

Metagenomics for epidemiological surveillance in one health

Edited by

Ratree Takhampunya, Yvonne-Marie Linton, Michael E. von Fricken
and Mel C. Melendrez

Published in

Frontiers in Microbiology
Frontiers in Public Health



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-2175-5
DOI 10.3389/978-2-8325-2175-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Metagenomics for epidemiological surveillance in one health

Topic editors

Ratree Takhampunya — Armed Forces Research Institute of Medical Science, Thailand

Yvonne-Marie Linton — Walter Reed Biosystematics Unit (WRBU), United States

Michael E. von Fricken — George Mason University, United States

Mel C. Melendrez — Anoka-Ramsey Community College, United States

Citation

Takhampunya, R., Linton, Y.-M., von Fricken, M. E., Melendrez, M. C., eds. (2023). *Metagenomics for epidemiological surveillance in one health*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2175-5

Table of contents

- 05 **Editorial: Metagenomics for epidemiological surveillance in One Health**
Michael E. von Fricken, Mel C. Melendrez, Yvonne-Marie Linton and Ratree Takhampunya
- 09 **Dissemination of Resistant *Escherichia coli* Among Wild Birds, Rodents, Flies, and Calves on Dairy Farms**
Rachel A. Hickman, Viktoria Agarwal, Karin Sjöström, Ulf Emanuelson, Nils Fall, Susanna Sternberg-Lewerin and Josef D. Järhult
- 18 **A Rapid, Whole Genome Sequencing Assay for Detection and Characterization of Novel Coronavirus (SARS-CoV-2) Clinical Specimens Using Nanopore Sequencing**
Maria T. Arévalo, Mark A. Karavis, Sarah E. Katoski, Jacquelyn V. Harris, Jessica M. Hill, Samir V. Deshpande, Pierce A. Roth, Alvin T. Liem and R. Cory Bernhards
- 31 **Metagenomic Investigation of Ticks From Kenyan Wildlife Reveals Diverse Microbial Pathogens and New Country Pathogen Records**
Koray Ergunay, Mathew Mutinda, Brian Bourke, Silvia A. Justi, Laura Caicedo-Quiroga, Joseph Kamau, Samson Mutura, Irene Karagi Akunda, Elizabeth Cook, Francis Gakuya, Patrick Omondi, Suzan Murray, Dawn Zimmerman and Yvonne-Marie Linton
- 47 **Metagenomic Next-Generation Sequencing Reveals the Profile of Viral Infections in Kidney Transplant Recipients During the COVID-19 Pandemic**
Xiangyong Tian, Wenjing Duan, Xiulei Zhang, Xiaoqiang Wu, Chan Zhang, Zhiwei Wang, Guanghui Cao, Yue Gu, Fengmin Shao and Tianzhong Yan
- 56 **Genomic and virologic characterization of samples from a shipboard outbreak of COVID-19 reveals distinct variants within limited temporospatial parameters**
Regina Z. Cer, Logan J. Voegtly, Bishwo N. Adhikari, Brian L. Pike, Matthew R. Lueder, Lindsay A. Glang, Francisco Malagon, Ernesto Santa Ana, James M. Regeimbal, Maria F. Potts-Szoke, Kevin L. Schully, Darci R. Smith and Kimberly A. Bishop-Lilly
- 67 **Metagenomic profiles of *Dermacentor* tick pathogens from across Mongolia, using next generation sequencing**
Doniddemberel Altantogtokh, Abigail A. Lilak, Ratree Takhampunya, Jira Sakolvaree, Nitima Chanarat, Graham Matulis, Betty Katherine Poole-Smith, Bazartseren Boldbaatar, Silas Davidson, Jeffrey Hertz, Buyandelger Bolorchimeg, Nyamdorj Tsogbadrakh, Jodi M. Fiorenzano, Erica J. Lindroth and Michael E. von Fricken
- 81 **The Remote Emerging Disease Intelligence—NETwork**
Nicole L. Achee and The Remote Emerging Disease Intelligence—NETwork (REDI-NET) Consortium

- 92 **Discovery of *Rickettsia* spp. in mosquitoes collected in Georgia by metagenomics analysis and molecular characterization**
Adam R. Pollio, Ju Jiang, Sam S. Lee, Jaykumar S. Gandhi, Brian D. Knott, Tamar Chunashvili, Matthew A. Conte, Shannon D. Walls, Christine E. Hulseberg, Christina M. Farris, Drew D. Reinbold-Wasson and Jun Hang
- 102 **Diagnostic efficiency of metagenomic next-generation sequencing for suspected spinal tuberculosis in China: A multicenter prospective study**
Yuan Li, Xiao-wei Yao, Liang Tang, Wei-jie Dong, Ting-long Lan, Jun Fan, Feng-sheng Liu and Shi-bing Qin
- 111 **Corrigendum: Diagnostic efficiency of metagenomic next-generation sequencing for suspected spinal tuberculosis in China: a multicenter prospective study**
Yuan Li, Xiao-wei Yao, Liang Tang, Wei-jie Dong, Ting-long Lan, Jun Fan, Feng-sheng Liu and Shi-bing Qin



OPEN ACCESS

EDITED AND REVIEWED BY

Axel Cloeckaert,
Institut National de Recherche pour
l'Agriculture, l'Alimentation et l'Environnement
(INRAE), France

*CORRESPONDENCE

Michael E. von Fricken
✉ mvonfric@gmu.edu

SPECIALTY SECTION

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

RECEIVED 22 March 2023

ACCEPTED 24 March 2023

PUBLISHED 31 March 2023

CITATION

von Fricken ME, Melendrez MC, Linton Y-M and
Takhampunya R (2023) Editorial: Metagenomics
for epidemiological surveillance in One Health.
Front. Microbiol. 14:1191946.
doi: 10.3389/fmicb.2023.1191946

COPYRIGHT

© 2023 von Fricken, Melendrez, Linton and
Takhampunya. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Editorial: Metagenomics for epidemiological surveillance in One Health

Michael E. von Fricken^{1*}, Mel C. Melendrez²,
Yvonne-Marie Linton^{3,4,5} and Ratree Takhampunya⁶

¹Department of Global and Community Health, College of Public Health, George Mason University, Fairfax, VA, United States, ²Department of Biology, Anoka-Ramsey Community College, Coon Rapids, MN, United States, ³Walter Reed Biosystematics Unit, Smithsonian Museum Support Center, Suitland, MD, United States, ⁴Department of Entomology, Smithsonian Institution – National Museum of Natural History (NMNH), Washington, DC, United States, ⁵Walter Reed Army Institute of Research, One Health Branch, Silver Spring, MD, United States, ⁶Department of Entomology, United States Army Medical Directorate—Armed Forces Research Institute of Medical Sciences (AFRIMS), Bangkok, Thailand

KEYWORDS

metagenomics, biosurveillance, One Health, next generation sequencing (NGS), vector-borne diseases

Editorial on the Research Topic

Metagenomics for epidemiological surveillance in One Health

The highest burden of endemic, neglected, and emerging zoonotic diseases occur in low and middle-income countries (LMICs) (Worsley-Tonks et al., 2022). However, limited resources, surveillance infrastructure and access to advanced molecular techniques in these regions can severely impact the rapid detection and characterization necessary to mitigate and respond to outbreak events. Metagenomics approaches continue to become less expensive, and now represent a potentially cost-effective approach to conduct biosurveillance and monitor molecular epidemiology (Ko et al., 2022). Compared to PCR, which only detects pathogen based on specific primers, metagenomic agnostic sequencing allows for the identification of known and unknown microbes (viruses, bacteria and protozoans) from a single sample (Govender et al., 2021). Forward-facing portable sequencing platforms such as the MinION can provide expansive knowledge of circulating pathogens from a wide array of field-collected samples, including arthropod vectors, soil and water samples and a variety of swabs and bloods from vertebrates, both in the field and in basic laboratory conditions (Gardy and Loman, 2018; Achee et al.). Using metagenomics approaches, detection of known, emergent, and emerging zoonotic pathogens of human and veterinary importance can be made available to decision-makers, researchers, and policymakers in near real-time, to inform effective containment or mitigation efforts to minimize public health impacts.

The nine publications presented here in this Research Topic capture a wide range of settings that metagenomics can be applied to through a One Health lens, including outbreak response, clinical management, biosurveillance, and disease monitoring within the human animal interface. Publication summaries can be found in Table 1.

TABLE 1 Summary of special edition publications as they relate to metagenomics, surveillance, and One Health.

Article title	Key findings
Dissemination of resistant <i>Escherichia coli</i> among wild birds, rodents, flies, and calves on dairy farms	Animal, arthropod, and environmental samples were collected on 57 Swedish dairy farms. Isolates from different sources on the same farm generally clustered together, with resistant <i>E. coli</i> detected from calves and scavenger animals with little genomic differences, suggesting interspecies transfer of pathogens. Transmission of <i>E. coli</i> between species highlights the potential risk of AMR spread in areas with low antimicrobial use.
Metagenomic next-generation sequencing reveals the profile of viral infections in kidney transplant recipients during the COVID-19 pandemic	NGS was utilized in kidney transplant recipients for the detection of viral infections during the COVID-19 pandemic. In 48/50 clinical samples, 15 types of viruses were detected, including cytomegalovirus, Torque teno virus, human alpha and beta herpesviruses, JC, BK, and WU polyomaviruses, primate bocaparvovirus 1, simian virus 12, and volepox virus.
A rapid, whole genome sequencing assay for detection and characterization of novel coronavirus (SARS-CoV-2) clinical specimens using nanopore sequencing	A modified SARS-CoV-2 nanopore sequencing assay was developed based on the ARTIC protocol for secluded, low resource settings. When analyzing six of these sequences for mutations, there were existing and unique changes in the sequence, with three occurring in the Spike protein. Rapid characterization of SARS-CoV-2 using this data may be informative for subsequent targeted control measures.
Metagenomic investigation of ticks from Kenyan wildlife reveals diverse microbial pathogens and new country pathogen records	Detected a variety of microbial species in ticks collected from the environment, wildlife, and domestic animals, using NGS. A major advantage of agnostic pathogen identification techniques resulted in the detection of several emerging zoonotic agents and resulted in the first ever detection of Jingmen tick virus in Kenya.
Metagenomic profiles of dermacentor tick pathogens from across Mongolia, using next generation sequencing	Dermacentor ticks ($n = 1,773$) from 15 provinces of Mongolia were screened using NGS. <i>Rickettsia</i> spp. was detected in 88.33% of screened tick pools. <i>Anaplasma</i> spp. and <i>Bartonella</i> spp. were detected in 3.18% and 0.79% of pools, respectively. <i>Anaplasma</i> spp. was only detected in ticks collected from livestock and this the first report of <i>B. melophagi</i> in ticks from Mongolia.
Genomic and virologic characterization of samples from a shipboard outbreak of COVID-19 reveals distinct variants within limited temporospatial parameters	Using whole-genome sequencing on COVID-19 positive swabs, produced 18 viral genomes, of which, seven were unique variants. High rates of vaccination may limit viral evolution therefore reducing the spread of COVID-19. During outbreaks, sequencing surveillance can prove vital toward understanding viral molecular epidemiology.
The remote emerging disease intelligence—NETwork	The REDI-NET consortium was established to improve pathogen surveillance in the United States, Kenya, and Belize utilizing a One Health approach. Metagenomic screening pathways are described in detailed frameworks between tiered laboratory systems, where analysis pipelines will allow for the formation of disease dashboards, risk maps, and models.
Discovery of <i>Rickettsia</i> spp. in mosquitoes collected in Georgia by metagenomics analysis and molecular characterization	Study used NGS to examine mosquitoes as potential vectors of <i>Rickettsia</i> spp. pathogens. Some 475 pools of <i>Aedes</i> , <i>Culex</i> , and <i>Culiseta</i> mosquitoes collected in Georgia from 2018 to 2019 were screened, 33 of which tested positive for rickettsial DNA. Genus-specific qPCR and multi-locus sequence typing identified a <i>Rickettsia</i> spp. closely related to <i>R. bellii</i> .
Diagnostic Efficiency of metagenomic next-generation sequencing for suspected spinal tuberculosis in China: A multicenter prospective study	NGS was used to detect pathogens in patients with suspected spinal tuberculosis (TB). A 100 patients were enrolled and pathogens were found in 82 patients. Patient findings included TB ($n = 37$) and 45 patients infected with other bacteria. The sensitivity of the NGS assay in patients with a spinal infection was higher than that of culture and pathological examination, but not statistically different than Xpert and T-SPOT.TB.

Metagenomics during outbreaks

During an outbreak investigation, Cer et al., showed the utilization of metagenomics by comparing genomic variants to measure human–human spread and mutation rates of the COVID-19 virus. Comparing two different outbreaks of COVID-19—one in a group prior to vaccine development and one in a group with high vaccination rates—a larger number of variants were observed within the unvaccinated group. The study highlighted further questions and discussion on the impacts of sampling bias, the impact of purifying selection in creating clonal viral populations in vaccinated individuals, and how that can inform and support vaccination efforts.

Pandemics such as the Ebola in 2014–2016 and COVID-19 have shown that global security is compromised when response requires advanced training, special tools, and established infrastructure to accurately survey and assess the outbreak situation, where delays ultimately cost lives. It was during the Ebola outbreak that portable sequencing technology (MinION) was piloted in the field to speed

up detection and public health response with a simpler, pared down protocol and analysis workflow. COVID-19 sequencing protocol using the ARTIC protocol from Illumina can be time consuming, limiting the potential use of molecular epidemiology for control measures. To reduce sequencing time during an outbreak, Arévalo et al., modified a COVID-19 whole genome sequencing assay that produced high quality metagenomics data relatively quickly in an isolated, low resource setting. Adaptation of current protocols to LMIC infrastructure and resources while providing a quick turnaround for results will be invaluable in controlling future outbreaks.

Clinical applications of NGS

Clinically, metagenomics can be used to identify unknown or rare pathogens that might go undetected using traditional methods. An example of this was outlined by Tian et al., to detect viral infections in kidney transplant recipients (KTRs) during the COVID-19 pandemic. An NGS approach detected rare viruses

in multiple sample types including bronchoalveolar lavage fluid (BALF), urine and blood. This method identified infections or co-infections in vulnerable immunosuppressed kidney transplant recipients, which was then used to inform treatment strategies.

Li et al., used NGS on clinical samples of patients who were suspected of having spinal tuberculosis (TB). Over half of the samples (45/82) tested had bacteria that was not *Mycobacterium tuberculosis*. This approach quickly and simultaneously confirmed patients with spinal infection if they had TB or non-TB infections. This approach also distinguished and identified other pathogens in these patients, which helped guide patient's clinical treatment. This study evaluated the efficacy and application of NGS as clinical diagnostic assay for suspected spinal TB infection in clinical laboratories.

Biosurveillance

Altantogtokh et al., tested pools of *Dermacentor nuttalli* ticks ($n = 1,773$) from 15 provinces of Mongolia using NGS. *Rickettsia* spp. was detected in 88.33% of screened tick pools, with Khentii aimag having the highest detection rate for *Rickettsia* spp. *Rickettsia raoultii* and *Rickettsia sibirica* were detected within this study, aligning with previous findings that applied NGS to livestock samples in Mongolia (Chaorattanakawee et al., 2022). *Anaplasma* spp. and *Bartonella* spp. were detected in 3.18 and 0.79% of pools, respectively, with *Anaplasma* spp. only detected in ticks collected from livestock, as seen elsewhere (von Fricken et al., 2020). This study represents the first detection of *Bartonella melophagi* in ticks tested from Mongolia.

Pollio et al., discovered a novel *Rickettsia* spp. in mosquitoes which shared a close identity with *Rickettsia bellii* based on data from NGS. While it is unknown if *R. bellii* is pathogenic in humans, detecting this microorganism in mosquitoes may hold value down the road should this disease emerge. The use of metagenomics to screen mosquito samples elsewhere in the world may hold value for the identification and characterization of potential pathogenic *Rickettsia* spp. within an atypical vector host.

Metagenomics at the human-animal interface

Human-livestock transmission is one mode for infection of zoonotic pathogens. The likelihood of infection is increased as animals pass pathogens between species similar to the spread of *Escherichia coli* in Hickman et al., This interspecies exchange is conducive to pathogen mutations, and passage of mobile genetic elements or plasmids via clonal or horizontal gene transfer, increasing the risk of spreading to humans potentially increasing the severity of infection, and resistance to antibiotic interventions.

Ergunay et al., used NGS to screen 75 blood-fed ticks (*Rhipicephalus* spp. and *Amblyomma* spp.) collected from wild and domestic animals in Kenya for pathogens. Fifty-six human or veterinary pathogenic bacterial species were detected including *Escherichia coli* (62.8%), *Proteus mirabilis* (48.5%), *Coxiella burnetii*

(45.7%), and *Francisella tularensis* (14.2%). Additional pathogens detected include fungal species, filarial pathogens, protozoa, and environmental and water/foodborne pathogens. Jingmen tick virus was detected within 13% of these samples, representing the first report of this virus in Kenya. Interestingly, in these wild and domestic animal blood-bags, human pathogens, including *Plasmodium falciparum* were detected, indicating back spill-over of pathogens of human pathogens to animals.

Finally, the adoption of a One Health focused surveillance strategy is described in detail by Achee et al., through the Remote Emerging Disease Intelligence—NETwork (REDI-NET), which has implemented MinION-based xenosurveillance strategies in the United States, Kenya and Belize. Comprehensive standard operating procedures (SOPs) were developed for the deployment of metagenomic surveillance at high-risk disease envelopes. The ultimate goal of REDI-NET is to apply agnostic pathogen detection to a variety of sample types including water, ticks, soil, and leeches across Gold- (reach-back laboratories with greater infrastructure) and Silver- (remote, field-forward laboratories) tiered laboratory systems. This study outlined the framework required for working with metagenomic data, efficient data pipeline strategies, and systematic processes for collection and testing of samples for deeper analysis.

This Research Topic highlights the diversity of metagenomic application across disciplines, geographical regions, and both research and clinical settings. The field of metagenomics is growing rapidly, with all papers in this Research Topic published within the past year. Utilizing metagenomics from a One Health perspective will expand surveillance and characterization of existing and new pathogens in humans and animals, which will likely continue to be adopted into existing systems as next generation sequencing becomes more accessible and cost-effective.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Chaorattanakawee, S., Wofford, R. N., Takhampunya, R., Katherine Poole-Smith, B., Boldbaatar, B., Lkhagvatseren, S., et al. (2022). Tracking tick-borne diseases in Mongolian livestock using next generation sequencing (NGS). *Ticks Tick Borne Dis.* 13, 101845. doi: 10.1016/j.ttbdis.2021.101845
- Gardy, J. L., and Loman, N. J. (2018). Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* 19, 9–20. doi: 10.1038/nrg.2017.88
- Govender, K. N., Street, T. L., Sanderson, N. D., and Eyre, D. W. (2021). Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis of diagnostic test accuracy studies. *J. Clin. Microbiol.* 59, e0291620. doi: 10.1128/JCM.02916-20
- Ko, K. K. K., Chang, K. R., and Nagarajan, N. (2022). Metagenomics-enabled microbial surveillance. *Nat. Microbiol.* 7, 486–496. doi: 10.1038/s41564-022-01089-w
- von Fricken, M. E., Quorollo, B. A., Boldbaatar, B., Wang, Y. W., Jiang, R. R., Lkhagvatseren, S., et al. (2020). Genetic diversity of *Anaplasma* and *Ehrlichia* bacteria found in *Dermacentor* and *Ixodes* ticks in Mongolia. *Ticks Tick Borne Dis.* 11, 101316. doi: 10.1016/j.ttbdis.2019.101316
- Worsley-Tonks, K. E. L., Bender, J. B., Deem, S. L., Ferguson, A. W., Fèvre, E. M., Martins, D. J., et al. (2022). Strengthening global health security by improving disease surveillance in remote rural areas of low-income and middle-income countries. *Lancet Glob. Health.* 10, e579–e584. doi: 10.1016/S2214-109X(22)00031-6



Dissemination of Resistant *Escherichia coli* Among Wild Birds, Rodents, Flies, and Calves on Dairy Farms

Rachel A. Hickman^{1,2*}, Viktoria Agarwal^{2,3}, Karin Sjöström⁴, Ulf Emanuelson⁴, Nils Fall⁴, Susanna Sternberg-Lewerin⁵ and Josef D. Järhult^{1*}

¹Department of Medical Sciences, Zoonosis Science Center, Uppsala University, Uppsala, Sweden, ²Department of Medical Biochemistry and Microbiology, Zoonosis Science Center, Uppsala University, Uppsala, Sweden, ³Institute of Environmental Engineering, Zürich, Switzerland, ⁴Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden, ⁵Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden

OPEN ACCESS

Edited by:

Michael E. von Fricken,
George Mason University,
United States

Reviewed by:

Clair L. Firth,
University of Veterinary Medicine
Vienna, Austria
Magdalena Zalewska,
University of Warsaw, Poland
Kees Veldman,
Wageningen Bioveterinary Research
(WBVR), Netherlands

*Correspondence:

Rachel A. Hickman
rachel.hickman@medsci.uu.se
Josef D. Järhult
josef.jarhult@medsci.uu.se

Specialty section:

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

Received: 17 December 2021

Accepted: 04 March 2022

Published: 01 April 2022

Citation:

Hickman RA, Agarwal V, Sjöström K,
Emanuelson U, Fall N,
Sternberg-Lewerin S and
Järhult JD (2022) Dissemination of
Resistant *Escherichia coli* Among
Wild Birds, Rodents, Flies, and
Calves on Dairy Farms.
Front. Microbiol. 13:838339.
doi: 10.3389/fmicb.2022.838339

Antimicrobial resistance (AMR) in bacteria in the livestock is a growing problem, partly due to inappropriate use of antimicrobial drugs. Antimicrobial use (AMU) occurs in Swedish dairy farming but is restricted to the treatment of sick animals based on prescription by a veterinary practitioner. Despite these strict rules, calves shedding antimicrobial resistant *Enterobacteriaceae* have been recorded both in dairy farms and in slaughterhouses. Yet, not much is known how these bacteria disseminate into the local environment around dairy farms. In this study, we collected samples from four animal sources (fecal samples from calves, birds and rodents, and whole flies) and two environmental sources (cow manure drains and manure pits). From the samples, *Escherichia coli* was isolated and antimicrobial susceptibility testing performed. A subset of isolates was whole genome sequenced to evaluate relatedness between sources and genomic determinants such as antimicrobial resistance genes (ARGs) and the presence of plasmids were assessed. We detected both ARGs, mobile genetic elements and low rates of AMR. In particular, we observed four potential instances of bacterial clonal sharing in two different animal sources. This demonstrates resistant *E. coli* dissemination potential within the dairy farm, between calves and scavenger animals (rodents and flies). AMR dissemination and the zoonotic AMR risk is generally low in countries with low and restricted AMU. However, we show that interspecies dissemination does occur, and in countries that have little to no AMU restrictions this risk could be under-estimated.

Keywords: antibiotic resistance, calves, rodents, flies, cross-species transfer, birds, dairy farms, livestock

INTRODUCTION

Any use of antimicrobials causes a selective pressure that favors resistant bacteria (Olesen et al., 2020). Hence, the global use and misuse of antibiotic drugs makes antimicrobial resistance (AMR) a growing problem, which is frequently reported in scientific literature and media outlets. The lack of treatment options due to antimicrobial resistance is most frequently discussed

in regard to human health. However, AMR is also a problem in animal health, particularly in regions where these drugs are used as growth promoters or supplements in animal feed (McEwen and Fedorka-Cray, 2002; McEwen and Collignon, 2018). Animal-related AMR contributes to the continuous AMR dissemination and transmission that can cause several potential risks to animals and humans, e.g., biosecurity risks along the food chain, including resistant bacteria in food products (Aarestrup et al., 2008; Carmo et al., 2018). Antimicrobial-resistant bacteria in food products from livestock usually originate from the normal microbiome of healthy animals or, where lack of control leads to slaughter of sick animals, pathogens that contaminate the meat or food products from these animals (Cho et al., 2018; Jans et al., 2018) or from other sources, e.g., fruits and vegetables where feces from animals, or human wastewater, have been used as fertilizer to cultivate these foods (Mesbah Zekar et al., 2017; Karlsson et al., 2021).

In the European Union, the antimicrobial use (AMU) in animals has been declining since 2011 (European Medicines Agency year, 2019). In Sweden, already low, the levels of AMU in Swedish animals and humans have also declined since 2011 (SVA, 2019). However, global AMU is still rising, one major driver is the increased global demand for animal protein that has been facilitated by the expansion of intensive farming (Tiseo et al., 2020). There is a large global demand for bovine dairy and meat products despite local fluctuations (Lhermie et al., 2018). Despite good animal management procedures, animals often at some point during their life contract an infection that requires treatment with antimicrobial drugs, i.e., antibiotics, anthelmintics, or antifungals. Within dairy farms the most common illness that requires antibiotic treatment is mastitis (Cheng and Han, 2020). In most countries, milk withdrawal periods apply to dairy cows that undergo treatment, where milk cannot be delivered for human consumption. Such milk is often fed to calves or discarded into drains or on manure heaps (Firth et al., 2021). This can influence the calves' microbiome and promote fecal shedding of resistant bacteria or lead to antibiotic drug residues selecting for AMR in the local environment (Duse et al., 2015).

From an international perspective, the Swedish dairy industry is small with a total of 510,340 cows in 2020, of which 303,390 for milk production (Jordbrukverket, 2020). According to Swedish legislation, antimicrobials are only available for treatment of animals on veterinary prescription. The major organic certification body in Sweden, KRAV,¹ follows the EU regulations on organic production and even go further in some areas concerning animal welfare. A previous study in 30 conventional and 30 organic dairy farms showed no apparent difference in AMR despite the differences in AMU regulation between organically and conventionally managed herds (Sjöström et al., 2020) and not significantly different use of injectable mastitis treatment (Olmos Antillón et al., 2020). In the current study, we wanted to further examine AMR in the local farm environment within these Swedish conventional

and organic dairy farms by extending the analyses to other potential sources as well as a more thorough genetic analysis of resistant *E. coli*.

MATERIAL AND METHODS

Sample Population and Collection

All samples were collected from 54 dairy farms across Sweden in 2017, the same farms as described in Sjöström et al. (2020), during the second sampling period of that study (27 organic and 27 conventional herds). A convenience sampling design was used for the original study (Sjöström et al., 2020). At each farm, up to 10 samples were collected, these samples were: five fecal swabs from healthy calves <2 months old; one swab from an indoor manure drainage site; one swab from the manure pit; one sample consisting of bird fecal droppings; one with rodent fecal droppings; and one collection of whole live flies picked from the fly tape in the barn. All calf samples were collected rectally with an Amie's charcoal culture swab (Copan diagnostics Inc., Murrieta, CA, United States) and the manure samples were collected with E-swabs (Copan diagnostics Inc., Murrieta, CA, United States), while fecal droppings from the bird and the rodent samples were placed in tubes with Amies agar gel with charcoal transport media and whole flies were crushed in tubes with Amies agar gel with charcoal transport media. All collected swab samples were immediately placed into transport tubes with Amies agar gel with charcoal transport media as well as other samples and stored at 4°C after collection and were continued to be stored at 4°C at the laboratory facility before sample processing begun.

Bacterial Isolation and Antimicrobial Susceptibility Testing

After collection all samples were subjected to indicator *E. coli* isolation within a week of their collection date at the Zoonosis Science Center, Uppsala University. All samples were diluted in 3 ml of 0.9% NaCl and, subsequently, 1 ml of 10-fold dilutions (10^{-2} and 10^{-4} for calf samples, 10^{-1} and 10^{-3} for other samples) were streaked on Petrifilm™ (3MTM, St Paul, MN, United States) Select *E. coli* count (SEC) plates (3M Microbiology Products) and cultured overnight at 42°C. Simultaneously, all samples were enriched in 4 ml of peptone water overnight at 37°C and subsequently streaked on cephalosporin CHROMagar C3G selection plates (CHROMagar, Paris, France) and incubated at 37°C overnight. From the Petrifilm SEC plates one random colony was collected and subcultured on Horse Blood Agar plates and identified as *E. coli* by morphology and the indole test. From the CHROMagar C3G plates, pink colonies were identified as *E. coli* by morphology and an indole test and were also subcultured. All isolates were antimicrobial susceptibility tested by the VetMIC™ GN-mo (version4; SVA, Uppsala, Sweden) broth microdilution microtiter panel of 13 antimicrobial substances (ampicillin, ceftazidime, ciprofloxacin, chloramphenicol, cefotaxime, florfenicol, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfamethoxazole, tetracycline, and trimethoprim) performed in the laboratory facility at the Zoonosis Science Center, Uppsala University.

¹www.krav.se

DNA Extraction and Whole Genome Sequencing

All isolates in this study were re-streaked onto Luria Bertani (LB) plates and incubated at 37°C for 18 h. Following culture, a check for potential contamination was done and 3 ml overnight cultures were made with one fresh colony into LB agar broth, incubated at 37°C for 18 h. DNA was extracted from the overnight culture using the Qiagen DNeasy blood and tissue kit in accordance with the manufacturer's instructions (Qiagen, Hilden, Germany). For all DNA extracts DNA concentration was calculated and performed using the dSDNA HS Assay kit on the Qubit Fluorometer (Thermo Fisher Scientific, Fair Lawn, NJ, United States) and OD_{260/280} in the range of 1.8–2.0 verified by the NanoDrop spectrophotometer (Thermo Fisher Scientific, Fair Lawn, NJ, United States). All DNA extracts were lyophilized and then shipped at room temperature to the Novogene sequencing facility (Novogene, Hong Kong, China) and sequenced on the Novoseq Illumina platform (Illumina, San Diego, CA, United States) which produced approximately 1 GB of 150 bp pair-end sequencing reads per isolate.

Antimicrobial Resistance Profiling and Genomic Analysis

All antimicrobial susceptibility test (AST) data were converted into either a susceptible, intermediate or resistant classification for each isolate and antibiotic drug in accordance to EUCAST clinical breakpoints (EUCAST, 2018). The use of clinical breakpoint for human medicine was due to our interest in the potential zoonotic threat, however not all tested antibiotic drugs had a clinical breakpoint. These data were then further processed in Python (3.8.3rc1 Documentation, 2020) using the matplotlib, pandas and seaborn packages. All processing of the whole genome sequences was done with open software with an in-house bioinformatics pipeline as described in Hickman et al. (2021). The genomic data processing was done on the Uppsala Multidisciplinary Centre for Advanced Computational cluster (UPPMAX). Using the described pipeline we generated a molecular report file (Supplementary File 1) to compile all detected antimicrobial resistance genes (ARGs), extra-chromosomal plasmids, sequence types by computationally multi-locus sequence typing (MLST), plasmids by plasmid multi-locus sequence typing (pMLST); and a Maximum Likelihood phylogenetic tree to assess isolate relatedness from the different farms and sample sources that was subjected to 100 times bootstraps, with confidence values included on the phylogenetic tree (Figure 1) For isolates from the same farm, where the same MLST sequence type was found and pMLST sequence type for the detected plasmids, further SNP analysis was done using Snippy version 4.0.5 (Seemann, 2015; Figure 2).

Overview of Study Isolate Collection

Forty-three *E. coli* isolates were included in this study, comprising all *E. coli* isolates that grew in the selective culturing (on C3G plates, 17 isolates). Furthermore, of the strains isolated from the non-selective culturing (on Petrifilm), the 26 isolates demonstrating resistance to the highest number of tested

antibiotics were included. This subset bacterial collection originated from 21 dairy farms where 11 farms had isolates from both a calf fecal sample and at least one environmental sample (from manure drain, manure pit, rodent feces or bird feces, and whole flies).

Isolates were compared for genomic relatedness and genomic determinants between each isolate source, when possible (1) between calves and rodent samples, (2) between calves and bird samples, (3) between calves and fly samples, (4) between bird and rodent samples, (5) between rodent and fly samples, and (6) between bird and fly samples. When possible, manure drain and manure pit samples were also compared as a reference, although there were only 4 and 2 isolates, respectively.

Ethical Approval

This study was done in accordance with the Swedish regulations, the competent authorities stated that no ethical permission was required for the sampling and the veterinarian handled the animals according to relevant ethical standards. All participating farmers were informed of the purpose and methods of the study, that participation was voluntary and anonymous and that they had the right to withdraw from the study at any time.

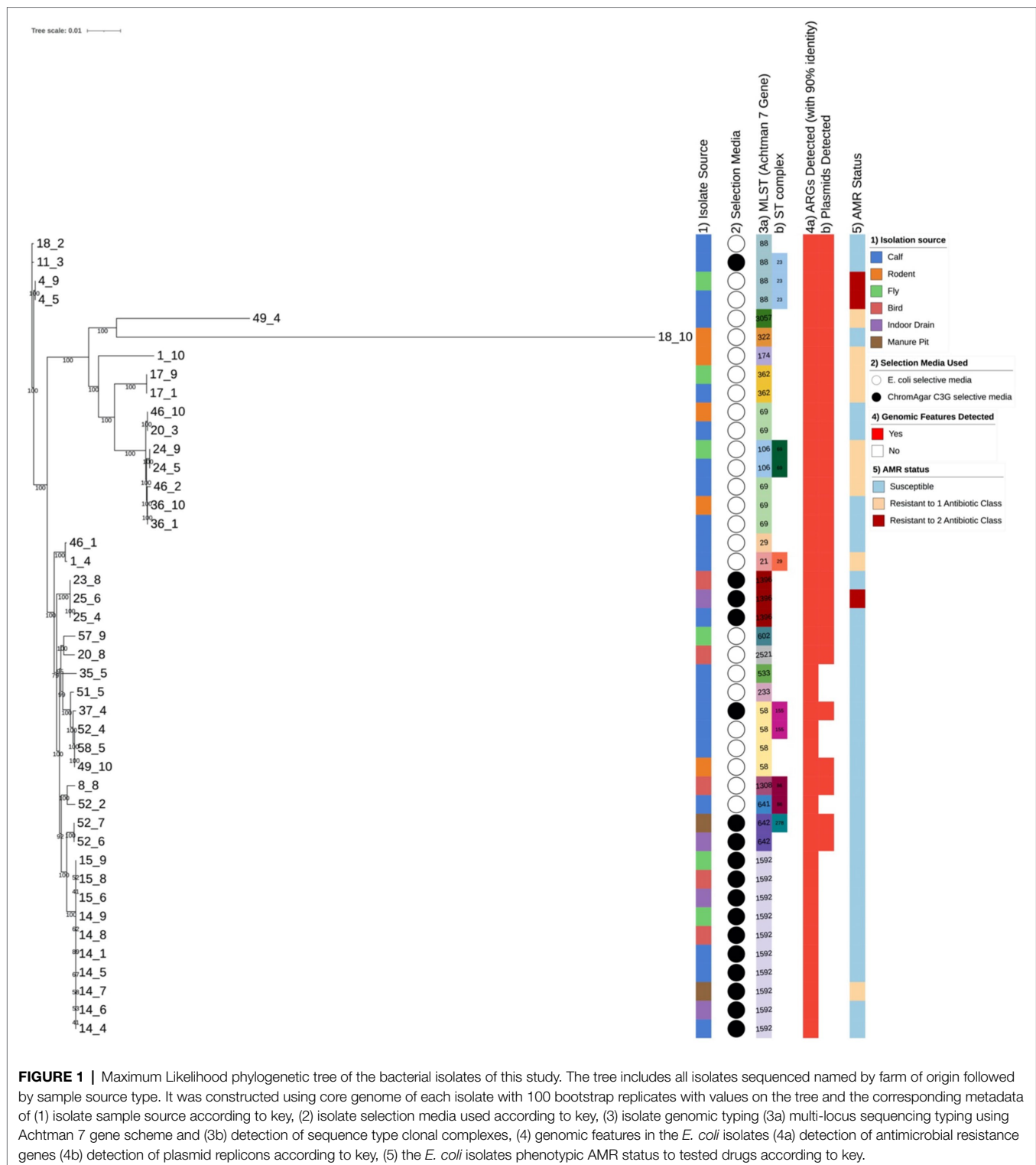
Data Availability

Raw sequence data can be obtained from the European Nucleotide Archive (ENA) under the project accession number PRJEB45447. All sequence data from computation workflow is compiled in **Supplementary File 1** and interspecies genomic difference data in **Supplementary File 2**.

RESULTS

Genomic Isolate Typing and Core Genome Analysis

Within our data we generally observed that isolates from the same farm clustered despite coming from different sources and the isolates from the same farm often had the same MLSTs (multi locus sequence types; Figure 1). This was clearly observed in farm 14 which had seven isolates sequenced: three from calves, and one each from the drain, well, bird, and fly sources, which all had Achtman 7 gene MLST ST-1592 (Figure 1—Part 3a and b). It is also important to note, that in a few cases different ST types were seen on the same farm. In the cases where different *E. coli* isolates ST were found within the same farm, we found the different ST *E. coli* isolate clustered together with those from other farms rather than with isolates from the farm where the isolate was acquired, e.g., calf isolate 46_1 from farm 46 which was ST-29 and was on a different branch of the phylogenetic tree in comparison to the other calf isolate 46_2 and the rodent sample that were both ST-69 (Figure 1). In farms 4, 17, and 24 samples from calves and flies shared the same ST, while in farms 36 and 46 calves and rodent samples shared the same ST, and in farm 15 bird and fly samples had the same ST.



Genomic Isolate and pMLST

The combining of MLST and pMLST results demonstrated interspecies sharing of clone types with the same mobile genetic elements. In seven farms, two or more sources shared the same MLST, when combined with the pMLST results there were four

farms where both the same MLST and pMLST were detected in isolates from different sources (**Figure 2**). Due to the low number of nucleotide differences between the two isolates and the fact that they also harbored the same plasmids, potential clonal sharing between the two host species is highly likely.





	Achtman 7 Gene MLST	ST Complex	pMLST Scheme	Genomic Variant		Isolates Differences		
				SNPs	INS	DEL	Complex	Total
Dairy Farm 4 	ST88	ST23 cplx	IncI1 - ST156 IncF - F2:A-B1	41	3	2	7	53
Dairy Farm 17 	ST362		IncF - F67-A-B16	15	1	1	-	17
Dairy Farm 24 	ST106	ST69 cplx	IncF - F-A-A-24	22	-	-	-	22
Dairy Farm 36 	ST69		IncF - F2:A-B16	47	-	1	3	51

FIGURE 2 | Bacterial isolates from two different sources in the same farm that share the same Achtman 7 Gene MLST, ST complex number is stated when available and pMLST types for plasmid incompatibility groups are given for IncF and IncI1 plasmids. Genomic variants differing among isolates are sorted by types (SNP = single nucleotide polymorphism, INS = genomic insertion, DEL = genomic deletion, Complex = genomic difference that could be comprised of single nucleotide polymorphism and or genomic insertion and or genomic deletion).

Phenotypic Antimicrobial Susceptibility Testing Coupled With Whole Genome Sequencing

The phenotypic susceptibility patterns in the isolates from calves and manure have been previously published (Sjöström et al., 2020). We plotted the AST results to each of the 13 drugs for each isolate according to their sample group (Supplementary Figure 1). From the group distribution plots we were unable to find correlations between groups, possibly due to the low number of isolates. Using EUCAST clinical breakpoints, we observed a generally low AMR prevalence; we did see that 21% of isolates (9/43) were resistant to one class of antibiotic drugs in the phenotypic AST data this was most frequently ampicillin or trimethoprim (Supplementary File 3). We also observed 7% of isolates (3/43) that were resistant to antibiotics from two different drug classes, the three isolates that were resistant to both ampicillin and trimethoprim (Supplementary File 3).

From our WGS results we were able to detect ARGs, AMR-related chromosomal mutations, plasmid replicons, virulence factors and plasmid types. In all isolates, 1 or more ARGs were detected and in 65% of isolates, extra-chromosomal plasmids were detected (Figure 1—Part 4a and b). In all sequenced isolates regardless of source, the *mdfA* gene was present; this gene encodes for the MdfA multi-drug efflux pump (Figure 3A). In the drug resistant (ampicillin and trimethoprim) isolate 4_5 (isolate source calf), we detected a plethora of ARGs (*blaTEM-1B*, *dfrA5*, *mdfA*, *tetA* and *sitABCD*) where *blaTEM-1B* (Seenama et al., 2019) and *dfrA5* (Thungapathra et al., 2002) are known to be responsible for the conferred resistance phenotype. In the drug resistant (ampicillin and trimethoprim) isolate 25_6 (isolate source drain), the only ARG was *mdfA* but further analysis in PointFinder (Zankari et al., 2017) revealed multiple chromosomal mutations

in *ampC*, *16S rrsB*, *16S rrsC*, *16S rrsH*, *23S*, *parC* and *pmrB* genes that might contribute to the resistance phenotype (Supplementary File 1). In the gentamicin resistant isolate 46_2 (isolate source calf), we detected the ARG *mdfA* and several RNA gene mutations *16S rrsC*, *16S rrsH*, *23S* that could confer gentamicin resistance via the mechanism of gentamicin binding to the 30S ribosome preventing protein synthesis (Eric Scholar, 2007), we also detected other chromosomal mutations in the *gyrA* and *pmrB* genes (Supplementary File 1).

In our isolate collection, we detected 15 different plasmid replicons (Figure 3B). From our pMLST results all detected plasmids belong to either the *incI1* or *incF* plasmid incompatibility groups (Figure 3C) both being clinically related groups due to the ARGs and virulence factors that can be carried in Enterobacterales (Mshana et al., 2013; Rozwandowicz et al., 2018).

DISCUSSION

Our study focused on *E. coli* isolates within the dairy farm environment where environmental samples and animal host samples were taken. From this we were able to see the difference in molecular characteristics of isolates from different sources and different farms as well as potential interspecies clone sharing.

In regards to phenotypic AMR, as reported in our earlier publication (Sjöström et al., 2020), the prevalence of AMR—especially MDR—was exceptionally low in our study compared to data from other countries (Van Boeckel et al., 2015). The isolate collection in the present study was heavily biased for antibiotic-resistant strains as we used isolates from selective culturing for beta-lactamase producing Gram negative bacteria and the most antibiotic-resistant isolates from non-selective culturing. Although stemming from a limited

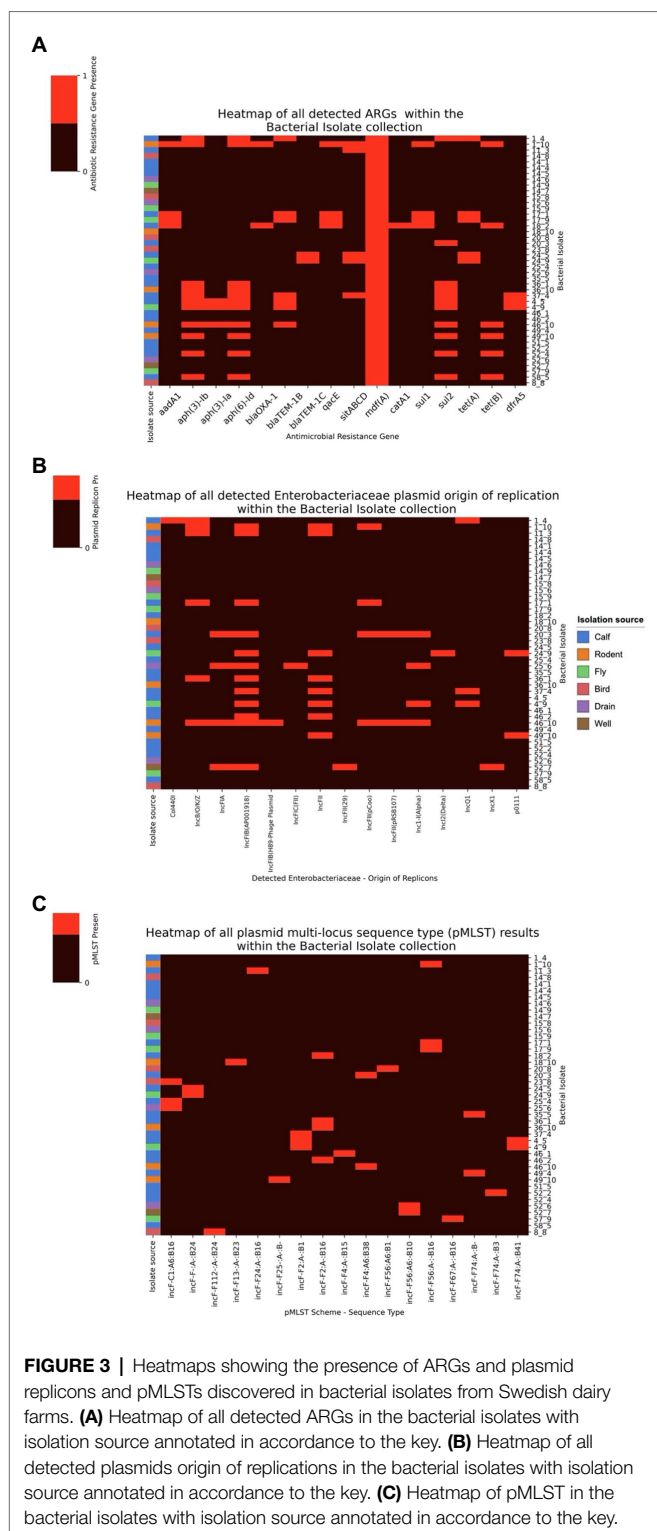


FIGURE 3 | Heatmaps showing the presence of ARGs and plasmid replicons and pMLSTs discovered in bacterial isolates from Swedish dairy farms. **(A)** Heatmap of all detected ARGs in the bacterial isolates with isolation source annotated in accordance to the key. **(B)** Heatmap of all detected plasmids origin of replications in the bacterial isolates with isolation source annotated in accordance to the key. **(C)** Heatmap of pMLST in the bacterial isolates with isolation source annotated in accordance to the key.

number of farms and based on a convenience sample, these data support the assumption that in countries with strict AMU, such as Sweden, there is a low baseline of AMR. The most frequently administered antibiotic classes for sick dairy animals in Sweden are beta-lactams, usually in the form of

penicillin G for the treatment of mastitis, or combinations of sulphonamide and trimethoprim (Olmos Antillón et al., 2020) although the latter is being phased out with new treatment recommendations. It is therefore unsurprising that this was the most frequently observed phenotypic and genotypic antimicrobial resistance in our data in regards to sulphonamide and trimethoprim resistance, we also observed high levels of ampicillin-resistant *E. coli* isolates observed within our data. Despite low baseline of AMR in our isolates from our phenotypic analysis we were able to cultivate 17 isolates on selective culture, which surprising had low to no resistance to 3rd generation cephalosporins. We suspect this may be due to modulations of copy number of ARGs within these isolates therefore demonstrating the rapid adaption to antibiotic perturbation within the dairy farm environment (Laehnemann et al., 2014).

