. Ticks Tick Borne Dis. 2013;4(3):171–80.

11. Rivera-Páez FA, Labruna MB, Martins TF, Perez JE, Castaño-Villa GJ, Ossa-López PA, et al. Contributions to the knowledge of hard ticks (Acari: Ixodidae) in Colombia. Ticks Tick Borne Dis [Internet]. 2018;9(1):57–66. Available from: https://doi.org/10.1016/j.ttbdis.2017.10.008

12. Arias-Giraldo LM, Muñoz M, Hernández C, Herrera G, Velásquez-Ortiz N, Cantillo-Barraza O, et al. Identification of blood-feeding sources in Panstrongylus, Psammolestes, Rhodnius and Triatoma using amplicon-based next-generation sequencing. Parasites and Vectors [Internet]. 2020;13(1):1–14. Available from: https://doi.org/10.1186/s13071-020-04310-z

13. Paez-Triana L, Muñoz M, Herrera G, Moreno-Perez D, Taffur-Gomez G, Montenegro D, et al. Genetic diversity and population structure of Rhipicephalus sanguineus s.l across different regions from Colombia. Parasit Vectors [Internet]. 2021;14(1):1–11. Available from: https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-021-04898-w

14. Moraes-Filho J, Krawczak FS, Costa FB, Soares JF, Labruna MB. Comparative evaluation of the vector competence of four South American populations of the rhipicephalus sanguineus group for the bacterium Ehrlichia canis, the agent of canine monocytic ehrlichiosis. PLoS One. 2015;10(9):1–16.

15. Shi J, Hu Z, Deng F, Shen S. Tick-Borne Viruses. Virol Sin [Internet]. 2018;33(1):21–43. Available from: https://doi.org/10.1007/s12250-018-0019-0

16. Yang Z, Wang H, Yang S, Wang X, Shen Q, Ji L, et al. Virome diversity of ticks feeding on domestic mammals in China. Virol Sin [Internet]. 2023;38(2):208–21. Available from: https://doi.org/10.1016/j.virs.2023.02.001

17. Kazimírová M, Thangamani S, Bartíková P, Hermance M, Holíková V, Štibrániová I, et al. Tick-borne viruses and biological processes at the tick-host-virus interface. Front Cell Infect Microbiol. 2017;7(JUL):1–21.

18. Bente DA, Forrester NL, Watts DM, McAuley AJ, Whitehouse CA, Bray M. Crimean-Congo hemorrhagic fever: History, epidemiology, pathogenesis, clinical syndrome and genetic diversity. Antiviral Res [Internet]. 2013;100(1):159–89. Available from: http://dx.doi.org/10.1016/j.antiviral.2013.07.006

19. Bourée P, Delaigue S, Ensaf A. Tick-borne encephalitis. Med Ther Pediatr. 2015;18(3):145–51.

20. Galindo I, Alonso C. African swine fever virus: A review. Viruses. 2017;9(5).

21. Krasteva S, Jara M, Frias-De-Diego A, Machado G. Nairobi Sheep Disease Virus: A Historical and Epidemiological Perspective. Front Vet Sci. 2020;7(July):1–15.

22. Wang G, Tian X, Peng R, Huang Y, Li Y, Li Z. Genomic and phylogenetic profiling of RNA of tick-borne arboviruses in Hainan Island, China. Microbes Infect [Internet]. 2023;105218. Available from: https://doi.org/10.1016/j.micinf.2023.105218

23. Tokarz R, Sameroff S, Leon MS, Jain K, Lipkin WI. Genome characterization of Long Island tick rhabdovirus, a new virus identified in Amblyomma americanum ticks. Virol J. 2014;11(1):1–5.

24. Xia H, Hu C, Zhang D, Tang S, Zhang Z, Kou Z, et al. Metagenomic profile of the viral communities in rhipicephalus spp. ticks from Yunnan, China. PLoS One. 2015;10(3):1–16.

25. Sameroff S, Tokarz R, Charles RA, Jain K, Oleynik A, Che X, et al. Viral Diversity of Tick Species Parasitizing Cattle and Dogs in Trinidad and Tobago. Sci Rep [Internet]. 2019;9(1):1–10. Available from: http://dx.doi.org/10.1038/s41598-019-46914-1

26. Temmam S, Chrétien D, Bigot T, Dufour E, Petres S, Desquesnes M, et al. Monitoring silent spillovers before emergence: a pilot study at the tick/human interface in thailand. Front Microbiol. 2019;10(OCT):1–18.

27. Gondard M, Temmam S, Devillers E, Pinarello V, Bigot T, Chrétien D, et al. RNA viruses of amblyomma variegatum and rhipicephalus microplus and cattle susceptibility in the French antilles. Viruses. 2020;12(2):1–23.

28. Orozco Orozco M, Gómez GF, Alzate JF, Isaza JP, Gutiérrez LA. Virome analysis of three Ixodidae ticks species from Colombia: A potential strategy for discovering and surveying tick-borne viruses. Infect Genet Evol. 2021;96.