Within our molecular results we were surprised to find the *mdfA* gene in all the sequenced isolates. We speculate the common presence of this gene is due to the resistance effect that it could produce against various chemical substances that could be present within the dairy farm environment (Edgar and Bibi, 1997; Bohn and Bouloc, 1998; Mine et al., 1998; Lewinson et al., 2003, 2004). In addition, we did observe a large variety of ARGs and AMR related chromosomal mutations despite low AMU. We speculate that this may be the results of antibiotics use periodically for treating sick animals, or a result of scavenger animals picking up ARGs from other sources (Vittecoq et al., 2016; Swift et al., 2019; Plaza-Rodríguez et al., 2021). Therefore, a part of the on-farm bacterial population will maintain these ARGs that will rise in the population when the relevant antibiotics are used. We saw a diversity in ARGs, which is likely to be a result of their dissemination by both clonal and horizontal gene transfer (HGT) mechanisms, while chromosomal mutations tend to be disseminated by clonal inheritance. In addition, within our sequenced isolates we detected extra-chromosomal plasmids showing that HGT could be occurring, with *incF* plasmid incapability groups being the most common. By whole genome sequencing we were able to characterize the isolates at higher resolution. However, due to the limited number of isolates from each farm that were whole genome sequenced we can see only a snap-shot of the ARGs, AMR-related chromosomal mutations and plasmids. Future studies to further explore our findings could involve collecting more independent source samples, e.g., other animal sources in the farm environments, selecting more bacterial isolates from each sample, and using direct sequencing by metagenomic methods to examine on genomic characteristics such as presence of ARGs and extra-chromosomal plasmids within the sample. With phenotypic AST the number of isolates that can be processed is limited and hence the combined phenotypic and genotypic characterization of isolates will not yield the same multitude of information.

The comparison of *E. coli* isolates from multiple sources showed that in 7 out of 11 farms (64%), there was sharing of MLSTs. The same MLSTs were found in both animal swabs or feces and manure collection drains and pits in the environment,

further supporting on-farm environmental spread of fecal bacteria from the animals, which was also observed by Massé et al. between calves and manure pits (Massé et al., 2021). On four farms (4, 17, 24, and 36) we saw both MLST and pMLST type match for *E. coli* isolates from different animal host samples. By whole genome sequencing we were able to see the genomic variation between the isolates; the low genomic variation observed between some isolate pairs constitutes strong evidence for cross-species *E. coli* potential clone sharing. This occurred on three farms between calf and fly sample and on one farm between calf and a rodent sample, demonstrating that clonal sharing of *E. coli* strains with extra-chromosomal plasmids does occur within the dairy farm environment such as between calves and other farm-based scavenger animals. We were unable to specify clone sharing due to the lack of set criteria for clonality by genomic differences in *E. coli* isolates unlike other bacterial species such as *Pseudomonas aeruginosa* (Lee et al., 2003).

We found previous reports of potential clone sharing between livestock cattle and birds (Fahim et al., 2019), as well as livestock cattle and deer (Singh et al., 2015). We did not obtain any samples from humans in the vicinity, such as livestock handlers. It would have been interesting to examine any potential zoonotic clone sharing but this could be an inclusion point in future studies. By whole genome sequencing rather than PCR methods we were not only able to see MLST types but due to the higher resolution from WGS we were also able to compare and observe genomic variation on a molecular level. Due to the limited number of available isolates for MLST, pMLST and genomic variation comparison, cross-sample dissemination rates could not be estimated. Nonetheless, we clearly demonstrate its occurrence and hope to further explore this on a larger scale to obtain a better picture of clone sharing potential and even investigate potential carriage in each investigated sample source to have a better idea of how AMR transmission occurs on dairy farms with low AMU.

CONCLUSION

In Sweden, a country with strictly regulated and comparatively low animal AMU, a low baseline phenotypic and genomic AMR was observed, supporting the notion that all AMU selects for AMR on dairy farms. Our results indicate on-farm transmission of *E. coli* clones between different host species

within the dairy farm environment, which has implications for the risk of environmental AMR spread.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

KS, UE, NF, SS-L, and JJ designed the study. KS collected the original samples. VA, KS, and RH did the laboratory work. RH analyzed the data and prepared figures and wrote the manuscript with critical revisions by JJ. RH and VA interpreted the data. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by the Swedish Research Council Vetenskapsrådet (to JJ, grant no. 2016-02606) and by Formas—The Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning grant no. 2014-281. The genomic data processing was enabled by resources in project [SNIC 2019/8-265 and SNIC 2019/8-11] provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

ACKNOWLEDGMENTS

We gratefully acknowledge the farmers participating in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.838339/full#supplementary-material>

REFERENCES

- Aarestrup, F. M., Wegener, H. C., and Collignon, P. (2008). Resistance in bacteria of the food chain: epidemiology and control strategies. *Expert Rev. Anti-Infect. Ther.* 6, 733–750. doi: 10.1586/14787210.6.5.733
- Bohn, C., and Bouloc, P. (1998). The *Escherichia coli* *cmlA* gene encodes the multidrug efflux pump Cmr/MdfA and is responsible for isopropyl- β -D-thiogalactopyranoside exclusion and spectinomycin sensitivity. *J. Bacteriol.* 180, 6072–6075. doi: 10.1128/JB.180.22.6072-6075.1998
- Carmo, L. P., Nielsen, L. R., Alban, L., da Costa, P. M., Schüpbach-Regula, G., and Magouras, I. (2018). Veterinary expert opinion on potential drivers and opportunities for changing antimicrobial usage practices in livestock in Denmark, Portugal, and Switzerland. *Front. Vet. Sci.* 5:29. doi: 10.3389/fvets.2018.00029
- Cheng, W. N., and Han, S. G. (2020). Bovine mastitis: risk factors, therapeutic strategies, and alternative treatments – a review. *Asian – Australas. J. Anim. Sci.* 33, 1699–1713. doi: 10.5713/AJAS.20.0156
- Cho, G.-S., Li, B., Rostalsky, A., Fiedler, G., Rösch, N., Igbinosa, E., et al. (2018). Diversity and antibiotic susceptibility of *Acinetobacter* strains From Milk powder produced in Germany. *Front. Microbiol.* 9:536. doi: 10.3389/fmicb.2018.00536
- Duse, A., Waller, K. P., Emanuelson, U., Unnerstad, H. E., Persson, Y., and Bengtsson, B. (2015). Risk factors for antimicrobial resistance in fecal

- Escherichia coli* from preweaned dairy calves. *J. Dairy Sci.* 98, 500–516. doi: 10.3168/JDS.2014-8432
- Edgar, R., and Bibi, E. (1997). MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. *J. Bacteriol.* 179, 2274–2280. doi: 10.1128/JB.179.7.2274-2280.1997
- Eric Scholar (2007). Gentamicin—an overview | ScienceDirect Topics. Available at: <https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/gentamicin> (Accessed August 13, 2021).
- EUCAST (2018). The European committee on antimicrobial susceptibility testing. Breakpoint tables for interpretation of MICs and zone diameters, version 8, 2018. Available at: http://www.eucast.org/clinical_breakpoints/ (Accessed August 20, 2021).
- European Medicines Agency year (2019). European Surveillance of Veterinary Antimicrobial Consumption (ESVAC) | European Medicines Agency Available at: <https://www.ema.europa.eu/en/veterinary-regulatory/overview/antimicrobial-resistance/european-surveillance-veterinary-antimicrobial-consumption-esvac#annual-report-on-sales-of-veterinary-antimicrobial-medicinal-products-section> (Accessed December 13, 2021).
- Fahim, K. M., Ismael, E., Khalefa, H. S., Farag, H. S., and Hamza, D. A. (2019). Isolation and characterization of *E. coli* strains causing intramammary infections from dairy animals and wild birds. *Int. J. Vet. Sci. Med.* 7, 61–70. doi: 10.1080/23144599.2019.1691378
- Firth, C. L., Kremer, K., Werner, T., and Käsbohrer, A. (2021). The effects of feeding waste milk containing antimicrobial residues on dairy calf health. *Pathogens* 10, 1–20. doi: 10.3390/pathogens10020112
- Hickman, R. A., Leangapichart, T., Lunha, K., Jiwakanon, J., Angkititakul, S., Magnusson, U., et al. (2021). Exploring antibiotic resistance burden in livestock, livestock handlers and contacts: a one health perspective. *Front. Microbiol.* 12:651461. doi: 10.3389/FMICB.2021.651461
- Jans, C., Sarno, E., Collineau, L., Meile, L., Stärk, K. D. C., and Stephan, R. (2018). Consumer exposure to antimicrobial resistant bacteria from food at Swiss retail level. *Front. Microbiol.* 9:362. doi: 10.3389/FMICB.2018.00362
- Jordbruksverket (2020). Lantbrukets djur i juni 2020 Slutlig statistik—Jordbruksverket. se. Available at: <https://jordbruksverket.se/5.38d764e917737562099cfa9c.html> (Accessed August 11, 2021).
- Karlsson, P. A., Tano, E., Jernberg, C., Hickman, R., Järhult, J., and Wang, H. (2021). Molecular characterization of multidrug resistant *Yersinia enterocolitica* from food-borne outbreaks in Sweden. *Front. Microbiol.* 12:664665. doi: 10.3389/FMICB.2021.664665
- Laehnemann, D., Peña-Miller, R., Rosenstiel, P., Beardmore, R., Jansen, G., and Schulenburg, H. (2014). Genomics of rapid adaptation to antibiotics: convergent evolution and scalable sequence amplification. *Genome Biol. Evol.* 6, 1287–1301. doi: 10.1093/GBE/EVU106
- Lee, T. W. R., Brownlee, K. G., Conway, S. P., Denton, M., and Littlewood, J. M. (2003). Evaluation of a new definition for chronic *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *J. Cyst. Fibros.* 2, 29–34. doi: 10.1016/S1569-1993(02)00141-8
- Lewinson, O., Adler, J., Poelarends, G. J., Mazurkiewicz, P., Driessen, A. J. M., and Bibi, E. (2003). The *Escherichia coli* multidrug transporter MdfA catalyzes both electrogenic and electroneutral transport reactions. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1667–1672. doi: 10.1073/PNAS.0435544100
- Lewinson, O., Padan, E., and Bibi, E. (2004). Alkalitolerance: a biological function for a multidrug transporter in pH homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14073–14078. doi: 10.1073/PNAS.0405375101
- Lhermie, G., Tauer, L. W., and Gröhn, Y. T. (2018). The farm cost of decreasing antimicrobial use in dairy production. *PLoS One* 13, e0194832. doi: 10.1371/JOURNAL.PONE.0194832
- Massé, J., Lardé, H., Fairbrother, J. M., Roy, J.-P., Francoz, D., Dufour, S., et al. (2021). Prevalence of antimicrobial resistance and characteristics of *Escherichia coli* isolates from fecal and manure pit samples on dairy farms in the province of Québec, Canada. *Front. Vet. Sci.* 8:654125. doi: 10.3389/FVETS.2021.654125
- McEwen, S. A., and Collignon, P. J. (2018). Antimicrobial resistance: a one health perspective. *Microbiol. Spectr.* 6. doi: 10.1128/microbiolspec.arba-0009-2017
- McEwen, S. A., and Fedorka-Cray, P. J. (2002). Antimicrobial use and resistance in animals. *Clin. Infect. Dis.* 34, S93–S106. doi: 10.1086/340246
- Mesbah Zekar, F., Granier, S. A., Marault, M., Yaici, L., Gassilloud, B., Manceau, C., et al. (2017). From farms to markets: gram-negative bacteria resistant to third-generation Cephalosporins in fruits and vegetables in a region of North Africa. *Front. Microbiol.* 8:1569. doi: 10.3389/FMICB.2017.01569
- Mine, T., Morita, Y., Kataoka, A., Mizushima, T., and Tsuchiya, T. (1998). Evidence for chloramphenicol/H⁺ antiport in cmr (MdfA) system of *Escherichia coli* and properties of the antiporter. *J. Biochem.* 124, 187–193. doi: 10.1093/oxfordjournals.jbchem.a022078
- Mshana, S. E., Hain, T., Domann, E., Lyamuya, E. F., Chakraborty, T., and Imirzalioglu, C. (2013). Predominance of *Klebsiella pneumoniae* ST14 carrying CTX-M-15 causing neonatal sepsis in Tanzania. *BMC Infect. Dis.* 13:466. doi: 10.1186/1471-2334-13-466
- Olesen, S. W., Lipsitch, M., and Grad, Y. H. (2020). The role of “spillover” in antibiotic resistance. *Proc. Natl. Acad. Sci. U. S. A.* 117, 29063–29068. doi: 10.1073/PNAS.2013694117/-/DCSUPPLEMENTAL
- Olmos Antillón, G., Sjöström, K., Fall, N., Sternberg Lewerin, S., and Emanuelson, U. (2020). Antibiotic use in organic and non-organic Swedish dairy farms: A comparison of three recording methods. *Front. Vet. Sci.* 7:568881. doi: 10.3389/FVETS.2020.568881
- Plaza-Rodríguez, C., Alt, K., Grobbel, M., Hammerl, J. A., Irrgang, A., Szabo, I., et al. (2021). Wildlife as sentinels of antimicrobial resistance in Germany? *Front. Vet. Sci.* 7:627821. doi: 10.3389/FVETS.2020.627821
- Rozwandowicz, M., Brouwer, M. S. M., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., et al. (2018). Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J. Antimicrob. Chemother.* 73, 1121–1137. doi: 10.1093/JAC/DKX488
- Seemann, T. (2015). Snippy: fast bacterial variant calling from NGS reads. Snippy: fast bacterial variant calling from NGS re. Available at: <https://github.com/tseemann/snippy> (Accessed June 15, 2020).
- Seenama, C., Thamlikitkul, V., and Ratthawongjirakul, P. (2019). Multilocus sequence typing and blaESBL characterization of extended-spectrum beta-lactamase-producing *Escherichia coli* isolated from healthy humans and swine in Northern Thailand. *Infect. Drug Resist.* 12, 2201–2214. doi: 10.2147/IDR.S209545
- Singh, P., Sha, Q., Lacher, D. W., Del Valle, J., Mosci, R. E., Moore, J. A., et al. (2015). Characterization of enteropathogenic and Shiga toxin-producing *Escherichia coli* in cattle and deer in a shared agroecosystem. *Front. Cell. Infect. Microbiol.* 5:29. doi: 10.3389/FMICB.2015.00029
- Sjöström, K., Hickman, R. A., Tepper, V., Olmos Antillón, G., Järhult, J. D., Emanuelson, U., et al. (2020). Antimicrobial resistance patterns in organic and conventional dairy herds in Sweden. *Antibiotics* 9:834. doi: 10.3390/antibiotics9110834
- SVA (2019). Rapport: Swedres-Svarm, 2019 SVA. Available at: <https://www.sva.se/vi-erbjuder/publikationer/rapport-swedres-svarm%2C-2018/c-28/c-83/p-1260> (Accessed December 14, 2021).
- Swift, B. M. C., Bennett, M., Waller, K., Dodd, C., Murray, A., Gomes, R. L., et al. (2019). Anthropogenic environmental drivers of antimicrobial resistance in wildlife. *Sci. Total Environ.* 649, 12–20. doi: 10.1016/J.SCITOTENV.2018.08.180
- Thungapathra, M., Amita, S., Sinha, K. K., Chaudhuri, S. R., Garg, P., Ramamurthy, T., et al. (2002). Occurrence of antibiotic resistance gene cassettes aac(6′)-Ib, dfrA5, dfrA12, and ereA2 in class I Integrons in non-O1, non-O139 *Vibrio cholerae* strains in India. *Antimicrob. Agents Chemother.* 46, 2948–2955. doi: 10.1128/AAC.46.9.2948-2955.2002
- Tiseo, K., Huber, L., Gilbert, M., Robinson, T. P., and Boeckel, T. P. Van (2020). Global trends in antimicrobial use in food animals from 2017 to 2030. *Antibiotics* 9, 918. doi: 10.3390/ANTIBIOTICS9120918
- Van Boeckel, T. P., Brower, C., Gilbert, M., Grenfell, B. T., Levin, S. A., Robinson, T. P., et al. (2015). Global trends in antimicrobial use in food animals. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5649–5654. doi: 10.1073/pnas.1503141112
- Vittecoq, M., Godreuil, S., Prugnolle, F., Durand, P., Brazier, L., Renaud, N., et al. (2016). Antimicrobial resistance in wildlife. *J. Appl. Ecol.* 53, 519–529. doi: 10.1111/1365-2664.12596
- Zankari, E., Allesøe, R., Joensen, K. G., Cavaco, L. M., Lund, O., and Aarestrup, F. M. (2017). PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J. Antimicrob. Chemother.* 72, 2764–2768. doi: 10.1093/jac/dkx217

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may

be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hickman, Agarwal, Sjöström, Emanuelson, Fall, Sternberg-Lewerin and Järhult. This is an open-access article distributed under the terms of the

Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Rapid, Whole Genome Sequencing Assay for Detection and Characterization of Novel Coronavirus (SARS-CoV-2) Clinical Specimens Using Nanopore Sequencing

OPEN ACCESS

Edited by:

Mel C. Melendrez,
Anoka-Ramsey Community College,
United States

Reviewed by:

Markus Antwerpen,
Bundeswehr Institute of Microbiology,
Germany
Jun Hang,
Walter Reed Army Institute
of Research, United States
Martina Rueca,
National Institute for Infectious
Diseases Lazzaro Spallanzani
(IRCCS), Italy

*Correspondence:

Maria T. Arévalo
maria.t.arevalo.civ@army.mil

Specialty section:

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

Received: 01 April 2022

Accepted: 09 May 2022

Published: 06 June 2022

Citation:

Arévalo MT, Karavis MA,
Katoski SE, Harris JV, Hill JM,
Deshpande SV, Roth PA, Liem AT and
Bernhards RC (2022) A Rapid, Whole
Genome Sequencing Assay for
Detection and Characterization of
Novel Coronavirus (SARS-CoV-2)
Clinical Specimens Using Nanopore
Sequencing.
Front. Microbiol. 13:910955.
doi: 10.3389/fmicb.2022.910955

Maria T. Arévalo^{1,2*}, Mark A. Karavis², Sarah E. Katoski², Jacquelyn V. Harris²,
Jessica M. Hill³, Samir V. Deshpande², Pierce A. Roth³, Alvin T. Liem³ and
R. Cory Bernhards²

¹ Defense Threat Reduction Agency, Aberdeen Proving Ground, MD, United States, ² United States Army Combat Capabilities Development Command Chemical Biological Center, Aberdeen Proving Ground, MD, United States, ³ DCS Corporation, Belcamp, MD, United States

A new human coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged at the end of 2019 in Wuhan, China that caused a range of disease severities; including fever, shortness of breath, and coughing. This disease, now known as coronavirus disease 2019 (COVID-19), quickly spread throughout the world, and was declared a pandemic by the World Health Organization in March of 2020. As the disease continues to spread, providing rapid characterization has proven crucial to better inform the design and execution of control measures, such as decontamination methods, diagnostic tests, antiviral drugs, and prophylactic vaccines for long-term control. Our work at the United States Army's Combat Capabilities Development Command Chemical Biological Center (DEVCOM CBC) is focused on engineering workflows to efficiently identify, characterize, and evaluate the threat level of any potential biological threat in the field and more remote, lower resource settings, such as forward operating bases. While we have successfully established untargeted sequencing approaches for detection of pathogens for rapid identification, our current work entails a more in-depth sequencing analysis for use in evolutionary monitoring. We are developing and validating a SARS-CoV-2 nanopore sequencing assay, based on the ARTIC protocol. The standard ARTIC, Illumina, and nanopore sequencing protocols for SARS-CoV-2 are elaborate and time consuming. The new protocol integrates Oxford Nanopore Technology's Rapid Sequencing Kit following targeted RT-PCR of RNA extracted from human clinical specimens. This approach decreases sample manipulations and preparation times. Our current bioinformatics pipeline utilizes Centrifuge as the classifier for quick identification of SARS-CoV-2 and RAMPART software for verification and mapping of reads to the full SARS-CoV-2 genome. ARTIC rapid sequencing results,

of previous RT-PCR confirmed patient samples, showed that the modified protocol produces high quality data, with up to 98.9% genome coverage at >1,000x depth for samples with presumably higher viral loads. Furthermore, whole genome assembly and subsequent mutational analysis of six of these sequences identified existing and unique mutations to this cluster, including three in the Spike protein: V308L, P521R, and D614G. This work suggests that an accessible, portable, and relatively fast sample-to-sequence process to characterize viral outbreaks is feasible and effective.

Keywords: nanopore sequencing, COVID-19, SARS-CoV-2, whole genome sequencing, whole genome assembly

INTRODUCTION

A new coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in Wuhan, China in 2019. It was quickly sequenced and identified as being related to severe acute respiratory syndrome (SARS) virus, with some homology to bat coronaviruses (Andersen et al., 2020; Lu et al., 2020; Wang et al., 2020; Wu et al., 2020). The disease caused by this novel coronavirus, COVID-19, was discovered in a cluster of pneumonia cases associated with a Huanan seafood market. At the time, the most common symptoms reported at the onset of illness were fever, cough, myalgia, and fatigue (Huang et al., 2020). The hospitalized patients all had pneumonia with acute respiratory distress syndrome (ARDS) as a common complication and a high fatality rate of 15% (Huang et al., 2020). Since then, the virus and disease have spread, causing a pandemic that has yet to be controlled.

SARS-CoV-2 has a single-stranded, positive sense RNA genome with a 5' cap and poly A tail (Romano et al., 2020). Its RNA genome has 14 open-reading frames (ORFs) that encode 16 non-structural proteins (Nsp1-16) that are involved in replication, and 4 structural proteins (spike—S, envelope—E, membrane—M, and nucleocapsid—N) that are assembled into the virion (Romano et al., 2020). When it comes to structural proteins, the trimeric S protein is particularly important because it mediates host cell receptor binding and entry. The S protein is also a main target of the neutralizing antibody response and thus, the majority of developing vaccine and antibody-based therapeutic approaches are directed against it (Korber et al., 2020). Monitoring changes in this protein will be particularly important because mutations in this protein may alter the phenotype of the virus, transmission, and effect efficacy of vaccines and other medical countermeasures that have been developed using strains that were identified and isolated in Wuhan, China early on in the pandemic.

Clinical isolates from around the globe have been sequenced, shared and published in databases such as GenBank and GISAID (Shu and McCauley, 2017). GISAID as an example, has received over 10 million genome sequence submissions as of April 12, 2022¹. Furthermore, GISAID has introduced a nomenclature system for major clades (GISAID, 2020). Classification is based on marker mutations from the early split of clades S and L,

evolution of L into V and G, and then G into GH, GR, and then GV. The current GISAID clades (GISAID, 2020) are shown in **Table 1**, with comparison to other classification schema, and including variants of concern as designated by the WHO (WHO, 2022).

The majority of whole genome sequencing for molecular epidemiology employ Illumina next-generation sequencing technology (Seth-Smith et al., 2019); this platform is currently considered the gold-standard for data quality and accuracy. However, Illumina equipment has a large footprint in terms of space, power consumption and requires a certified technician for set up and maintenance, making the technology less operable in the field and less obtainable in remote regions of the world where new pathogens of interest may emerge. Oxford Nanopore Technology's (ONT, Oxford, United Kingdom) MinION sequencers, which rely on use of nanopores for sequencing, are hand-held portable devices that are accessible and easy to set up anywhere, without an ONT technician. Thus, the MinION handheld sequencers are an attractive alternative technology for rapid and fieldable deployment. Moreover, the ARTIC Network has been developing end-to-end protocols utilizing this technology to sequence RNA viruses that include Ebola, influenza, and more recently SARS-CoV-2. The original ARTIC SARS-CoV2 protocol was released in early January 2020, enabling sequencing in different countries of the world, early on in the pandemic (Tyson et al., 2020). It has since become a widely used approach and more recently, a head-on comparison of ARTIC sequencing assays with Illumina versus nanopore sequencing showed these yielded similar results with respect to coverage and identification of variants (Charre et al., 2020).

The ARTIC Network's SARS-CoV-2 protocol for nanopore sequencing relies on direct amplification of the reverse-transcribed viral genome using a tiled, multiplexed, primer approach. The primer scheme is based on GenBank accession MN908947, released shortly after identification of the virus (Artic-Network, 2020). The protocol is highly sensitive, making it possible to sequence viruses directly from clinical samples. The protocol has also been adopted by investigators worldwide; primer sets have been published and are available commercially as a full set. However, while effective and highly sensitive, the ARTIC SARS-CoV-2 protocol is elaborate and time-consuming. In this study, we describe a modified ARTIC process that decreases sample manipulation and preparation times, resulting in high quality data that can be used downstream for viral genome assembly and analyses. The new protocol was evaluated

¹<https://www.gisaid.org/>

TABLE 1 | SARS-CoV-2 GISAID clade classifications, corresponding pango lineage, and variants of concern.

GISAID clade	Pango lineage	Marker variants	Variants of concern
S	A	C8782T, T28144C includes NS8-L84S	
L	B	C241, C3037, A23403, C8782, G11083, G26144, T28144 (early clade markers in WIV04- GISAID reference sequence)	
V	B.2	G11083T, G26144T NSP6-L37F + NS3-G251V	
G	B.1	C241T, C3037T, A23403G includes S-D614G	
GH	B.1.*	C241T, C3037T, A23403G, G25563T includes S-D614G + NS3-Q57H	Beta (B.1.3151)
GR	B.1.1.1	C241T, C3037T, A23403G, G28882A includes S-D614G + N-G204R	Gamma (P.1 or B.1.1.28.1)
GV	B.1.177	C241T, C3037T, A23403G, C22227T includes S-D614G + S-A222V	
GRY	B.1.1.7	C241T, C3037T, 21765-21770del, 21991-21993del, A23063T, A23403G, G28882A includes S-H69del, S-V70del, S-Y144del, S-N501Y + S-D614G + N-G204R	Alpha
GK	B.1.617.2	C241T, C3037T, A23403G, C22995A S-D614G + S-T478K	Delta
GRA	B.1.1.529	A67V, del69-70, T95I, del142-144, Y145D, del211, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	Omicron (includes BA.1-BA.5 or B.1.1.529.1-B.1.1.529.5, XE-recombinant BA.1/BA.2)

using a panel of CoV-2 positive and negative clinical samples as previously diagnosed by RT-PCR assays.

MATERIALS AND METHODS

Cohort

Twenty positive samples and twenty negative samples were received from Justin T. Bacca's group at the University of New Mexico. The samples were tested by their reference lab *via* EUA cleared RT-PCR assays and were provided as TRIzol-inactivated samples (2 parts Trizol to 1 part sample in VTM).

RNA Extractions

The total RNA was extracted from the TRIzol-inactivated clinical samples by using the Direct-zol RNA MicroPrep Kit (catalog number R2060) from Zymo Research. The manufacturer's instructions were followed with the exception of the elution volume being doubled from a volume of 15 to 30 μ L. After the RNA was extracted, the concentration and quality of the RNA was determined by Nanodrop analysis.

Library Preparations

The RNA was converted to cDNA and amplified using a targeted approach developed by the ARTIC Network. The ARTIC "nCoV-2019 sequencing protocol" (Quick, 2020) was followed precisely, except for a couple steps as described and can be visualized in **Figure 1A**. The ARTIC's V3 primer panel (Tyson et al., 2020), consisting of two pools of primer pairs with 98 primers in each pool were used in the amplification of the cDNA (ARTIC nCoV-2019 V3 Panel, 500rxn, Catalog number 10006788) from Integrated DNA Technologies (Coralville, IA, United States). Two PCR reactions per sample were prepared, one using the first set of the primer pool (primer pool #1)

and another using the second primer set (primer pool #2). During the PCR amplification step, 10 μ L of cDNA was used instead of 2.5 μ L for each reaction to maximize the amount of product going into the amplification. Thirty cycles of amplification was performed using the Q5 Hot Start High Fidelity DNA Polymerase (M0493L, New England Biolabs/NEB, Ipswich, MA, United States) and Applied Biosystems GeneAmp 9700 thermocycler. After amplification, the two ARTIC PCR reactions per sample are combined together, and this is followed by an AMPure XP (A63880, Beckman-Coulter, Indianapolis, IN, United States) bead DNA clean-up step. The concentration of the eluted DNA was determined by Qubit analysis with the Qubit dsDNA HS Assay Kit (Q33231, Thermo Fisher Scientific, Waltham, MA, United States). Using the standard protocol, the amplicons were prepared for barcode ligation using the Ultra II End Prep reactions included in the NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (Catalog # E7180S), barcoded using the NEBNext Ultra II Ligation Module (Catalog # E7595S, NEB) with ONT's Native Barcoding Expansion 1–12 kit (EXP-NBD104). This barcoding process adds 42 min of preparation time. The ARTIC amplicons can then be pooled together for multiplexed runs, another bead clean-up is performed, and the library preparation is completed using the T4 DNA Ligase included in the NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing kit and the Ligation Sequencing Kit (SQK-LSK109, ONT). As an alternative, less time-consuming, and more streamlined approach to the standard ARTIC v3 protocol for preparing multiplexed libraries for nanopore sequencing, we used the Rapid Barcoding Kit (SQK-RBK004, ONT) for one-step, transposase-based fragmentation and barcoding of the two, pooled ARTIC PCR reactions for each sample (**Figure 1B**). The rapid adapter (RAP) is then added to the barcoded amplicons, and the final library is prepared for sequencing on the MinION. Finally, for sequencing and analysis

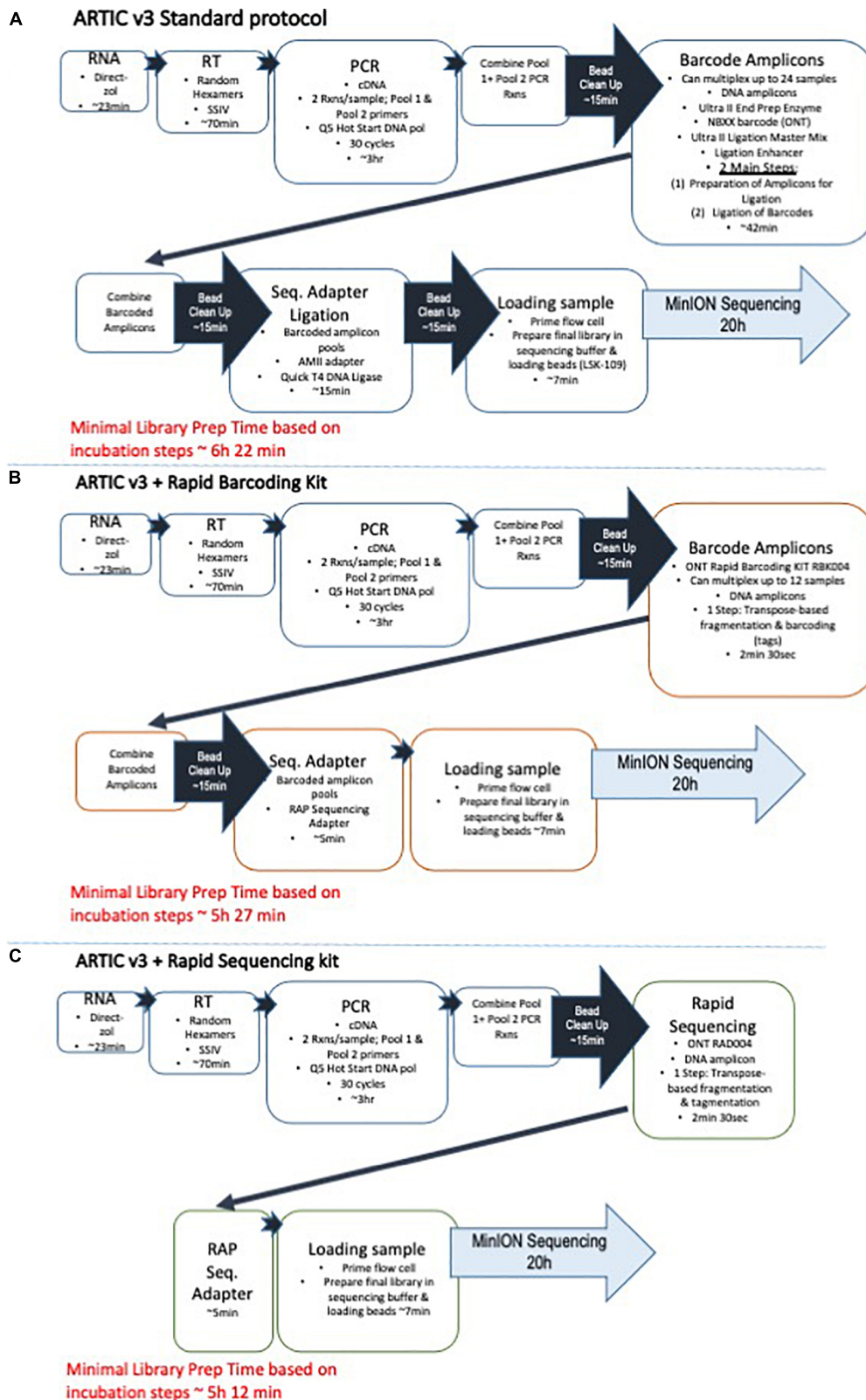


FIGURE 1 | Artic protocol and modified versions. Schematics for the COVID-19 sequencing workflows are shown starting with (A) the standard Artic protocol with v3 primers and followed by (B) the Artic protocol modified for use with the Rapid Barcoding Sequencing kit for multiplexed samples, and (C) the Artic protocol modified for use with the Rapid Sequencing kit for rapid sequencing of individual samples.

of single samples, we pooled the two PCR reactions for each sample, performed an AMPure bead clean-up, and then used the Rapid Sequencing Kit (SQK-RAD004, ONT) for one-step, transposase-based fragmentation. The process is completed by addition of the RAP sequencing adapter and the final library is prepared for sequencing (**Figure 1C**).

Sequencing

All sequencing runs were performed using either a MinION connected to a MinIT (or the Mk1C with MinKNOW 19.12 software). Each run was performed using a MinION flow cell (FLO-MIN 106 R9 version; Mk 1Spot-ON). For experiments using barcoded samples, four samples were run per flow cell. Samples that were not barcoded were run individually; one per flow cell. Prior to every run, the flow cells were assessed for the amount of total active nanopores available for sequencing, as per manufacturer's protocol. Live basecalling with high accuracy (ONT Guppy 3.2.10) was selected for the run if the final concentration of the library was less than 10 ng/μL. If the final library concentration was too large, the basecalling would lag behind and significantly extend the time of the run. For these samples, high-accuracy basecalling was performed after the run was complete (ONT Guppy 3.2.10). For each run, the sequencing time was set to end at 20 h.

Analysis

After basecalling, passed fastq files were processed by an in-house metagenomic pipeline (**Figure 2**, panel 1 outline). If samples were barcoded, demultiplexing was performed using qcat (ONT, version 1.0.1) with the minimum barcode quality set to 10. Centrifuge (Johns Hopkins University; Baltimore, MD, United States) was then run to align reads using Centrifuge's pre-indexed database (h + p + v + c.tar.gz) to determine

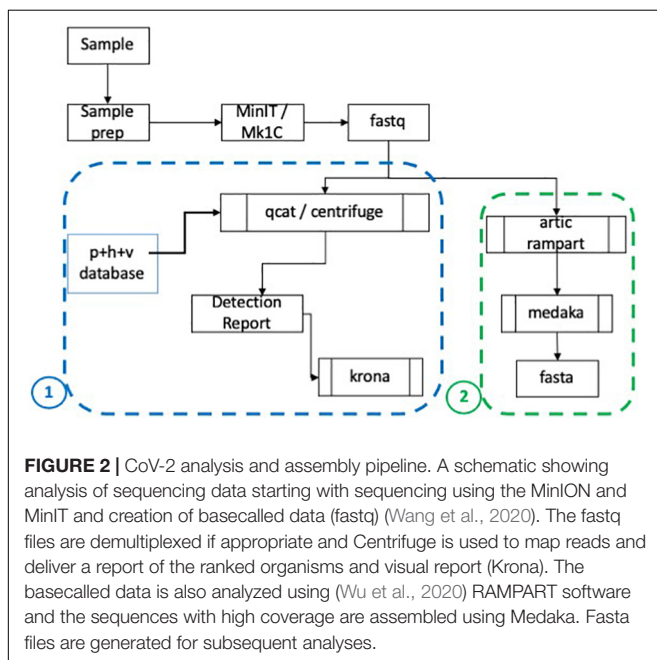
the organisms present in each sample. The Centrifuge database includes human, prokaryotic and viral genomes, and has been updated to include 106 SARS-CoV-2 genomes². A report containing the top ten organisms sorted by number of reads aligned, and excluding human hits, was generated as shown in **Table 2**. The data was also visually compiled using Krona (National Biodefense Analysis and Countermeasures Center; Frederick, MD, United States) to produce a visual and interactive report as shown in analysis pipeline schematic (**Figure 2**).

The ARTIC Network's RAMPART (Read Assignment, Mapping, and Phylogenetic Analysis in Real Time) software was downloaded with instructions from: https://hub.docker.com/r/ontresearch/artic_rampart. RAMPART was used to align passed reads to the SARS-CoV-2 genome (Wuhan-Hu-1 isolate; Accession MN908947) and provided read mapping statistics and visual representations of coverage across the genome.

Additionally, bioinformatics tools for ARTIC³ version artic 1.1.3 were downloaded and installed inside an anaconda ver 4.8.3⁴ environment to facilitate reference based genome assembly. Assemblies with at least 89% coverage of the genome with sequencing depth of 10X were selected for whole genome assembly. Assembly was performed using a medaka⁵ based workflow for our analysis (**Figure 2**, panel 2 outline). The assemblies were submitted to GenBank and accession numbers are provided in **Table 3**. These sequences are also provided in a fasta file as **Supplementary Material**. Based on feedback from the GenBank submission process, we had to make corrections to P12 and P15 assemblies. There was a frameshift in P12 and upon comparison with the reference sequence at position 26,655, we found that there was an extra T after a series of 5 Ts. Since nanopore sequencing can result in errors in homopolymer regions (Watson and Warr, 2019), we corrected the assembly by removing the extra T. For P15, we found a deletion of one nucleotide in P15 sequence TTTCTTCAC was causing a frameshift and stop codons in the translation as compared to the reference sequence of TTGGCTTCAC at position 3011. This was also attributed to being error due the homopolymer region and was corrected with the addition of a T to TTTTCTTCAC.

Classification and Mutational Analyses

A FASTA file containing the six assembled genomes was uploaded to the CoVServer app (A*STAR Bioinformatics Institute, Singapore) located on the GISAID site. The GISAID reference strain, hCoV-10/Wuhan/WIV04/2019, was used as the reference strain for comparison. The app computes and provides a list of variations and mutations in the genome. It also provides clade classification as per the GISAID classification scheme. CoV-GLUE (Singer et al., 2020), as enabled by GISAID, was used to look at the frequency of mutations in CoV-2 as observed in GISAID sequences. CoV-GLUE contains database of reported CoV-2 amino acid replacements, insertions, and deletions.



²<https://ccb.jhu.edu/software/centrifuge/index.shtml>

³<https://github.com/artic-network/fieldbioinformatics>

⁴www.anaconda.com

⁵<https://github.com/nanoporetech/medaka>

TABLE 2 | Pilot study evaluating ARTIC-based assays for sequencing SARS-CoV-2 from clinical specimens.

Sample ID	RT-PCR Diagnosis	Seq Assay	Centrifuge				RAMPART					
			Total Reads	# CoV-2 Reads	# Unique CoV-2 Reads	% CoV-2 Reads	#CoV-2 Reads	Median Length	% CoV-2 Genome	Depth > 10x	Depth > 100x	Depth > 1,000x
BEI RNA	n/a	ARTIC v3	179,391	170,400	168,872	95.0	119,188	500	97.3	99.8	99.8	83.1
BEI RNA	n/a	ARTIC + Rapid	2,256,855	1,896,808	1,896,808	84.0	2,148,026	290	95.2	99.9	99.8	99.8
P02	+	ARTIC v3	547,474	483,403	470,900	88.3	327,577	490	77.74	98	92.9	72
P02	+	ARTIC + Rapid Barcoding	24,000	15,507	15,507	64.6	ND	ND	ND	ND	ND	ND
P02	+	ARTIC + Rapid	443,368	370,676	370,676	83.6	418,818	270	94.46	99.1	95.8	80.6
P03	+	ARTIC v3	344,910	73,885	73,867	21.4	29,429	300	15.24	56.5	20.6	15.8
P03	+	ARTIC + Rapid Barcoding	11,000	1,011	1,011	9.2	ND	ND	ND	ND	ND	ND
P03	+	ARTIC + Rapid	645,017	490,929	490,929	76.1	545,826	270.00	84.62	48.20	41.40	30.10
P04	+	ARTIC v3	543,005	119,657	118,894	22.0	39,939	310	12.2	73.1	22.6	15.8
P04	+	ARTIC + Rapid Barcoding	13,000	1,597	1,597	12.3	ND	ND	ND	ND	ND	ND
P04	+	ARTIC + Rapid	861,734	121,254	120,491	14.1	748,263	260.00	86.21	64.90	53.80	35.90
N01	-	ARTIC v3	568,266	59,246	59,246	10.4	10,482	300	3.02	82.6	22.5	4.1
N01	-	ARTIC + Rapid Barcoding	32,000	–	–	0.0	ND	ND	ND	ND	ND	ND
N01	-	ARTIC + Rapid	578,916	446,225	443,662	77.1	495,874	280	85.66	62	58.8	43.8
Neg. Ctl 1/Water	n/a	ARTIC + Rapid	25,173	7,487	7,487	29.7	8,336	190	33.11	1.5	1.4	1.3
Neg. Ctl 2/Water	n/a	ARTIC + Rapid	12,131	–	–	0.0	0	-	0	0	0	0

Nextclade⁶ was used to generate CoV-2 clade assignments. Phylogenetic Assignment of Named Global Outbreak LINEages (pangolin, version v2.3.5, lineages version 2021-03-16⁷) was used to assign genome sequences to global CoV-2 lineages (Rambaut et al., 2020). After P12 and P15 sequences were corrected, the analyses were performed again on all the sequences using the most current versions of each of the apps: GISAID CoVsurver, Nextclade v.14.1, and Pangolin v4.0.6.

RESULTS

The goal of the study was to test the ARTIC CoV-2 protocol and make modifications to simplify and streamline the approach to make it faster and more accessible in its employment. We first performed a pilot, proof of concept study using a small set of previously diagnosed clinical samples. Three positive samples (P02, P03, P04) and one negative (N01), as previously diagnosed *via* EUA-cleared RT-PCR assays, were tested in this pilot study. CoV-2 RNA that was obtained from BEI Resources (NR-52285, from isolate USA-WA1/2020, Accession MN985325.1) was used as? A positive control for the assay and using the standard ARTIC protocol, 95 and 97% of 170K reads mapped to the CoV-2 genome as analyzed by Centrifuge and RAMPART, respectively (Table 2). Of note, the depth of coverage at over 100X was 99.8%, while at over 1,000X it was 83%. Over 344K total reads were obtained from each of the barcoded positive samples, but only a fraction of them mapped to CoV-2 (21–22% for P03 and P04, 88% for P02) as determined by analysis using Centrifuge and (12, 15, and 78% for P03, P04, and P02) by RAMPART. In this sample set, P02 had the most CoV-2 mapped reads (327,577), and depth of coverage of 98% at 10X as determined by RAMPART analysis. P03 and P04 had less CoV-2 mapped reads, with less of the genome covered (56.5–73.31%) at 10X depth of coverage. Unexpectedly, N01 also had several CoV-2 specific reads: 59,246 and 10,482 by Centrifuge and RAMPART, respectively. Next, we used the same amplified PCR products, but barcoded and completed sequencing library preparation using the Rapid Barcoding Kit. While this method identified CoV-2 specific reads in the positive

samples by Centrifuge analysis, the number of reads returned averaged 60-fold less in comparison to the standard ARTIC v3 protocol. The ARTIC with Rapid Barcoding Kit approach did not identify any reads for the N01 sample. Finally, amplified PCR sample from the pilot samples were tested individually (and with no barcoding) *via* the ARTIC with Rapid Sequencing approach. This approach generated the most CoV-2 specific reads for all the positive samples, including the control RNA from BEI, especially as analyzed by the RAMPART pipeline. In addition, a greater percentage of the reads were mapped to CoV-2 by RAMPART (range of 84–95%) for P02–P04. The depth of coverage at over 10X for these ARTIC with Rapid sequencing positive samples were 99, 48, and 65% for P02, P03, and P04. Finally, N01 as prepared by the ARTIC with Rapid sequencing approach yielded a high number of reads (446,225 and 495,874 by Centrifuge and RAMPART, respectively) with a 62% depth of coverage at over 10X. Negative control water samples were also amplified *via* PCR and prepared using the Rapid Kit to generate background/baseline levels to expect from this assay. Two samples returned 0 and 7,487 reads by Centrifuge analysis and 0 and 8,336 by RAMPART of analysis with a depth of coverage of 1.5% at 10X depth of coverage for the negative sample that returned reads. Since the pilot study revealed that the ARTIC with Rapid sequencing approach was promising due to reduction in steps, time, and increase in CoV-2 specific reads and coverage, additional positive and negative samples were prepared and sequenced using the ARTIC with Rapid sequencing approach.

In total, 20 positive samples (P01–P20) and 12 negative samples (N01–N12) were sequenced using the ARTIC with Rapid sequencing approach (Table 4). The number of reads that were specifically mapped to CoV-2 as determined by Centrifuge ranged from 1,511 to 4,538,653, with a median of 150,000 in the positive sample set. The sample with the least reads was P10, which was below the estimated baseline of the assay (Table 4 and Figure 3A), but the other positive samples generated at least 15.8 K reads. For a subset of samples that were analyzed by both Centrifuge and RAMPART pipelines, CoV-2 reads ranged from 15,807 to 4,538,653 (median 336,264) using Centrifuge and 18,607–5,065,568 (median 482,322) using RAMPART. This indicated that the RAMPART pipeline was able to map more CoV-2 reads than Centrifuge (Table 4 and Figure 3B). In

⁶clades.nextstrain.org

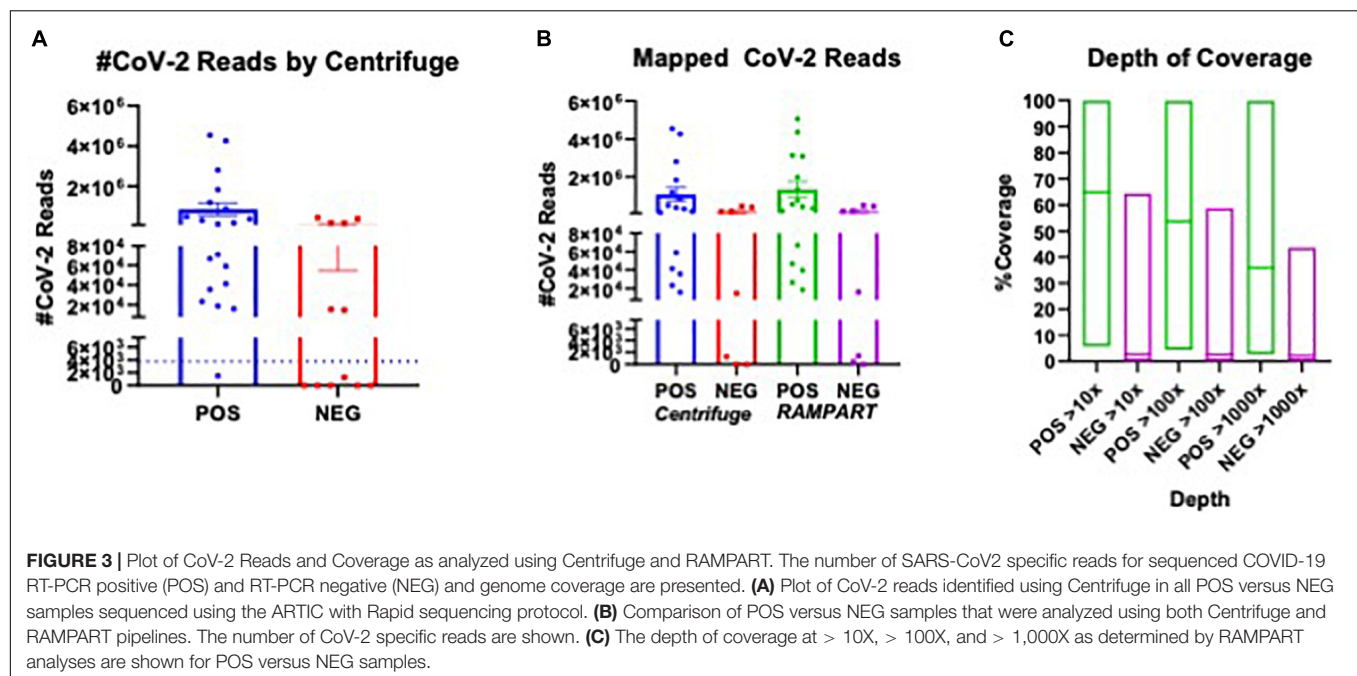
⁷https://pangolin.cog-uk.io/

TABLE 3 | Assemblies with GenBank accession numbers.

ID	Seq assay	Assembly	Accession #
BEI RNA	ARTIC v3	CoV2_Isk109/ARTIC/medaka MN908947.3	ON311289
BEI RNA	ARTIC + Rapid	CoV2_rad004/ARTIC/medaka MN908947.3	ON311149
P02	ARTIC v3	1P2_Isk109/ARTIC/medaka MN908947.3	ON311288
P02	ARTIC + Rapid	1P2_rad004/ARTIC/medaka MN908947.3	ON310862
P11	ARTIC + Rapid	P11_rad004/ARTIC/medaka MN908947.3	ON310894
P12 fixed	ARTIC + Rapid	P12_rad004 (organism = Severe acute respiratory syndrome coronavirus 2) (isolate = P12) reference assembly to MN908947.3	ON398848
P14	ARTIC + Rapid	P14_rad004/ARTIC/medaka MN908947.3	ON310966
P15 fixed	ARTIC + Rapid	P15_rad004 (organism = Severe acute respiratory syndrome coronavirus 2) (isolate = P15) reference assembly to MN908947.3	ON398955
P18	ARTIC + Rapid	P18_rad004/ARTIC/medaka MN908947.3	ON311005

TABLE 4 | Sequencing of previously diagnosed clinical specimens by ARTIC with rapid sequencing approach.

Sample ID	RT-PCR Diagnosis	Seq Assay	Centrifuge							RAMPART		
			Total Reads	# CoV-2 Reads	# Unique CoV-2 Reads	% CoV-2 Reads	#CoV-2 Reads	Median Length	% CoV-2 Genome	Depth > 10x	Depth > 100x	Depth > 1,000x
P01	+	ARTIC + Rapid	462,000	71,138	71,138	15.4	ND	ND	ND	ND	ND	ND
P02	+	ARTIC + Rapid	443,368	370,676	370,676	83.6	418,818	270	94.46	99.1	95.8	80.6
P03	+	ARTIC + Rapid	645,017	490,929	490,929	76.1	545,826	270	84.62	48.2	41.4	30.1
P04	+	ARTIC + Rapid	861,734	121,254	120,491	14.1	748,263	260	86.21	64.9	53.8	35.9
P05	+	ARTIC + Rapid	44,838	18,779	18,779	41.9	ND	ND	ND	ND	ND	ND
P06	+	ARTIC + Rapid	5,136,573	4,260,587	3,808,126	82.9	4,361,013	280	84.9	99.8	99.8	98.9
P07	+	ARTIC + Rapid	1,523,922	1,189,946	1,189,171	78.1	1,331,034	290	87.34	62.2	61.3	56.2
P08	+	ARTIC + Rapid	701,282	66,782	66,782	9.5	ND	ND	ND	ND	ND	ND
P09	+	ARTIC + Rapid	2,245,950	59,114	59,114	2.6	66,862	340	2.98	57.6	42.2	18
P10	+	ARTIC + Rapid	16,691	1,511	1,511	9.1	ND	ND	ND	ND	ND	ND
P11	+	ARTIC + Rapid	3,417,719	2,816,093	2,724,325	82.4	3,087,854	290	90.35	99.8	98.3	91.5
P12	+	ARTIC + Rapid	861,623	301,852	293,972	35.0	339,009	280	60.65	89	80.2	54.5
P13	+	ARTIC + Rapid	360,276	23,244	23,244	6.5	26,335	280	7.31	7.1	6.8	4.3
P14	+	ARTIC + Rapid	3,437,488	875,159	853,443	25.5	3,136,063	290	91.23	99.9	99.8	98.2
P15	+	ARTIC + Rapid	2,100,544	1,822,242	1,754,580	86.8	1,982,939	290	94.4	97.4	96.3	82.9
P16	+	ARTIC + Rapid	237,504	15,807	15,807	6.7	18,607	280	7.83	5.7	4.4	2.7
P17	+	ARTIC + Rapid	526,781	178,745	174,792	33.9	198,615	300	37.7	25.6	24	26
P18	+	ARTIC + Rapid	5,724,659	4,538,653	4,460,075	79.3	5,065,568	280	88.49	99.9	99.8	99.8
P19	+	ARTIC + Rapid	162,340	35,574	35,574	21.9	39,440	290	24.29	38.2	34.5	7.7
P20	+	ARTIC + Rapid	286,145	41,196	41,196	14.4	46,575	230	16.28	44.1	21.4	4.6
N01	-	ARTIC + Rapid	578,916	446,225	443,662	77.1	495,874	280	85.66	62	58.8	43.8
N02	-	ARTIC + Rapid	121,813	15,313	15,313	12.6	ND	ND	ND	ND	ND	ND
N03	-	ARTIC + Rapid	67,918	–	–	0.0	–	–	0	0	0	0
N04	-	ARTIC + Rapid	141,075	1,273	1,273	0.0	1,407	290	1	2	1.9	0
N05	-	ARTIC + Rapid	649,918	–	–	0.0	431	340	0.07	1.30	1.20	0.00
N06	-	ARTIC + Rapid	168,575	–	–	0.0	ND	ND	ND	ND	ND	ND
N07	-	ARTIC + Rapid	111,549	–	–	0.0	ND	ND	ND	ND	ND	ND
N08	-	ARTIC + Rapid	30,365	–	–	0.0	ND	ND	ND	ND	ND	ND
N09	-	ARTIC + Rapid	91,432	14,671	14,656	16.0	16,219	280	17.74	2.60	2.50	2.10
N10	-	ARTIC + Rapid	574,708	183,764	183,764	32.0	208,541	280	36.29	33.90	31.70	25.50
N11	-	ARTIC + Rapid	395,322	166,603	166,603	42.1	189,251	290	47.87	64.40	56.50	33.70
N12	-	ARTIC + Rapid	512,171	386,424	386,424	75.4	441,824	270	82.26	50.30	49.70	43.30



addition, RAMPART analysis indicated that coverage varied in these positive samples from as low as 5.7% to as high as 99.9% (median 64.9%) at > 10X depth, 4.4–99.8% (median 53.8%) at > 100X depth, and 2.7–99.8% at > 1,000X depth (Table 4 and Figure 3C). RAMPART also offers graphical representations of coverage, plotting the number of reads that map to each specific region of the CoV-2 genome (Figure 4). P09 is shown as a graphical representative of a sample with coverage just below the median (Figure 4A), while P14 represents a sample with high coverage of 99.9% (Figure 4B), both at > 10X depth of coverage.

Some unexpected results were encountered upon analysis of negative samples N01–N12. Besides the N01 sample in the pilot study, six other negative samples had reads that mapped to CoV-2 ranging from 1,273–446,225 as per analysis with Centrifuge (Table 4 and Figure 3A). Upon more careful inspection of RAMPART figures and statistics, N04 and N09 were ruled as negative because of the low number of reads (1,407 and 16,219), low depth of coverage at > 10X (2 and 2.6%), and just 2 peaks or areas of coverage on the map of the CoV-2 genome (Figure 4D for N04, Table 4). In contrast, sequencing of clinical specimens N01, N10, N11, and N12 resulted in 189,251–495,874 mapped reads, 36.29–62% depth of coverage at > 10X, and multiple areas of coverage on the CoV-2 genome map (Figures 4E,F for N11 and N12, Table 4). N03 (Figure 4C) is shown in comparison as a negative clinical specimen that results in 0 CoV-2 specific reads when sequenced while clinical specimens P09 (Figure 4A) and P14 (Figure 4B) are shown as positive specimens with different percentages and areas of coverage.

Next, six positive samples with relatively high genome coverage were chosen for assembly using the Wuhan-Hu-1 isolate (Accession MN908947.3) as the reference sequence. Once the sequences were successfully assembled, we used CoVsurver for clade classification and mutational analyses (Table 5). For

comparison to work by others, we also included Nextstrain clade and Pango lineage classifications. As a control and reference point, we also assembled sequences generated using the CoV-2 RNA (Isolate USA-WA1/2020) from BEI that were generated using the standard ARTIC v3 protocol versus the ARTIC with Rapid modification method. The comparison of the reference strain from BEI using the standard ARTIC v3 method versus the ARTIC with Rapid sequencing modification gave us the same results when performing mutational analyses using CoVsurver, as well as other lineage and classification analyses.

As expected, this isolate was assigned to GISAID clade S, and contained the NS8 L84S mutation that is one of the markers of this clade (see Table 1). We also performed this comparison for the P02 assemblies where sequencing was performed using the same standard and modified approaches. We expected similar results, but found that while the percentage of nucleotides missing (Table 5) were higher for the P02 sequenced using the ARTIC with the Rapid modification versus the standard assay, these nucleotides missing may be in 5', 3', or other untranslated regions as the CoVsurver report showed only four amino acids were missing, and these were located in the NSP6 region (Supplementary Material). However, there were 26 amino acid deletions in the Spike region for the P02 sequenced using the standard ARTIC v3 assay (with LSK-109 kit). These deletions in the P02 sequenced using the standard ARTIC v3 assay may have affected the classification of this sample into clade G. Mutations in spike at D614G and P521R were also found.

For the six genomes assembled from clinical specimens that were sequenced using the ARTIC with Rapid Sequencing protocol, we identified the Spike protein D614G mutation. P02, P12, P14, P15, and P18 were classified in clade GH. P11 was classified into clade GR, which diverged after GH, and deviates from GH in that it carries the N R203K mutation, but not the

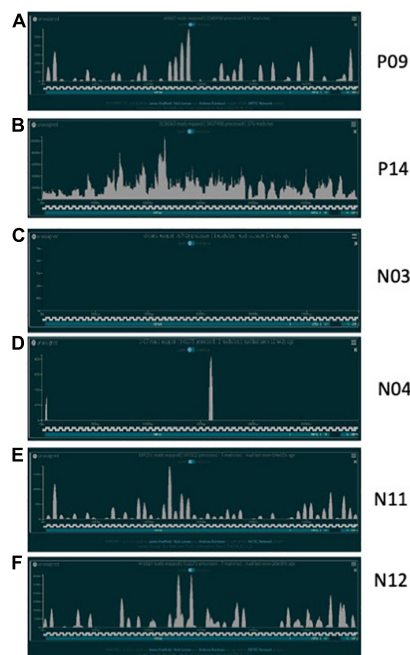


FIGURE 4 | CoV-2 Reads mapped over reference genome. Representative plots showing the number of CoV-2 reads mapping over specific regions of a CoV-2 reference genome after sequencing using the ARTIC with Rapid sequencing protocol and RAMPART analyses. **(A)** P09 is shown as positive sample just below the median coverage for the positive sample cohort; **(B)** P14 is representative of high coverage samples; **(C)** N03 is a negative sample with zero CoV-2 reads identified; **(D)** N04 is a negative sample with possible non-specific amplification or contamination; **(E,F)** are two RT-PCR negative samples with 50–60% CoV-2 genome coverage following sequencing.

NS3 Q57H mutation (see **Tables 1, 5**). Overall, the number of mutations as compared to the WIV04 reference isolate were minimal ranging from 3 to 7 amino acid changes (0.05–0.10%). There were a number of mutations that had been previously observed by others, more notably, the Spike D614G, Nsp12 P323L, and NS3 Q57H mutations. There were also five mutations that were unique to these samples when we first analyzed them using CoVsurver on October 20, 2020: Spike P521R, NS3 G76S, N P364L, Nsp13 V209I, and Nsp15 D212V. Of note, the Spike P521R mutation was observed in 3 of these assemblies.

DISCUSSION

It is not surprising that an RNA virus would mutate over time, especially during an extended period of human-to-human transmission as has been the case with the pandemic caused by SARS-CoV-2. One important mutation that arose early on was the D614G mutation in the Spike protein, and this mutation quickly becoming prevalent upon its introduction or emergence in different areas of the world (Korber et al., 2020). Of note, while the rate of mutation in the CoV-2, including in the Spike protein was low at the time, the D614G mutation was caused by a single nucleotide mutation from A-to-G at

position 23,403 in the Wuhan reference strain (Korber et al., 2020). Our sequence assemblies all carried this substitution D614G in the Spike protein, and the P323L mutation in the Nsp12 (RNA-dependent RNA polymerase) that frequently co-evolves with it (Coppee et al., 2020). GISAID and others began to track the mutation in March 2020, and the clade carrying the substitution was designated as G. Within that month, the variant went from being present in 10% of global sequences, to 67% of global sequences (Korber et al., 2020). Given the speed at which the D614G variant was able to spread, higher Ct counts in patients (Korber et al., 2020), and higher infectivity of VSV-pseudotyped virions (Korber et al., 2020; Li et al., 2020), the mutation may offer a fitness advantage that makes the virus more infectious. Furthermore, other important variants with increased transmissibility continued to emerge that carried the D614G mutation, along with a number of other mutations. These variants emerged in the fall of 2020 and were reported by the United Kingdom (B.1.1.7, WHO designation Alpha, GISAID Clade GRY), South Africa (B.1.351), and Brazil (P.1, WHO designation Gamma) (CDC, 2021a). We also found a P521R mutation in the Spike RBD in three of our six assemblies. The GISAID site reports that this mutation was first reported in March 2020 as found in an Israeli strain: hCoV-19/Israel,CVL-n2487/2020. Additional analyses on sequences containing this mutation as reported by CoV-GLUE confirm the mutation's presence in New Mexico and Arizona from strains collected in April 2020 to June 2020. Skipping a year forward, and there was the emergence and eventual global dominance of the WHO label Delta variant (B.1.617.2, GISAID clade GK), that carried 15 substitutions or deletions in the Spike protein including the aforementioned D614G (CDC, 2021b). Once again, this newer variant was observed to have increased transmissibility, with reduction in neutralization by post-vaccination sera was observed. By the November 2021, South Africa had reported and identified yet a new variant that contained at least 30 amino acid substitutions, 3 deletions and an insertion in the Spike protein alone. This new variant was designated as the Omicron variant of concern by the WHO (B.1.1.529, GISAID clade GRA) and while vaccines are still considered to be effective in preventing severe illness and hospitalizations, break-through infections are observed (CDC, 2021c).

Besides the D614G and P323L mutations that as previously mentioned have a tendency to co-evolve, another frequent mutation observed in our assembled genomes was ORF3a or NS3 Q57H (P02, P14, P15, P18, and P12 once corrected). Coppée et al. reported this mutation as the fourth frequent mutation observed in European populations after D164G, P323L, and L84S as accessed in sequences collected through April 17, 2020. The Q57H mutation was frequent in samples from France and Belgium, but not Italy and Spain (Coppee et al., 2020). By November 2020, the introduction of a GISAID clade GH virus with this G57H mutation had lead to a fourth wave of CoV-2 infections in Hong Kong (Chu et al., 2021). While they did not find enhanced replication kinetics or increased induction of cytokines/chemokines by this virus (Chu et al., 2021), a different study showed that the Q57H mutation increased the intraviral

TABLE 5 | Mutation analysis of assembled sequences using CoVsurver and comparison to Pangolin and Nextclade classifications.

ID	Seq Assay	Query	%N	Length (nt)	Length (aa)	#Muts	%Muts	Comment	Unique Mut	Existing Mut	GISSAID Clade	Pango lineage	Nextstrain Clade
BEI RNA	ARTIC v3	CoV2_Isk109/ARTIC /medaka MN908947.3	0.00%	29,903	9,710	1	0.01%			NS8_L84S	S	A	19B
BEI RNA	ARTIC + Rapid	CoV2_rad004/ARTIC /medaka MN908947.3	0.00%	29,903	9,710	1	0.01%			NS8_L84S	S	A	19B
P02	ARTIC v3	1P2_Isk109/ARTIC /medaka MN908947.3	1.18%	29,903	9,710	4	0.04%	Stretches of NNNs (1.18% of overall sequence).		NSP12_P323L, Spike_D614G, Spike_D138Y, Spike_P521R	G	B.1.473	20A
P02	ARTIC + Rapid	1P2_rad004/ARTIC /medaka MN908947.3	2.91%	29,903	9,710	6	0.06%	Stretches of NNNs (2.91% of overall sequence).		NSP12_P323L, Spike_D614G, Spike_D138Y, Spike_P521R, NS3_Q57H, NS3_G76S	GH	B.1.473	20A
P11	ARTIC + Rapid	P11_rad004/ARTIC /medaka MN908947.3	0.00%	29,903	9,710	6	0.06%			NSP12_P323L, NSP12_H613Y, NSP14_A274S, Spike_D614G, N_G204R, N_R203K	GR	B.1.1	20B
P12	ARTIC + Rapid	P12_rad004/ARTIC /medaka MN908947.3	15.51%	29,904	9,534	6	0.06%	Long stretches of NNNs (15.51% of overall sequence). Insertion of 1 nucleotide(s) found at refpos 26653 (FRAMESHIFT). M without BLAST coverage. NSP3 has 103 Δs, NSP4 has 22 Δs; NSP16 has 80 Δs		NSP3_V477F, NSP12_P323L, Spike_D614G, Spike_P521R, NS3_Q57H, M_R44S	GH	B.1	20A
P14	ARTIC + Rapid	P14_rad004/ARTIC /medaka MN908947.3	0.00%	29,903	9,710	5	0.05%			NSP2_T85I, NSP12_P323L, Spike_D614G, NS3_Q57H, N_P364L	GH	B.1	20C
P15	ARTIC + Rapid	P15_rad004/ARTIC /medaka MN908947.3	3.30%	29,902	9,710	10	0.10%	Stretches of NNNs (3.30% of overall sequence). Gap of 1 nucleotide(s) found at refpos 3013 (FRAMESHIFT). NSP3 has 248Δs	NSP3_L98I	NSP3_A99S, NSP3_T1456I, NSP3_V477F, NSP12_P323L, Spike_D614G, Spike_V308L, Spike_P521R, NS3_Q57H, N_R209I	GH	B.1	20A
P18	ARTIC + Rapid	P18_rad004/ARTIC /medaka MN908947.3	0.00%	29,903	9,710	6	0.06%		NSP15_D212V	NSP2_T85I, NSP12_P323L, NSP13_V209I, Spike_D614G, NS3_Q57H	GH	B.1	20C
P12 fixed	ARTIC + Rapid	P12_rad004 (organism = Severe acute respiratory syndrome coronavirus 2) (isolate = P12) reference assembly to MN908947.3	15.51%	29,903	9,534	6	0.06%	Long stretches of NNNs (15.51% of overall sequence). NSP3 has 103 Δs, NSP4 has 22 Δs; NSP16 has 80 Δs		NSP3_V477F, NSP12_P323L, Spike_D614G, Spike_P521R, NS3_Q57H, M_R44S	GH	B.1	20A
P15 fixed	ARTIC + Rapid	P15_rad004 (organism = Severe acute respiratory syndrome coronavirus 2) (isolate = P15) reference assembly to MN908947.3	3.30%	29,903	9,710	10	0.10%	Stretches of NNNs (3.30% of overall sequence). NSP3 has 248Δs		NSP3_A99S, NSP3_T1456I, NSP3_V477F, NSP3_L98F, NSP12_P323L, Spike_D614G, Spike_V308L, Spike_P521R, NS3_Q57H, N_R209I	GH	B.1	20A

protein affinities (Wu et al., 2021). Moreover, while Q57 was not involved in protein-binding interfaces, Q57H was a hotspot for protein-interactions (Wu et al., 2021). The emergence of Q57H earlier on during the pandemic as reported here and by others, and more recent reemergence associated with a wave of infections may suggest an advantage to the virus. Besides the Q57H mutation, we also found a mutation in NS3 that was unique to our subset: G76S.

Other unique mutations to our assemblies included those in the multidomain, multifunctional Nsp3 protein (Santerre et al., 2020) in sample P15: A99S and L98I or L98F. However, because the Nsp3 region in this assembly had significant gaps, it is difficult to ascertain that these mutations are real. In one of our better assemblies, using sample P18, we found mutations in the helicase Nsp13 V209I and endoribonuclease Nsp15 NSP15D212V.

A subset of negative samples, as determined by PCR analyses returned hits that mapped to SARS-CoV-2 as analyzed by Centrifuge and RAMPART. A caveat of this study is that we did not have detailed PCR assay information with the samples, including which PCR test was used or Ct values. However, the PCR tests with EUA at the time from April to May 2020 consisted of single to three target site assays (FDA, 2022). Since the ARTIC primers span the entire genome, the assay is likely more sensitive, and is likely to pick up samples that were found to be negative by the assays. For clinical specimens N01, N10, N11, and N12 we had 189,251–495,874 mapped reads, and multiple areas of coverage on the CoV-2 genome map (Figure 4), and thus it is possible that these specimens were in fact positive, but missed by the one to three-target PCR tests. At the same time, because of the high number of primers/targets and amplification cycles that we used, there was also a chance for non-specific amplification as may have been the case with N04 and N09, which have low read numbers, and very little coverage of the CoV-2 genome. Thus, we don't think that N04 and N09 are positive for CoV-2. On the other hand, positive sample P10 seemed to generate very few reads, and even fewer CoV-2 specific reads similar to subject N04. Because we don't have corresponding PCR information, including Ct value, we can't rule out if this was a false positive, a sample with a very low viral load, or a sample that got degraded after processing and shipment to our lab.

Non-specific or contaminating reads that we observed in the samples as reported by Centrifuge mapped to *Escheria coli*, *Shigella boydii*, *Shigella flexneri*, *Shigella phage*, and *Escheria phage Mu* as examples. These reads may be related to contaminating bacterial DNA derived from the extraction kit or library preparation kits. These hits were associated with a low percentage of unique reads (e.g., *E. coli* 6.5%, *Escheria phage Mu* 0.1%). There were also reads that failed to map using Centrifuge, and these could have been reads associated with the host (human). However, human/host reads can now be filtered out during sequencing using “adaptive sequencing” from MinKNOW and selecting for depletion of human sequences. This is an approach that may be used in future studies.

In short, the work presented shows that it is possible to streamline the Artic v3 protocol for nanopore sequencing, saving

an hour and 10 min in sample and library preparation time (Figure 1), and still acquire data for sensitive and confident identification of CoV-2 in human nasopharyngeal swab samples. For a subset of samples, presumably for those with ample viral loads, whole genome assembly was possible. Future studies will investigate what additional methods can be used to improve results, for example, use of primers that generate longer reads might improve results. One caveat of the current approach is that the Rapid Sequencing Kit was used to cleave the 400 b amplicons generated using the ARTIC v3 primer panel. This could result in failed sequences since the shorter reads may be misidentified as adapters during acquisition or may be flagged as failed reads by the basecaller. However, from the samples that were assembled, we were able to find and confirm deviations that were previously reported by others, and that were consistent with geographic location and period in which these samples were collected. Moreover, we were able to identify a few mutations that were unique to these samples. This work suggests that accessible, portable, and relatively fast sample-to-sequence processes can be effectively used to characterize viral outbreaks. Processes like this are needed to bring sequencing and characterization to the initial sites of emerging outbreaks, even if they are occurring in remote regions, and will help us be better prepared to respond. There are additional modifications that could make this approach more field-friendly. One modification would be to move from RNA extraction kits that use Trizol and require centrifugation with benchtop instruments to kits that use a different lysis buffer and can use either a magnetic, RNA-binding bead-based method or a syringe-based silica-based column method. Adding an automated software package that can perform analysis in real time as the sample is being sequenced would also help expedite the process by providing faster identification of the sample. If the sample of interest is positive, additional sequencing time can be used to generate enough reads for further characterization and assembly. Another time-saving modification could be focusing on a single area of interest such as the spike protein gene to simplify the library preparation process and reduce analysis time. Although this would further streamline the approach, it would not be as powerful as whole genome sequencing that would provide further insight as to how the whole virus is evolving. Furthermore, as this virus continues to mutate and affect the global population, an all-hands/all-methods approach to surveillance may be needed to finally get ahead of the curve.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI GenBank repository (<https://www.ncbi.nlm.nih.gov/nucleotide/>), with accession numbers provided in Table 3.

AUTHOR CONTRIBUTIONS

MA and RB conceived the study, design, and experiments. MA, MK, SK, and JVH performed experiments. JMH, SD,

PR, and AL performed bioinformatic analysis and assembly of the sequences. MA performed classification and mutational analyses. MA prepared the manuscript with input, review, and contributions provided by MK, SK, JVH, JMH, SD, PR, AL, and RB. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Joint Program Executive Office for Chemical, Biological, Radiological and Nuclear Defense (JPEO-CBRND).

REFERENCES

- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452.
- Artic-Network (2020). *Artic Network SARS-CoV-2*. Oxford: Oxford Nanopore Technologies.
- CDC (2021a). *SARS-CoV-2 Variants 2021*. Available online at: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html> (accessed March 16, 2021).
- CDC (2021b). *SARS-CoV-2 Variant Classifications and Definitions 2021*. Available online at: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html> (accessed December 01, 2021).
- CDC (2021c). *Science Brief: Omicron (B.1.1.529) Variant 2021*. Available online at: <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-omicron-variant.html> (accessed January 28, 2021).
- Charre, C., Ginevra, C., Sabatier, M., Regue, H., Destras, G., Brun, S., et al. (2020). Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol.* 6:veaa075. doi: 10.1093/ve/veaa075
- Chu, D. K. W., Hui, K. P. Y., Gu, H., Ko, R. L. W., Krishnan, P., Ng, D. Y. M., et al. (2021). Introduction of ORF3a-Q57H SARS-CoV-2 Variant Causing Fourth Epidemic Wave of COVID-19, Hong Kong, China. *Emerg. Infect. Dis. J.* 27, 1492–1495. doi: 10.3201/eid2705.210015
- Coppee, F., Lechien, J. R., Declèves, A. E., Tafforeau, L., and Saussez, S. (2020). Severe acute respiratory syndrome coronavirus 2: virus mutations in specific European populations. *New Microbes New Infect.* 36:100696. doi: 10.1016/j.nmni.2020.100696
- FDA (2022). *In Vitro Diagnostics EUAs - Molecular Diagnostic Tests for SARS-CoV-2*. Available online at: <https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas-molecular-diagnostic-tests-sars-cov-2> (accessed April 14, 2022).
- GISAID (2020). *Clade and Lineage Nomenclature Aids in Genomic Epidemiology Studies of Active hCoV-19 Viruses*. Munich: GISAID.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506.
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182, 812–827.e19.
- Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., et al. (2020). The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182, 1284–1294.e9.
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- Quick, J. (2020). *nCoV-2019 Sequencing Protocol 2020*. Available online at: https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w?version_warning=no (accessed January 20, 2020).
- Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407.

ACKNOWLEDGMENTS

We would like to thank Justin Bacca, Robert Taylor, and Darrell Dinwiddie at the University of New Mexico for kindly providing the clinical samples used in this study. We would also like to thank Adina Doyle for reviewing this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.910955/full#supplementary-material>

- Romano, M., Ruggiero, A., Squeglia, F., Maga, G., and Berisio, R. (2020). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells* 9:1267. doi: 10.3390/cells9051267
- Santerre, M., Arjona, S. P., Allen, C. N., Shcherbik, N., and Sawaya, B. E. (2020). Why do SARS-CoV-2 NSPs rush to the ER? *J. Neurol.* 268, 2013–2022. doi: 10.1007/s00415-020-10197-8
- Seth-Smith, H. M. B., Bonfiglio, F., Cuénod, A., Reist, J., Egli, A., and Wüthrich, D. (2019). Evaluation of rapid library preparation protocols for whole genome sequencing based outbreak investigation. *Front. Public Health* 7:241. doi: 10.3389/fpubh.2019.00241
- Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro. Surveill.* 22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494
- Singer, J. B., Gifford, R. J., Cotten, M., and Robertson, D. L. (2020). CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. *Preprints* 2020:2020060225.
- Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., et al. (2020). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* [Preprint]. doi: 10.1101/2020.09.04.283077
- Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., et al. (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* 92, 667–674. doi: 10.1002/jmv.25762
- Watson, M., and Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. *Nat. Biotechnol.* 37, 124–126.
- WHO (2022). *Tracking SARS-CoV-2 variants 2022*. Geneva: WHO
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269.
- Wu, S., Tian, C., Liu, P., Guo, D., Zheng, W., Huang, X., et al. (2021). Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions. *J. Med. Virol.* 93, 2132–2140. doi: 10.1002/jmv.26597

Conflict of Interest: JMH, PR, and AL were employed by DCS Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Arévalo, Karavis, Katoski, Harris, Hill, Deshpande, Roth, Liem and Bernhards. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Metagenomic Investigation of Ticks From Kenyan Wildlife Reveals Diverse Microbial Pathogens and New Country Pathogen Records

Koray Ergunay^{1,2,3,4,*†}, Mathew Mutinda^{5†}, Brian Bourke^{1,2,4}, Silvia A. Justí^{1,2,4}, Laura Caicedo-Quiroga^{1,2,4}, Joseph Kamau⁶, Samson Mutura⁶, Irene Karagi Akunda⁶, Elizabeth Cook⁷, Francis Gakuya⁸, Patrick Omondi⁸, Suzan Murray⁹, Dawn Zimmerman^{1,2,4,10} and Yvonne-Marie Linton^{1,2,4}

¹ Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States,

² One Health Branch, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States, ³ Department of Medical Microbiology, Virology Unit, Faculty of Medicine, Hacettepe University, Ankara, Turkey, ⁴ Department of Entomology, Smithsonian Institution, National Museum of Natural History (NMNH), Washington, DC, United States,

⁵ Kenya Wildlife Services Corporation, Nairobi, Kenya, ⁶ One Health Centre, Institute of Primate Research (IPR), Nairobi, Kenya, ⁷ International Livestock Research Institute (ILRI), Nairobi, Kenya, ⁸ Wildlife Research and Training Institute (WRTI), Naivasha, Kenya, ⁹ Global Health Program, Smithsonian Conservation Biology Unit, Fort Royal, VA, United States,

¹⁰ Department of Epidemiology of Microbial Disease, Yale School of Public Health, New Haven, CT, United States

OPEN ACCESS

Edited by:

Jens Andre Hammerl,
Bundesinstitut für Risikobewertung,
Germany

Reviewed by:

Jifei Yang,
Lanzhou Veterinary Research Institute
(CAAS), China
Komal Jain,
Columbia University, United States

*Correspondence:

Koray Ergunay
ekoray@hacettepe.edu.tr

[†]These authors share first authorship

Specialty section:

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

Received: 29 April 2022

Accepted: 27 May 2022

Published: 01 July 2022

Citation:

Ergunay K, Mutinda M, Bourke B,
Justí SA, Caicedo-Quiroga L,
Kamau J, Mutura S, Akunda IK,
Cook E, Gakuya F, Omondi P,
Murray S, Zimmerman D and
Linton Y-M (2022) Metagenomic
Investigation of Ticks From Kenyan
Wildlife Reveals Diverse Microbial
Pathogens and New Country
Pathogen Records.
Front. Microbiol. 13:932224.
doi: 10.3389/fmicb.2022.932224

Focusing on the utility of ticks as xenosurveillance sentinels to expose circulating pathogens in Kenyan drylands, host-feeding ticks collected from wild ungulates [buffaloes, elephants, giraffes, hartebeest, impala, rhinoceros (black and white), zebras (Grévy's and plains)], carnivores (leopards, lions, spotted hyenas, wild dogs), as well as regular domestic and Boran cattle were screened for pathogens using metagenomics. A total of 75 host-feeding ticks [*Rhipicephalus* (97.3%) and *Amblyomma* (2.7%)] collected from 15 vertebrate taxa were sequenced in 46 pools. Fifty-six pathogenic bacterial species were detected in 35 pools analyzed for pathogens and relative abundances of major phyla. The most frequently observed species was *Escherichia coli* (62.8%), followed by *Proteus mirabilis* (48.5%) and *Coxiella burnetii* (45.7%). *Francisella tularemia* and Jingmen tick virus (JMTV) were detected in 14.2 and 13% of the pools, respectively, in ticks collected from wild animals and cattle. This is one of the first reports of JMTV in Kenya, and phylogenetic reconstruction revealed significant divergence from previously known isolates and related viruses. Eight fungal species with human pathogenicity were detected in 5 pools (10.8%). The vector-borne filarial pathogens (*Brugia malayi*, *Dirofilaria immitis*, *Loa loa*), protozoa (*Plasmodium* spp., *Trypanosoma cruzi*), and environmental and water-/food-borne pathogens (*Entamoeba histolytica*, *Encephalitozoon intestinalis*, *Naegleria fowleri*, *Schistosoma* spp., *Toxoplasma gondii*, and *Trichinella spiralis*) were detected. Documented viruses included human mastadenovirus C, Epstein-Barr virus and bovine herpesvirus 5, Trinbago virus, and Guarapuava tymovirus-like virus 1. Our findings confirmed that host-feeding ticks are an efficient sentinel for xenosurveillance and demonstrate clear potential for wildlife-livestock-human pathogen transfer in the Kenyan landscape.

Keywords: tick, metagenomics, wildlife, pathogen, surveillance

INTRODUCTION

Tick-borne infections account for a major portion of all vector-borne diseases in many developed and underdeveloped countries, with significant health, economic, and food security consequences (Rochlin and Toledo, 2020). The incidence of tick-borne infections is increasing due to environmental changes at both local and global scales, resulting in changing habitats and geographical repartition of ticks, with subsequent novel pathogen exposure in naive populations (Ogden et al., 2013).

Ticks can transmit a diverse array of microbial pathogens including viruses, bacteria, and protozoans to humans and susceptible animals (Rochlin and Toledo, 2020). Tick-borne infections in humans are of zoonotic origin, with pathogens maintained in natural cycles involving various domestic and wild animal hosts (Wikel, 2018). Timely identification of the circulating pathogens and assessment of potential public health threats rely on effective surveillance, especially considering that most zoonoses originate from wildlife (Woolhouse and Gowtage-Sequeria, 2005; Miller et al., 2013). Many tick-borne agents of concern might already be circulating in wildlife awaiting detection. Microbial pathogens may be overlooked until the emergence of a case cluster or a local epidemic. An effective surveillance strategy should be based on surveying a wide range of pathogens and hosts, prior to the spread and symptomatic disease. The availability of metagenomic sequencing platforms and approaches offers great advantages and possibilities to explore microbial diversity (Brinkmann et al., 2016). Metagenomic investigations in tick vectors collected from wildlife hosts provide opportunities to obtain spatial and temporal microbial genome data as well as from tick-borne pathogens, and are likely to facilitate predictions on possible emergents.

The Republic of Kenya is a large sub-Saharan country in eastern Africa, with an area of 580,367 square kilometers and a population of over 47 million (Kenya National Bureau of Statistics, 2021). In Kenya, the pastoralist population make up around 20% of the population with an agricultural land use of nearly 50% (State of Wildlife Conservancies in Kenya Report, 2016). National parks, nature reserves, and conservancies account for about 11% of the Kenyan landmass with increased human and livestock populations in peripheries (Welde et al., 1989; Ogutu et al., 2016). These trends facilitate interactions between livestock, humans, and wildlife in a complex environment, increasing the opportunity for pathogen spill-over events (Okal et al., 2020). Several tick-borne bacterial, viral, and parasitic agents have been documented in Kenya, including agents of anaplasmosis, ehrlichiosis, rickettsiosis, Crimean-Congo hemorrhagic fever (CCHF), babesiosis, and theileriosis (Omondi et al., 2017; Morrison et al., 2020; Okal et al., 2020; Chiuya et al., 2021; Getange et al., 2021; Peter et al., 2021), with occasional reporting of novel pathogens (Mwamuye et al., 2017). However, reports from wildlife are relatively rare and often focus on well-documented bacterial agents. This study aims to screen ticks collected from wildlife hosts using an unbiased metagenomic approach, to investigate a broad spectrum of circulating microbial pathogens in Kenya dryland ecosystems.