29. Gómez GF, Isaza JP, Segura JA, Alzate JF, Gutiérrez LA. Metatranscriptomic virome assessment of

Rhipicephalus microplus from Colombia. Ticks Tick Borne Dis [Internet]. 2020;11(5):101426. Available from: https://doi.org/10.1016/j.ttbdis.2020.101426

30. Máttar S, Parra M. Detection of Antibodies To Anaplasma, Bartonella and Coxiella in Rural Inhabitants of the Caribbean Area of Colombia. RevMVZ Córdoba. 2006;11(2):781–9.

31. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med [Internet]. 2015;7(1):1–13. Available from: http://dx.doi.org/10.1186/s13073-015-0220-9

32. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol [Internet]. 2016;17(1):1–11. Available from: http://dx.doi.org/10.1186/s13059-016-1103-0

33. Claro IM, Ramundo MS, Coletti TM, da Silva CAM, Valenca IN, Candido DS, et al. Rapid viral metagenomics using SMART-9N amplification and nanopore sequencing. Wellcome Open Res. 2021;6:241.

34. Kopylova E, Noé L, Touzet H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211–7.

35. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721–9.

36. Breitwieser FP, Salzberg SL. Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics. 2020;36(4):1303–4.

37. Wickham H. Elegant Graphics for Data Analysis: ggplot2. Applied Spatial Data Analysis with R. 2008. 21–54 p.

38. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Watson M, editor. PLoS One. 2013 Apr;8(4):e61217.

39. Lahti L, Shetty S. Tools for microbiome analysis in R. Version 2.1.28. 2017.

40. Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, et al. Genome Detective: An automated system for virus identification from high-throughput sequencing data. Bioinformatics. 2019;35(5):871–3.

41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

42. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen CY, et al. Proksee: In-depth characterization and visualization of bacterial genomes. Nucleic Acids Res. 2023;51(W1):W484–92.

43. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.

44. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics [Internet]. 2012 Apr 15 [cited 2021 Mar 3];28(8):1166–7. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts091

45. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Teeling E, editor. Mol Biol Evol [Internet]. 2020 May 1 [cited 2021 Mar 3];37(5):1530–4. Available from: https://academic.oup.com/mbe/article/37/5/1530/5721363

46. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1).

47. Walker J, Keinrans J, Horak I. The Genus Rhipicephalus (Acari, Ixodidae): A Guide to the Brown Ticks of the World. Cambridge University Press; 2000.

48. Couper L, Swei A. Tick microbiome characterization by next-generation 16S rRNA amplicon sequencing. J Vis Exp. 2018;2018(138):1–10.

49. Ergunay K, Mutinda M, Bourke B, Justi SA, Caicedo-Quiroga L, Kamau J, et al. Metagenomic Investigation of Ticks From Kenyan Wildlife Reveals Diverse Microbial Pathogens and New Country Pathogen Records. Front Microbiol. 2022;13(July):1–16.

50. Pascoe EL, Nava S, Labruna MB, Paddock CD, Levin ML, Marcantonio M, et al. Predicting the northward expansion of tropical lineage Rhipicephalus sanguineus sensu lato ticks in the United States and its implications for medical and veterinary health. PLoS One [Internet]. 2022;17(8 August):1–20. Available from: http://dx.doi.org/10.1371/journal.pone.0271683

51. Gómez M, Martinez D, Muñoz M, Ramírez JD. microbiome / virome : new strategies for controlling arboviral transmission ? Parasit Vectors [Internet]. 2022;518:406–13. Available from: https://doi.org/10.1186/s13071-022-05401-9

52. Baidaliuk A, Miot EF, Lequime S, Moltini-Conclois I, Delaigue F, Dabo S, et al. Cell-Fusing Agent Virus Reduces Arbovirus Dissemination in Aedes aegypti Mosquitoes In Vivo Artem. J Virol. 2019;93(18):e00705-19.

53. Romo H, Kenney JL, Blitvich BJ, Brault AC. Restriction of Zika virus infection and transmission in

Aedes aegypti mediated by an insect-speci fi c fl avivirus. Emerg Microbes Infect [Internet]. 2018; Available from: http://dx.doi.org/10.1038/s41426-018-0180-4

54. Paraskevopoulou S, Käfer S, Zirkel F, Donath A, Petersen M, Liu S, et al. Viromics of extant insect orders unveil the evolution of the flavi-like superfamily. Virus Evol. 2021;7(1):1–16.

55. Dinçer E, Hacioğlu S, Kar S, Emanet N, Brinkmann A, Nitsche A, et al. Survey and characterization of jingmen tick virus variants. Viruses. 2019;11(11).

56. Pascoal J de O, Siqueira SM de, Maia R da C, Juan Szabó MP, Yokosawa J. Detection and molecular characterization of Mogiana tick virus (MGTV) in Rhipicephalus microplus collected from cattle in a savannah area, Uberlândia, Brazil. Ticks Tick Borne Dis [Internet]. 2019;10(1):162–5. Available from: https://doi.org/10.1016/j.ttbdis.2018.10.002

57. Wang ZD, Wang B, Wei F, Han SZ, Zhang L, Yang ZT, et al. A New Segmented Virus Associated with Human Febrile Illness in China. N Engl J Med. 2019;380(22):2116–25.

58. Zhang X, Wang N, Wang Z, Liu Q. The discovery of segmented flaviviruses: implications for viral emergence. Curr Opin Virol [Internet]. 2020;40:11–8. Available from: https://doi.org/10.1016/j.coviro.2020.02.001

59. Maruyama SR, Castro-Jorge LA, Ribeiro JMC, Gardinassi LG, Garcia GR, Brandão LG, et al. Characterisation of divergent flavivirus NS3 and NS5 protein sequences detected in Rhipicephalus microplus ticks from Brazil. Mem Inst Oswaldo Cruz. 2014;109(1):38–50.

60. Ternovoi VA, Protopopova E V., Shvalov AN, Yu. Kartashov M, Bayandin RB, Tregubchak T V., et al. Complete Coding Genome Sequence for a Novel Multicomponent Kindia Tick Virus Detected from Ticks Collected in Guinea. bioRxiv. 2020;4–9.

61. Valles SM, Chen Y, Firth AE, Guérin DMA, Hashimoto Y, Herrero S, et al. ICTV virus taxonomy profile: Iflaviridae. J Gen Virol. 2017;98(4):527–8.

62. Daveu R, Hervet C, Sigrist L, Sassera D, Jex A, Labadie K, et al. Sequence diversity and evolution of a group of iflaviruses associated with ticks. Arch Virol [Internet]. 2021;166(7):1843–52. Available from: https://doi.org/10.1007/s00705-021-05060-8

63. René-Martellet M, Minard G, Massot R, Van Tran V, Valiente Moro C, Chabanne L, et al. Bacterial microbiota associated with Rhipicephalus sanguineus (s.l.) ticks from France, Senegal and Arizona. Parasites and Vectors. 2017;10(1):1–10.

64. Bratuleanu BE, Temmam S, Munier S, Chrétien D, Bigot T, van der Werf S, et al. Detection of Phenuiviridae, Chuviridae Members, and a Novel Quaranjavirus in Hard Ticks From Danube Delta. Front Vet Sci. 2022;9(April):1–9.

65. Guo L, Ma J, Lin J, Chen M, Liu W, Zha J, et al. Virome of Rhipicephalus ticks by metagenomic analysis in Guangdong, southern China. Front Microbiol. 2022;13(August):1–9.

66. Schultz MJ, Frydman HM, Connor JH, Infectious E, Laboratories D. Dual Insect specific virus infection limits Arbovirus Replication in Aedes mosquito cells Michaela. Virology. 2021;406–13.

67. Goenaga S, Kenney JL, Duggal NK, Delorey M, Ebel GD, Zhang B, et al. Potential for Co-Infection of a Mosquito-Specific Flavivirus , Nhumirim Virus , to Block West Nile Virus Transmission in Mosquitoes. 2015;5801–12.

68. Zhang Y, Hu B, Agwanda B, Fang Y, Wang J, Kuria S, et al. Viromes and surveys of RNA viruses in camel-derived ticks revealing transmission patterns of novel tick-borne viral pathogens in Kenya. Emerg Microbes Infect. 2021;10(1):1975–87.

Supporting information

Supplementary figures and tables available in: Tesis-Figuras

Fig S1. Geographic locations where samples were collected. The map was made with ArcGIS Online (Basemap: Elevation/World_Hillshade https://bit.ly/3vVQ1lL; Sources: Esri, Airbus DS, USGS, NGA, NASA, CGIAR, N Robinson, NCEAS, NLS, OS, NMA, Geodatastyrelsen, Rijkswaterstaat, GSA, Geoland, FEMA, Intermap and the GIS user community.

Fig S2. Analysis of diversity in the species identified through the NCBI virus database. A. Alpha Diversity for sexes and departments. B. Beta Diversity for sexes. C. Beta Diversity for departments. Significant *p-value*: * <0.05.

Table S1. Metadata of the samples used, including location and sex

Table S2. Accession number database virus with tick species as host

Table S3. Accession number database NCBI virus

Table S4. p-values for species comparing sexes and departments

Contribuciones específicas del estudiante dentro de la investigación llevada a cabo. Señalar las actividades en las cuáles estuvo involucrado activamente el estudiante:

La concepción del proyecto de investigación	(X)
El diseño del estudio	(X)
La adquisición de los datos a través de la experimentación	(X)
Análisis e interpretación de los datos	(X)
Elaboración del borrador del artículo	(X)
Revisión y aprobación definitiva de la versión que se presenta	(X)