MATERIALS AND METHODS

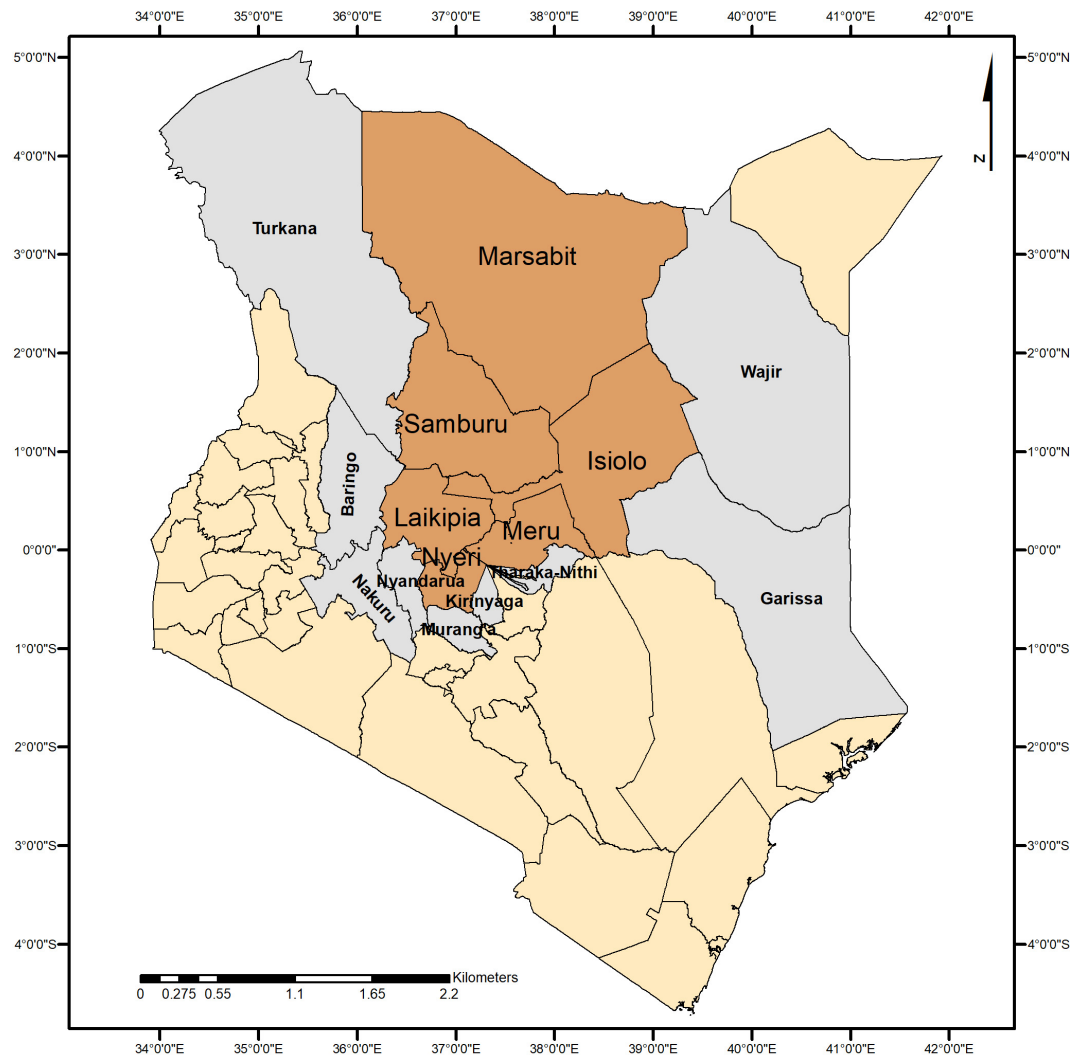
Tick Specimens

Tick specimens were collected by the Kenya Wildlife Service (KWS) from animals in wildlife conservancy sites in Isiolo county, Kalama, Lewa, Ol Pejeta, Samburu, Sangare gardens, Sarara, Solio ranch and Oljogi, between October 2013 and April 2019 (Figure 1). The host animals included black rhinoceros (*Diceros bicornis*), buffalo (*Syncerus caffer*), elephant (*Loxodonta africana*), giraffe (*Giraffa camelopardalis*), Grévy's zebra (*Equus grevyi*), hartebeest (*Alcelaphus buselaphus*), impala (*Aepyceros melampus*), leopard (*Panthera pardus*), lion (*Panthera leo*), plains zebra (*Equus quagga*), spotted hyena (*Crocuta crocuta*), white rhinoceros (*Ceratotherium simum*), and wild dog (*Lycaon pictus*), as well as farmed Boran (*Bos indicus*) and cattle (*Bos taurus*) (Table 1). The Boran and cattle were kept in ranches without fencing, in direct contact with wildlife. The ticks were collected opportunistically from wild animals immobilized by the KWS for veterinary intervention, and ticks were kept in 70% ethanol and stored at -80°C prior to processing. Information on the collection site, date, and the host was recorded. Ticks were morphologically identified as *Rhipicephalus* spp. or *Amblyomma* spp. and pooled according to individual host and genera, prior to pooling and DNA extraction (Table 1).

Metagenomic Sequencing and Molecular Testing

Tubes containing tick pools (1–8 ticks/pool) were dipped in liquid nitrogen and homogenized using beads. Nucleic acid purification was conducted using the RNeasy Mini Kit (Qiagen, Hilden, Germany), and complementary DNA (cDNA) synthesis with random hexamers was performed using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Hennigsdorf, Germany), as per the manufacturer's recommendations. Tick nucleic acid extraction and cDNA conversion were carried out at IPR laboratories in Kenya, and cDNAs were sent to WRBU (United States) for further testing.

The cDNA was quantified using Quant-iT OliGreen ssDNA Assay Kit (ThermoFisher Scientific, MA, United States) on a Fluoroskan FL instrument (ThermoFisher Scientific, MA, United States). Libraries were prepared using KAPA HyperPlus Kits (Roche, CA, United States), as recommended by the manufacturer, with fragmentation step of 20 min at 35°C , ligation step of 90 min, and 20 PCR cycles. Library quantification and quality control were performed using the TapeStation 4,200 Automated Electrophoresis instrument (Agilent Technologies, VA, United States). Excess adapter dimer and small fragments were removed using KAPA pure beads (Roche, CA, United States). Unbiased metagenomic sequencing was performed on the NovaSeq platform (Illumina, CA, United States) (PE 2×150) at the Walter Reed Army Institute of Research (WRAIR). Samples were run on one lane of an S4 flow cell with XP workflow designed to maximize the overall raw reads output. Read quality was assessed using fastqc (Wingett and Andrews, 2018), and adapter trimming



County	Species (N)
Meru	Grevy's Zebra (7)
	Cattle (4)
	Hyena (4)
	Lion (3)
	Buffalo (1)
	Giraffe (1)
	Hartebeest (1)
	Impala (1)
	White Rhino (1)
Laikipia	Buffalo (3)
	Lion (2)
	Wild Dog (2)
	Black Rhino (1)
	Elephant (1)
	Hartebeest (1)
Samburu	Leopard (1)
	Elephant (5)
Isiolo	Lion (2)
Isiolo	Elephant (2)
Marsabit	Cattle (1)
Marsabit	Boran (1)
Nyeri	Plains Zebra (1)

FIGURE 1 | Map indicating the counties where ticks were collected and sampled when feeding on domestic and wild animals in Kenya.

TABLE 1 | Overview of the tick specimens and their hosts included in the study.

Host	Pool Code	Filtered Read Count	Pooled Individuals	Identification	Ticks from Host
Elephant (n:7, 15.2%)	#71	8,122	1♀	<i>Rhipicephalus</i> spp.	8/75 (10.6%)
	#65	6,426	1♂		
	#52	564,490	2♂		
	#48	71,616	1♀		
	#42	4,882	1♂		
	#41	3,294	1♂		
	#6	31,338	1♂		
Lion (n:7, 15.2%)	#53	1,631,344	1♀	<i>Rhipicephalus</i> spp.	7/75 (9.3%)
	#47	2,851,375	1♀		
	#45	1,350	1♀		
	#43	216,970	1♀		
	#23	15,276	1♂		
	#22	918	1♂		
	#10	84,976	1♀		
Grévy's zebra (n:7, 15.2%)	#83	144,564	1♂	<i>Rhipicephalus</i> spp.	7/75 (9.3%)
	#78	3,552	1♀		
	#70	441,284	1♂		
	#69	28,338	1♂		
	#67	1,544	1♂		
	#25	5,296	1♀		
	#13	2,548	1♂		
Buffalo (n:4, 8.7%)	#84	43,042	5 (3♀2♂)	<i>Rhipicephalus</i> spp.	21/75 (28%)
	#58	33,668	1♂		
	#9	34,960	7 (4♀3♂)		
	#8	6,186	8 (5♀3♂)		
Cattle (n:4, 8.7%)	#31	8,596	1♂	<i>Rhipicephalus</i> spp.	10/75 (13.3%)
	#19	598	1♀		
	#11	247,466	7♂		
	#3	501,204	1♀		
Hyena (n:4, 8.7%)	#79	186	1♂	<i>Amblyomma</i> spp.	4/75 (5.3%)
	#68	156,604	1♂		
	#36	3,494	1♂		
	#12	1,182	1♂		
Plains Zebra (n:3, 6.5%)	#66	1,010	1♂	<i>Amblyomma</i> spp.	3/75 (4%)
	#40	3,698	1♂		
	#35	4,752	1♀		
Hartebeest (n:2, 4.3%)	#57	155,402	1♂	<i>Rhipicephalus</i> spp.	2/75 (2.6%)
	#54	1,013,892	1♂		
Wild dog (n:2, 4.3%)	#60	602,012	3♀	<i>Rhipicephalus</i> spp.	4/75 (5.3%)
	#56	37,670	1♂		
Boran (n:1, 2.1%)	#30	23,186	4 (2♀2♂)	<i>Rhipicephalus</i> spp.	4/75 (5.3%)
Impala (n:1, 2.1%)	#55	600	1♀	<i>Rhipicephalus</i> spp.	1/75 (1.3%)
Giraffe (n:1, 2.1%)	#20	52,486	1♂	<i>Rhipicephalus</i> spp.	1/75 (1.3%)
Leopard (n:1, 2.1%)	#81	439,608	1♀	<i>Rhipicephalus</i> spp.	1/75 (1.3%)
B. Rhino (n:1, 2.1%)	#21	92,750	1♂	<i>Rhipicephalus</i> spp.	1/75 (1.3%)
W. Rhino (n:1, 2.1%)	#72	74,580	1♂	<i>Rhipicephalus</i> spp.	1/75 (1.3%)
Total: 46 pools					75 ticks

and sequence filtering were performed using Trimmomatic (Bolger et al., 2014).

Jingmen tick virus (JMTV) screening was performed by a specific nested PCR assay, targeting the viral NS5-like protein located on the genome segment 1, as described

previously (Yu et al., 2020). Amplified products were visualized on 1% agarose gels, stained with GelGreen nucleic acid stain (Biotium, California, United States), and visualized after electrophoresis in a SmartDoc 2.0 blue light imaging system (Accuris Instruments, New Jersey, United States). PCR products

were purified with ExoSapIT Express (Applied Biosystems, California, United States) using the manufacturer's instructions and prepared for sequencing with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, California, United States), with the PCR primers used for the second nested PCR step for JMTV (Yu et al., 2020). Capillary sequencing was done on an ABI 3,730 automated sequencer (PE Applied BioSystems) at the Laboratories of Analytic Biology, Smithsonian Institution—National Museum of Natural History.

Bioinformatics and Phylogenetic Analysis

Raw data from the Illumina NovaSeq sequencing runs were uploaded to the CZ-ID platform (formerly ID-Seq), a publicly accessible cloud-based storage and analysis platform for metagenomic pathogen detection (Kalantar et al., 2020), and were analyzed using the built-in pipeline (version 6.8). The pipeline is open source and incorporates various steps for validation, which include selected host, human and barcode adaptor removal, quality assessment, alignment, assembly, and taxonomic identification using the National Center for Biotechnology Information (NCBI) nucleotide and protein databases. Pathogens were considered as significant when ≥ 2 nucleotide reads or contigs were verified in the final output. Relative abundances were calculated in samples with bacterial reads $\geq 10^2$ and expressed as taxon read count/total bacteria read count. Reads and contigs from viruses, parasites, and tick-borne bacteria were manually reviewed with the taxon identity confirmed using BLAST (Altschul et al., 1990).

Viral reads were also recovered using the VirIdAl pipeline (Budkina et al., 2021). This pipeline merged paired-end reads to increase sequence length after which sequences were adapter-trimmed and quality-filtered using fastp (Chen et al., 2018). Sequences with a complexity of $<30\%$, with an average PHRED quality score of <20 , and with <36 bases were removed. Host sequences were also removed with Bowtie2 (Langmead and Salzberg, 2012), using the most closely related available tick reference genome (*Rhipicephalus sanguineus*, ASM133396v1). Data were then clustered using vsearch (Rognes et al., 2016) to recover “centroid” sequences formed with a 0.9 identity threshold. These sequences then entered a two-step alignment process. First, sequences were blasted against both viral NT and NR databases at high sensitivity (e -value 10^{-3}) to give “virus-like” sequence matches. These were then passed to a second step alignment to the complete NCBI NT and NR databases (e -value 10^{-10}) for final classification. We used CZ-ID for the initial workup and VirIdAl on samples with detectable virus signal. Outputs from both pipelines were combined for producing final contigs, following duplicate removal.

Virus genome assemblies and Sanger sequencing data analyses were carried out using Geneious Prime version 2022.0.2¹. Nucleotide and deduced amino acid sequence alignments and pairwise comparisons were generated using CLUSTAL W (Thompson et al., 1994). MEGA11 was used for estimating the

optimal substitution model on individual alignments and to infer evolutionary history according to the Bayesian information criteria (Tamura et al., 2021).

The raw reads from tick pools tested in this study are freely available in <https://idseq.net/> under the project name “KWS ticks” and in National Center for Biotechnology Information (NCBI) biosample database with the ID: SUB11147880.

RESULTS

A total of 75 adult ticks collected from 15 animal species were sequenced in 46 pools. Most of the pools (84.8%) comprised single ticks, whereas multiple ticks (c. 2–8 individuals) were screened in 7 of the 46 pools (15.2%) (Table 1). Majority of the samples comprised *Rhipicephalus* spp. (73/75, 97.3%) and *Amblyomma* spp. (2/75, 2.7%). Host sequence data were utilized to confirm the genus-level identification in the tick pools. The average raw reads count of 9,313,080 was recorded per sample, whereas it was noted as 190,915 following the low-quality read and host removal.

Bacteria

Of the 46 pools sequenced, 35 tick pools were analyzed for pathogenic bacterial species and relative abundances of major phyla. Eleven pools were omitted from the analysis due to the relatively low number of total bacteria reads ($<10^2$). Members of the phylum *Proteobacteria* were abundant in the bacterial composition of the tick pools, comprising 51.6–97.5% of the total bacterial reads in 17 (48.6%) pools, regardless of the host species (Supplementary Table 1). *Actinobacter* and *Firmicutes* species provided the major source of bacterial reads in 6 (17.1%) and 2 (5.7%) pools, respectively. Predominance of a single bacterial species (with reads constituting over 50%) within a phylum was observed in certain pools, namely, *Proteus mirabilis* (pools #6, #48, #52, #53, #56, and #60), *Clostridium botulinum* (pools #3, #69, and #84), *Coxiella burnetii* (pool #21), and *Enterobacter cloacae* (pool #78). Other than these instances, no particular tendency in relative bacterial abundances or host species was noted.

A total of 56 bacterial species documented as human and/or animal pathogens were identified across the 35 pools examined (Figure 2). The most frequently detected pathogen was *E. coli* (22/35, 62.8%), followed by *P. mirabilis* (17/35, 48.5%) and *C. burnetii* (16/35, 45.7%). *C. burnetii* is the etiologic agent of Q fever, characterized by febrile and respiratory symptoms. Infections are zoonotic and are initiated by inhalation of particles contaminated by infected animals, most commonly sheep, goats, and cattle, with ticks implicated in transmission (Körner et al., 2021). *E. coli* and *P. mirabilis* are Gram-negative bacteria classified in the family *Enterobacteriaceae* alongside other potential pathogenic species in the genera *Klebsiella*, *Enterobacter*, *Citrobacter*, *Salmonella*, *Shigella*, and *Serratia*. Many are found in the normal gut flora of humans and animals, and others are widely distributed in soil and water. Commonly associated with infections of the urinary and gastrointestinal tracts,

¹<https://www.geneious.com>

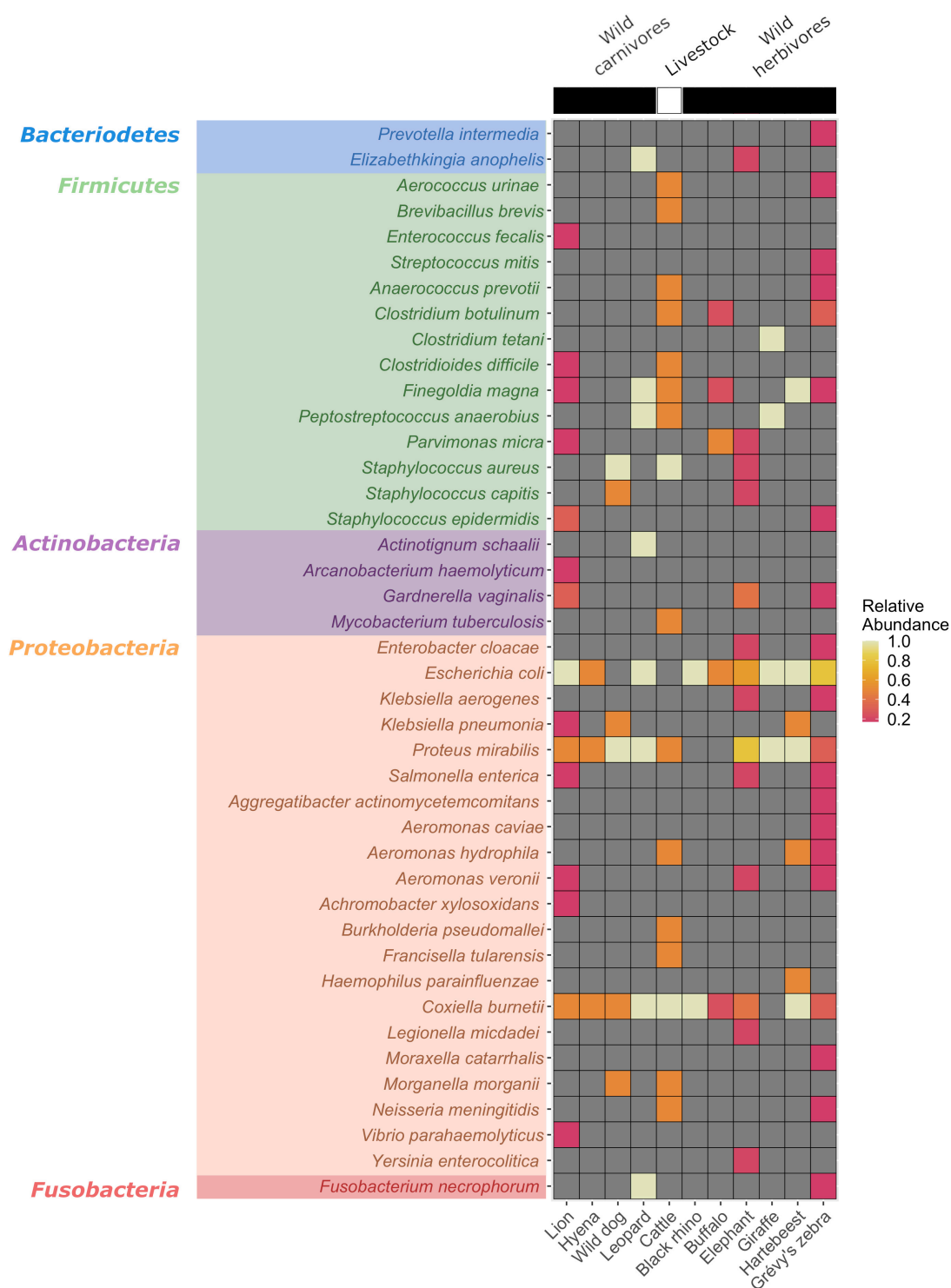


FIGURE 2 | Heatmap of the detected bacterial pathogens according to the hosts. Pathogen detection rates were calculated as positive tick pools divided by the number of hosts examined.

they can also produce severe and life-threatening, as well as nosocomial, infections (Mairi et al., 2018). Other Gram-negative bacteria of the genera *Achromobacter*, *Aeromonas*,

Aggregatibacter, *Hemophilus*, *Morganella*, *Moraxella*, *Neisseria*, *Vibrio*, and *Yersinia* were detected in this dataset, with varying prevalences (Figure 2).

TABLE 2 | Overview of non-bacterial pathogens detected in tick pools.

	Species	Prevalence	Host ¹	Pool Code	Read Count
Fungi	<i>Aspergillus flavus</i>	1/46 (2.1%)	Hartebeest (1/2)	#54	4
	<i>Candida parapsilosis</i>	2/46 (4.3%)	Hartebeest (1/2)	#54	7
			Lion (1/7)	#23	2
	<i>Cladophialophora bantiana</i>	1/46 (2.1%)	Lion (1/7)	#10	2
	<i>Cladophialophora carrionii</i>	1/46 (2.1%)	Leopard (1/1)	#81	4
	<i>Cryptococcus neoformans</i>	1/46 (2.1%)	Leopard (1/1)	#81	2
	<i>Cryptococcus gattii</i>	1/46 (2.1%)	Hartebeest (1/2)	#54	2
	<i>Pneumocystis jirovecii</i>	1/46 (2.1%)	Lion (1/7)	#47	3
	<i>Rhizopus microsporus</i>	1/46 (2.1%)	Leopard (1/1)	#81	2
Parasites	<i>Brugia malayi</i>	2/46 (4.3%)	Leopard (1/1)	#81	2
			Lion (1/7)	#47	4
	<i>Dirofilaria immitis</i>	1/46 (2.1%)	Elephant (1/7)	#52	2
	<i>Encephalitozoon intestinalis</i>	1/46 (2.1%)	Lion (1/7)	#47	2
	<i>Entamoeba histolytica</i>	1/46 (2.1%)	Hyena (1/4)	#68	2
	<i>Loa loa</i>	1/46 (2.1%)	Elephant (1/7)	#52	2
	<i>Naegleria fowleri</i>	1/46 (2.1%)	Elephant (1/7)	#52	3
	<i>Plasmodium falciparum</i>	3/46 (6.5%)	Lion (3/7)	#43,#47,#53	8,2,4
	<i>Plasmodium malariae</i>	1/46 (2.1%)	Hartebeest (1/2)	#54	3
	<i>Plasmodium vivax</i>	1/46 (2.1%)	Hartebeest (1/2)	#54	2
	<i>Schistosoma haematobium</i>	1/46 (2.1%)	Cattle (1/4)	#3	2
	<i>Schistosoma mansoni</i>	4/46 (8.7%)	Hartebeest (2/2)	#54,#57	2,3
			Lion (1/7)	#47	5
			Elephant (1/7)	#52	4
	<i>Toxoplasma gondii</i>	2/46 (4.3%)	Leopard (1/1)	#81	2
			Wild dog (1/1)	#60	2
	<i>Trichinella spiralis</i>	6/46 (13%)	Lion (2/7)	#47,#53	4,736
			Elephant (1/7)	#52	54
			Leopard (1/1)	#81	4
			Grévy's zebra (1/7)	#83	7
			Hartebeest (1/2)	#54	8
	<i>Trypanosoma cruzi</i>	3/46 (6.5%)	Grévy's zebra (1/7)	#70	2
			Leopard (1/1)	#81	2
			Lion (1/7)	#47	2
Viruses	Human mastadenovirus C	10/46 (21.7%)	Grévy's zebra (2/10)	#67,#83	4,6
			Leopard (1/10)	#81	22
			Elephant (3/10)	#6,#65,#71	2,2,4
			Lion (3/10)	#10,#22,#43	2,4,4
			Wild Dog (1/10)	#56	2
	Bovine alphaherpesvirus 5	4/46 (8.7%)	Buffalo (1/3)	#84	34
			Leopard (1/3)	#81	24
			Cattle (1/3)	#3	72
			Grévy's zebra (1/3)	#69	3
			Lion (1/1)	#47	1660
	Epstein-Barr virus	1/46 (2.1%)	Elephant (2/7)	#41,#52	2, n.a.
	Jingmen tick virus	6/46 (13%)	Cattle (1/4)	#11	2
			Lion (1/7)	#10	2
			Hartebeest (1/2)	#54	n.a.
			Wild Dog (1/2)	#60	n.a.

¹ Positive/total number of the species.

n.a.: not applicable due to detection by specific PCR.

Of special note is the identification of *Francisella tularensis* and *Bacillus anthracis*, two zoonotic agents classified as Category A pathogens (National Institutes of Allergy and Infectious Diseases, 2018),

owing to their significant public health threat and bioterrorism potential. *F. tularensis* was detected in 5 tick pools (5/35, 14.2%, read count range: 2–60) collected from cattle, lions, an elephant, and a Grevy's zebra, whereas *B. anthracis* was identified in

a single pool (1/35, 2.8%, read count: 22) collected from an elephant. A Gram-negative facultative intracellular bacterium, *F. tularensis*, causes tularemia, an anthroponosis that occurs via contact with wild animals, arthropod bites, or environment through respiratory, oral, or cutaneous routes (Maurin, 2020). Symptomatic disease may present with distinct clinical forms and can be mild or severe, depending on the entry route. As *Ixodidae* ticks are common arthropod vectors, detection of *F. tularensis* in tick pools indicates probable activity in the region. *B. anthracis* is an endospore-forming Gram-positive rod, causing anthrax in livestock and humans. In humans, cutaneous, respiratory, or gastrointestinal infections have been described, which can be fatal without proper treatment (Finke et al., 2020). A related bacteria, *Bacillus cereus*, mostly associated with gastrointestinal symptoms, was also detected (Figure 2).

We identified other *Firmicutes* such as *Burkholderia pseudomallei*, an opportunistic environmental bacterial pathogen of humans and animals, causing melioidosis (Limmathurotsakul et al., 2016). Moreover, pathogenic *Vibrio*, including *Vibrio cholerae*, were also detected. Facultative anaerobic cocci of the *Streptococcus* and *Staphylococcus* genera and anaerobic cocci and bacilli of various genera including *Clostridium* spp. were also observed. Most of the listed bacteria are ubiquitous in soil and gut microbiota of various animals as well as humans. Interestingly, *C. botulinum*, producer of the botulinum neurotoxin, was detected in 4 tick pools and emerged as the single predominant bacterial species in 3 of these pools, collected from cattle, buffalo, and Grevy's zebra. Documented pathogens within the *Actinomycetota* phylum include *Corynebacterium*, *Gardnerella*, and *Mycobacterium* species. *Fusobacterium necrophorum*, classified in the phylum *Fusobacteriota*, is observed as the only pathogen outside the aforementioned bacterial phyla (Holm et al., 2016; Figure 2).

Fungi

Eight fungal species listed as human pathogens were detected in 5 tick pools (10.8%) with low number of reads (range: 2–7). *Candida parapsilosis* was the most frequent pathogen (4.3%) identified in ticks collected from a hartebeest and a lion (Table 2). It is among the major causative agents of hospital-acquired invasive fungal disease, with severe outcomes in critically ill patients (Tóth et al., 2019). Not being an obligate pathogen, *C. parapsilosis* can be found in domestic animals, insects, soil, and marine environments, as well as a commensal of the human skin. Similarly, the remaining fungi are mostly opportunistic pathogens frequently found in the environment. They are associated with transcutaneous inoculation (chromoblastomycosis, *Cladophialophora* spp.) or diagnosed as opportunistic pathogens in immunocompromised hosts, with the exception of *Cryptococcus* spp., and are mainly disseminated by inhalation of the infectious particles from environmental sources (Gushiken et al., 2021).

Parasites

A total of 14 protozoan or nematode species listed as human pathogens were detected in 13 pools (28.2%), mostly in low to moderate number of reads (range: 2–736) (Table 2). The most

frequent of these was *Trichinella spiralis*, detected in 6 pools (13%) collected from various hosts. It is a nematodal parasite of many carnivorous and omnivorous animals, where infections occur by ingestion of muscle tissue with encysted larvae (Ribicich et al., 2020). Human infections called trichinellosis or trichinosis are due to the consumption of undercooked meat. In most hosts, adult worms present in the intestine may continue to produce

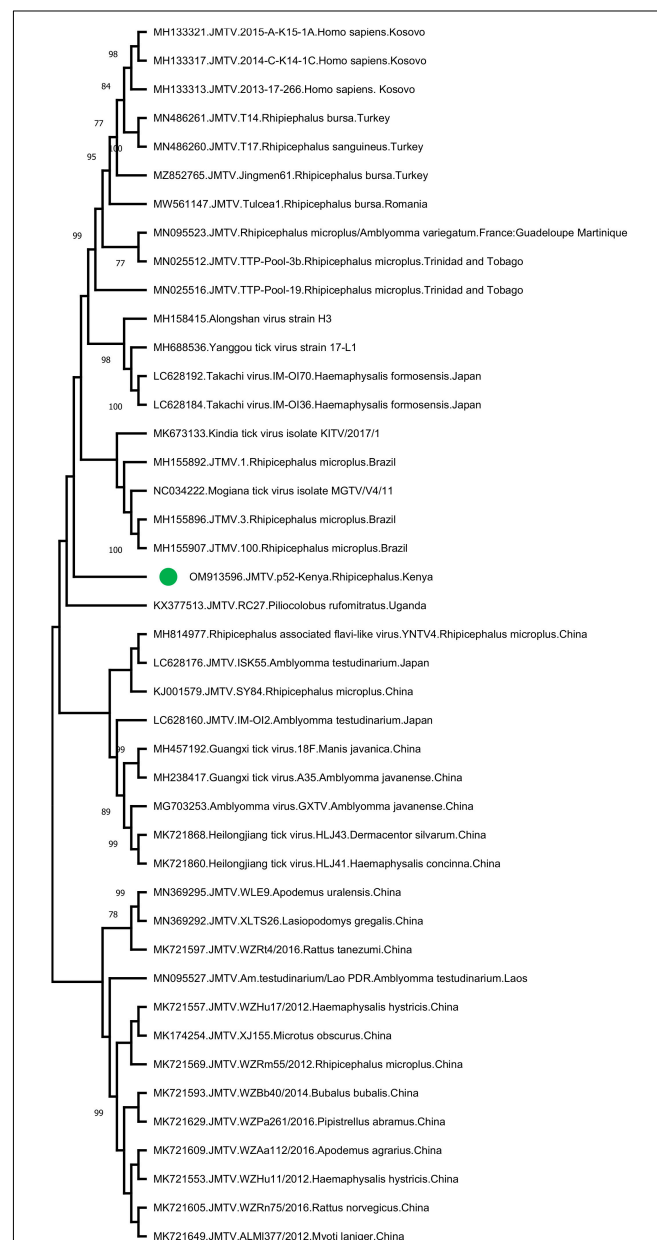


FIGURE 3 | The maximum likelihood analysis of the Jingmen tick virus (JMTV) partial segment 1 sequences (366 nucleotides). The tree is constructed using a gamma-distributed Kimura 2-parameter model for 500 replications. JMTVs included in the analysis are indicated by GenBank accession number, name, isolate/strain identifier, host, and detection region. JMTV sequence characterized in this study is marked. Bootstrap values greater than 75 are displayed.

larvae for prolonged periods. Other than *Trichinella*, *Schistosoma mansoni*, *Trypanosoma cruzi*, and *Plasmodium* spp. were also detected, with prevalences of 6.5–8.7% (Table 2). However, co-detection of several parasites in particular pools [such as pool #52 from an elephant (*T. spiralis*, *S. mansoni*, *Naegleria fowleri*, *Loa loa*, *Dirofilaria immitis*) and pool #47 from a lion (*T. spiralis*, *S. mansoni*, *T. cruzi*, *Brugia malayi*, *Encephalitozoon intestinalis*)] strongly suggests an environmental origin for most of the identified pathogens. Moreover, human *Plasmodium falciparum* sequences detected most likely indicate prior feeding of the ticks on infected humans, evidencing the potential for cross-species zoonotic pathogen transfer from wildlife to humans via tick bites in African dryland ecosystems (Cable et al., 2017).

Viruses

We recovered 4 viruses with significant human or veterinary health impact (Table 2). Human mastadenovirus C was most frequent viral pathogen, detected in 10 tick pools (10/46, 21.7%). Commonly referred as the adenoviruses, mastadenoviruses comprise a separate genus in the Adenoviridae family. They are non-enveloped viruses with double-stranded DNA genomes. Human mastadenovirus C is among the 51 currently described species of the genus, for which humans and other mammals such as bats, bovids, canine, equids, caprids and suids serve as natural hosts (Harrach et al., 2019). Symptomatic infections in humans may result in acute respiratory illness of variable severity, as well as conjunctivitis and gastroenteritis, which are usually mild.

Bovine alphaherpesvirus 5 (BHV-5) was detected in 4 tick pools (8.7%), including one collected from cattle (pool #3, Table 2). It is an enveloped double-stranded DNA virus classified in the *Varicellovirus* genus of the *Alphaherpesvirinae* subfamily. Similar to other alphaherpesviruses, BHV-5 establishes latency in the central nervous system of the exposed animals and is excreted in ocular, nasal and genital secretions upon reactivation (Del Médico Zajac et al., 2010). It causes severe meningoencephalitis in cattle, with high mortality in young

calves. Sporadic cases and outbreaks have been reported in several countries in South America, Europe, and Asia. Another herpes virus, Epstein-Barr virus (EBV), was also detected in a tick pool collected from a lion (Table 2). It is another globally prevalent human herpesvirus and the causative agent of infectious mononucleosis. Originally discovered from African Burkitt's lymphoma cells, EBV mainly infects B lymphocytes, may cause reactivating infections, and is associated with particular autoimmune syndromes (Houen and Trier, 2021).

Jingmen tick virus is the only virus where tick-borne transmission appears as the main mode of spread in the study. JMTV sequences were detected in a total of 6 pools (13%) (Table 2). Metagenomic screening initially revealed 3 positive pools (from an elephant, cattle, and lion), whereas 3 additional pools were subsequently identified using JMTV-specific PCR amplification. Initially described from China, JMTV is the index of a new group of viruses (called the Jingmenvirus group) tentatively classified within *Flaviviridae* family (Temmam et al., 2019). JMTV possesses an RNA genome of 4 segments, where two segments encoding for the non-structural proteins are related to those from flaviviruses (Qin et al., 2014). It is reported as a human pathogen causing tick-borne infections presenting with mild to severe febrile diseases (Jia et al., 2019). We recovered a 366-nucleotide segment of the JMTV NS5-like protein-coding region by sequence analysis of the amplified product in pool #52 (GenBank accession: OM913596). The sequence displayed up to 94.62% nucleotide and 99.23% deduced amino acid identities to previously characterized JMTVs. Maximum likelihood analysis showed that the sequence is phylogenetically distinct from all previously described JMTV clades from various regions, as well as from related viruses such as Mogiana tick virus (Figure 3). The closest relative appeared as the JMTV-RC27, identified in a plasma sample of an eastern or Tana River red colobus monkey (*Piliocolobus rufomitratus*) from Uganda.

In addition to those listed above, we identified other viruses not currently associated with disease in vertebrates. In 8 (17.3%)

TABLE 3 | Viruses further detected in tick pools.

	Pool	Read Count	Host	Contig Length	Genome Location ¹	Similarity ¹	
						nt	aa
Trinbago virus	3	18	Cattle	338 bp	12823-13160	93.7%	97.3%
	11	14	Cattle	436 bp	12042-12477	91.9%	95.1%
	43	13	Lion	334 bp	10488-10821	87.4%	85.5%
	53	2	Lion	217 bp	12680-12896	71.4%	70.4%
	57	2	Hartebeest	230 bp	13839-14068	92.6%	94.7%
	70	3	Grévy's zebra	246 bp	6414-6659	94.3%	95.0%
	83	3	Grévy's zebra	470 bp	925-1394	85.1%	85.8%
	81	46	Leopard	1141 bp	4679-5819	89.8%	93.6%
Guarapuava tymovirus-like 1 virus	83	32	Grévy's zebra	486 bp	4657-5142	79.2%	90.1%
	53	78	Lion	1802 bp	3564-5348	82.4%	91.6%
	52	2	Elephant	336 bp	3872-4207	81.2%	88.2%
	43	1238	Lion	5898 bp	172-6069	80.4%	89.7%

¹ According to Trinbago virus isolate TTP-Pool-4 (MN025505) and Guarapuava tymovirus-like 1 isolate 3 (MH155881). bp: base pairs, nt: nucleotide, aa: amino acid.

of the pools, sequences with the highest nucleotide and deduced amino acid identities to Trinbago virus (TBOV) were detected. TBOV is a recently described virus, distantly related to *Pestivirus* genus of the family *Flaviviridae* (Sameroff et al., 2019). The TBOV sequences identified in ticks vary in size and cover

different parts of the viral genome without overlaps (Table 3 and Supplementary Table 2). The longest contig (1,141 base pairs), originating from pool 81 and covering approximately 10% of the genome, was used for phylogeny construction. In the maximum likelihood analysis, this sequence remained distinct and shared

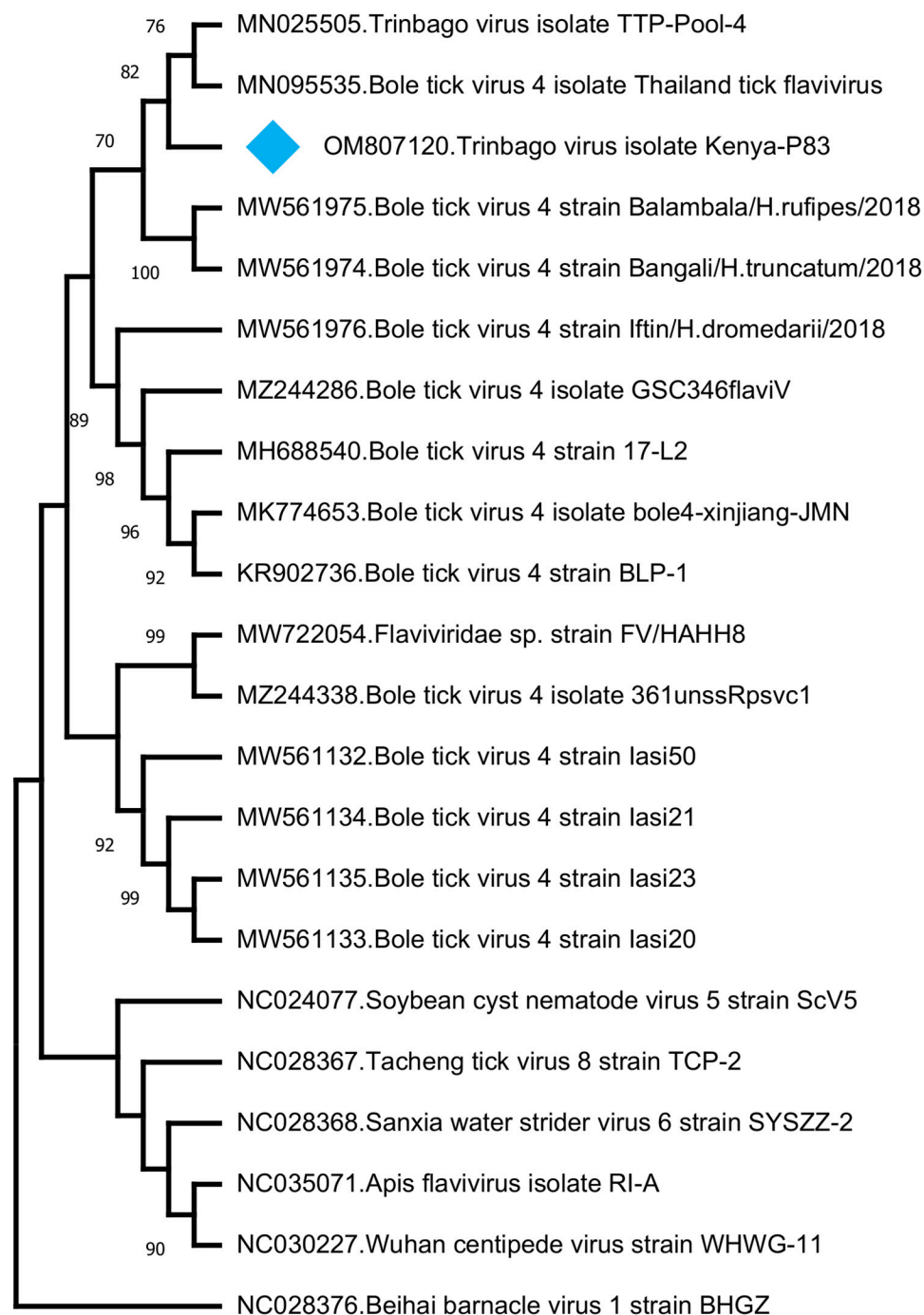


FIGURE 4 | The maximum likelihood analysis of the Trinbago virus partial genome sequences (1,141 nucleotides). The tree is constructed using a gamma-distributed Kimura 2-parameter model for 500 replications. Viruses included in the analysis are indicated by GenBank accession number, name and isolate/strain identifier. Trinbago virus isolate Kenya-P83 characterized in this study is marked (GenBank accession: OM807120). Beihai barnacle virus 1 strain BHGZ is included as the outgroup.

a common ancestor with the initial TBOV isolate (identified in *R. sanguineus* ticks from Trinidad and Tobago) and closely related Bole tick virus 4 isolate Thailand tick flavivirus (**Figure 4**).

We further identified Guarapuava tymovirus-like virus (GTLV) sequences in 4 (8.7%) pools and pairwise comparisons revealed the highest identities to GTLV-1 (**Table 3** and **Supplementary Table 3**). GTLVs have been recently described in ticks and appear as highly divergent viruses within the order *Tymovirales* (Souza et al., 2018). Partially overlapping contigs and a complete coding sequence could be obtained from the pools (tentatively named as GTLV-1 isolate Kenya-P43), with no further effort to confirm the non-coding 5' and 3' ends of the viral genome.

Similar to GTLVs, two open reading frames (ORFs) are recognized in the GTLV-1 isolate Kenya-P43 genome, encoding for the ORF1 polyprotein (nucleotides 1–5031) and viral capsid (nucleotides 5,041–5,898) (Souza et al., 2018). Analysis of the deduced amino acid sequences of these viral proteins revealed several functional motifs pertaining to intracellular virus replication (**Table 4**). Phylogeny reconstruction revealed clustering of GTLV-1 and 2 viruses and a distinct separation of GTLV-1 isolate Kenya-P4 (**Figure 5**).

DISCUSSION

Our cross-sectional metagenomic investigation of ticks collected from wildlife species revealed a broad spectrum of bacterial, fungal, parasitic, and viral pathogens. We identified tick-borne pathogens *F. tularemia* and JMTV in 14.2 and 13% of the tick pools, respectively, with both agents observed in ticks collected from wild animals and cattle. Currently, tularemia is considered a global re-emerging zoonotic disease and *F. tularemia* as a potential agent of biological warfare, due to its ease of aerosol dissemination, low infectious dose, and potential fatal outcomes (Maurin, 2015). Other than the evidence of probable exposure in patients with undifferentiated febrile illness (Njeru et al., 2017), *F. tularemia* has not previously been reported from Kenya. Our findings indicate that it is present in potential vectors in livestock and wildlife, requiring further screening and specific diagnosis in cases presenting with compatible symptoms.

The JMTV is a recently described virus associated with tick-borne diseases in humans. Initially discovered in *Rhipicephalus microplus* ticks from China, the virus has been shown to be

distributed in many countries in Eurasia and the Americas, and also in a variety of non-tick hosts including cattle, rodents, bats, and humans (Guo et al., 2020). Human infections were initially documented in Crimean-Congo hemorrhagic fever cases from Kosovo (Emmerich et al., 2018) and later confirmed in China (Jia et al., 2019). Moreover, genetically related viruses with similar genomic structure—some also associated with human disease (Alongshan virus) have been described—are collectively referred to as the Jingmen virus group (Shi et al., 2015; Ladner et al., 2016; Kholodilov et al., 2020, 2021). The first documentation of a Jingmen virus in the African continent was a JMTV variant genome sequence (RC27) in Uganda, detected in plasma collected from a red colobus monkey (*P. rufomitratus*), a critically endangered primate (Ladner et al., 2016). Hereby, we report the detection of JMTV in ticks from Kenya as well as from Africa. Interestingly, phylogenetic reconstructions based on a relatively short segment of the virus genome showed significant divergence from previously known JMTVs and related viruses, likely to represent variations due to geographical segregation (**Figure 3**). Further screening and sequencing efforts will elucidate JMTV genome diversity in Kenya as well as the African continent. As of this writing, JMTV was reported in ticks from two other pastoralist-dominated areas (Ogola et al., 2022).

We further detected another pathogenic bacteria, *C. burnetii* in 45.7% of the tick pools and in all host species except for a giraffe (**Figure 2**). It is also observed as one of the frequently identified bacteria in our tick cohort. *C. burnetii* is known to display a wide host range and can replicate in many mammalian, avian, and reptilian species, as well as arthropods (Yon et al., 2019). However, the role of wildlife in the Q fever epidemiology in livestock and humans is yet to be established (Körner et al., 2020). Moreover, despite evidence for vector potential, the role of ticks in transmission still remains controversial. The presence of *C. burnetii* in Kenya is well established, with detections in various cohorts of ticks and their hosts, including those from wildlife and at the wildlife-livestock interface (Knobel et al., 2013; Ndeereh et al., 2017; Koka et al., 2018; Getange et al., 2021). It demonstrates a widespread circulation in the country and is likely to be a causative agent of undiagnosed febrile disease in pastoral communities. Our findings confirm previous reports and suggest wildlife-livestock interactions to contribute further to *C. burnetii* transmission.

Other bacterial agents with high pathogenicity identified in the study include *B. anthracis* and *C. botulinum*. Anthrax is among the top priority zoonotic diseases in Kenya (Nderitu et al., 2021). Infections have been reported in over 30% of Kenyan wildlife conservancies, with the majority occurring in dry seasons and primarily affecting herbivore species including buffalo, black and white rhinos, and elephants (Gachohi et al., 2019). In this study, we detected *B. anthracis* in a single tick (2.8%) collected feeding on an elephant, demonstrating ongoing activity in wildlife. Botulinum toxin produced by the anaerobic rod *C. botulinum* causes severe poisoning, frequently occurring due to oral intake. *C. botulinum* is ubiquitously found in soil and aquatic sediments that serve as an environment for sporulation (Espelund and Klaveness, 2014). Botulinum toxin produced by *C. botulinum* can intoxicate and kill various animal species upon

TABLE 4 | Functional motifs in the Guarapuava tymovirus-like 1 virus isolate Kenya-P43 genome.

Motif	Domain Accession	Location ¹
Viral methyltransferase	pfam01660	54–326
UL36 large tegument protein	PHA03247	458–623
Tymovirus endopeptidase	pfam05381	658–742
Viral helicase	pfam01443	836–1071
RNA-dependent RNA polymerase	pfam00978	1364–1585
Tymovirus capsid protein	pfam00983	1784–1959

¹According to polyprotein deduced amino acid sequence (OM807119).

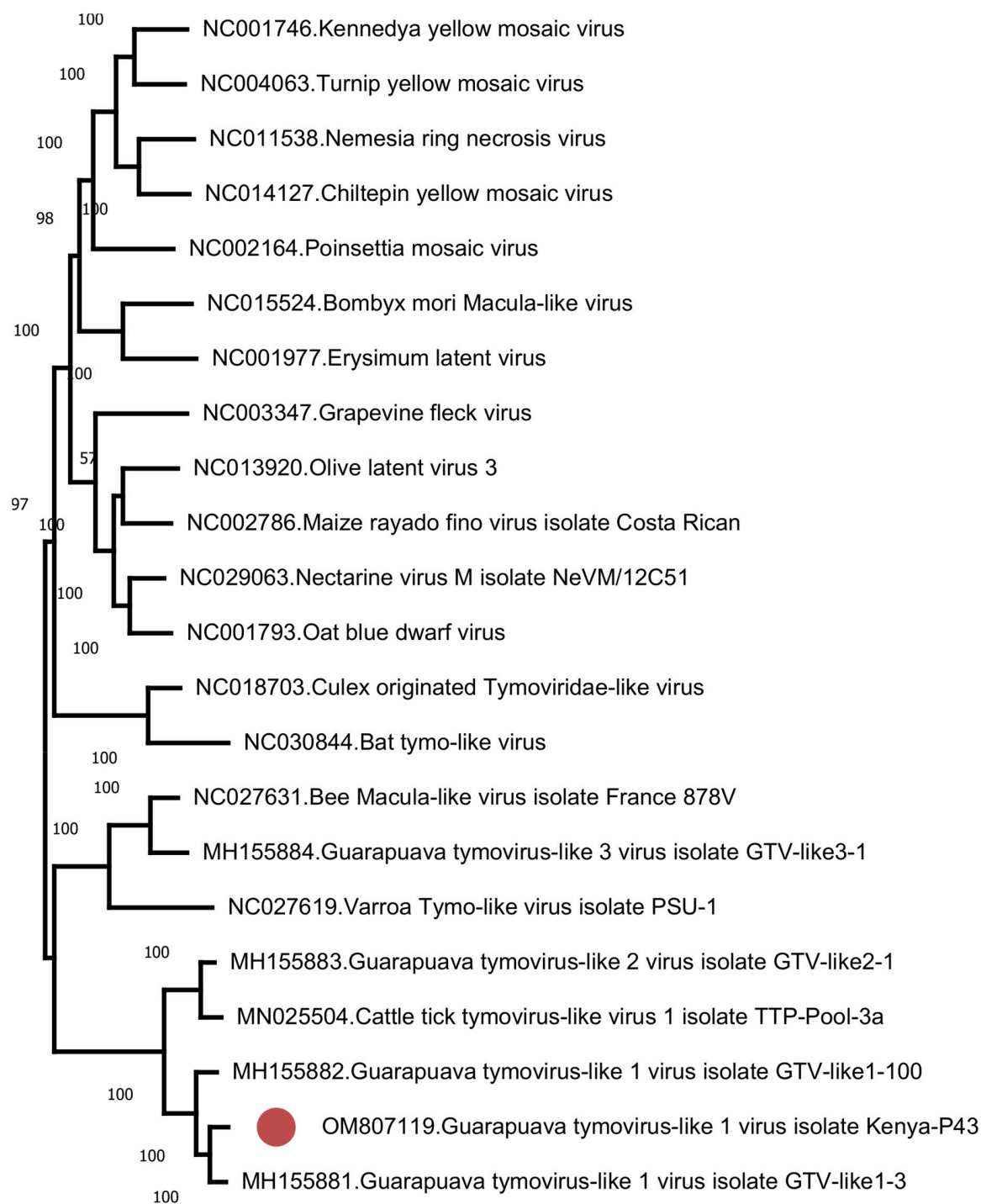


FIGURE 5 | The maximum likelihood analysis of the Guarapuava tymovirus-like virus ORF sequences (5,898 nucleotides). The tree is constructed using the general time reversible (GTR) model, gamma distributed with invariant sites (G + I) model for 500 replications. Guarapuava tymovirus-like 1 virus isolate Kenya-P43 characterized in this study is marked (GenBank accession: OM807119). Bootstrap values greater than 50 are displayed.

entering their food webs. It has been detected in soil specimens and has caused documented outbreaks in Kenya (Smith et al., 1979; Nightingale and Ayim, 1980; Yamakawa et al., 1990). We identified *C. botulinum* in a total of 4 pools (8.7%) collected from cattle, buffalo, and endangered Grevy's zebra. In this study, the source is likely to be environmental in origin. However, the presence of *C. botulinum* as the single bacterial species in 3 pools suggests that it can readily predominate and pose a threat to human and animal health without particular precautions.

In addition to the tick-borne or highly pathogenic species discussed above, we detected many species of bacteria in the tick pools, where *E. coli* and *P. mirabilis* constitute the most frequent species, identified in 62.8% and 48.5% of the pools, respectively. Furthermore, several other species of *Enterobacteriaceae*, *Staphylococci*, *Streptococci*, and *Clostridia* spp. other than *C. botulinum* were present in tick pools (Figure 2). These bacteria are either ubiquitously found in soil or are present in gut or skin microbiota of several host species including humans, some being consistently shed in feces. Although they are widespread in nature, they can be opportunistic pathogens in specific hosts in various settings (such as nosocomial infections) (Lax and Gilbert, 2015) or can cause zoonotic infections (such as *B. pseudomallei*). Similarly, we identified several fungi including *Aspergillus*, *Candida*, *Cladophialophora*, *Cryptococcus*, *Pneumocystis*, and *Rhizopus* species, found widespread in nature, that mostly cause opportunistic infections. However, *Cryptococcus neoformans* and *Cryptococcus gattii*, detected in the study in 4.2% of the pools, are documented to produce severe infections in immunocompetent individuals as well (May et al., 2016; Hsu et al., 2022). Infections and genotypes of *Cryptococcus* species have been previously reported from Kenya (Kangogo et al., 2015).

Our metagenomic analysis displayed several medically important parasites, most of which are endemic in the region (Table 2). The vector-borne filarial pathogens (*B. malayi*, *D. immitis*, and *L. Loa*) and protozoa (*Plasmodium* spp. and *T. cruzi*) detected in ticks probably represent blood meals from infected hosts. The environmental and water-/food-borne pathogens *Entamoeba histolytica*, *Encephalitozoon intestinalis*, *Naegleria fowleri*, *Schistosoma* spp., *Toxoplasma gondii*, and *Trichinella spiralis* were detected, which indicate, along with the diversity of parasites detected in single ticks, contamination from environmental sources. These findings demonstrate a marked potential for wildlife-livestock-human pathogen transfer. This is also supported by the frequent detection of adenovirus and herpesvirus that cause human or cattle infections (Table 2).

We further described two recently characterized viruses with currently unknown health impact, namely, GTLV-1 and TBOV, and documented their first reporting from the African continent. These viruses were prevalent in ticks with co-detection in three pools (Table 3). TBOV was initially discovered in Trinidad and Tobago and was present in all *Rhipicephalus* and *Amblyomma* spp. screened in the region (Sameroff et al., 2019). The TBOV genome shares significant similarities to Bole tick virus 4, a tick-associated virus discovered in China (Shi et al., 2015). Particular non-structural proteins from both viruses share similarities to Pestiviruses in the family *Flaviviridae*, albeit

forming a distinct phylogenetic clade suggesting a novel genus (Sameroff et al., 2019). TBOV has been documented in a wide range of tick species, suggesting acquisition from an unknown vertebrate host. Since *Flaviviridae* family includes several tick-borne viral pathogens, potential association of TBOV with vertebrate infections or its impact on vector capacity for tick-borne flaviviruses requires further investigation. Virome analyses performed in ticks collected from camels in Kenya showed Bole tick virus 4 in three *Hyalomma* tick species and virus exposure in hosts (Zhang et al., 2021). We further detected a local GTLV-1 isolate and analyzed its complete coding region (Table 4). Similar to TBOV, GTLVs including GTLV 1, 2, and 3 are novel viruses described recently in *R. microplus* ticks from Brazil, tentatively named due to their limited genome similarities to members of the order *Tymovirales* (Souza et al., 2018). Tymoviruses are plant viruses associated with mosaic disease where arthropods facilitate spread as mechanical vectors (Lefkowitz et al., 2018). Due to their divergent genomes, GTLVs have been proposed as a novel family, with currently unknown pathogenicity in invertebrate hosts (Souza et al., 2018).

Finally, particular limitations or shortcomings of the study must be addressed. Due to the sampling approach and storage conditions, we were unable to provide a species-level identification of ticks. Moreover, the sequencing runs produced a relatively lower number of reads in the tick pools, which is probably due to ethanol storage, a frequently used approach to preserve entomological specimens. For metagenomic investigations, we used a straightforward cDNA-based approach, without prior treatment to enrich particular targets such as viruses. Nevertheless, given the breadth of microbial pathogens detected, our approach has been successful even in ethanol-stored specimens, enabling the identification of many human and animal pathogens, some having been documented for the first time in the region. We also observed ticks as a promising sentinel to monitor pathogen circulation in domestic-wildlife interfaces in the Kenyan landscape. Due to their natural life cycle that may involve several host animals, they provide information not only on tick-borne agents but from hosts and the environment as well. Despite the costs and requirement of a significantly equipped laboratory infrastructure and trained personnel, metagenome-based investigations are capable of producing crucial information on the identification of circulating pathogens, prioritization of targets for surveillance, or management of mitigation efforts.

In conclusion, we detected several microbial pathogens by a metagenomic approach in ticks collected from animals at the livestock-wildlife interfaces in Kenya. Tick-borne pathogens JMTV and *F. tularensis* were documented in Kenya, as well TBOV and GTLV-1, with currently unexplored impact on vertebrate pathogens.

DATA AVAILABILITY STATEMENT

The raw reads from tick pools tested in this study are available in <https://idseq.net/> under the project name "KWS ticks", and in the

National Library of Medicine - National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov>) Biosample and Sequence Read Archive (SRA) under accession numbers 28626882–28626931.

AUTHOR CONTRIBUTIONS

KE: laboratory analysis, data analysis and interpretation, and manuscript drafting. MM: wildlife specimen collection, transport, and initial processing. BB: laboratory analysis and data analysis and interpretation. SJ: laboratory analysis and data analysis. LC-Q: laboratory analysis and data analysis. JK: wildlife specimen collection, transport, and initial processing. SaM: wildlife specimen collection, transport, and initial processing. IA: wildlife specimen collection, transport, and initial processing. EC: wildlife specimen collection, transport, and initial processing. FG: wildlife specimen collection, transport, and initial processing. PO: wildlife specimen collection, transport, and initial processing. SuM: study design. DZ: study design and data interpretation. Y-ML: study design and conception, data interpretation, and manuscript drafting.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brinkmann, A., Nitsche, A., and Kohl, C. (2016). Viral metagenomics on blood-feeding arthropods as a tool for human disease surveillance. *Int. J. Mol. Sci.* 17:1743. doi: 10.3390/ijms17101743
- Budkina, A. Y., Korneenko, E. V., Kotov, I. A., Kiselev, D. A., Artyushin, I. V., Speranskaya, A. S., et al. (2021). Utilizing the VirIdAl pipeline to search for viruses in the metagenomic data of bat samples. *Viruses* 13:2006. doi: 10.3390/v13102006
- Cable, J., Barber, I., Boag, B., Ellison, A. R., Morgan, E. R., Murray, K., et al. (2017). Global change, parasite transmission and disease control: lessons from ecology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160088. doi: 10.1098/rstb.2016.0088
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chiuya, T., Masiga, D. K., Falzon, L. C., Bastos, A. D. S., Fèvre, E. M., and Villinger, J. (2021). Tick-borne pathogens, including crimean-congo haemorrhagic fever virus, at livestock markets and slaughterhouses in western Kenya. *Transbound. Emerg. Dis.* 68, 2429–2445. doi: 10.1111/tbed.13911
- Del Médico, Zajac, M. P., Ladelfa, M. F., Kotsias, F., Muylkens, B., Thiry, J., et al. (2010). Biology of bovine herpesvirus 5. *Vet. J.* 184, 138–145. doi: 10.1016/j.tvjl.2009.03.035
- Emmerich, P., Jakupi, X., von Pössel, R., Berisha, L., Halili, B., Günther, S., et al. (2018). Viral metagenomics, genetic and evolutionary characteristics of Crimean-Congo hemorrhagic fever orthonairovirus in humans kosovo. *Infect. Genet. Evol.* 65, 6–11. doi: 10.1016/j.meegid.2018.07.010
- Espelund, M., and Klaveness, D. (2014). Botulism outbreaks in natural environments - an update. *Front. Microbiol.* 5:287. doi: 10.3389/fmicb.2014.00287
- Finke, E. J., Beyer, W., Loderstädt, U., and Frickmann, H. (2020). Review: the risk of contracting anthrax from spore-contaminated soil - a military medical perspective. *Eur. J. Microbiol. Immunol. (Bp)*. 10, 29–63. doi: 10.1556/1886.2020.00008
- Gachohi, J., Gakuya, F., Lekolool, I., Osoro, E., Nderitu, L., Munyua, P., et al. (2019). Temporal and spatial distribution of anthrax outbreaks among Kenyan wildlife, 1999–2017. *Epidemiol. Infect.* 147:E249.
- Getange, D., Bargul, J. L., Kanduma, E., Collins, M., Bodha, B., Denge, D., et al. (2021). Ticks and tick-borne pathogens associated with dromedary camels (*Camelus dromedarius*) in Northern Kenya. *Microorganisms* 9:1414. doi: 10.3390/microorganisms9071414
- Guo, J. J., Lin, X. D., Chen, Y. M., Hao, Z. Y., Wang, Z. X., Yu, Z. M., et al. (2020). Diversity and circulation of Jingmen tick virus in ticks and mammals. *Virus. Evol.* 6:veaa051. doi: 10.1093/ve/veaa051
- Gushiken, A. C., Saharia, K. K., and Baddley, J. W. (2021). Cryptococcosis. *Infect. Dis. Clin. North Am.* 35, 493–514.
- Harrach, B., Tarjan, Z. L., and Benko, M. (2019). Adenoviruses across the animal kingdom: a walk in the zoo. *FEBS Lett.* 593, 3660–3673. doi: 10.1002/1873-3468.13687
- Holm, K., Bank, S., Nielsen, H., Kristensen, L. H., Prag, J., and Jensen, A. (2016). The role of *Fusobacterium necrophorum* in pharyngotonsillitis - a review. *Anaerobe* 42, 89–97.
- Houen, G., and Trier, N. H. (2021). Epstein-barr virus and systemic autoimmune diseases. *Front. Immunol.* 11:587380.
- Hsu, E., Webster, S. M., and Nanes, M. (2022). Disseminated Cryptococcosis in an immunocompetent host presenting as osteomyelitis and leading to adrenal insufficiency. *Am. J. Med. Sci.* 363, 75–79. doi: 10.1016/j.amjms.2020.12.007
- Jia, N., Liu, H. B., Ni, X. B., Bell-Sakyi, L., Zheng, Y. C., Song, J. L., et al. (2019). Emergence of human infection with Jingmen tick virus in China: a retrospective study. *EBioMedicine* 43, 317–324. doi: 10.1016/j.ebiom.2019.04.004
- Kalantar, K. L., Carvalho, T., de Bourcy, C. F. A., Dimitrov, B., Dingle, G., Egger, R., et al. (2020). IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* 9:giaa111. doi: 10.1093/gigascience/giaa111
- Kangogo, M., Bader, O., Boga, H., Wanyoike, W., Folba, C., Worasilchai, N., et al. (2015). Molecular types of *Cryptococcus gattii*/*Cryptococcus neoformans* species complex from clinical and environmental sources in Nairobi, Kenya. *Mycoses* 58, 665–670. doi: 10.1111/myc.12411
- Kenya National Bureau of Statistics (2021). Available online at: <https://www.knbs.or.ke> (accessed February 27, 2022).

All authors contributed, read and approved the submitted manuscript.

FUNDING

This study was funded through the Armed Forces Health Surveillance Division – Global Emerging Infectious Disease Surveillance (AFHSD-GEIS) Project P0031_21_WR (to Y-ML). Material contained within this publication has been reviewed by the Walter Reed Army Institute of Research and the Kenyan Wildlife Service. There is no objection to its presentation and/or publication. The opinions or assertions contained in this study are the private views of the authors and are not to be construed as official or as reflecting true views of the Department of the Army, Navy or the Department of Defense.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.932224/full#supplementary-material>

- Kholodilov, I. S., Belova, O. A., Morozkin, E. S., Litov, A. G., Ivannikova, A. Y., Makenov, M. T., et al. (2021). Geographical and tick-dependent distribution of flavi-like Alongshan and Yanggou tick viruses in Russia. *Viruses* 13:458. doi: 10.3390/v13030458
- Kholodilov, I. S., Litov, A. G., Klimentov, A. S., Belova, O. A., Polienko, A. E., Nikitin, N. A., et al. (2020). Isolation and characterisation of Alongshan virus in Russia. *Viruses* 12:362. doi: 10.3390/v12040362
- Knobel, D. L., Maina, A. N., Cutler, S. J., Ogola, E., Feikin, D. R., Junghae, M., et al. (2013). *Coxiella burnetii* in humans, domestic ruminants, and ticks in rural western Kenya. *Am. J. Trop. Med. Hyg.* 88, 513–518. doi: 10.4269/ajtmh.12-0169
- Koka, H., Sang, R., Kutima, H. L., and Musila, L. (2018). *Coxiella burnetii* detected in tick samples from pastoral communities in Kenya. *Biomed. Res. Int.* 2018:8158102.
- Körner, S., Makert, G. R., Mertens-Scholz, K., Henning, K., Pfeffer, M., Starke, A., et al. (2020). Uptake and fecal excretion of *Coxiella burnetii* by *Ixodes ricinus* and *dermacentor marginatus* ticks. *Parasit. Vectors* 13:75. doi: 10.1186/s13071-020-3956-z
- Körner, S., Makert, G. R., Ulbert, S., Pfeffer, M., and Mertens-Scholz, K. (2021). The Prevalence of *Coxiella burnetii* in hard ticks in Europe and their role in Q fever transmission revisited - a systematic review. *Front. Vet. Sci.* 8:655715. doi: 10.3389/fvets.2021.655715
- Ladner, J. T., Wiley, M. R., Beitzel, B., Auguste, A. J., Dupuis, A. P., Lindquist, M. E., et al. (2016). A multicomponent animal virus isolated from mosquitoes. *Cell Host. Microbe* 20, 357–367. doi: 10.1016/j.chom.2016.07.011
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lax, S., and Gilbert, J. A. (2015). Hospital-associated microbiota and implications for nosocomial infections. *Trends. Mol. Med.* 21, 427–432. doi: 10.1016/j.molmed.2015.03.005
- Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith, D. B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* 46, 708–717. doi: 10.1093/nar/gkx932
- Limmathurotsakul, D., Golding, N., Dance, D. A., Messina, J. P., Pigott, D. M., Moyes, C. L., et al. (2016). Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. *Nat. Microbiol.* 1:15008.
- Mairi, A., Pantel, A., Sotto, A., Lavigne, J. P., and Touati, A. (2018). OXA-48-like carbapenemases producing *Enterobacteriaceae* in different niches. *Eur. J. Clin. Microbiol. Infect. Dis.* 37, 587–604. doi: 10.1007/s10096-017-3112-7
- Maurin, M. (2015). *Francisella tularensis* as a potential agent of bioterrorism? *Expert. Rev. Anti Infect Ther.* 13, 141–144. doi: 10.1586/14787210.2015.986463
- Maurin, M. (2020). *Francisella tularensis*. tularemia and serological diagnosis. *Front. Cell Infect Microbiol.* 10:512090. doi: 10.3389/fcimb.2020.512090
- May, R. C., Stone, N. R. H., Wiesner, D. L., Bicanic, T., and Nielsen, K. (2016). *Cryptococcus*: from environmental saprophyte to global pathogen. *Nat. Rev. Microbiol.* 14, 106–117. doi: 10.1038/nrmicro.2015.6
- Miller, R. S., Farnsworth, M. L., and Malmberg, J. L. (2013). Diseases at the livestock–wildlife interface: status, challenges, and opportunities in the United States. *Prev. Vet. Med.* 110, 119–132. doi: 10.1016/j.prevetmed.2012.11.021
- Morrison, W. I., Hemmink, J. D., and Toye, P. G. (2020). *Theileria parva*: a parasite of African buffalo, which has adapted to infect and undergo transmission in cattle. *Int. J. Parasitol.* 50, 403–412. doi: 10.1016/j.ijpara.2019.12.006
- Mwamuye, M. M., Kariuki, E., Omondi, D., Kabii, J., Odongo, D., Masiga, D., et al. (2017). Novel Rickettsia and emergent tick-borne pathogens: a molecular survey of ticks and tick-borne pathogens in Shimba Hills National Reserve, Kenya. *Ticks Tick Borne Dis.* 8, 208–218. doi: 10.1016/j.ttbdis.2016.09.002
- National Institutes of Allergy and Infectious Diseases (2018). *NIAID Emerging Infectious Diseases/ Pathogens*. Available online at: <https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens> (accessed February 17, 2022)
- Ndeereh, D., Muchemi, G., Thaiyah, A., Otiende, M., Angelone-Alasaad, S., and Jowers, M. J. (2017). Molecular survey of *Coxiella burnetii* in wildlife and ticks at wildlife-livestock interfaces in Kenya. *Exp. Appl. Acarol.* 72, 277–289. doi: 10.1007/s10493-017-0146-6
- Nderitu, L. M., Gachohi, J., Otieno, F., Mogoa, E. G., Muturi, M., Mwatondo, A., et al. (2021). Spatial clustering of livestock Anthrax events associated with agro-ecological zones in Kenya, 1957–2017. *BMC Infect. Dis.* 21:191. doi: 10.1186/s12879-021-05871-9
- Nightingale, K. W., and Ayim, E. N. (1980). Outbreak of botulism in Kenya after ingestion of white ants. *Br. Med. J.* 281, 1682–1683. doi: 10.1136/bmj.281.6256.1682-a
- Njeru, J., Tomaso, H., Mertens, K., Henning, K., Wareth, G., Heller, R., et al. (2017). Serological evidence of *Francisella tularensis* in febrile patients seeking treatment at remote hospitals, northeastern Kenya, 2014–2015. *New Microbes New Infect.* 19, 62–66. doi: 10.1016/j.nmni.2017.05.015
- Ogden, N. H., Mechai, S., and Margos, G. (2013). Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. *Front. Cell Infect Microbiol.* 3:46. doi: 10.3389/fcimb.2013.00046
- Ogola, E. O., Kopp, A., Bastos, A. D. S., Slothouwer, I., Marklewitz, M., Omoga, D., et al. (2022). Jimgen tick virus in ticks from Kenya. *Viruses* 14:1041. doi: 10.3390/v14051041
- Ogutu, J. O., Piepho, H. P., Said, M. Y., Ojwang, G. O., Njino, L. W., Kifugo, S. C., et al. (2016). Extreme wildlife declines and concurrent increase in livestock numbers in Kenya: what are the causes? *PLoS One* 11:e0163249. doi: 10.1371/journal.pone.0163249
- Okal, M. N., Odhiambo, B. K., Otieno, P., Bargul, J. L., Masiga, D., Villinger, J., et al. (2020). Anaplasma and theileria pathogens in cattle of lambwe valley, kenya: a case for pro-active surveillance in the wildlife-livestock interface. *Microorganisms* 8:1830. doi: 10.3390/microorganisms8111830
- Omondi, D., Masiga, D. K., Fielding, B. C., Kariuki, E., Ajamma, Y. U., Mwamuye, M. M., et al. (2017). Molecular detection of tick-borne pathogen diversities in ticks from livestock and reptiles along the shores and adjacent islands of Lake Victoria and Lake Baringo, Kenya. *Front. Vet. Sci.* 4:73. doi: 10.3389/fvets.2017.00073
- Peter, S. G., Kariuki, H. W., Aboge, G. O., Gakuya, D. W., Maingi, N., and Mulei, C. M. (2021). Prevalence of ticks infesting dairy cattle and the pathogens they harbour in smallholder farms in Peri-Urban Areas of Nairobi, Kenya. *Vet. Med. Int.* 2021:9501648. doi: 10.1155/2021/9501648
- Qin, X. C., Shi, M., Tian, J. H., Lin, X. D., Gao, D. Y., He, J. R., et al. (2014). A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *Proc. Natl. Acad. Sci. USA.* 111, 6744–6749. doi: 10.1073/pnas.1324194111
- Ribicich, M. M., Fariña, F. A., Aronowicz, T., Ercole, M. E., Bessi, C., Winter, M., et al. (2020). A review on Trichinella infection in South America. *Vet. Parasitol.* 285:109234.
- Rochlin, I., and Toledo, A. (2020). Emerging tick-borne pathogens of public health importance: a mini-review. *J. Med. Microbiol.* 69, 781–791. doi: 10.1099/jmm.0.001206
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584
- Sameroff, S., Tokarz, R., Charles, R. A., Jain, K., Oleynik, A., Che, X., et al. (2019). Viral diversity of tick species parasitizing cattle and dogs in Trinidad and Tobago. *Sci. Rep.* 9:10421. doi: 10.1038/s41598-019-46914-1
- Shi, M., Lin, X. D., Vasilakis, N., Tian, J. H., Li, C. X., Chen, L. J., et al. (2015). Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the Flaviviridae and related viruses. *J. Virol.* 90, 659–669. doi: 10.1128/JVI.02036-15
- Smith, D. H., Timms, G. L., and Refai, M. (1979). Outbreak of botulism in Kenyan nomads. *Ann. Trop. Med. Parasitol.* 73, 145–148. doi: 10.1080/00034983.1979.11687241
- Souza, W. M., Fumagalli, M. J., Torres Carrasco, A. O., Romeiro, M. F., Modha, S., Seki, M. C., et al. (2018). Viral diversity of Rhipicephalus microplus parasitizing cattle in southern Brazil. *Sci. Rep.* 8:16315. doi: 10.1038/s41598-018-34630-1
- State of Wildlife Conservancies in Kenya Report (2016). Available online at: <https://kwakenya.com/download/state-of-wildlife-conservancies-in-kenya-report/> (accessed February 27, 2022).

- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120
- Temmam, S., Bigot, T., Chrétien, D., Gondard, M., Pérot, P., Pommelet, V., et al. (2019). Insights into the host range, genetic diversity, and geographical distribution of jingmenviruses. *mSphere* 4, e645–e619. doi: 10.1128/mSphere.00645-19
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Tóth, R., Nosek, J., Mora-Montes, H. M., Gabaldon, T., Bliss, J. M., Nosanchuk, J. D., et al. (2019). *Candida parapsilosis*: from genes to the bedside. *Clin. Microbiol. Rev.* 32:e00111-18. doi: 10.1128/CMR.00111-18
- Welde, B. T., Chumo, D. A., Reardon, M. J., Waema, D., Smith, D. H., Gibson, W. C., et al. (1989). Epidemiology of Rhodesian sleeping sickness in the Lambwe Valley, Kenya. *Ann. Trop. Med. Parasitol.* 83, 43–62. doi: 10.1080/00034983.1989.11812409
- Wikel, S. K. (2018). Ticks and tick-borne infections: complex ecology, agents, and host interactions. *Vet. Sci.* 5:60. doi: 10.3390/vetsci5020060
- Wingett, S. W., and Andrews, S. (2018). FastQ screen: a tool for multi-genome mapping and quality control. *F1000Res.* 7:1338. doi: 10.12688/f1000research.15931.2
- Woolhouse, M. E., and Gowtage-Sequeria, S. (2005). Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* 11, 1842–1847. doi: 10.3201/eid1112.050997
- Yamakawa, K., Kamiya, S., Yoshimura, K., Nakamura, S., and Ezaki, T. (1990). *Clostridium botulinum* in the soil of Kenya. *Ann. Trop. Med. Parasitol.* 84, 201–203. doi: 10.1080/00034983.1990.11812457
- Yon, L., Duff, J. P., Ågren, E. O., Erdélyi, K., Ferroglio, E., Godfroid, J., et al. (2019). Recent changes in infectious diseases in European wildlife. *J. Wildl. Dis.* 55, 3–43. doi: 10.7589/2017-07-172
- Yu, Z. M., Chen, J. T., Qin, J., Guo, J. J., Li, K., Xu, K. Y., et al. (2020). Identification and characterization of Jingmen tick virus in rodents from Xinjiang, China. *Infect. Genet. Evol.* 84:104411. doi: 10.1016/j.meegid.2020.104411
- Zhang, Y., Hu, B., Agwanda, B., Fang, Y., Wang, J., and Kuria, S. (2021). Viromes and surveys of RNA viruses in camel-derived ticks revealing transmission patterns of novel tick-borne viral pathogens in Kenya. *Emerg. Microbes Infect.* 10, 1975–1987. doi: 10.1080/22221751.2021.1986428

Conflict of Interest: MM was employed by Kenya Wildlife Services Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ergunay, Mutinda, Bourke, Justi, Caicedo-Quiroga, Kamau, Mutura, Akunda, Cook, Gakuya, Omondi, Murray, Zimmerman and Linton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Metagenomic Next-Generation Sequencing Reveals the Profile of Viral Infections in Kidney Transplant Recipients During the COVID-19 Pandemic

Xiangyong Tian^{1†}, Wenjing Duan^{2†}, Xiulei Zhang³, Xiaoqiang Wu¹, Chan Zhang¹, Zhiwei Wang¹, Guanghui Cao¹, Yue Gu⁴, Fengmin Shao⁴ and Tianzhong Yan^{1*}

¹ Department of Urology, Henan Provincial Clinical Research Center for Kidney Disease, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, China, ² Department of the Clinical Research Center, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, China, ³ Microbiology Laboratory, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, China, ⁴ Department of Nephrology, Henan Provincial People's Hospital, Henan Provincial Key Laboratory of Kidney Disease and Immunology, Henan Provincial Clinical Research Center for Kidney Disease, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, China

OPEN ACCESS

Edited by:

Binwu Ying,
West China Hospital, Sichuan
University, China

Reviewed by:

Zisis Kozlakidis,
International Agency For Research On
Cancer (IARC), France
José Eduardo Levi,
University of São Paulo, Brazil

*Correspondence:

Tianzhong Yan
ytz460@hotmail.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Infectious Diseases-Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Public Health

Received: 03 March 2022

Accepted: 20 June 2022

Published: 11 July 2022

Citation:

Tian X, Duan W, Zhang X, Wu X,
Zhang C, Wang Z, Cao G, Gu Y,
Shao F and Yan T (2022)
Metagenomic Next-Generation
Sequencing Reveals the Profile of Viral
Infections in Kidney Transplant
Recipients During the COVID-19
Pandemic.
Front. Public Health 10:888064.
doi: 10.3389/fpubh.2022.888064

Background: To study the clinical application of metagenomic next-generation sequencing (mNGS) in the detection of viral infections in kidney transplant recipients (KTRs) during the COVID-19 pandemic.

Methods: Using mNGS technology, 50 human fluid samples of KTRs were detected, including 20 bronchoalveolar lavage fluid (BALF) samples, 21 urine samples and 9 blood samples. The detected nucleic acid sequences were compared and analyzed with the existing viral nucleic acid sequences in the database, and the virus infection spectrum of KTRs was drawn.

Results: The viral nucleic acids of 15 types of viruses were detected in 96.00% (48/50) of the samples, of which 11 types of viruses were in BALF (95.00%; 19/20), and the dominant viruses were *torque teno virus* (TTV) (65.00%; 13/20), *cytomegalovirus* (CMV) (45.00%; 9/20) and *human alphaherpesvirus 1* (25.00%; 5/20). 12 viruses (95.24%, 20/21) were detected in the urine, and the dominant viruses were TTV (52.38%; 11/21), *JC polyomavirus* (52.38%; 11/21), *BK polyomavirus* (42.86%; 9/21), CMV (33.33%; 7/21) and *human betaherpesvirus 6B* (28.57%; 6/21). 7 viruses were detected in the blood (100.00%, 9/9), and the dominant virus was TTV (100.00%; 9/9). Four rare viruses were detected in BALF and urine, including *WU polyomavirus*, *primate bocaparvovirus 1*, *simian virus 12*, and *volepox virus*. Further analysis showed that TTV infection with high reads indicated a higher risk of acute rejection ($P < 0.05$).

Conclusions: mNGS detection reveals the rich virus spectrum of infected KTRs, and improves the detection rate of rare viruses. TTV may be a new biomarker for predicting rejection.

Keywords: metagenomic next-generation sequencing, kidney transplantation, infection, virus, *torque teno virus*

INTRODUCTION

COVID-19 has rapidly escalated into a global pandemic, with more than 276.4 million cumulative cases and five million deaths worldwide as of December 2021 (1–3). The threats of the constantly mutating coronavirus continue to emerge. Variant strains with higher morbidity, stronger transmissibility, broader epidemic potential and higher mortality have been identified with the help of the development of genetic sequencing technology over the past 2 years (4–6). Although the worldwide epidemic of COVID-19 has imposed great challenges and heavy burdens on global public health, the work of transplantation clinicians has never halted (7, 8). That is because organ transplantation offers the greatest hope of survival and functional recovery for patients with irreversible end-stage organ failure. As far as the end-stage kidney disease patients are concerned, renal transplantation is the optimal treatment. And with the widespread application of potent immunosuppressive agents and the improvement of organ preservation techniques, the one-year survival rate of kidney transplantation has increased to more than 90% (9). However, with the long-term use of large amounts of immunosuppressants, the immune function of kidney transplant recipients (KTRs) is obviously impaired, increasing the chance of postoperative infection. Therefore, KTRs represent a population with an increased risk for COVID-19 and other pathogens infection, especially occult viral infections such as BK polyomavirus and cytomegalovirus, in which the outcomes of the infections are worse and in severe cases the infections may lead to graft loss and even death (10, 11).

KTRs can be simultaneously infected by multiple viruses. And their symptoms induced by infections sometimes are difficult to differentiate from rejections and drug application. Viral infections can't be identified by routine cultures. While detection of the viral genome by polymerase chain reaction (PCR) has several limitations, including difficulties in identifying multiple infections and low-throughput: only one species can be detected at a time, which causes the challenges for clinical treatment strategies and prognosis. Therefore, rapid determination of the presence of viral infections and effective improvement of accuracy and detection rates are urgent needs in the field of transplant infections.

Metagenome next-generation sequencing (mNGS) is an emerging method for pathogen identification. Since its successful use in the detection of new pathogenic infections in 2008, mNGS has gradually realized the transition from laboratory to clinical applications (12, 13). This culture-independent technique allows for rapid and accurate sequence detection of pathogenic microorganisms (including bacteria, fungi, viruses, and parasites) without bias by directly targeting nucleic acids in clinical samples. mNGS showed promising values in the rapid diagnosis of clinical infections, and can be applied in transplantation (14).

During the COVID-19 pandemic, factors of occult viral infections in KTRs are likely to be influenced by the COVID-19 epidemic in different ways and the management is more complicated, especially by telemedicine. However, few studies on this topic are currently available. Given this global background, in order to have a more definitive and

comprehensive understanding of the viral pathogens following renal transplantation, we used mNGS in this study to identify virus spectrum of KTRs with symptoms from infection during this special period, hoping to provide a basis for improving clinical management strategies.

MATERIALS AND METHODS

Study Design and Population

This study was performed in Henan Provincial People's Hospital, a tertiary teaching hospital in Zhengzhou, China. KTRs hospitalized with clinical symptoms and relevant signs of infection or unexplained fever, from May 2020 to May 2021, were enrolled. After recording demographic and clinical details, multifarious body fluid samples depending upon the site of infection at different stages were collected from enrolled KTRs, which were tested for viral infection profile by mNGS. Additional data on treatment, response to treatment, outcomes, and any relevant follow-up data were also collected. The results were reviewed by at least two clinicians to discriminate infection from colonization and contamination.

This study was approved by the Ethics Committee of the Henan Provincial People's Hospital [(2021)213], and all data were anonymised prior to analysis. The study was conducted in compliance with the Declaration of Helsinki.

Sample Collection

The corresponding specimens were collected from each KTR enrolled in the study according to their symptoms. The exclusion criteria were consistent for all patients, namely: (1) patients with a previous history of multiple organ transplantation; (2) patients with samples sent for testing who were clearly considered contaminated; (3) kidney transplant patients with positive pregnancy tests. KTRs, who presented the most common pulmonary infection manifestations, including fever, cough, expectoration, shortness of breath, chest tightness, dyspnoea etc., were performed with chest CT for confirmation, and then the bronchoalveolar lavage fluids (BALF) samples collected during fiberoptic bronchoscopy were sent to laboratory for mNGS. Urinary tract infection is the second most common infection among KTRs, mainly manifests as lower urinary tract symptoms, including urinary frequency, urgency, pain, and a burning sensation during urination, with or without fever. KTRs with urinary tract infection were confirmed by routine urine test and quantitative urine culture, and then the clean midstream urine samples were sent to laboratory for mNGS. Peripheral blood samples were collected for mNGS from KTRs with fever of unknown origin or both of the above-mentioned infections. A 5-mL sample was taken per BALF or peripheral blood specimen. A 50-mL sample was taken per clean midstream urine specimen. All samples were collected according to standard operating procedures in accordance with the rules of aseptic technique, and were transported to the sequencing laboratory by cold chain in time.

Metagenomic Next-Generation Sequencing and Data Analysis

Nucleic Acid Extraction: TIANamp Micro DNA Kit (DP316, TIANGEN BIOTECH, Beijing, China) was employed for the process of DNA extraction. Nucleic acid extraction were performed according to the manufacturer's operational guidebooks. DNA extraction was conducted for all samples.

Library Construction and Sequencing: The total DNA or cDNA was subjected to library construction through an end-repair method. A specific tag sequence was introduced at the end of each library. The library concentration was determined by Qubit 4.0 nucleic acid fluorescence quantitative analyzer (Q33226, Thermo Fisher, USA) and Qubit® dsDNA HS Assay Kit (Q32854, Thermo Fisher, USA). Agilent 2100 Bioanalyzer (G2939BA, Agilent, USA) was used to evaluate the DNA concentration and fragment size in the library to be sequenced for the quality control of the DNA libraries. DNA nanospheres were prepared by one-step DNB Kit (1000025076, Huada Zhizao, China). Quality qualified libraries were sequenced by MGISEQ-200 platform.

Bioinformatic Analysis: After removing low-quality (< 35 bp) and low-complexity reads according to prinseq (version 0.20.4), and computational subtracting the human host sequences mapped to the human reference genome (hg38) from the sequencing data by Burrows-Wheeler Alignment (0.7.10-r789), high-quality sequences were generated (15). The remaining non-host sequences were matched and classified with dedicated viral databases which were downloaded from the National Center for Biotechnology Information (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>) and other public databases. So far, more than 4,000 viral genomes were contained in the integrated classification reference databases. The mapped data were processed for advanced data analysis. Lists of suspected pathogenic viruses were produced, which included the numbers of strictly mapped reads, coverage rate, and depth (16). The clinical diagnosis was determined by considering all the clinical manifestations, possible pathogens identified by mNGS and other laboratory tests together.

Statistical Analysis

Continuous data were expressed as the mean \pm standard deviation (SD). Counting data were expressed as the number of cases with percentage (%). The Chi-square test was conducted for comparing the rate of low concentration group and high concentration group, rejection group and non-rejection group. Data analyses were performed using Statistical Package for the Social Sciences (SPSS) version 24.0 statistical software (IBM SPSS, Chicago, IL, USA) (17). All *P*-values were two-sided, and statistical significance was defined as *P* < 0.05.

RESULTS

Characteristics of the Patients

A total of 39 KTRs were investigated in the present study. A total of 50 samples were collected from these KTRs at different infection phases, including 20 BALF samples, 21 urine samples and 9 blood samples. Abstractions of patients' demographic

TABLE 1 | Baseline characteristics of participants.

Characteristics	Cases (n = 39)
Gender	
Male	22 (56.41%)
Female	17 (43.59%)
Age (years)	39.49 \pm 10.42
Pre-transplant dialysis durations (months)	25.51 \pm 32.92
Comorbidity	
Hypertension	32 (82.05%)
Hypertension and Diabetes	1 (2.56%)
Immunosuppressant	
Tac+MMF+Pred	32 (82.05%)
Tac+MPS+Pred	7 (17.95%)

Data were provided as n / percentage (%) or mean \pm standard deviation. Tac, tacrolimus; MMF, mycophenolate mofetil; MPS, mycophenolate sodium; Pred, prednisone.

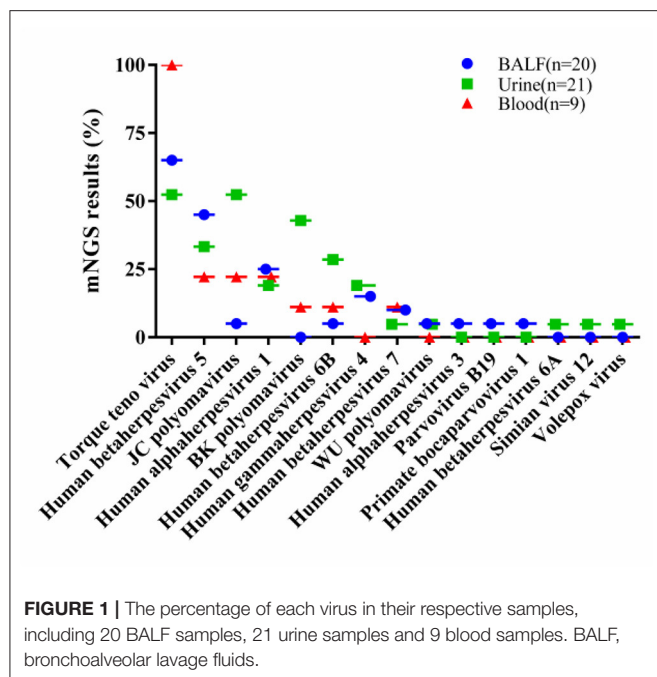
TABLE 2 | Baseline characteristics of patients.

Characteristics	Samples (n = 50)
Body temperature	
Normal	21 (42.00%)
$\geq 37.3^{\circ}\text{C}$	29 (58.00%)
Serum creatinine (umol/L)	148.26 \pm 84.22
Lymphocyte count ($\times 10^9/\text{L}$)	0.80 \pm 0.57
Lymphocyte ratio (%)	14.64 \pm 12.09
Acute rejection	
Yes	17 (34.00%)
No	33 (66.00%)
Tacrolimus blood concentration	
$\geq 8 \text{ ng/ml}$	13 (26.00%)
< 8 ng/ml	37 (74.00%)

Data were provided as n / percentage (%) or mean \pm standard deviation. Normal lymphocyte count: $(1.1\text{--}3.2) \times 10^9/\text{L}$; Normal lymphocyte ratio: 20–50%.

and clinical information were collected, including age, sex, dialysis durations, comorbidity, infection signs, application of immunosuppression, laboratory examinations, as summarized in **Tables 1, 2**. The median age of all KTRs was 39.49 (range 20–58) years. The clinical signs on physical examination of the KTRs were heterogeneous and non-specific. Mean values of lymphocyte count and lymphocyte proportion were both below the normal range.

The same renal transplant team performed all the surgical procedures and postoperative management together. Routine standard of care and post-transplant medication including immunosuppressive drug therapy was administered in accordance to center standard. The standard immunosuppressive protocol included induction with anti-thymocyte globulin (ATG), followed by maintenance immunosuppressive regimen consisting of tacrolimus, mycophenolate mofetil (MMF) / mycophenolate sodium (MPS) and prednisone. The dosage of tacrolimus was weight-based (0.05 mg/kg twice daily) started



at the time of transplantation and then adjusted according to close monitoring to maintain tacrolimus blood concentrations within the therapeutic range (6–8 ng/ml) to ensure efficacy and safety.

Diagnostic Performance of mNGS in Three Sample Types

Viral nucleic acids were detectable in 48 of 50 samples (96.00%), for a total of 15 virus types. Of these, 19 (95.00%) of the 20 BALF samples were positive for viral nucleic acids, for a total of 11 virus types, of which the top five were torque teno virus (TTV) in 13 cases (65.00%), human betaherpesvirus 5 (Cytomegalovirus, CMV) in 9 cases (45.00%), human alphaherpesvirus 1 in 5 cases (25.00%), human gammaherpesvirus 4 (Epstein-Barr virus, EBV) in 3 cases (15.00%), and human betaherpesvirus 7 in 2 cases (10.00%). Mixed viral infections were observed in 11 cases (55.00%).

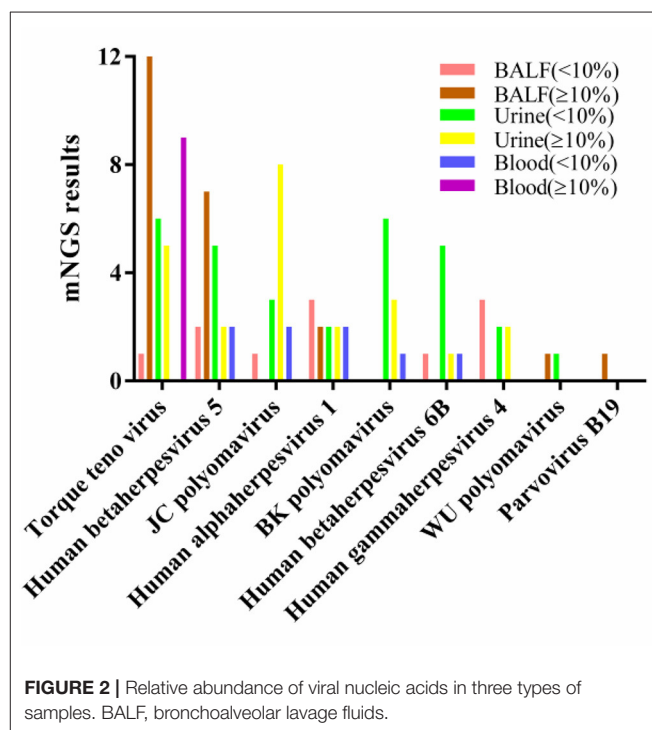
Among the 21 urine samples, viral nucleic acids were detected in 20 cases (95.24%) with 12 virus types, of which the top five were TTV in 11 cases (52.38%), JC polyomavirus (JCV) in 11 cases (52.38%), BK polyomavirus (BKV) in 9 cases (42.86%), CMV in 7 cases (33.33%), and human betaherpesvirus 6B in 6 cases (28.57%). Mixed viral infections were observed in 18 cases (85.71%).

Viral nucleic acids were detected in all 9 samples of peripheral blood (100.00%) with 7 virus types, namely, 9 cases of TTV (100.00%), 2 cases each of human alphaherpesvirus 1, CMV and JCV (22.22%), 1 case each of BKV, human betaherpesvirus 6B and human betaherpesvirus 7 (11.11%). Four cases were mixed viral infections (44.44%) (Figure 1).

TABLE 3 | Relationship between tacrolimus concentration and viral infection.

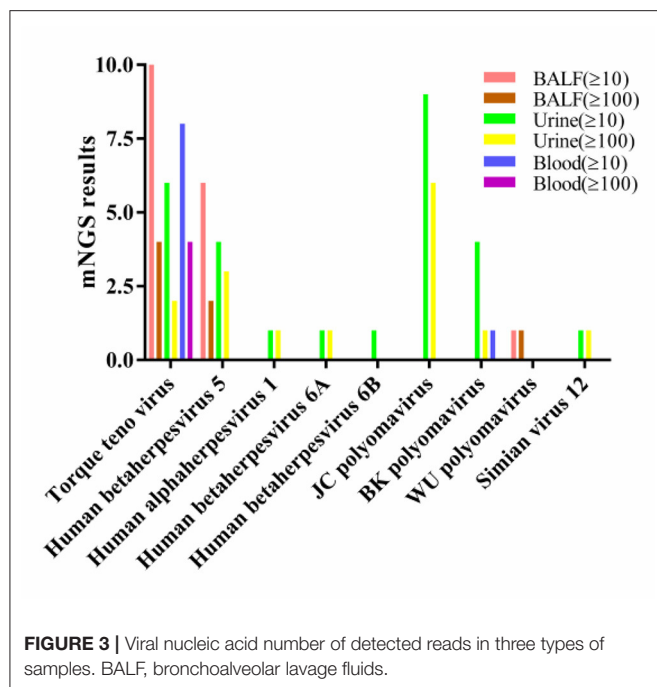
	LCG	HCG	χ^2 -value	P-value
All cases	37	13		
Reads ≥ 10	25 (67.57%)	11 (84.62%)	0.670	0.413
Reads ≥ 100	13 (35.14%)	7 (53.85%)	1.403	0.236
TTV positive cases	22 (59.46%)	12 (92.31%)	3.380	0.066
Reads ≥ 10	9 (24.32%)	5 (38.46%)	0.954	0.329
Reads ≥ 100	6 (16.22%)	4 (30.77%)	0.526	0.468

Data were provided as n / percentage (%). LCG, low concentration group; HCG, high concentration group; TTV, torque teno virus.



Relative Abundance of Viral Nucleic Acids in Three Types of Samples

We used 10% as the threshold to screen out virus types with high relative abundance, and the results showed that there were 5 virus types in the alveolar lavage fluid, namely, TTV ($n = 12$), CMV ($n = 7$), human alphaherpesvirus 1 ($n = 2$), WU polyomavirus ($n = 1$), and parvovirus B19 ($n = 1$). Urine had 8 virus types, namely, JCV ($n = 8$), TTV ($n = 5$), BKV ($n = 3$), CMV ($n = 2$), human alphaherpesvirus 1 ($n = 2$), EBV ($n = 2$), human betaherpesvirus 6A ($n = 1$) and human betaherpesvirus 6B ($n = 1$). There was only TTV in peripheral blood ($n = 9$) (Table 3). The only virus whose relative abundance exceeded 10% in all three samples was TTV (Figure 2).



Reads of Viral Nucleic Acids Detected in Three Types of Samples

In BALF, 14 cases (70.00%) had a total viral nucleic acid reads ≥ 10 and 6 cases (30.00%) had nucleic acid reads ≥ 100 . The viruses with reads ≥ 100 were TTV ($n = 4$), CMV ($n = 2$), and WU polyomavirus ($n = 1$).

In urine, 14 cases (66.67%) had a total viral nucleic acid reads ≥ 10 , 10 cases (47.62%) had nucleic acid reads ≥ 100 . The viruses with reads ≥ 100 were JCV ($n = 6$), CMV ($n = 3$), TTV ($n = 2$), BKV ($n = 1$), human alphaherpesvirus 1 ($n = 1$), human betaherpesvirus type 6A ($n = 1$), and simian virus 12 ($n = 1$).

In peripheral blood, 8 cases (88.89%) had a total viral nucleic acid reads ≥ 10 and 4 cases (44.44%) had nucleic acid reads ≥ 100 , which were TTV ($n = 4$).

The only virus with a viral nucleic acid reads of more than 100 in all three samples was TTV (Figure 3).

mNGS Improves the Detection of Rare Viruses

mNGS significantly improved the detection rate of rare or uncommon viruses. Four rare viruses were detected in this study: WU polyomavirus, simian virus 12, volepox virus and primate bocavirus type 1. Among them, simian virus 12 and volepox virus were detected once in urine samples. Primate baculovirus type 1 was detected once in alveolar lavage fluid. WU polyomavirus was detected in one KTR's alveolar lavage sample and another KTR's urine sample, respectively.

Nucleic acid reads were less than 10 for all viruses except for WU polyomavirus in one BALF sample and simian virus 12 in one urine sample, which had nucleic acid reads greater than 100.

TABLE 4 | Relationship between viral infection and rejection after renal transplantation.

	Rejection group	Non-rejection group	χ^2 -value	P-value
All cases	17	33		
Reads ≥ 10	16 (94.12%)	20 (60.61%)	4.698	0.030
Reads ≥ 100	10 (58.82%)	10 (30.30%)	3.803	0.051
TTV positive cases	14 (82.35%)	20 (60.61%)	1.542	0.214
Reads ≥ 10	12 (70.59%)	11 (33.33%)	6.269	0.012
Reads ≥ 100	8 (47.06%)	2 (6.06%)	9.364	0.002

Data were provided as n / percentage (%). TTV, torque teno virus.

Relationship Between Tacrolimus Concentration and Viral Infection

We divided all samples into low concentration group (LCG, $n = 37$) and high concentration group (HCG, $n = 13$) according to whether the tacrolimus concentration was ≥ 8 ng/ml, and found that there were 25 (67.57%) and 11 (84.62%) cases with total viral nucleic acid reads ≥ 10 in the LCG and the HCG, respectively. There were 13 (35.14%) and 7 (53.85%) cases with viral nucleic acid reads ≥ 100 in the two groups, respectively. No statistical differences were found between the two groups ($P > 0.05$).

We focused on the infection of the TTV in the LCG and HCG, and found that the nucleic acid of the virus was detected in 22 (59.46%) and 12 (92.31%) cases, respectively, with number of detected reads ≥ 10 in 9 (24.32%) and 5 (38.46%) cases, and ≥ 100 in 6 (16.22%) and 4 (30.77%) cases, respectively, without statistical difference between the two groups ($P > 0.05$) (Table 3).

Relationship Between Viral Infection and Rejection After Renal Transplantation

We compared the viral infections in the rejection group ($n = 17$) and the non-rejection group ($n = 33$) and found that the total nucleic acid reads ≥ 10 was detected in 16 (94.12%) and 20 (60.61%) cases in the two groups, respectively, with statistically significant difference between the two groups ($\chi^2 = 4.698$, $P = 0.030$). The patients number of detected reads ≥ 100 was 10 cases (58.82%) and 10 cases (30.30%) in the two groups, respectively, and there was no statistically significant difference between the two groups ($P > 0.05$).

We compared the TTV infection in the rejection and non-rejection groups and found that TTV nucleic acid reads were detected in 14 (82.35%) and 20 (60.61%) cases, respectively, with no statistical difference between the two groups ($P > 0.05$). Reads ≥ 10 were detected in 12 (70.59%) and 11 (33.33%) cases, respectively, with statistical difference in comparison ($\chi^2 = 6.269$, $P = 0.012$). There were 8 (47.06%) and 2 (6.06%) cases with reads ≥ 100 , respectively, with a statistical difference in comparison ($\chi^2 = 9.364$, $P = 0.002$) (Table 4).

Treatments and Outcomes

All patients were given empirical anti-infective therapy upon admission, and the dose of immunosuppressive drugs was reduced after definitive infection, i.e., MMF or EC-MPS

was reduced or discontinued directly, and tacrolimus was maintained in small doses or reduced appropriately. Ganciclovir or penciclovir was also administered. All patients were cured with anti-infective treatment and combination therapy.

Among those with pulmonary infections, one case resulted in eventual failure of the transplanted kidney due to infection and rejection, and the function of the transplanted kidney was affected in two patients due to infection and rejection. Among those with urinary tract infections, 8 cases had recurrent urinary tract infections the function of the transplanted kidney was affected in 1 case.

DISCUSSION

The current world is still going through a rough patch for the outbreak of COVID-19. It causes public health concerns especially in the field of transplantation. KTRs undergoing post-transplant immunosuppressive therapy are at the risk of infection (18). Although many transplant practitioners have studied one or more pathogenic infections in KTRs, studies revealing their viral profile have not been seen. Fortunately, mNGS technology, with the characteristics of fast detection speed, high sensitivity and wide coverage, is able to effectively compensate for the deficiency of traditional culture and PCR, particularly offer a very significant practical advantage for KTRs (19–21). Since traditional culture cannot detect viruses and PCR is low-throughput, the advantage of mNGS for virus detection is highlighted. It directly extracts all viral nucleic acid fragments in the specimen, compares the reference sequences in the specific database with the specimen sequences, analyzes them by intelligent algorithms to obtain viruses in the specimen that have the same sequences as various reference pathogens, avoiding the missed detection of difficult-to-identify viruses (22). The findings of this research shed new light onto the viral infection profile in KTRs during the COVID-19 pandemic.

mNGS has helped researchers and clinicians solve many intractable diseases since its initial clinical application. As tested by Miao et al., the sensitivity of mNGS (50.7%) was higher than traditional method (35.2%). They considered that mNGS could yield a higher sensitivity for pathogen identification and be less affected by prior antibiotic exposure, thereby emerging as a promising technology for detecting infectious diseases (20). Jerome et al. (23) performed mNGS among 40 febrile returning travelers for the pathogenic diagnosis, avoiding the missed detection by traditional methods, which indicated that mNGS had the potential to be an all-in-one rapid diagnostic testing. Palacios et al. (12) reported that 3 patients who received the same donor organ died of high fever within 4–6 weeks, but the results of traditional culture were negative. With the help of mNGS, the patients were clearly diagnosed as arenavirus infection, which revealed the mystery of pathogenic microorganisms. Gazzani reported a case of fatal disseminated cowpox virus infection in an adolescent renal transplant recipient, which was greatly assisted by mNGS (24).

In our study, viral nucleic acids were detected in 96.00% of the samples, involving 15 virus types. Urine positivity

rate was 95.24% with 12 virus types, and the predominant viruses were TTV, JCV, BKV, CMV, and human betaherpesvirus 6B. The virus-positive rate in urine samples was 95.24%, and a total of 12 viruses were detected, among which the predominant viruses were TTV, JCV, BKV, CMV, and human betaherpesvirus 6B. 95.00% BALF was positive for a total of 11 virus types, with the predominant viruses being TTV, CMV, human alphaherpesvirus 1, EBV, and human betaherpesvirus 7. Peripheral blood was 100.00% positive for 7 virus types, with the predominant virus being TTV. All peripheral blood samples were virus-positive, and TTV was the predominant one among the 7 viruses detected. It is estimated that 5–8% of KTRs are infected with BKV in the first 3 years after transplant, which can lead to nephropathy, impaired kidney function and graft loss (25–28). In this study, BKV fragments were detected in only 9 urine samples and 1 blood sample, and only 1 urine sample had a number of detected reads more than 100. Accordingly, it is hypothesized that the relatively low positive rate of BKV and viral load in infected KTRs is probably due to the reduction of immunosuppression. Herpesvirus infections, especially CMV infection, have also been well studied (27, 29, 30). The number of studies on JCV infection is also increasing (31, 32). We also detected TTV (33–35) and parvovirus B19 (36–38), which are still relatively rare in the field of transplantation. Four rare virus types, WU polyomavirus, primate bocavirus type 1, simian virus 12, and vole pox virus, have also been detected and have not been reported in KTRs before, so we cannot be sure if they caused the disease, but these patients eventually recovered with treatment. Therefore, mNGS significantly improved the detection rate of common and rare viruses in KTRs, and the application of mNGS in the renal transplant is worth promoting.

Next, we studied the nucleic acid read length and relative abundance of the viruses in each sample, and found that the top virus in BALF and blood was TTV, while in urine, the top virus was JCV, followed by TTV. Notably, only TTV was positive in all three types of samples with nucleic acid reads ≥ 100 and relative abundance $\geq 10\%$, which was completely unexpected. The detection rate of the TTV was 66.00% in all samples, and 100.00% in blood samples, which is consistent with the findings of previous studies (33, 34, 39). In addition to the detection in these types of samples, detection of the TTV in cerebrospinal fluid has also been reported (40), suggesting that the TTV may be widely present in human body fluids and may be particularly evident in immunosuppressed population. An Australian cross-sectional study noted that TTV was detectable in the plasma of 93% of KTRs, suggesting that TTV may be a novel marker for immunosuppressive therapy for KTRs (39). Another study has also identified TTV as a predictor of the level of immunosuppression and infection after solid organ transplantation (41, 42). We compared TTV infection in the LCG and HCG. Although the difference did not achieve statistical significance, we believe that this was due to the very small sample size. The TTV positivity rate was lower in the LCG than in the HCG (59.46 vs. 92.31%) and consistent results were obtained in cases with nucleic acid reads ≥ 10 and ≥ 100 . It is suggested that the higher the tacrolimus concentration is, the more susceptible

the KTRs are to be infected by the virus and higher the viral load is.

Further, we compared viral infections in the rejection and non-rejection groups, and the rates of viral infections with nucleic acid reads ≥ 10 were 94.12 and 60.61% in the two groups, respectively, which were statistically different ($P < 0.05$), suggesting that the rejection group is more likely to lead to viral infections. Specifically for TTV, we found that the infection rate was higher in the rejection group than in the non-rejection group (82.35 vs. 60.61%). Although the difference was not statistically significant in the primary analysis, the trend was obvious. Moreover, the stratified analysis of viral nucleic acid read lengths showed that the infected cases was significantly more in the rejection group than in the non-rejection group (70.59 vs. 33.33% for acid reads ≥ 10 , and 47.06 vs. 6.06% for reads ≥ 100), indicating a higher proportion of KTRs with TTV load in the rejection group. These results suggested that TTV may be a potential biomarker for predicting renal transplant rejection, which was also tentatively validated by Strassl et al. (43).

Despite the great value of mNGS in infectious diseases, especially rare infectious diseases, there are still many practical problems in its clinical application. More than 99% of the reads generated by sample sequencing are from human hosts (44), while microorganisms represent only a small percentage. And sequencing all nucleic acids reduces the sensitivity of pathogen identification, making it difficult to distinguish between colonizing, background and pathogenic bacteria among the various species detected (45). But it is possible to deplete host nucleic acids by certain methods (46, 47), and reducing the human-derived nucleic acid sequence proportion can increase the amount of microbial data and improve sensitivity to some extent. In any case, the determination of mNGS results requires a combination of nucleic acid fragment counts, clinical presentation, other laboratory results and background microorganisms.

In conclusion, we revealed the viral profile of KTRs by mNGS technology. Certain viruses infection such as TTV may be a reflection of the degree of immunity in KTRs, as well as a

potential biomarker for predicting rejection. Therefore, mNGS is recommend as a routine testing for KTRs to make real and lasting benefits for health and healthcare.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ebi.ac.uk/metagenomics~ERP136978>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Henan Provincial People's Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XT and WD drafted the manuscript. XT and XZ carried out the statistical analysis. XW, CZ, ZW, GC, and YG interpreted data. XT, WD, FS, and TY participated in the design of the study and coordination. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Project of Science and Technology of Henan Province (Grant No. 202102310438), the 23456 Talent Project Foundation of Henan Provincial People's Hospital (Grant No. ZC23456127), Joint construction project of Henan Medical Science and technology research plan (Grant No. LHGJ20210042), and Foundation of Henan Educational Committee (Grant No. 22A320012).

REFERENCES

- World Health Organization. WHO coronavirus disease (COVID-19) dashboard. Available online at: <https://covid19.who.int>.
- Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. (2020) 382:1708–20. doi: 10.1056/NEJMoa2002032
- White NJ, Strub-Wourgaft N, Faiz A, Guerin PJ. WHO COVID-19 therapeutic guidelines – Authors' reply. *Lancet*. (2021) 398:118. doi: 10.1016/S0140-6736(21)01327-1
- Callaway E. Delta coronavirus variant: scientists brace for impact. *Nature*. (2021) 595:17–8. doi: 10.1038/d41586-021-01696-3
- Charre C, Ginevra C, Sabatier M, Regue H, Destras G, Brun S, et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol*. (2020) 6:veaa075. doi: 10.1093/ve/veaa075
- Yuan M, Huang D, Lee CD, Wu NC, Jackson AM, Zhu X, et al. Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science*. (2021) 373:818–23. doi: 10.1126/science.abb1139
- Reddy S, Dumbill R, Akhtar MZ, Udupa V, Storey BM, Gilbert J, et al. Transplant programmes in areas with high SARS-CoV-2 transmission. *Lancet*. (2020) 396:1395–96. doi: 10.1016/S0140-6736(20)32209-1
- Pan L, Zeng J, Yang H. Challenges and countermeasures for organ donation during the SARS-CoV-2 epidemic: the experience of Sichuan Provincial People's Hospital. *Intensive Care Med*. (2020) 46:844–45. doi: 10.1007/s00134-020-05978-8
- Hart A, Smith JM, Skeans MA, Gustafson SK, Stewart DE, Cherikh WS, et al. OPTN/SRTR 2015 annual data report: kidney. *Am J Transplant*. (2017) 17:21–116. doi: 10.1111/ajt.14124
- Hirsch HH, Knowles W, Dickenmann M, Passweg J, Klimkait T, Mihatsch MJ, et al. Prospective study of polyomavirus type BK replication and

- nephropathy in renal-transplant recipients. *N Engl J Med.* (2002) 347:488–96. doi: 10.1056/NEJMoa020439
11. Reischig T, Kacer M, Hes O, Machova J, Nemcova J, Lysak D, et al. Cytomegalovirus prevention strategies and the risk of BK polyomavirus viremia and nephropathy. *Am J Transplant.* (2019) 19:2457–67. doi: 10.1111/ajt.15507
 12. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med.* (2008) 358:991–8. doi: 10.1056/NEJMoa073785
 13. Perlejewski K, Popiel M, Laskus T, Nakamura S, Motooka D, Stokowy T, et al. Next-generation sequencing (NGS) in the identification of encephalitis-causing viruses: Unexpected detection of human herpesvirus 1 while searching for RNA pathogens. *J Virol Methods.* (2015) 226:1–6. doi: 10.1016/j.jviromet.2015.09.010
 14. Goldberg B, Sichtig H, Geyer C, Ledebner N, Weinstock GM. Making the leap from research laboratory to clinic: challenges and opportunities for next-generation sequencing in infectious disease diagnostics. *MBio.* (2015) 6:e01888–15. doi: 10.1128/mBio.01888-15
 15. Abugessaisa I, Noguchi S, Hasegawa A, Harshbarger J, Kondo A, Lizio M, et al. FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Sci Data.* (2017) 4:170107. doi: 10.1038/sdata.2017.107
 16. Li N, Cai Q, Miao Q, Song Z, Fang Y, Hu B. High-throughput metagenomics for identification of pathogens in the clinical settings. *Small Methods.* (2020) 5:2000792. doi: 10.1002/smt.202000792
 17. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA.* (1998) 279:281–6. doi: 10.1001/jama.279.4.281
 18. Sethi S, Peng A, Najjar R, Vo A, Jordan SC, Huang E. Infectious complications in tocilizumab-treated kidney transplant recipients. *Transplantation.* (2021). doi: 10.1097/TP.0000000000003512
 19. Barlow G, Nathwani D, Williams E, Ogston S, Winter J, Jones M, et al. Reducing door-to-antibiotic time in community-acquired pneumonia: controlled before-and-after evaluation and cost-effectiveness analysis. *Thorax.* (2007) 62:67–74. doi: 10.1136/thx.2005.056689
 20. Miao Q, Ma Y, Wang Q, Pan J, Zhang Y, Jin W, et al. Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clin Infect Dis.* (2018) 67:S231–40. doi: 10.1093/cid/ciy693
 21. Parize P, Muth E, Richaud C, Gratigny M, Pilimis B, Lamamy A, et al. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. *Clin Microbiol Infect.* (2017) 23:574.e1–e6. doi: 10.1016/j.cmi.2017.02.006
 22. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* (2008) 40:1413–15. doi: 10.1038/ng.259
 23. Jerome H, Taylor C, Sreenu VB, Klymenko T, Filipe ADS, Jackson C, et al. Metagenomic next-generation sequencing aids the diagnosis of viral infections in febrile returning travellers. *J Infect.* (2019) 79:383–8. doi: 10.1016/j.jinf.2019.08.003
 24. Gazzani P, Gach JE, Colmenero I, Martin J, Morton H, Brown K, et al. Fatal disseminated cowpox virus infection in an adolescent renal transplant recipient. *Pediatr Nephrol.* (2017) 32:533–6. doi: 10.1007/s00467-016-3534-y
 25. Dadhania D, Snopkowski C, Ding R, Muthukumar T, Chang C, Aull M, et al. Epidemiology of BK virus in renal allograft recipients: independent risk factors for BK virus replication. *Transplantation.* (2008) 86:521–8. doi: 10.1097/TP.0b013e31817c6447
 26. Zanutto E, Allesina A, Barreca A, Sidoti F, Gallo E, Bottino P, et al. Renal allograft biopsies with polyomavirus BK nephropathy: turin transplant center, 2015–19. *Viruses.* (2020) 12:1047. doi: 10.3390/v12091047
 27. Srivastava A, Bodnar J, Osman F, Jorgenson MR, Astor BC, Mandelbrot DA, et al. Serum albumin level before kidney transplant predicts post-transplant BK and possibly cytomegalovirus infection kidney. *Int Rep.* (2020) 5:2228–37. doi: 10.1016/j.ekir.2020.09.012
 28. Nicleleit V, Singh HK, Dadhania D, Cornea V, El-Husseini A, Castellanos A, et al. The 2018 Banff Working Group classification of definitive polyomavirus nephropathy: a multicenter validation study in the modern era. *Am J Transplant.* (2021) 21:669–80. doi: 10.1111/ajt.16189
 29. Hellemans R, Abramowicz D. Cytomegalovirus after kidney transplantation in 2020: moving towards personalized prevention. *Nephrol Dial Transplant.* (2020) 12:1–7. doi: 10.1093/ndt/gfaa249
 30. Pellett Madan R, Hand J. AST Infectious Diseases Community of Practice. Human herpesvirus 6, 7, and 8 in solid organ transplantation: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. *Clin Transplant.* (2019) 33:e13518. doi: 10.1111/ctr.13518
 31. Wiegley N, Walavalkar V, Auja H, Chen LX, Huang Y, Lee BK, et al. Clinicopathologic Characteristics of JC Virus Nephropathy in Kidney Transplant Recipients. *Transplantation.* (2021) 105:1069–76. doi: 10.1097/TP.0000000000003363
 32. Lindner JM, Cornacchione V, Sathe A, Be C, Srinivas H, Riquet E, et al. Human Memory B Cells Harbor Diverse Cross-Neutralizing Antibodies against BK and JC Polyomaviruses. *Immunity.* (2019) 50:668–76.e5. doi: 10.1016/j.immuni.2019.02.003
 33. Doberer K, Schiemann M, Strassl R, Haupenthal F, Dermuth F, Görzer I, et al. Torque teno virus for risk stratification of graft rejection and infection in kidney transplant recipients-A prospective observational trial. *Am J Transplant.* (2020) 20:2081–90. doi: 10.1111/ajt.15810
 34. Uhl P, Heilos A, Bond G, Meyer E, Böhm M, Puchhammer-Stöckl E, et al. Torque teno viral load reflects immunosuppression in paediatric kidney-transplanted patients-a pilot study. *Pediatr Nephrol.* (2021) 36:153–62. doi: 10.1007/s00467-020-04606-3
 35. Focosi D, Maggi F. Torque teno virus monitoring in transplantation: the quest for standardization. *Am J Transplant.* (2019) 19:1599–601. doi: 10.1111/ajt.15194
 36. Egbuna O, Zand MS, Arbin A, Menegus M, Taylor J. A cluster of parvovirus B19 infections in renal transplant recipients: a prospective case series and review of the literature. *Am J Transplant.* (2006) 6:225–31. doi: 10.1111/j.1600-6143.2005.01139.x
 37. Eid AJ, Posfay-Barbe KM. AST infectious diseases community of practice. Parvovirus B19 in solid organ transplant recipients. *Am J Transplant.* (2009) 9:S147–50. doi: 10.1111/j.1600-6143.2009.02905.x
 38. Bentata Y. Parvovirus B19 in kidney transplantation: key points and essential pitfalls to know. *Infect Dis.* (2021) 53:404–8. doi: 10.1080/23744235.2021.1893379
 39. Davis JS, Chu G, Pathinayake P, Jones D, Giffard P, Macera L, et al. Seroprevalence of Torque Teno Virus in hemodialysis and renal transplant patients in Australia: a cross-sectional study. *Transpl Infect Dis.* (2020) 22:e13400. doi: 10.1111/tid.13400
 40. Erdem G, Kaptan I, Sharma H, Kumar A, Aylward SC, Kapoor A, et al. Cerebrospinal fluid analysis for viruses by metagenomic next-generation sequencing in pediatric encephalitis: not yet ready for prime time? *J Child Neurol.* (2021) 36:350–6. doi: 10.1177/0883073820972232
 41. Elwasif SM, Denewar AA, Khreba N, Sheashaa H. Torque teno virus polymerase chain reaction titer: a promising immunometry. *Exp Clin Transplant.* (2021). doi: 10.6002/ect.2020.0303. [Epub ahead of print].
 42. Kotton CN. Torque teno virus: predictor of infection after solid organ transplant? *J Infect Dis.* (2018) 218:1185–7. doi: 10.1093/infdis/jiy384
 43. Strassl R, Doberer K, Rasoul-Rockenschaub S, Herkner H, Görzer I, Kläger JP, et al. Torque teno virus for risk stratification of acute biopsy-proven alloreactivity in kidney transplant recipients. *J Infect Dis.* (2019) 219:1934–9. doi: 10.1093/infdis/jiz039
 44. Kline KA, Lewis AL. Gram-positive uropathogens, polymicrobial urinary tract infection, and the emerging microbiota of the urinary tract. *Microbiol Spectr.* (2016) 4:4. doi: 10.1128/microbiolspec.UTI-0012-2012
 45. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. *Annu Rev Pathol.* (2019) 14:319–38. doi: 10.1146/annurev-pathmechdis-012418-012751
 46. Hasan MR, Rawat A, Tang P, Jithesh PV, Thomas E, Tan R, et al. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J Clin Microbiol.* (2016) 54:919–27. doi: 10.1128/JCM.03050-15
 47. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, et al. A method for selectively enriching microbial

DNA from contaminating vertebrate host DNA. *PLoS ONE*. (2013) 8:e76096. doi: 10.1371/journal.pone.0076096

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tian, Duan, Zhang, Wu, Zhang, Wang, Cao, Gu, Shao and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

EDITED BY

Mel C. Melendrez,
Anoka-Ramsey
Community College,
United States

REVIEWED BY

Ting-Yu Yeh,
Auxergen Inc., United States
David Hugh Evans,
University of Alberta,
Canada

*CORRESPONDENCE

Kimberly A. Bishop-Lilly
kimberly.a.bishop-lilly.civ@mail.mil

†PRESENT ADDRESSES

Bishwo N. Adhikari,
United States Department of Agriculture
(USDA), Animal and Plant Health Inspection
Service, Plant Germplasm Quarantine
Program, Laurel, MD, United States
Matthew R. Lueder,
Department of Pathology and Laboratory
Medicine, The Children's Hospital of
Philadelphia, Philadelphia, PA, United States
Lindsay A. Glang,
Illumina Inc., San Diego, CA, United States

SPECIALTY SECTION

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

RECEIVED 03 June 2022

ACCEPTED 05 July 2022

PUBLISHED 10 August 2022

CITATION

Cer RZ, Voegtly LJ, Adhikari BN, Pike BL,
Lueder MR, Glang LA, Malagon F, Ana ES,
Regeimbal JM, Potts-Szoke MF, Schully KL,
Smith DR and Bishop-Lilly KA (2022)
Genomic and virologic characterization of
samples from a shipboard outbreak of
COVID-19 reveals distinct variants within
limited temporospatial parameters.
Front. Microbiol. 13:960932.
doi: 10.3389/fmicb.2022.960932

COPYRIGHT

© 2022 Cer, Voegtly, Adhikari, Pike, Lueder,
Glang, Malagon, Ana, Regeimbal, Potts-
Szoke, Schully, Smith and Bishop-Lilly. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Genomic and virologic characterization of samples from a shipboard outbreak of COVID-19 reveals distinct variants within limited temporospatial parameters

Regina Z. Cer¹, Logan J. Voegtly^{1,2}, Bishwo N. Adhikari^{1,3†},
Brian L. Pike⁴, Matthew R. Lueder^{1,2†}, Lindsay A. Glang^{1,2†},
Francisco Malagon^{1,2}, Ernesto Santa Ana⁴,
James M. Regeimbal⁴, Maria F. Potts-Szoke⁴, Kevin L. Schully⁵,
Darci R. Smith⁶ and Kimberly A. Bishop-Lilly^{1*}

¹Department of Genomics and Bioinformatics, Biological Defense Research Directorate, Naval
Medical Research Center, Fort Detrick, MD, United States, ²Leidos, Reston, VA, United States,

³Defense Threat Reduction Agency, Ft. Belvoir, VA, United States, ⁴Department of Operations,
Biological Defense Research Directorate, Naval Medical Research Center, Fort Detrick, MD, United
States, ⁵The Austere Environments Consortium for Enhanced Sepsis Outcomes (ACESO), Biological
Defense Research Directorate, Naval Medical Research Center, Fort Detrick, MD, United States,

⁶Department of Microbiology and Immunology, Biological Defense Research Directorate, Naval
Medical Research Center, Fort Detrick, MD, United States

Early in the pandemic, in March of 2020, an outbreak of COVID-19 occurred aboard the aircraft carrier USS Theodore Roosevelt (CVN-71), during deployment in the Western Pacific. Out of the crew of 4,779 personnel, 1,331 service members were suspected or confirmed to be infected with SARS-CoV-2. The demographic, epidemiologic, and laboratory findings of service members from subsequent investigations have characterized the outbreak as widespread transmission of virus with relatively mild symptoms and asymptomatic infection among mostly young healthy adults. At the time, there was no available vaccination against COVID-19 and there was very limited knowledge regarding SARS-CoV-2 mutation, dispersal, and transmission patterns among service members in a shipboard environment. Since that time, other shipboard outbreaks from which data can be extracted have occurred, but these later shipboard outbreaks have occurred largely in settings where the majority of the crew were vaccinated, thereby limiting spread of the virus, shortening duration of the outbreaks, and minimizing evolution of the virus within those close quarters settings. On the other hand, since the outbreak on the CVN-71 occurred prior to widespread vaccination, it continued over the course of roughly two months, infecting more than 25% of the crew. In order to better understand genetic variability and potential transmission dynamics of COVID-19 in a shipboard environment of immunologically naïve, healthy individuals, we performed whole-genome sequencing and virus culture from eighteen COVID-19-positive swabs collected over the course of one week. Using the unique variants identified in those genomes, we detected seven discrete groups of individuals within the population aboard CVN-71

infected with viruses of distinct genomic signature. This is in stark contrast to a recent outbreak aboard another U.S. Navy ship with >98% vaccinated crew after a port visit in Reykjavik, Iceland, where the outbreak lasted only approximately 2 weeks and the virus was clonal. Taken together, these results demonstrate the utility of sequencing from complex clinical samples for molecular epidemiology and they also suggest that a high rate of vaccination among a population in close communities may greatly reduce spread, thereby restricting evolution of the virus.

KEYWORDS

shipboard outbreak, COVID-19, genomic characterization, aircraft carrier, molecular epidemiology

Introduction

COVID-19, which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has emerged as a public health crisis all over the world. Early in the pandemic, in March of 2020, an outbreak of COVID-19 occurred aboard the aircraft carrier USS Theodore Roosevelt during deployment to the Western Pacific. Out of 4,779 ship crew personnel, 1,331 service members (27.9%) were suspected or confirmed to be infected with SARS-CoV-2 (Alvarado et al., 2020; Kasper et al., 2020). Subsequent investigations into this outbreak have characterized it as widespread transmission of virus with relatively mild symptoms and asymptomatic infection among mostly young and healthy adults (Alvarado et al., 2020; Kasper et al., 2020; Payne et al., 2020). This was significant because at the time of this outbreak, data from SARS-CoV-2 viral infections were primarily ascertained from older and immunocompromised components of the population, and there was not much information from outbreaks among younger, healthy populations living in close quarters. Additionally, at that time, vaccination against COVID-19, a key element to limiting the spread of SARS-CoV-2, was not available. Since then, there have been documented studies of SARS-CoV-2 transmission dynamics in the close quarter settings that are characteristic of the U.S. military (Letizia et al., 2020, 2021a,b). To understand how the SARS-CoV-2 virus is transmitted, mutated, and dispersed among unvaccinated service members in a shipboard environment, this retroactive study examines viral characterization data from nasal swab samples obtained from 18 individuals with confirmed or suspected infection. Analysis of these SARS-CoV-2 genomes demonstrates the existence of multiple genetic variants of the virus within the same population over a very short length of time.

The shipboard environment is a unique and challenging environment, in a number of ways, particularly when it comes to infectious disease control. Although it has been 2 years since the pandemic began, there are very limited genomic analyses performed on viral samples from shipboard outbreaks in general, including both military and commercial vessels. Literature

searches were performed in PubMed, the National Center of Biotechnology Information, and Google Scholar for articles published in English between 2019 and April 25, 2022, with keyword combination of “SARS-CoV-2” or “genomic characterization” and “ship.” This search resulted in a few COVID-19 outbreak reports on the Diamond Princess cruise ship in Japan in February 2020 (Sekizuka et al., 2020; Yeh and Contreras, 2021), a fishing vessel that departed from Seattle, Washington, in May 2020 with 122 COVID cases (Addetia et al., 2020), the USNS COMFORT which was deployed in New York City to assist the inpatient health care capacity in New York City and later had 13 cases arise onboard (Lalani et al., 2021), the USS Ronald Reagan (CVN-76; Mullinax et al., 2022), a Navy vessel off the coast of Iceland (Servies et al., 2022), as well as follow-up reviews of some of these outbreaks (Batista et al., 2020; Vicente et al., 2021). These studies, however, mainly focused on demographic, epidemiologic, and general phenotypic characterization of COVID positive samples and hardly on genomic characterization of the pathogen itself. Fortunately, not every vessel will experience an outbreak, as has been demonstrated by the USS Harry S. Truman Strike Group deployed from Norfolk, VA, in November 2019, where there was no outbreak and the ship returned to its home port in June 2020 with zero COVID-19 cases (Bigornia, 2021), but understanding how a novel pathogen may spread and diverge in such an environment may help to develop effective countermeasure strategies.

A limited number of studies of shipboard outbreaks with viral genetic analyses conducted exist, but these include: the Diamond Princess Cruise ship study that analyzed SARS-CoV-2 genomes from 28 individuals, the fish vessel outbreak study with 39 genomes sequenced out of 122 individuals infected, and the Navy vessel off the coast of Iceland, from which 18 samples were sequenced. Widespread vaccination and high rates of natural infection have dramatically limited our ability to examine the natural history of SARS-CoV-2 transmission and mutation within an immunologically naïve population in very close quarters. Here, we present a retrospective study that is the first report of the Theodore Roosevelt (CVN-71) outbreak that includes the

application of viral genome sequencing with virologic and epidemiological data to study pathogen variations arising within an immunologically naïve population confined to close quarters. Despite the outbreak occurring in an isolated shipboard environment where it might be expected that one viral strain might rapidly multiply, we identified distinct variations occurring within the cohort in a very short timeframe. Based on some unique genetic variations identified in those genomes, we were able to categorize the subjects into seven groups with viruses of distinct genomic signatures that build upon one another as sequential mutations in a very short time, demonstrating possible transmission chains. The implications of constant and rapid evolution of SARS-CoV-2 could be relevant in efforts to halt transmission chains and to enable a more targeted approach to disease control in a shipboard environment for this, and potential future, pandemics.

Materials and methods

Study design, sample collection, and RT-PCR

Eighteen SARS-CoV-2-positive samples from the COVID-19 outbreak aboard the USS Theodore Roosevelt (CVN-71) were samples of convenience that were randomly selected from symptomatic individuals that presented to sick call aboard the ship on March 29 and March 30, 2020 (Figure 1), de-identified, and sent for viral culture and genome sequencing. These samples were collected as part of the ship's outbreak response, which involved testing of suspected COVID-19 cases and close contacts as previously reported (Kasper et al., 2020). Briefly, nasopharyngeal (NP) swab specimens were collected from individuals using viral transport medium (VTM) swab kits. Samples were processed with either the Qiagen QIAamp Viral RNA Mini Kit or the Roche MagNA Pure 96 instrument for automated nucleic acid extraction per the manufacturer's instructions. The presence of SARS-CoV-2 infection was determined by the Seegene Allplex 2019-nCoV assay test kit (Seegene Technologies) or by the Centers for Disease Control and Prevention (CDC) emergency-use-authorization (EUA) assay, each using primers targeting two sites in the nucleocapsid gene, N1 and N2.

Virus culture

The viral component of each sample was cultured under Biosafety Level 3 (BSL-3) conditions and simultaneously sequenced from primary material under nonhuman subject research determination PJT 20-08. Patient samples were cultured for SARS-CoV-2 using a standard plaque assay on Vero cells in six-well plates as well as by a cytopathic effect (CPE) assay on Vero cells in T-25 cm² flasks. For the plaque assay, duplicate wells were infected with 0.2 ml aliquots from a 1:2 and serial 10-fold dilutions in Minimum Essential

Medium (MEM), followed by an hour incubation at 37°C with 5% CO₂ to allow virus adsorption to occur. After incubation, cells were overlaid with MEM containing 0.5% agar supplemented with 5% heat-inactivated fetal bovine serum (FBS) and 1% penicillin/streptomycin and incubated for 72 h at 37°C with 5% CO₂. Cells were fixed in 10% formalin prior to staining with crystal violet and plaque counting. For the CPE assay, cells were seeded in T-25 cm² flasks and each flask was infected with 0.5 ml aliquots from a 1:2, 1:5, and 1:10 dilution in MEM, followed by a 1 h incubation at 37°C with 5% CO₂. After incubation, 5 ml of MEM was supplemented with 5% heat-inactivated FBS, 1% penicillin/streptomycin was added, and the flasks were incubated for 5 days at 37°C with 5% CO₂. CPE was monitored daily and post 5 days, supernatant was passed onto fresh cells to allow additional time to amplify.

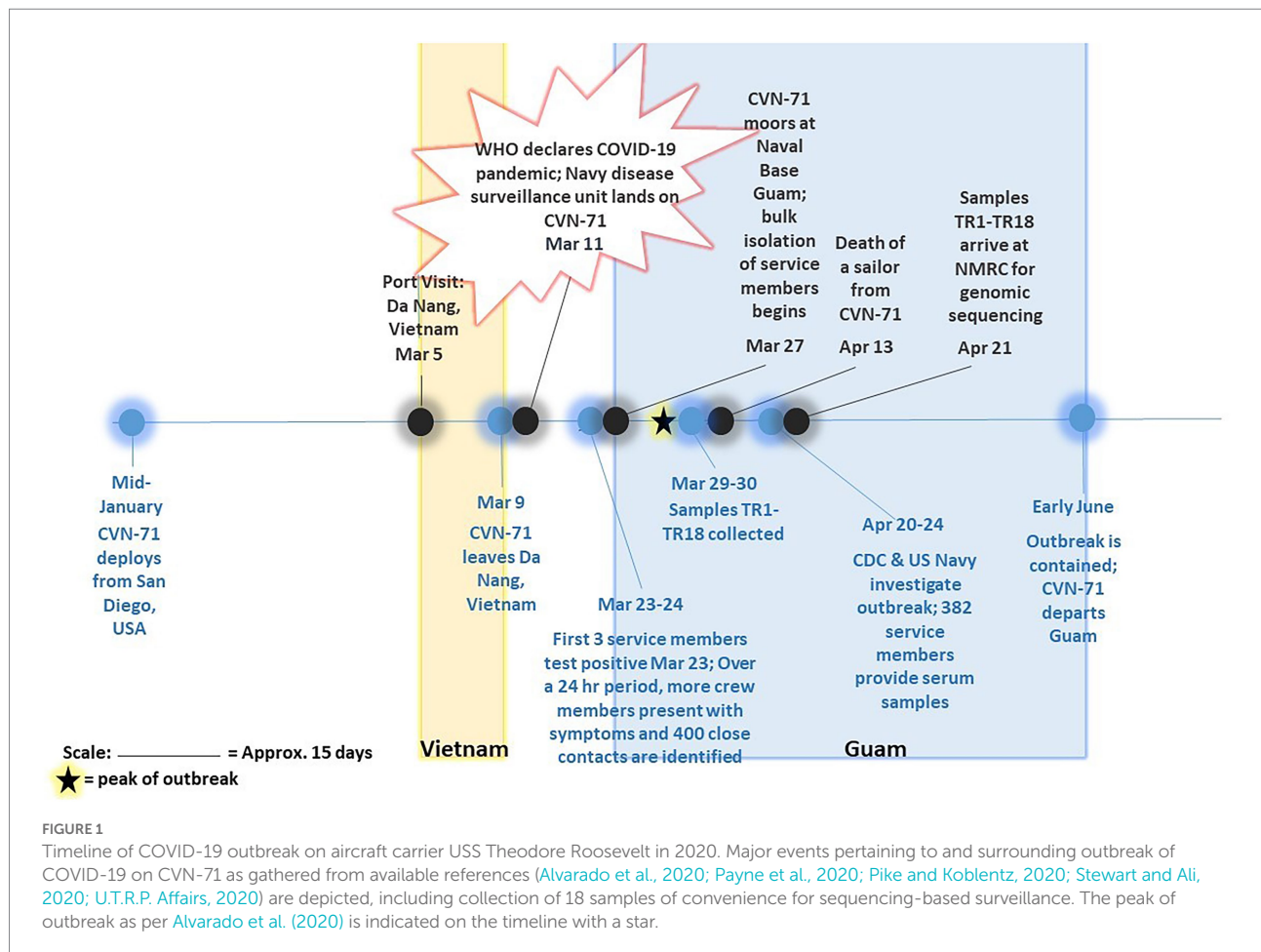
Library preparation and genome sequencing

RNA was extracted from 0.25 ml of VTM using 0.75 ml of TRIzol LS reagent (Invitrogen) according to the manufacturer's protocol. RNA concentration was measured using Qubit RNA High-Sensitivity assay (Thermo Fisher Scientific) prior to use in the ARTIC v3 nCoV-2019 Sequencing protocol (Quick, 2020) with the exception of one sample (i.e., TR1) that required additional sequencing, for which the YouSeq v2 SARS-CoV-2 Coronavirus NGS Library preparation kit was used. In that case, the YouSeq reverse transcriptase was replaced with SuperScript IV (Thermo Fisher Scientific). Complementary DNA (cDNA) was amplified using multiplex PCR and either the associated ARTIC primer pools or YouSeq primer pools. Samples prepared *via* the ARTIC protocol were cleaned using 1x AMPure XP beads (Beckman Coulter) and resuspended in nuclease-free molecular grade water. Sequencing libraries were completed following the QiaSeq FX protocol (Qiagen). Libraries were checked for quality using an Agilent Bioanalyzer High-sensitivity kit (Agilent) and quantitated using the Qubit DNA High-Sensitivity assay (Thermo Fisher Scientific) prior to sequencing using Illumina MiSeq v3 2×300 chemistry (Illumina).

Bioinformatic analyses

Raw sequencing reads were processed using Viral Amplicon Illumina Workflow (VAIW, version 1; Batista et al., 2020). Briefly, the reads were first assessed for quality (Q20) and trimmed using bbduk.¹ Then, resulting paired reads were merged using bbmerge and aligned to the Wuhan reference genome (NC_045512.2) using bbmap (Bushnell, 2014). ARTIC or YouSeq primers were trimmed from sequence ends

¹ BBTools User Guide.



using align_trim (ARTIC pipeline). The consensus genome was generated and Single-Nucleotide Variants (SNVs) were determined using samtools mpileup (Li et al., 2009) and iVar (Intrahost Variant Analysis of Replicates; Grubaugh et al., 2019) with a minimum coverage of 10x and minimum nucleotide frequency of 30%. Global lineage was determined using Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLin v3.1.19). A maximum-likelihood tree was generated using 16 representative sequences, the Wuhan Hu-1 reference, and the 18 samples collected for this study (herein called TR samples) using MAFFT (Katoh et al., 2002) and IQ-tree 2 ML (GTR + G; Minh et al., 2020). Resulting trees were visualized using FigTree (Rambaut, 2012). The 16 representative sequences were obtained by selecting a subset of B.1.1 reference genomes from Global Initiative on Sharing All Influenza Data repository (GISAID accessed, April 8, 2022) by filtering for genome completeness, high coverage, and collection date availability up to April 1, 2020, to match the timing of the sample collection. These samples were further reduced by removing sequences that contain stretches of ambiguous or unsequenced bases (represented as “N” in reference sequences) and by clustering with mmseqs (v13.45111; Hauser et al., 2016) to remove duplicates. The

remaining samples were processed through nextclade (v10.0.10) to identify SNVs present in the samples. Final representatives, which had the earliest collection date and the same core SNVs as the TR samples, were chosen, thereby reducing the dataset from 5,730 samples to 16 samples, representing 1,415 identical samples.

Results

PCR testing

Eighteen COVID-positive surveillance samples of convenience from patients who presented to sick call aboard the Theodore Roosevelt over a 2-day period between March 29 and 30, 2020, at the peak of the outbreak (Alvarado et al., 2020; Figure 1) were sent for viral isolation and sequencing (Table 1). Three samples, TR16, TR17, and TR18, were initially identified as close contacts of other cases on the ship and subsequently tested within a pooled testing format and found negative on March 24 or March 25, 2020. However, when tested again on March 30, 2020, they became COVID positive. Most of the cases had very low real-time PCR cycle

TABLE 1 Metadata of samples included in study.

Sample ID	Referral source	Sample collection date	COVID-19 test		Ct value		
			Date	Result	N1	N2	RNaseP
TR1	Sick call	3/30/2020	3/30/2020	Positive	26.69	30.06	19.28
TR2	Sick call	3/30/2020	3/30/2020	Positive	23.23	25.3	20.28
TR3	Sick call	3/29/2020	3/30/2020	Indeterminate	31.32	35.11	24.35
TR4	Sick call	3/30/2020	3/30/2020	Positive	15.53	15.55	26
TR5	Sick call	3/30/2020	3/30/2020	Positive	UND	13.2	24.14
TR6	Sick call	3/30/2020	3/30/2020	Positive	19.59	19.82	23.55
TR7	Missing data	3/29/2020	3/30/2020	Positive	20.02	21.74	18.54
TR8	Sick call	3/30/2020	3/30/2020	Positive	10	9.13	22.31
TR9	Sick call	3/30/2020	3/30/2020	Positive	23.96	24.12	24.69
TR10	Sick call	3/30/2020	3/30/2020	Positive	UND	12	19.17
TR11	Sick call	3/30/2020	3/30/2020	Positive	20.36	22.53	24.21
TR12	Sick call	3/29/2020	3/30/2020	Positive	UND	12.13	24.39
TR13	Sick call	3/30/2020	3/30/2020	Positive	18.69	19.01	22.72
TR14	Sick call	3/30/2020	3/30/2020	Positive	16.34	17.65	22.62
TR15	Sick call	3/30/2020	3/30/2020	Positive	10.1	10.4	22.26
TR16*	Close contacts	3/24/2020	3/24/2020	Negative	N/A (pooled)	N/A (pooled)	
	Sick call	3/30/2020	3/30/2020	Positive	14.5	15.57	25.42
TR17*	Close contacts	3/24/2020	3/25/2020	Negative	N/A (pooled)	N/A (pooled)	
	Sick call	3/30/2020	3/30/2020	Positive	16.7	19.12	25.25
TR18*	Close contacts	3/25/2020	3/26/2020	Negative	N/A (pooled)	N/A (pooled)	
	Sick call	3/30/2020	3/30/2020	Positive	18.67	20.31	24.55

N/A = not applicable; UND = undetermined; *Identified as close contacts of other cases on the ship and thus tested within a pooled testing format and found negative 5–6 days prior before becoming positive on March 30, 2020.

threshold (CT value ~10) to relatively low CT values (<20) indicating high viral loads with the exceptions of two samples: TR1 (N1: 26.69 and N2: 30.06) and TR3 (N1: 31.32 and N2: 35.11), for which PCR test result for the latter was indeterminate.

Virus culture

Aliquots of VTM from each individual were assessed by CPE assay on Vero cells in T-25 cm² flasks that were monitored for signs of growth and by standard plaque assay in six-well plates to determine the infectious virus titer. Three samples, TR1, TR11, and TR13, were negative by both CPE and plaque assays. Two samples, TR2 and TR9, were positive for CPE but negative by plaque assay. The remaining 13 samples were positive by both CPE and plaque (with titers ranging from 2.9 to 5.3 log₁₀ plaque-forming units per ml PFU/ml) assays (Table 2). Sample TR3, which was “indeterminate” by PCR assay, was determined to have a midrange titer of 4.5 log₁₀ PFU/ml in the plaque assay. This suggests that other factors than low titer may have caused the indeterminate PCR result and that indeterminate PCR results may still merit genome sequencing in an outbreak investigation, particularly if taken from a symptomatic patient and/or a patient with epidemiology info such as a known exposure.

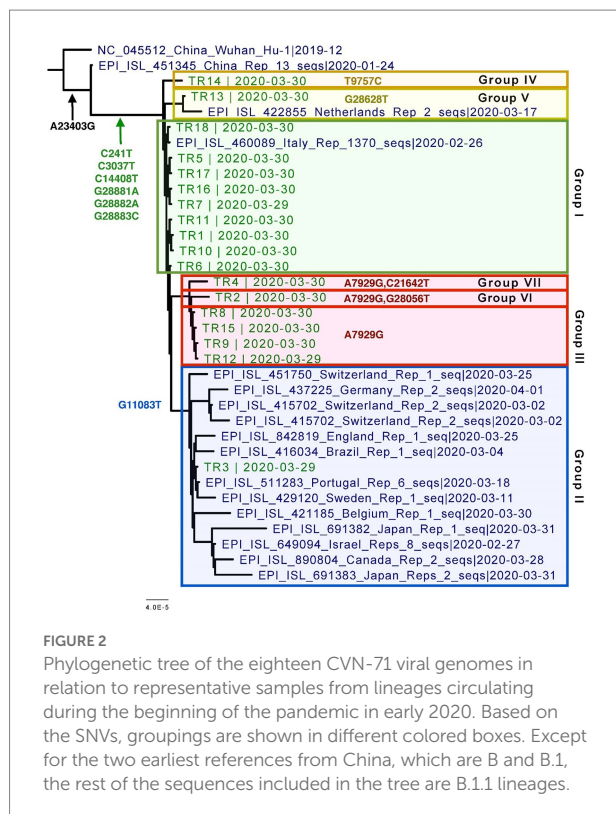
Genome sequencing and assembly

Despite differences in Ct values and viral titers, including some samples that did not grow in cell culture at all, coding complete viral genomes were achieved from all eighteen samples, indicating that even if the samples did not contain viable virus any longer, they did contain sufficient viral nucleic acids. Sample TR1, which had higher Ct values (N1: 26.69, N2:30.06) than TR11 and TR13 negative by both assays, was sequenced using two different and complementary SARS-CoV-2 amplicon sequencing strategies, ARTIC and YouSeq. The additional sequencing protocol applied to this sample generated more sequence data from the 5′ and 3′ noncoding regions than for other samples, resulting in a slightly longer consensus genome length of 29,801 nt as opposed to 29,782 nt in other samples. Another sample with a higher CT value, sample TR3 (N1:31.32, N2: 35.11), required multiple library preparations to yield a full viral genome.

Overall, all of the 18 genomes were very similar to each other; all of them belonged to Pangolin lineage B.1.1, NextStrain clade 20B, and GISAID clade GR (Table 2; Figure 2). They all contained mutations: A23403G (D614G), C241T (5′UTR), C3037T (F924F), C14408T (P4715L), G2881A (R203K), G2882A (R203R), and G2883C (G204R). This is in contrast to one specific report on viral genome sequence analysis from cases on the Diamond Princess, in which 8 of 28 cases were identical to the original Wuhan WIV-4 sequence (Yeh and Contreras, 2021). This

TABLE 2 Virus cultivation results and genome sequencing statistics.

Sample ID	Titer (PFU/ml)	Cytopathic effect (CPE)	Number of raw sequencing reads	Consensus genome length (bp)	Virus lineage/nextStrain clade/GISAID clade
TR1	Negative	Negative	5,606,273	29,801	B.1.1/20B/GR
TR2	Negative	Positive	7,792,070	29,782	B.1.1/20B/GR
TR3	4.5 log ₁₀	Positive	9,483,672	29,782	B.1.1/20B/GR
TR4	4.4 log ₁₀	Positive	2,217,062	29,782	B.1.1/20B/GR
TR5	5.3 log ₁₀	Positive	6,332,426	29,782	B.1.1/20B/GR
TR6	3.6 log ₁₀	Positive	7,750,816	29,782	B.1.1/20B/GR
TR7	2.9 log ₁₀	Positive	1,327,290	29,782	B.1.1/20B/GR
TR8	4.5 log ₁₀	Positive	4,580,614	29,782	B.1.1/20B/GR
TR9	Negative	Positive	8,195,212	29,782	B.1.1/20B/GR
TR10	3.7 log ₁₀	Positive	4,096,140	29,782	B.1.1/20B/GR
TR11	Negative	Negative	4,375,622	29,782	B.1.1/20B/GR
TR12	4.1 log ₁₀	Positive	2,102,416	29,782	B.1.1/20B/GR
TR13	Negative	Negative	3,039,242	29,782	B.1.1/20B/GR
TR14	3.6 log ₁₀	Positive	10,141,716	29,782	B.1.1/20B/GR
TR15	5.3 log ₁₀	Positive	3,130,260	29,782	B.1.1/20B/GR
TR16	3.4 log ₁₀	Positive	1,316,262	29,782	B.1.1/20B/GR
TR17	3.5 log ₁₀	Positive	6,894,320	29,782	B.1.1/20B/GR
TR18	5.3 log ₁₀	Positive	3,130,260	29,782	B.1.1/20B/GR



constellation of multiple variations in common among the samples is not surprising since the samples were all collected within 2 days and from a closely knit community. Based on the phylogenetic tree placement, viral sequences from 18 samples can

be categorized into seven discrete groups. Group I consisted of genomes from the three samples (TR16, TR17, and TR18) which were collected from close contacts and tested negative initially on March 24 and March 25. Group 1 also included viruses from samples TR1 and TR11, which were negative by both CPE and plaque assays. Sample TR3, which had the highest CT value among the 18 samples with indeterminate PCR result, produced the only genome belonging to Group II. Genomes from samples TR2 and TR9, which were positive samples for CPE but too low titer to quantify, belonged to Groups VI and III, respectively. Finally, sample TR13 that was negative by both CPE and titer assays produced a genome that belonged to Group V.

Specifically, these groups are based on the presence or absence of six high-quality single-nucleotide variations (SNVs): A7929G (K2555R in ORF1a), R3164R (T9757C in ORF1a), L37F (G11083T in ORF1a), A27V (C21642T in S), A55S (G28056T in ORF8), and A119S (G28628T in N; Figure 3; Table 3). These six SNVs are not any of the five characteristic mutations of B.1.1: P314L (ORF1b), D614G (S), S84L (ORF8), R203K (N), and G204R (N). Briefly, one sample each was found to have a non-synonymous SNV in the S gene that encodes spike glycoprotein (TR4), ORF8 (TR2), or the nucleocapsid phosphoprotein gene (TR13). Notably, the A7929G SNV in nsp3 that results in a non-synonymous mutation K2555R was shared among 33% of the samples (n = 6, TR2, TR4, TR8, TR9, TR12, and TR14) and this was the only SNV found in common with any other SNV in a single genome. This is in contrast with the findings from the Diamond Princess cruise ship study where the G11083T SNV in nsp6 that results in an L37F mutation was observed in 14% (4/28) of the samples spread during shipboard quarantine and arose through *de novo* RNA recombination under positive selection pressure

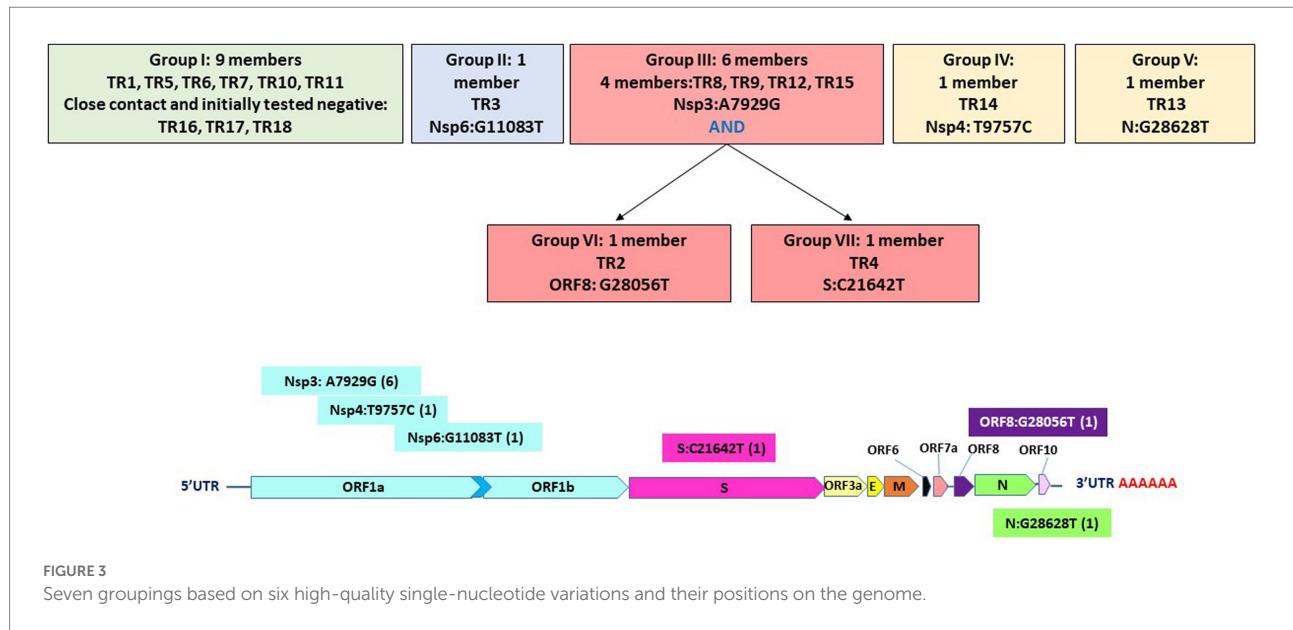


TABLE 3 Single nucleotide variations and effects.

Variant	# of samples	Group	Coverage [§]	Allele frequency (%)	Gene	Protein product affected	synonymous/nonsynonymous	Amino acid change	Biological significance
A7929G	6	III, VI, VII	8495 [§]	99.7 [§]	ORF1a	nsp3	Non-synonymous	K2555R	No report found
T9757C	1	IV	15913	98.2	ORF1a	nsp4	synonymous	Not applicable	Not applicable
G11083T	1	II	17012	99.9	ORF1a	nsp6	Non-synonymous	L37F	Associated with asymptomatic SARS-CoV-2 infection. Linked to viral hypotoxicity (Wang et al., 2020; Sun et al., 2022)
C21642T	1	VII	7117	99.9	S	spike	Non-synonymous	A27V	Located on N-terminal domain; no significant effects (Ferrareze et al., 2021)
G28056T	1	VI	5222	95.4	ORF8	ORF8 protein	Non-synonymous	A55S	Destabilizes the RNA binding domain (Rahman et al., 2021)
G28628T	1	V	6174	100	N	Nucleocapsid phosphoprotein	Non-synonymous	A119S	N:A119S is one of the five lineage-defining SNVs that distinguish P.2 (B.1.1.28.2) sequences from all other B.1.1.28 sequences available in Brazil (Resende et al., 2021)

N/A = not applicable.

[§]Number of reads supporting alternative (mutant) allele.

[§]Average of six samples with read coverage (3253, 5954, 7559, 9790, 11566, 12847).

(Yeh and Contreras, 2021). In the Theodore Roosevelt (CVN-71) shipboard outbreak, we observed L37F in only 5% (1/18) of the group. It is important to note that the K2555R mutation in ORF1a appears to be unique to the Theodore Roosevelt shipboard outbreak as compared to other published B.1.1 genomes from that period.

Despite all these distinct genomic signatures, sample bias in data availability from different parts of the world was an inevitable issue that made the confident assignment of the geographic origin of the outbreak impossible. In this case, B.1.1 lineage reference genomes from samples collected through April 1, 2020, in GISAID were mainly from Europe (4,058), followed by North America (625), Asia (580), Oceania (284), South America (131), and Africa (25). We found certain parts of Asia to be undersampled in SARS-CoV-2 genome sequencing. For instance, at the time of these genomes being sequenced, specifically looking at GISAID submission dates through June 12, 2020, to allow time for sample processing and submission post sample collection, we found that there were 46,636 SARS-CoV-2 genomes available in GISAID, without filtering, and of those, only 3,875 (8.31%) were from Asia. If we applied filtering for complete genomes and low depth of coverage, 31,729 total genomes were available, of that, 2,796 (8.81%) were from Asia. In other words, whether with or without filtering, the entire continent of Asia was contributing less than 10% of the reference genomes for SARS-CoV-2 worldwide, despite the outbreak having been observed there first (Table 4) and there being ~20% of COVID cases worldwide from that area, demonstrating strong biases toward sequencing from samples taken in Europe and Oceania in general. Specifically analyzing data by country indicates undersampling in the Philippines and oversampling in Vietnam. That is, within Asia, there were only 17 reference SARS-CoV-2 genomes available in GISAID for the Philippines although it reported 26,420 COVID cases (i.e., 0.06%), whereas Vietnam submitted 48 genomes (14.37%) out of 334 COVID cases reported at the time (Figure 4). Due to this unevenness in viral genome sampling from different geographic regions around the world at that time and the possibility that the databases might therefore lack

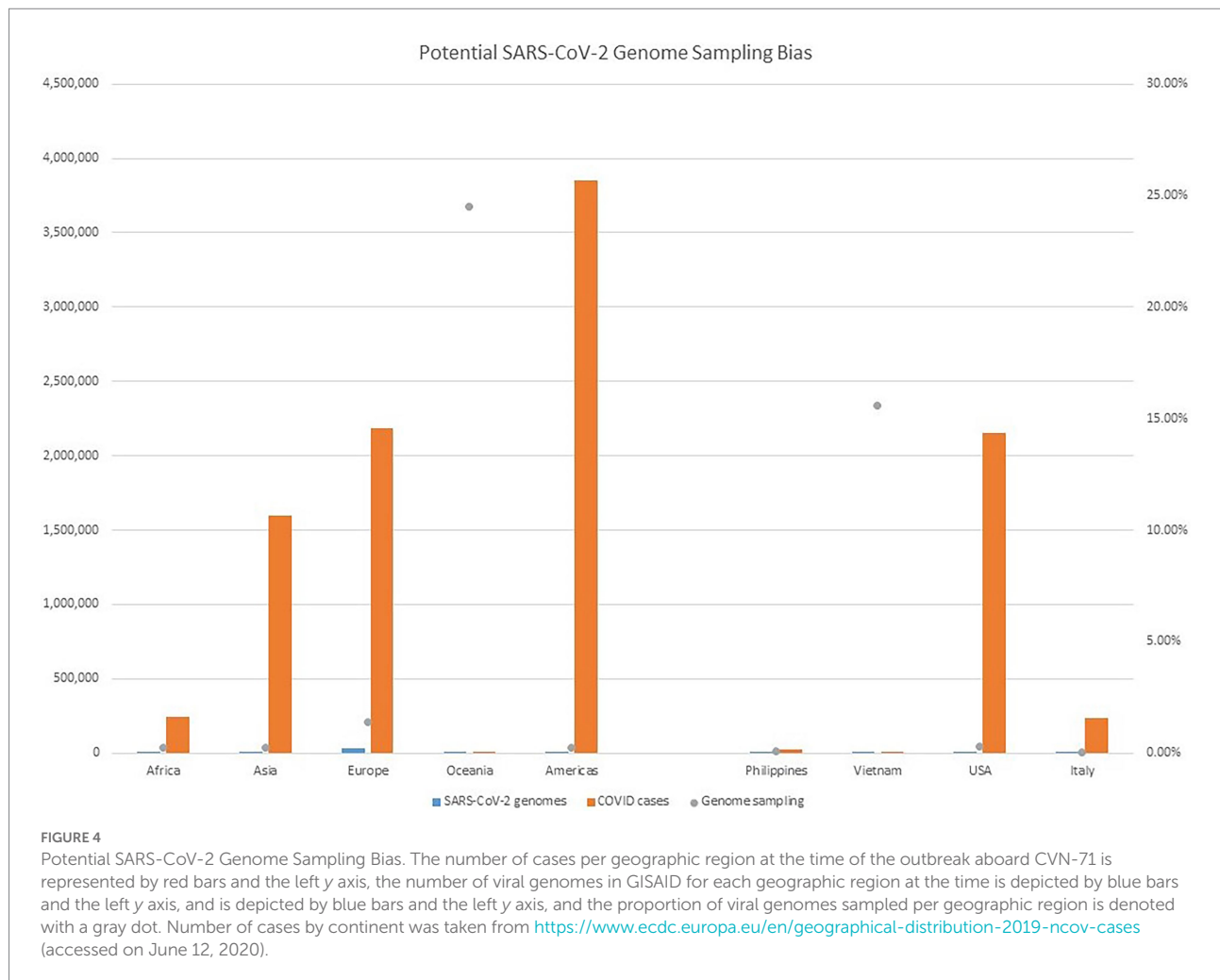
enough representative sequences for a given region, speculation on the origin of the outbreak based on sequencing-based surveillance on the Theodore Roosevelt (CVN-71) was not attempted.

Discussion

Herein, we have presented a unique genetic and virologic investigation of a shipboard outbreak of an emerging viral pathogen. The 18 viral genomes from servicemen and servicewomen onboard the Theodore Roosevelt (CVN-71) were sequenced from samples taken in a span of 2 days only approximately 1 week post recognition of the outbreak on the ship, and represented only a small fraction of the total cases. Yet, they contained variations that allowed them to be separated into seven distinct groups. One particular variant, A7929G, a non-synonymous variation in nsp3, was favored among this sample set occurring in 33% of the cohort. Nsp3 is an essential component of SARS-CoV-2 replication (Lei et al., 2018) and is considered a potential target for antiviral drugs (Baez-Santos et al., 2015). It is interesting to note that despite the close timeline of outbreaks between Diamond Princess Cruise ship (February 10–25, 2020) and the Theodore Roosevelt ship outbreak (March 24, 2020), the most common mutation differed between them: L37F and K2555R being the most common, respectively. In fact, it is interesting to note that the K255R mutation appears to be unique to the Theodore Roosevelt shipboard outbreak. In addition, we have demonstrated a sampling bias in terms of viral genome sequencing in different regions around the world at the time, a bias that would somewhat limit conclusions that could be drawn regarding the potential source of infection. It is important to remember that these are potentially important caveats to the interpretation of data regarding predominance of a given lineage over another in certain geographic regions, as well as the origin of the virus in this particular shipboard outbreak.

TABLE 4 SARS-CoV-2 reference genomes and cases by geographic region at the time of viral genome sequencing from Theodore Roosevelt outbreak (up to June 12, 2020).

		GenBank/NCBI			GISAID	
	Geographic region	Total number of COVID cases	Total number of SARS-CoV-2 genomes	Genome sampling %	Total number of SARS-CoV-2 genomes	Genome sampling %
By continent	Africa	243,125	539	0.22	469	0.19
	Asia	1,594,208	3,780	0.24	3,875	0.24
	Europe	2,187,307	29,911	1.37	29,980	1.37
	Oceania	8,796	2,157	24.52	2,152	24.47
	Americas	3,848,098	10,149	0.26	10,160	0.26
By country	Philippines	26,420	17	0.06	17	0.06
	Vietnam	334	52	15.57	48	14.37
	United States	2,150,000	5,973	0.28	8,688	0.40
	Italy	237,000	108	0.05	147	0.06



Recently, sequencing-based surveillance was applied to another Navy shipboard outbreak of COVID-19, and the findings were rather different in that more recent study, there was much less genetic variation observed. In fact, in that more recent study, of the 18 samples sequenced, 16 samples were 100% identical to each other. The genome of the 17th sample was found to be only one nucleotide different from the sixteen samples and the 18th sample did not produce a sufficient consensus genome to be analyzed in this manner (Servies et al., 2022). The most salient difference between these two outbreaks is that the outbreak on CVN-71 occurred prior, in an unvaccinated crew, whereas the other outbreak occurred more recently among a highly vaccinated crew. Taken together, these results suggest that a high rate of vaccination among a population in close quarters may greatly reduce spread, thereby restricting evolution of the virus; an important reminder for future pandemics. In fact, a recent analysis published online prior to peer review investigates the potential relationship between vaccination rates in various countries as compared to viral mutation rate, and this preprint suggests that purifying selection pressure on the

spike gene of SARS-CoV-2 may increase with increasing vaccination rate.² Further study is clearly warranted to augment our collective understanding of the biology of SARS-CoV-2 and its interplay with the human immune system in support of effective countermeasure development.

The viral genetic data derived from this study, when combined with the epidemiological data, demonstrate possible transmission chains and provide new information as to how quickly a virus may begin to diverge in a contained, immunologically naïve population. Our results demonstrate that genetic variations can occur constantly and rapidly over a relatively short period and that these mutations may be useful for tracking transmission chains. This is important because as novel viruses such as SARS-CoV-2 evolve rapidly after infection the mutations may affect virulence, infectivity, and transmissibility. In addition, this study demonstrates how sequencing-based surveillance, whether it be targeted

² <https://doi.org/10.1101/2021.08.08.21261768>

sequencing as in this study, or whether it be unbiased shotgun sequencing, can be used for molecular epidemiological purposes to support force health protection decision-making, such as to track and halt transmission chains with additional protective measures. Finally, our study confirms the need to accumulate more sequence data from this outbreak to better trace the viral genome evolution and associate the changes with epidemiological data and clinical symptoms. A better understanding of this, and other similarly isolated SARS-CoV-2 outbreaks, will aid in the preparation for containment of future shipboard outbreaks.

Data availability statement

The datasets generated and analyzed for this study can be found in the NCBI GenBank at: <https://www.ncbi.nlm.nih.gov/nuccore/>, accession MW130903-MW130920.

Author contributions

BP, EA, and MP-S performed sample collection and COVID-19 testing for shipboard surveillance. KS and DS performed virus culture. LG and FM performed genome sequencing. LV, RC, ML, EA, JR, and KB-L performed data analyses. RC, LV, BA, BP, and KB-L wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Armed Forces Health Surveillance Division (AFHSD), Global Emerging Infections Surveillance (GEIS) Branch, ProMIS IDs P0013_AH_01.01 to KB-L as well as Navy Work Unit Number (WUN) A1417.

References

- Addetia, A., Crawford, K. H. D., Dingens, A., Zhu, H., Roychoudhury, P., Huang, M. L., et al. (2020). Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with a high attack rate. *J. Clin. Microbiol.* 58:20. doi: 10.1128/JCM.02107-20
- Alvarado, G. R., Pierson, B. C., Teemer, E. S., Gama, H. J., Cole, R. D., and Jang, S. S. (2020). Symptom characterization and outcomes of sailors in isolation After a COVID-19 outbreak on a US aircraft carrier. *JAMA Netw. Open* 3:e2020981. doi: 10.1001/jamanetworkopen.2020.20981
- Baez-Santos, Y. M., St John, S. E., and Mesecar, A. D. (2015). The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antivir. Res.* 115, 21–38. doi: 10.1016/j.antiviral.2014.12.015
- Batista, B., Dickenson, D., Gurski, K., Kebe, M., and Rankin, N. (2020). Minimizing disease spread on a quarantined cruise ship: A model of COVID-19 with asymptomatic infections. *Math. Biosci.* 329:108442. doi: 10.1016/j.mbs.2020.108442
- Bigornia, V. E. (2021). U.S. navy aircraft carrier prevents outbreak at sea in midst of COVID-19. *Mil. Med.* 186, 178–180. doi: 10.1093/milmed/usab107
- Bushnell, B. (2014). BBMAP v38.93: A Fast, Accurate, Splice-Aware Aligner. Available at: <https://sourceforge.net/projects/bbmap>
- Ferrareze, P. A. G., Franceschi, V. B., Mayer, A. M., Caldana, G. D., Zimmerman, R. A., and Thompson, C. E. (2021). E484K as an innovative phylogenetic event for viral evolution: genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. *Infect. Genet. Evol.* 93:104941. doi: 10.1016/j.meegid.2021.104941
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., et al. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* 20:8. doi: 10.1186/s13059-018-1618-7
- Hauser, M., Steinegger, M., and Soding, J. (2016). MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* 32, 1323–1330. doi: 10.1093/bioinformatics/btw006

Acknowledgments

We would like to acknowledge CDR Benjamin Espinosa for thoughtful feedback on this manuscript and we would also like to acknowledge all the service members aboard the USS Theodore Roosevelt for their service.

Conflict of interest

LV, ML, LG, and FM were employed by the company Leidos.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of Defense, Department of the Navy, nor the U.S. Government. Several of the authors are U.S. Government employees. This work was prepared as part of their official duties. Title 17 U.S.C. § 105 provides that “Copyright protection under this title is not available for any work of the United States Government.” Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties.

- Kasper, M. R., Geibe, J. R., Sears, C. L., Riegodedios, A. J., Luse, T., Von Thun, A. M., et al. (2020). An outbreak of Covid-19 on an aircraft carrier. *N. Engl. J. Med.* 383, 2417–2426. doi: 10.1056/NEJMoa2019375
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Lalani, T., Lee, T. K., Laing, E. D., Ritter, A., Cooper, E., Lee, M., et al. (2021). SARS-CoV-2 infections and serologic responses Among military personnel deployed on the UNS COMFORT to new York City During the COVID-19 pandemic. *Open Forum Infect. Dis.* 8:ofaa654. doi: 10.1093/ofid/ofaa654
- Lei, J., Kusov, Y., and Hilgenfeld, R. (2018). Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antivir. Res.* 149, 58–74. doi: 10.1016/j.antiviral.2017.11.001
- Letizia, A. G., Ge, Y., Goforth, C. W., Weir, D. L., Lizewski, R., Lizewski, S., et al. (2021a). SARS-CoV-2 Seropositivity among US marine recruits attending basic training, United States, spring-fall 2020. *Emerg. Infect. Dis.* 27:732. doi: 10.3201/eid2704.204732
- Letizia, A. G., Ge, Y., Vangeti, S., Goforth, C., Weir, D. L., Kuzmina, N. A., et al. (2021b). SARS-CoV-2 seropositivity and subsequent infection risk in healthy young adults: a prospective cohort study. *Lancet Respir. Med.* 9, 712–720. doi: 10.1016/S2213-2600(21)00158-2
- Letizia, A. G., Ramos, I., Obla, A., Goforth, C., Weir, D. L., Ge, Y., et al. (2020). SARS-CoV-2 transmission among marine recruits during quarantine. *N. Engl. J. Med.* 383, 2407–2416. doi: 10.1056/NEJMoa2029717
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). Genome project data processing, The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Mullinax, R., Krug, A., Harvey, K., Wilde, C., Nzegwu, U., Wilcox, C., et al. (2022). Effectiveness of a COVID-19 preventive sequestration strategy: deployment of the aircraft carrier USS Ronald Reagan (CVN-76). *Disaster Med. Pub. Health Prep.* doi: 10.1017/dmp.2022.66
- Payne, D. C., Smith-Jeffcoat, S. E., Nowak, G., Chukwuma, U., Geibe, J. R., Hawkins, R. J., et al. (2020). SARS-CoV-2 infections and serologic responses from a sample of U.S. navy service members - USS Theodore Roosevelt, April 2020. *MMWR Morb. Mortal. Wkly Rep.* 69, 714–721. doi: 10.15585/mmwr.mm6923e4
- Pike, B.L., and Koblenz, G.D. (2020). Lessons From The Roosevelt: A Call For Improving The U. S. Navy's Preparedness For Biological Threats, War on the Rocks, Metamorphic Media LLC.
- Quick, J. (2020). nCoV-2019 sequencing protocol, protocols.io.
- Rahman, M. S., Islam, M. R., Alam, A., Islam, I., Hoque, M. N., Akter, S., et al. (2021). Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *J. Med. Virol.* 93, 2177–2195. doi: 10.1002/jmv.26626
- Rambaut, A. (2012). FigTree v1. 4.4: a graphical viewer of phylogenetic trees. Available at: <http://tree.bio.ed.ac.uk/software/figtree/>
- Resende, P. C., Bezerra, J. F., Teixeira Vasconcelos, R. H., Arantes, I., Appolinario, L., Mendonca, A. C., et al. (2021). Severe acute respiratory syndrome coronavirus 2 P.2 lineage associated with reinfection case, Brazil, June–October 2020. *Emerg. Infect. Dis.* 27, 1789–1794. doi: 10.3201/eid2707.210401
- Sekizuka, T., Itokawa, K., Kageyama, T., Saito, S., Takayama, I., Asanuma, H., et al. (2020). Haplotype networks of SARS-CoV-2 infections in the diamond princess cruise ship outbreak. *Proc. Natl. Acad. Sci. U. S. A.* 117, 20198–20201. doi: 10.1073/pnas.2006824117
- Servies, T. E., Larsen, E. C., Lindsay, R. C., Jones, J. S., Cer, R. Z., Voegtly, L. J., et al. (2022). Notes from the field: outbreak of COVID-19 Among a highly vaccinated population aboard a U.S. navy ship After a port visit - Reykjavik, Iceland, July 2021. *MMWR* 71, 279–281. doi: 10.15585/mmwr.mm7107a5
- Stewart, P., and Ali, I. (2020). U.S. sailor from coronavirus-hit aircraft carrier dies after contracting virus, Reuters.
- Sun, X., Liu, Y., Huang, Z., Xu, W., Hu, W., Yi, L., et al. (2022). SARS-CoV-2 non-structural protein 6 triggers NLRP3-dependent pyroptosis by targeting ATP6AP1. *Cell Death Differ.* 29, 1240–1254. doi: 10.1038/s41418-021-00916-7
- U.T.R.P. Affairs (2020). Sailors Contribute During Vietnam Port Call, Commander, U.S. 7th Fleet; official web site for the U.S. 7th Fleet, Defense Media Activity - WEB.mil.
- Vicente, D., Maves, R., Elster, E., and Shwayhat, A. (2021). U.S. Navy's response to a shipboard coronavirus outbreak: considerations for a medical management plan at sea. *Mil. Med.* 186, 23–26. doi: 10.1093/milmed/usaa455
- Wang, R., Chen, J., Hozumi, Y., Yin, C., and Wei, G. W. (2020). Decoding asymptomatic COVID-19 infection and transmission. *J. Phys. Chem. Lett.* 11, 10007–10015. doi: 10.1021/acs.jpclett.0c02765
- Yeh, T. Y., and Contreras, G. P. (2021). Viral transmission and evolution dynamics of SARS-CoV-2 in shipboard quarantine. *Bull. World Health Organ.* 99, 486–495. doi: 10.2471/BLT.20.255752



OPEN ACCESS

EDITED BY

Axel Cloeckaert,
Institut National de recherche pour
l'agriculture, l'alimentation et
l'environnement (INRAE), France

REVIEWED BY

Algimantas Paulauskas,
Vytautas Magnus University,
Lithuania
Qingli Niu,
Lanzhou Veterinary Research Institute
(CAAS), China
Mónica Nunes,
Instituto de Biologia e Tecnologia
Experimental (iBET), Portugal

*CORRESPONDENCE

Michael E. von Fricken
mvonfric@gmu.edu

SPECIALTY SECTION

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

RECEIVED 17 May 2022

ACCEPTED 28 June 2022

PUBLISHED 10 August 2022

CITATION

Altantogtokh D, Lilak AA, Takhampunya R,
Sakolvaree J, Chanarat N, Matulis G,
Poole-Smith BK, Boldbaatar B, Davidson S,
Hertz J, Bolorchimeg B, Tsogbadrakh N,
Fiorenzano JM, Lindroth EJ and von
Fricken ME (2022) Metagenomic profiles of
Dermacentor tick pathogens from across
Mongolia, using next generation
sequencing.
Front. Microbiol. 13:946631.
doi: 10.3389/fmicb.2022.946631

COPYRIGHT

© 2022 Altantogtokh, Lilak, Takhampunya,
Sakolvaree, Chanarat, Matulis, Poole-Smith,
Boldbaatar, Davidson, Hertz, Bolorchimeg,
Tsogbadrakh, Fiorenzano, Lindroth and von
Fricken. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Metagenomic profiles of *Dermacentor* tick pathogens from across Mongolia, using next generation sequencing

Doniddemberel Altantogtokh¹, Abigail A. Lilak²,
Ratree Takhampunya³, Jira Sakolvaree³, Nitima Chanarat³,
Graham Matulis², Betty Katherine Poole-Smith³,
Bazartseren Boldbaatar⁴, Silas Davidson^{3,5}, Jeffrey Hertz⁶,
Buyandelger Bolorchimeg¹, Nyamdorj Tsogbadrakh¹,
Jodi M. Fiorenzano⁶, Erica J. Lindroth³ and
Michael E. von Fricken^{2*}

¹National Center for Zoonotic Diseases, Ulaanbaatar, Mongolia, ²Department of Global and
Community Health, George Mason University, Fairfax, VA, United States, ³Department of
Entomology, US Army Medical Directorate of the Armed Forces Research Institute of Medical
Sciences (USAMD-AFRIMS), Bangkok, Thailand, ⁴School of Veterinary Medicine, Mongolian
University of Life Sciences, Ulaanbaatar, Mongolia, ⁵Department of Chemistry and Life Science, US
Military Academy, West Point, NY, United States, ⁶Naval Medical Research Unit TWO (NAMRU-2),
Sembawang, Singapore

Tick-borne diseases are a major public health concern in Mongolia. Nomadic pastoralists, which make up ~26% of Mongolia's population, are at an increased risk of both tick bite exposure and economic loss associated with clinical disease in herds. This study sought to further characterize tick-borne pathogens present in *Dermacentor* ticks ($n=1,773$) sampled in 2019 from 15 of Mongolia's 21 aimags (provinces). The ticks were morphologically identified and sorted into 377 pools which were then screened using Next-Generation Sequencing paired with confirmatory PCR and DNA sequence analysis. *Rickettsia* spp. were detected in 88.33% of pools, while *Anaplasma* spp. and *Bartonella* spp. were detected in 3.18 and 0.79% of pools, respectively. Khentii had the highest infection rate for *Rickettsia* spp. (76.61%; CI: 34.65–94.79%), while Arkhangai had the highest infection rate for *Anaplasma* spp. (7.79%; CI: 4.04–13.72%). The exclusive detection of *Anaplasma* spp. in tick pools collected from livestock supports previous work in this area that suggests livestock play a significant role in disease maintenance. The detection of *Anaplasma*, *Bartonella*, and *Rickettsia* demonstrates a heightened risk for infection throughout Mongolia, with this study, to our knowledge, documenting the first detection of *Bartonella melophagi* in ticks collected in Mongolia. Further research deploying NGS methods is needed to characterize tick-borne pathogens in other endemic tick species found in Mongolia, including *Hyalomma asiaticum* and *Ixodes persulcatus*.

KEYWORDS

next generation sequencing, *Dermacentor*, Mongolia, tick-borne disease, *Rickettsia*, *Bartonella*, *Anaplasma*, surveillance

Introduction

Ticks and the pathogens they carry pose a significant threat to both human and animal health. This holds true in Mongolia, where an estimated 26% of the population continues to live a nomadic pastoral lifestyle and 37% of households own livestock (Odontsetseg et al., 2009; Boldbaatar et al., 2017; Barnes et al., 2020). These populations spend prolonged periods of time moving herds through tick habitats, resulting in a heightened risk for exposure to ticks and tick-borne diseases (TBDs). The Mongolian economy is also likely impacted by the effects of TBDs, where an estimated 67 million heads of livestock are present within the country¹ and roughly 18% of the nation's GDP comes from animal-related products (Odontsetseg et al., 2009). In neighboring China, an estimated \$70 million every year is lost due to the impact of tick-borne disease impacts on small mammal production (Yin and Luo, 2007).

Ticks gathered in Mongolia have previously tested positive for various TBDs, including *Anaplasma* spp., *Borrelia* spp., Crimean-Congo hemorrhagic fever, *Ehrlichia* spp., *Rickettsia* spp., and tick-borne encephalitis virus (Moore et al., 2018; Voorhees et al., 2018; Černý et al., 2019; von Fricken et al., 2020a). Rickettsial diseases are of particular concern due to high rates of severe illness and death in previously healthy individuals, (Aung et al., 2014; Biggs, 2016; von Fricken et al., 2018). A previous study by our team found that 20% of humans and livestock animals in Mongolia have had past exposure to *Rickettsia* spp., with variations observed by geographic location (von Fricken et al., 2018). This also held true when examining previous exposure to *Anaplasma* spp., which was detected in 37% of nomadic herders and over 40% of livestock (von Fricken et al., 2018). We also have detected *Anaplasma ovis* infection rates as high as 80% in sheep and 69% in goats, which aligns with what has previously been detected in ticks from the same region (Ochirkhuu et al., 2017; von Fricken et al., 2018, 2020a; Enkhtaivan et al., 2019; Fischer et al., 2020; Chaorattanakawee et al., 2022). Anaplasmosis in livestock can result in anoxia, abortions, infertility, significant weight loss, and even death, all of which can impact economic security in pastoralist communities.

Dermacentor ticks are the most common and one of the more important ticks of medical and veterinary concern within Mongolia due to their wide geographic range and the pathogens they carry (Černý et al., 2019). Ticks collected from southern and central aimags have previously had high pool positivity rates (>80%) for *Rickettsia* spp., with molecular detections of *R. raoultii*, *R. sibirica mongolitimonae*, and *R. sibirica* reported (Fischer et al., 2020; von Fricken et al., 2020b). In contrast, a study of pathogens within ticks collected from aimags of central Mongolia found lower overall levels of *Anaplasma* spp. within *Dermacentor* ticks, although the infectivity rates increased substantially when specifically examining ticks removed from livestock (von Fricken

et al., 2020a). Additional pathogens have been detected within *Dermacentor* ticks collected from Mongolian aimags include *Babesia caballi*, *B. equi*, *Borrelia afzelii*, *Candidatus* Midichloria sp., *Candidatus* Neoehrlichia mukurensis, *Theileria equi*, and *T. orientalis* (Battsetseg et al., 2001; Javkhlan et al., 2014; Fischer et al., 2020). In neighboring countries, pathogens reported from *Dermacentor* spp. ticks include *Babesia venatorum*, *Borrelia miyamotoi*, *Brucella* spp., *Francisella tularensis* subsp. *holarctica*, *Rickettsia aeschlimannii*, and the Far Eastern genotype of tick-borne encephalitis virus (Zhang et al., 2008; Wei et al., 2016; He et al., 2018; Yin et al., 2018; Huang et al., 2020; Gao et al., 2021; Jiao et al., 2021).

The potential threat tick-borne diseases present to both the Mongolian population and its growing ecotourism industry is substantial, given the high rates of various pathogens reported in previous tick survey studies (Černý et al., 2019; Fischer et al., 2020; von Fricken et al., 2020b). Improved molecular characterization of TBDs within Mongolia may help inform future preventative measures for locals and visitors, while also establishing baseline sequence data to monitor evolution over time. The variety of tick-borne pathogens found within Mongolia complicates attempts to fully characterize pathogens found in samples collected within the country. Our research group has recently used an analytical workflow on livestock blood samples from three aimags in Mongolia, initially applying next-generation sequencing (NGS) to obtain a snapshot of pathogen groups present, followed by conventional PCR and Sanger sequencing for confirmation and species characterization (Chaorattanakawee et al., 2022). In this study, we deploy Next-Generation Sequencing for on *Dermacentor* ticks collected from a wide geographic range of Mongolia to further our understanding of tick-borne pathogens in this region.

Materials and methods

Dermacentor ticks were collected from the environment (questing) and off domestic animals from 15 aimags across Mongolia in 2019 (Uvs, Khovd, Govi-Altai, Zavhan, Khuvsgul, Arkhangai, Bayankhongor, Arkhangai, Uvurkhangai, Bulgan, Tuv, Dundgovi, Khentii, Dornogovi, Sukhbaatar, and Dornod; Figure 1). Adult ticks were morphologically identified as *D. nuttalli* or to the genus level as *Dermacentor* spp. by entomologists using local keys (Boldbaatar and Byambaa, 2015). In total, 7,275 ticks were collected and sorted into 1,489 pools according to location and collection source (environment vs. animal). Of these pools, 377 pools of adult stage ticks, representing pools from all sampled provinces, were selected for analysis by next-generation sequencing, including 51 pools collected from livestock (Tables 1, 2). Whole ticks in 250 µl of ATL buffer were punctured with a fine tip under a stereomicroscope to release the tissue from the hard chitin exoskeleton prior to adding 2 mg/ml of Proteinase K solution. Samples were then incubated at 55°C overnight. A total volume of 250 µl homogenized solution was then used for DNA extraction on the QIAasymphony® SP instrument with QIAasymphony® DSP

¹ <https://1212.mn/>

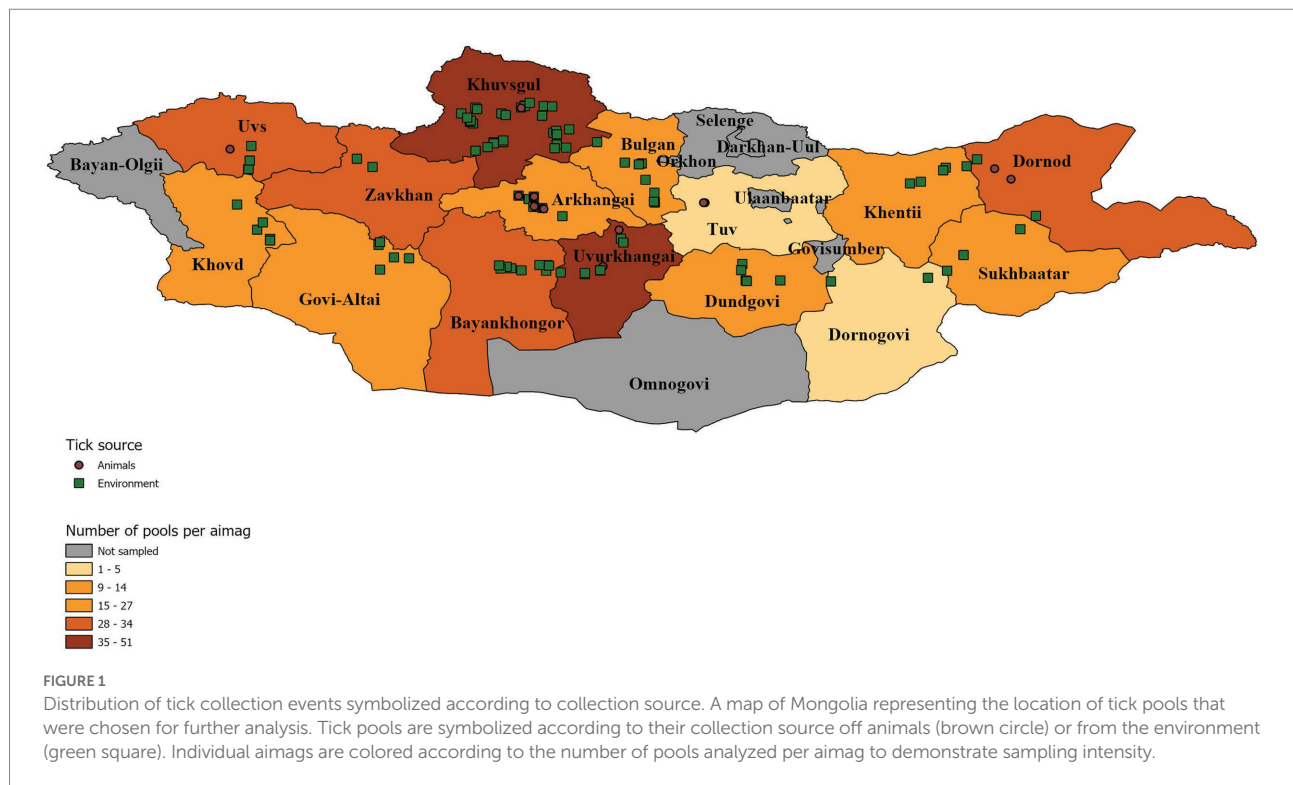


TABLE 1 Samples selected for pathogen screening by next-generation sequencing (NGS) in this study.

Provinces	Number of tick pools	Number of ticks	Number of tick pools selected for NGS	Number of ticks in pools selected for NGS
Arkhangai	61	305	27	135
Bayankhongor	101	496	29	143
Bulgan	14	56	14	56
Dornod	122	599	34	159
Dornogovi	5	18	5	18
Dundgovi	47	226	27	126
Govii-Altai	135	675	24	120
Khentii	8	17	8	17
Khovd	162	810	26	130
Khuvsgul	223	1,050	51	232
Sukhbaatar	140	693	27	133
Tuv	4	21	4	21
Uvs	162	810	29	145
Uvurkhangai	124	594	42	188
Zavhan	181	905	30	150
Total	1,489	7,275	377	1773

TABLE 2 Pool positivity rate by collection source [% and (95% CI)].

Pathogens	Sources (% Infection, 95% CI)			
	Animals (N=51)	Environment (N=307)	Rock and bush (N=19)	Total
<i>Rickettsia</i>	41 [80.4% (69.5, 91.3%)]	275 [89.6% (86.2, 93.0%)]	17 [89.5% (75.7, 103.0%)]	333
<i>Anaplasma</i>	12 [23.5% (11.9, 35.2%)]	0	0	12
<i>Bartonella</i>	1 [1.96% (-1.8, 5.8%)]	2 [0.7% (-0.3, 1.6%)]	0	3

DNA Mini Kit using Tissue LC 200 DSP protocol (Qiagen, Hombrechtikon, Switzerland). The DNA was eluted in 50 µl of ATE buffer and stored at -20°C until use.

Bacterial 16S DNA amplification

Nested PCR was performed as described in Chaorattanakawee et al., 2022 to amplify both the V1-V6 region and V3-V4 region of the bacterial 16S rDNA. Each round of PCR included both an ultrapure DNA/RNA-free water negative control and a mock DNA extraction control. The nested PCR amplicon products were isolated using AMPure XP magnetic beads and the quality of the products was assessed as previously described (Chaorattanakawee et al., 2022). Amplicon products were stored at -20°C until further analysis.

Library preparation and sequencing

The Nextera XT Index Kit v2 (Illumina) was used for index PCR to attach the dual indices and Illumina sequencing adapters to purified 16S amplicons as previously described. Each batch of indexing reactions included a DNA/RNA-free water as a negative control. The index PCR products were cleaned using AMPure XP beads, followed by library purity analysis using the QIAxcel Advanced System (Qiagen). The index libraries were then quantified using the Qubit dsDNA HS Assay Kit (Invitrogen). Libraries were denatured with NaOH according to the manufacturer's protocol (Illumina). Sequencing was performed using the MiSeq Reagent Kit V3 with the Illumina Miseq System. A 10% PhiX internal control (Illumina) was included in each low-diversity library run.

NGS data analysis

Sequence reads produced by the Illumina MiSeq system were processed using the CLC Genomics workbench (v 11.0.1) and CLC microbial genomics module (v 3.0; Qiagen, Aarhus A/S1), which included merging paired reads, primer sequence removal, low read sample removal, and chimeric sequence removal. The filtered sequences were then clustered into operational taxonomic unites (OTUs) using a threshold of 97% sequence identity and the reference OTU database downloaded from the Greengenes database (v 13.8) and SILVA 16S (v 132). Pathogen reads detected in the negative controls represented cross-contamination and were used to subtract respective reads detected in samples.

Pathogen characterization by PCR and sanger sequencing

To confirm the detection of pathogens and the taxonomic assignment as indicated by NGS analysis, PCR and DNA

sequencing were conducted on NGS samples with read counts above a set threshold. The assays and gene targets for selected pathogens (*Anaplasma*, *Bartonella*, *Rickettsia*, *Coxiella*) were detailed previously in Takhampunya et al. (2019). PCR amplification products were cleaned using the ExoSAP-IT kit (Applied Biosystem), followed by cycle-sequencing and sequencing using the SeqStudio Genetic Analyzer (Applied Biosystems), as previously described (Chaorattanakawee et al., 2022). Sample sequences were assembled using Sequencher (v 5.1, Gene Codes Corp.) and aligned with GenBank reference sequences using the MUSCLE codon alignment program. Maximum likelihood phylogenetic trees were constructed for each bacterial target gene using MEGA 6.

Mapping

ArcGIS Pro (v 2.8.0, ESRI) was used for spatial visualizations of data, including tick collection events, tick collection source, and pathogen detection. The map layer of Mongolia and its delineated aimags was accessed from ESRI.²

Statistical analysis

To estimate the probability of pathogen detection within the pooled samples, prevalence rates, maximum likelihood estimates (MLE) and minimum infection rates (MIR) were calculated, which is standard when analyzing pooled tick data. The MLE and MIR estimates were conducted in Excel with the use of the CDC's Mosquito Surveillance Software tool which calculate point and confidence intervals using pooled data that take into account individual pool sample sizes to estimate infection rates.³

Results

Detection of *Rickettsia*, *Anaplasma*, and *Bartonella*

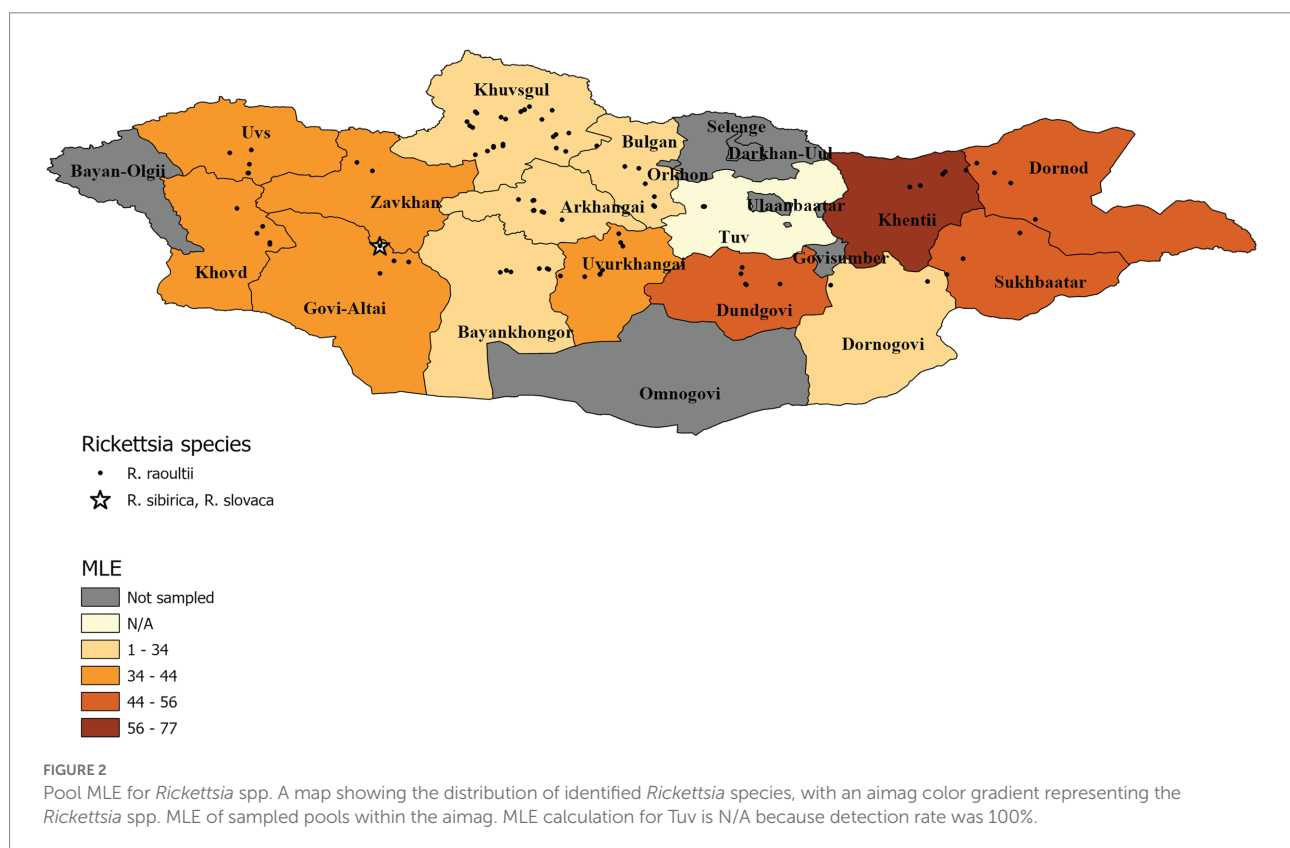
The summary results for the *Rickettsia* spp., confirmed through qPCR analysis and DNA sequencing, are presented in Table 3 and Figure 2. Overall, *Rickettsia* spp. were detected in tick pools from all aimags sampled, with 88% of pools testing positive (333/377). The highest *Rickettsia* spp. pool detection rate was seen in Tuv (100%) followed by: Dornod (97%) Dundgovi (96%) and Sukhbaatar (96%), while the Bulgan aimag showed the lowest pool positivity rate (57%). Maximum likelihood estimates (MLE) found an average prevalence of 37.30% (95%CI: 33.50–41.01%), where

² https://services.arcgis.com/P3ePLMYs2RVChkJs/arcgis/rest/services/MNG_Boundaries_2018/FeatureServer

³ <https://www.cdc.gov/westnile/resourcepages/mosqSurvSoft.html>

TABLE 3 Maximum likelihood estimates of *Rickettsia* spp. by region based on confirmatory results including 95% confidence intervals.*Rickettsia* spp.

Province	Positive pools	Total number of ticks	MLE		
			Point	Low	High
Arkhangai	20/27 (74%)	135	23.7	14.9	33.3
Bayankhongor	24/29 (83%)	143	31.2	20.0	42.4
Bulgan	8/14 (57%)	56	18.3	9.2	30.2
Dornod	33/34 (97%)	159	55.4	34.4	70.7
Dornogovi	4/5 (80%)	18	33.4	12.1	54.9
Dundgovi	26/27 (96%)	126	51.5	31.1	67.1
Govi-Altai	22/24 (92%)	120	39.2	23.7	52.9
Khentii	7/8 (88%)	17	76.6	34.7	94.8
Khovd	24/26 (92%)	130	40.1	24.7	53.7
Khuvsgul	41/51 (80%)	232	31	22.6	39.4
Sukhbaatar	26/27 (96%)	133	48.7	29.2	63.9
Tuv	4/4 (100%)	21	N/A	N/A	N/A
Uvs	27/29 (93%)	145	41.4	26.2	54.7
Uvurkhangai	39/42 (93%)	188	43.9	31	55.1
Zavhan	28/30 (93%)	150	41.8	26.6	55.0
Total	333/377 (88%)	1773	37.3	33.5	41.0



Dornod aimag had the highest MLE of 55.40% (95%CI: 34.38–70.67%) and a MIR of 20.75% (95%CI: 14.45–27.06%) and Bulgan had the lowest MLE of 18.34% (95%CI: 9.21–30.23%) with a MIR

of 14.29% (95%CI: 5.12–23.45%). In general, higher MLEs were found in tick pools collected from eastern and western aimags of Mongolia, with lower MLEs seen in central aimags (Figure 2).

TABLE 4 Maximum likelihood estimates of *Anaplasma* spp. by region based on confirmatory results including 95% confidence intervals.*Anaplasma* spp.

Province	Positive pools	Total number of ticks	MLE		
			Point	Low	High
Arkhangai	9/27 (33%)	135	7.8	4.0	13.7
Bayankhongor	0/29 (0%)	143	–	–	–
Bulgan	0/14 (0%)	56	–	–	–
Dornod	0/34 (0%)	159	–	–	–
Dornogovi	0/5 (0%)	18	–	–	–
Dundgovi	0/27 (0%)	126	–	–	–
Govi-Altai	0/24 (0%)	120	–	–	–
Khentii	0/8 (0%)	17	–	–	–
Khovd	0/26 (0%)	130	–	–	–
Khuvsgul	0/51 (0%)	232	–	–	–
Sukhbaatar	0/27 (0%)	133	–	–	–
Tuv	0/4 (0%)	21	–	–	–
Uvs	2/29 (7%)	145	1.4	0.4	4.8
Uvurkhangai	1/42 (2%)	188	0.5	0.1	2.9
Zavhan	0/30 (0%)	150	–	–	–
Total	12/377 (3%)	1773	0.7	0.4	1.2

TABLE 5 Maximum likelihood estimates of *Bartonella* spp. by region based on confirmatory results including 95% confidence intervals.*Bartonella* spp.

Province	Positive pools	Total number of ticks	MLE		
			Point	Low	High
Arkhangai	3/27 (11%)	135	2.3	0.8	6.4
Bayankhongor	0/29 (0%)	143	–	–	–
Bulgan	0/14 (0%)	56	–	–	–
Dornod	0/34 (0%)	159	–	–	–
Dornogovi	0/5 (0%)	18	–	–	–
Dundgovi	0/27 (0%)	126	–	–	–
Govi-Altai	0/24 (0%)	120	–	–	–
Khentii	0/8 (0%)	17	–	–	–
Khovd	0/26 (0%)	130	–	–	–
Khuvsgul	0/51 (0%)	232	–	–	–
Sukhbaatar	0/27 (0%)	133	–	–	–
Tuv	0/4 (0%)	21	–	–	–
Uvs	0/29 (0%)	145	–	–	–
Uvurkhangai	0/42 (0%)	188	–	–	–
Zavhan	0/30 (0%)	150	–	–	–
Total	3/377 (0.8%)	1773	0.2	0.0	0.5

Summary results for the *Anaplasma* spp., confirmed through PCR and DNA sequencing of the tick pools, are presented in Table 4. Pools were found to have an overall positivity rate of 3.18% for *Anaplasma* spp. (12/377), with only ticks sampled from Arkhangai (33% of pools), Uvs (7% of pools) and Uvurkhangai (2% of pools) testing positive. MLE found an average prevalence of 0.69% (95%CI: 0.39–1.19%), with Arkhangai having the

highest MLE of 7.79% (95%CI: 4.04–13.72%) and MIR of 6.67% (95%CI: 2.46–10.87%). In contrast, Uvurkhangai had an MLE of 0.53% (95%CI: 0.09–2.87%) with a MIR of 0.53% (95%CI: 0–1.57%).

Table 5 summarizes the results for the *Bartonella* spp., confirmed through PCR and DNA sequencing of the tick pools. The overall *Bartonella* spp. pool positivity rate was found to be 0.79% (3/377),

TABLE 6 Identity and Genbank accession numbers for *Anaplasma*, *Bartonella* and *Rickettsia* spp. from pooled samples *Dermacentor* spp.

Target Gene	Organism	Location	GenBank#	Identity		
gltA	<i>Bartonella schoenbuchii</i>	Arkhangai	OM281134-OM281135	99.21% AJ564635.1		
	<i>B. melophagi</i>	Arkhangai	OM281136	100% AY692475.1		
	<i>Rickettsia raoultii</i>	Arkhangai, Bayankhongor, Dornogovi, Dornod, Dundgovi, Govi-Altai, Khentii, Khovd, Khuvsgul, Sukhbaatar, Tuv, Uvs, Uvurkhangai, Zavhan	OM281112; OM281162; OM281137-OM281146; OM281148-OM281155; OM281157; OM281160-OM281168; OM281170-OM281171; OM281173-OM281177; OM281179-OM281185; OM281187-OM281192; KU961538;	100% MT178337.1		
		<i>R. raoultii</i>	Dornod, Khovd, Khuvsgul, Uvs	OM281156; OM281158; OM281159; OM281169; OM281178; OM281186	100% OK638145.1	
		<i>R. sibirica/R. slovacica</i>	Govi-Altai	OM281147	100% MG811709.1;	
		<i>ompA</i>	<i>R. raoultii</i>	Bayankhongor, Dornod, Govi-Altai, Khentii, Khovd, Khuvsgul, Sukhbaatar, Uvs, Uvurkhangai	OM281193-OM281217	100% MK726326.1
	16S rRNA	<i>Anaplasma ovis</i>	Arkhangai	OM320148-OM320155	100% MN266936.1	
		<i>A. capra</i>				
		<i>A. centrale</i>				
<i>A. marginale</i>						
<i>A. ovis</i>		Uvs	OM320157	100% MN266936.1		
groEL	<i>A. ovis</i>	Arkhangai, Uvs	OM281118-OM281120; OM281122-OM281128	99.69% MT268377.1		
	<i>A. ovis, A. centrale, A. marginale</i>	Arkhangai	OM28121	99.39% MT268375.1; 92.05 KY305559.1		
	<i>A. ovis</i>	Uvs	OM281232	100% MH292916.1		
	<i>A. ovis</i>	Arkhangai	OM281229-OM281231	100% MH292916.1		
	16S rRNA	Coxiella endosymbiont of <i>Dermacentor marginatus</i>	Arkhangai, Dornod, Dornogovi, Khentii, Khuvsgul, Sukhbaatar, Uvurkhangai	OM333168-OM333184	99.44% MZ047981.1	

with Arkhangai being the only region with positive pools (3/27 pools). Maximum likelihood estimates (MLE) found an overall prevalence of 0.17% (95% CI: 0–0.06–0.50%). Arkhangai had a MLE of 2.33% (95% CI: 0.78–6.37%) and a MIR of 2.22% (95% CI: 0–4.71%). A full list of sequence accession numbers by gene target and microorganism can be found in Table 6.

Pathogen detection by tick source

The pathogen pool positivity rate by tick collection source is detailed in Table 2. *Rickettsia* spp. was detected in 80.4% (95% CI 69.5, 91.3) of tick pools removed from livestock animals, with Tuv having the highest pool positivity rate (100%) and Khuvsgul having the lowest pool positivity rate (33.3%; Table 3). The *Rickettsia* spp. infection rate in ticks collected from different sources (animal vs. environment including from rock and bush) was compared using Chi-square test and no significant difference was found (Chi-square = 2.7685, $df = 1$, value of $p = 0.09614$). Of note, *Anaplasma* spp. was only detected in tick pools collected from animals, with a pool positivity rate of 23.5% (95% CI 11.9, 35.2). Arkhangai had the highest level of pool positivity, with 47.4% of tick pools collected from animals having *Anaplasma* spp. DNA present (Table 1). *Bartonella* spp. was detected in 1.96% (95%

CI -1.8, 5.8) of pools of ticks removed from animals and 0.7% (95% CI -0.3, 1.6) of tick pools collected from the environment. As discussed above, all three pools testing positive for *Bartonella* spp. came from the Arkhangai aimag, with one pool representing ticks collected from animals (5.26% of animal tick pools from Arkhangai (Table 2), and the other two pools being ticks collected from the environment (25% of environmental tick pools from Arkhangai).

Pathogen species confirmation

DNA sequencing allowed for pathogen species confirmation of pools testing positive for the various bacterial groups. Within *Rickettsia* spp. positive tick pools, the *gltA* and *ompA* sequences were analyzed, with the summarizing maximum likelihood (ML) tree presented in Figure 3. As shown in the ML tree, most pools were identified as having *R. raoultii* ($n = 332$), although one environmental tick pool from Govi-Altai had a pathogen identified as *R. sibirica/Rickettsia slovacica* (100% sequence identity). *Anaplasma* species were identified by analyses of both the 16S rDNA and *groEL*, placing *Anaplasma*-positive pools within the *A. capra*, *A. centrale*, *A. marginale*, and *A. ovis* group, with all *Anaplasma*-positive pools eventually being grouped within the *A. ovis* group ($n = 12$, Figure 4). Finally, ML *gltA* gene

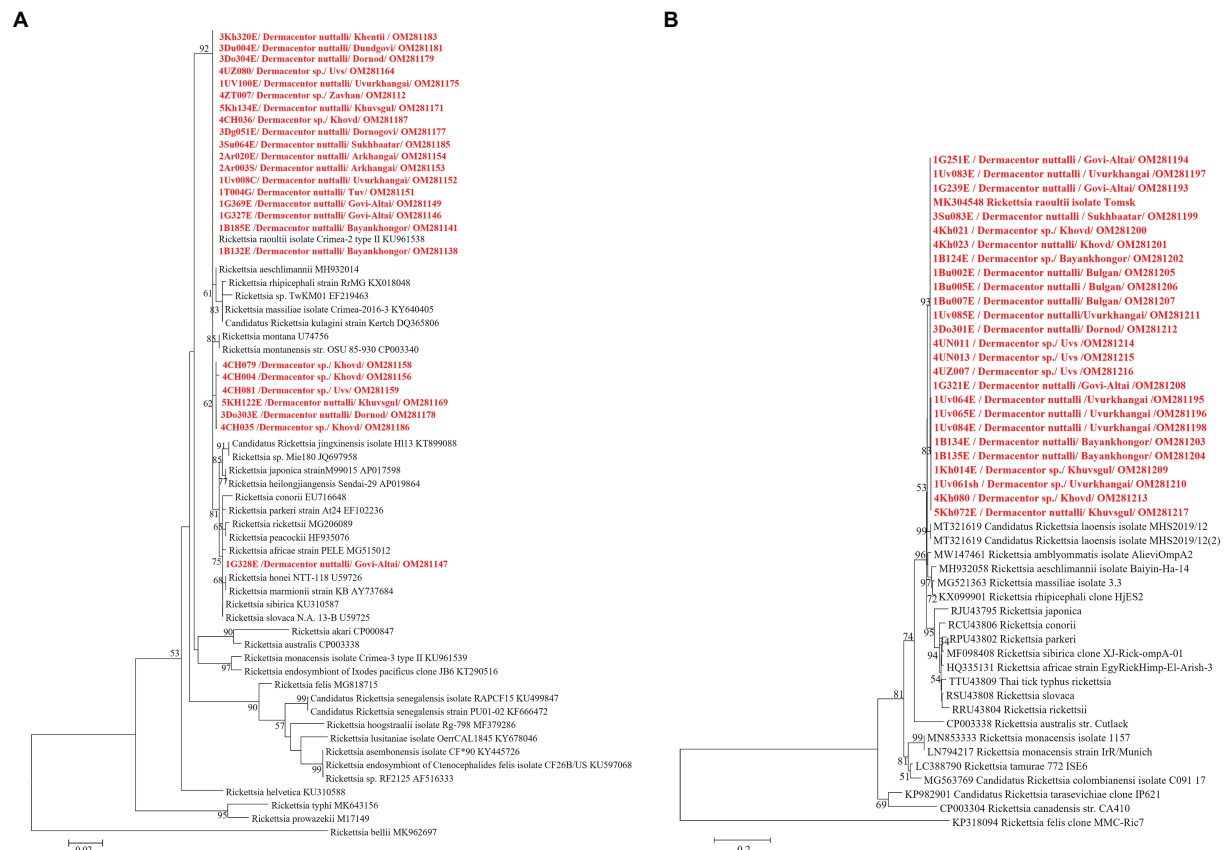


FIGURE 3

Maximum likelihood (ML) tree was constructed from *gltA* gene (A) and *ompA* gene (B) of *Rickettsia* spp. using T92+G model with 1,000 bootstrap replicates (>50% are shown on each node). Sequences of tick samples in this study are shown in red letters.

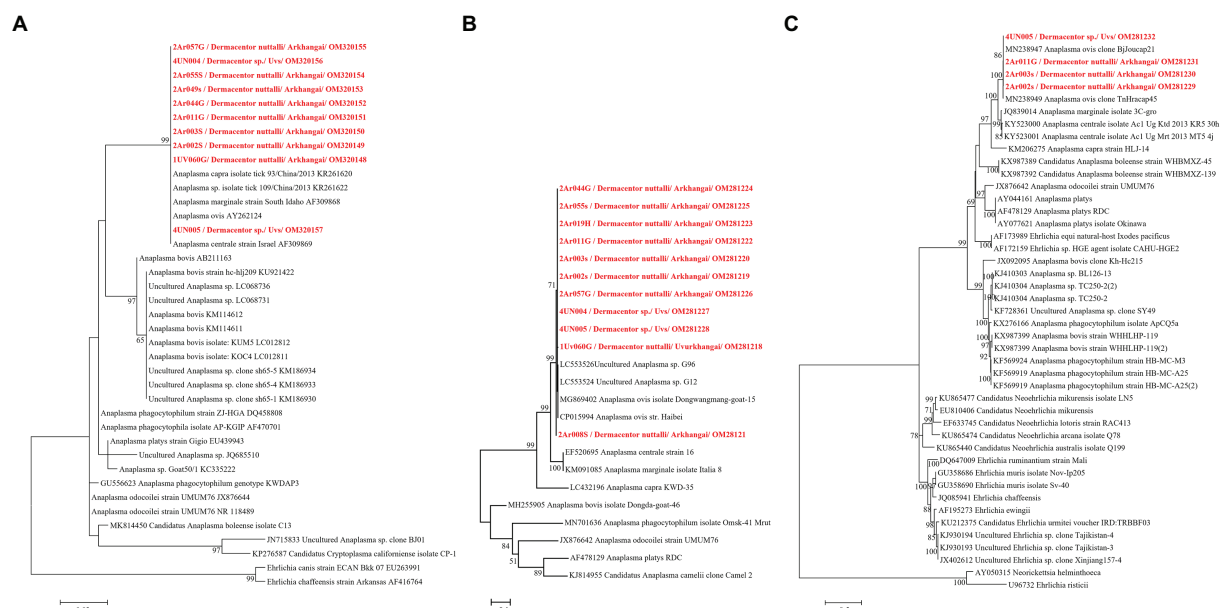


FIGURE 4

ML tree was constructed from 16S rRNA (A) and *groEL* genes using K9+G (B) and TN93+G+I (C) models, respectively, with 1,000 bootstrap and value over 50% are indicated on each node.

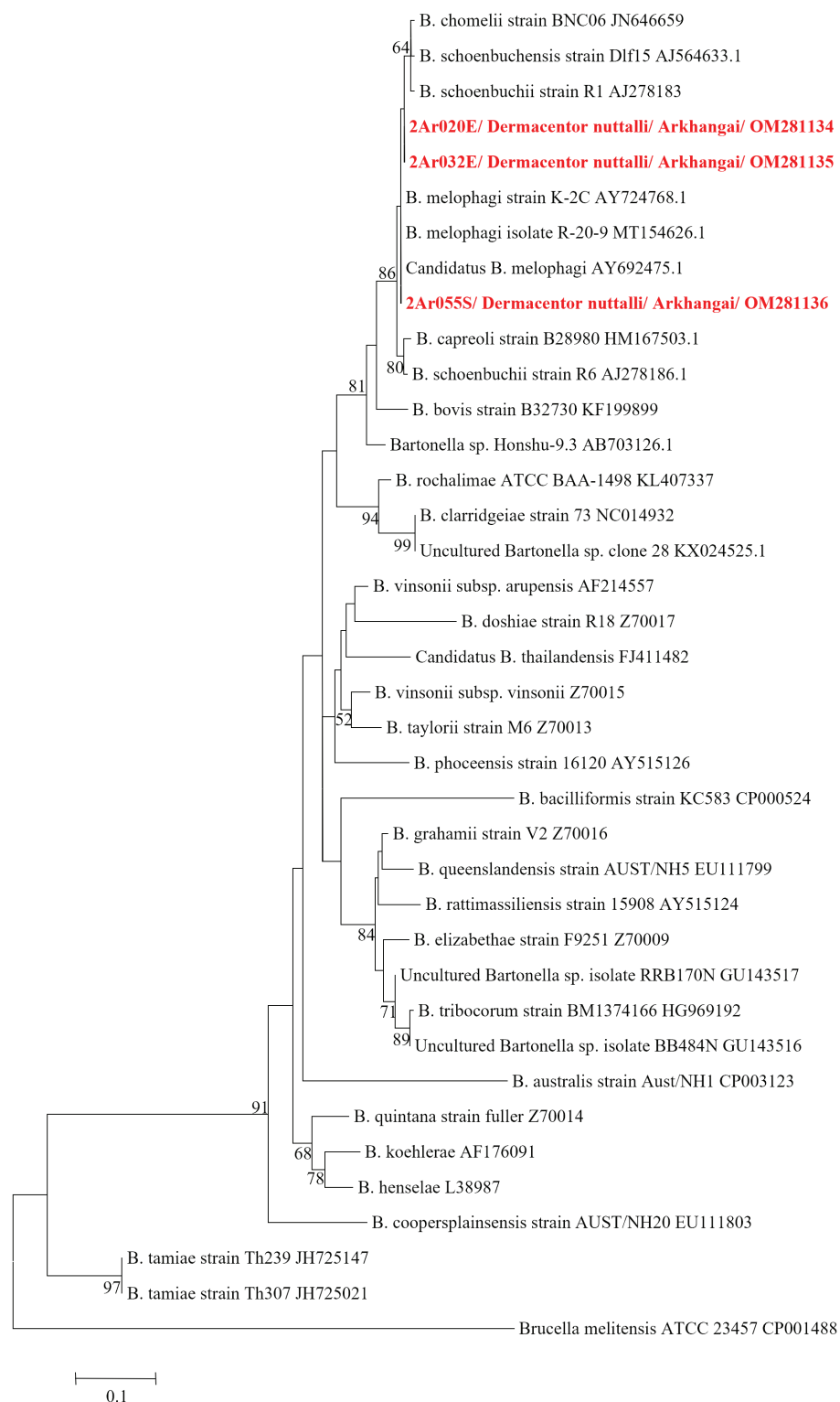
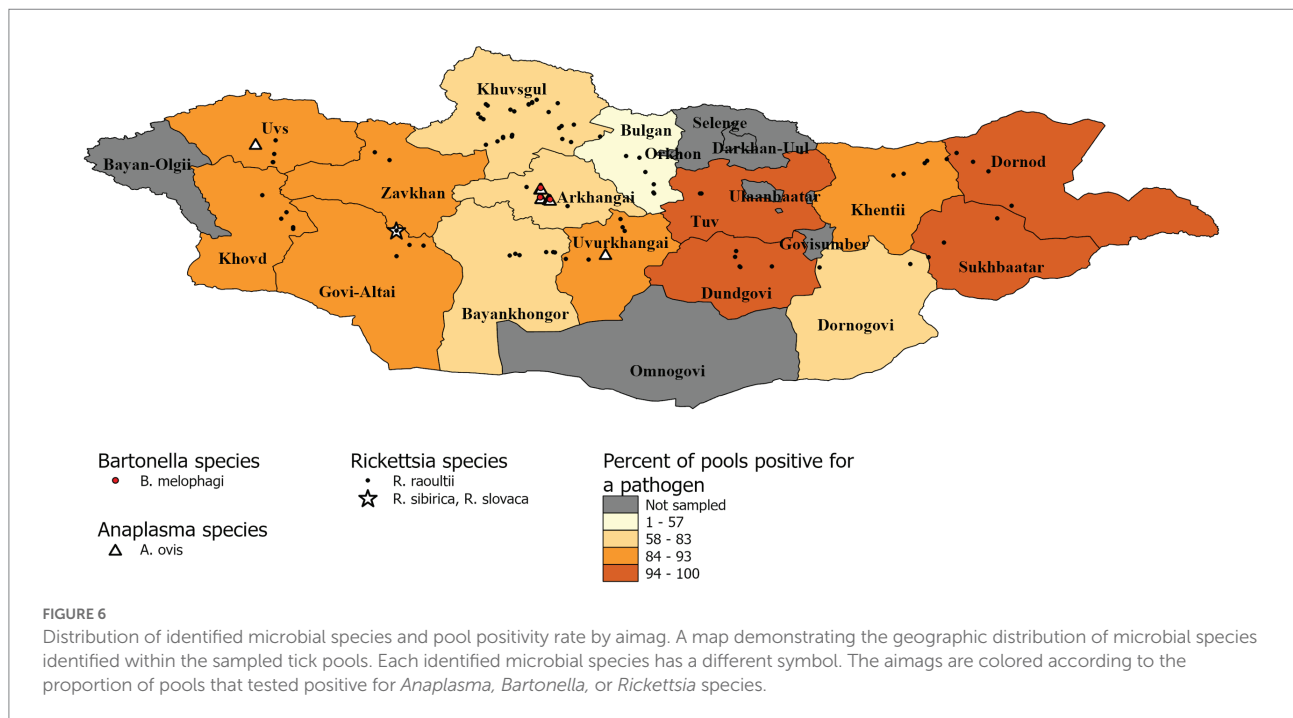


FIGURE 5

ML tree constructed from *gltA* gene of *Bartonella* spp. using T92+G model with 1,000 bootstrap replicates (>50% value are shown on each node). Sequence of tick samples in this study are indicated in red letters.

analysis of the three pools that tested positive for *Bartonella* spp. identified the species as *Bartonella melophagi* (Figure 5). Figure 6 summarizes the geographic distribution of identified microbial

species as well as the proportion of pools within an aimag that tested positive for *Anaplasma*, *Bartonella*, or *Rickettsia* species. A higher proportion of tick pools tested positive for a pathogen in



the eastern and western part of Mongolia, which was largely driven by high detection rates of *R. raoultii*.

Discussion

This study continues previous work describing the microbial diversity found within *Dermacentor* ticks of Mongolia, applying next-generation sequencing to ticks collected from a wider geographic range. Findings from this study reiterate that *Rickettsia* spp., specifically *R. raoultii*, are highly prevalent across Mongolia, with aimag pools having a positivity rate of above 50%. While previous work has documented high pool positivity levels for *Rickettsia* spp. in *Dermacentor* ticks from southern and central aimags of Mongolia (Fischer et al., 2020; von Fricken et al., 2020b), this study has a much wider geographic range and represents, to our knowledge, the first time NGS methods have been applied to testing *Dermacentor* ticks in Mongolia. The finding of higher MLEs and pool-positivity rates in the eastern part of Mongolia may indicate a potential hotspot for *Dermacentor* tick-related *Rickettsia* spp. exposure, warranting future human and animal serological studies in these areas. Of note, the total MLE in this study (37.30%; 95% CI: 33.50–41.01) is similar to a previous work that sampled ticks from five southern aimags, where the MLE for *Dermacentor* spp. was 33.2% (95% CI: 30.1–36.2; von Fricken et al., 2020b). Additionally, zero larvae and few nymphs were found across 191 geographically distinct collection events spread out through 15 aimags, which we believe is suggestive of *Dermacentor* ticks spending earlier life cycle stages underground in rodent burrows, given harsh dry winter seasons common in Mongolia. This theory is also supported by

high levels of *Rickettsia* detection in rodent reservoirs, where 17/18 *Meriones unguiculatus* (Mongolian Gerbil) tested positive for *Rickettsia* DNA, with this rodent commonly found across Mongolia (Pulscher et al., 2018). When paired with evidence of transovarial transmission of *Rickettsia* in *Dermacentor* ticks from Mongolia (Moore et al., 2018), it is not surprising to observe such high detection rates across this wide geographic range. The *Rickettsia* species identified through NGS analysis include *R. raoultii* and *R. sibirica/R. slovaca*, which aligns with previous reports of *R. raoultii* and *R. sibirica* from *Dermacentor* spp. collected from the Omnogovi, Dornogovi, Govi-Altai, Khovd, Khentii and Bayankhongor aimags (Fischer et al., 2020; von Fricken et al., 2020b). Infections with *R. raoultii* typically manifest with eschars and lymphadenopathy, although severe cases have been reported with pulmonary edema (Li et al., 2018). Similarly, *R. sibirica* subspecies present with non-specific flu-like symptoms accompanied by a rash and eschar, with more severe complications such as disseminated intravascular coagulation, renal failure, and neurological symptoms (Nouchi et al., 2018). Therefore, the detection of these pathogens across a wide geographic distribution of ticks within Mongolia represent a major public health threat that is likely under reported in pastoral communities, due to limited access to healthcare in rural regions and low treatment-seeking behaviors within this population (Lkhagvatseren et al., 2019).

While only seen in three aimags (Arkhangai, Uvs, and Uvurkhangai), *Anaplasma* spp. was still detected in 3.18% of pools overall, eventually being identified as *A. ovis*. This *Anaplasma* species causes anaplasmosis in sheep, goats, and wildlife ruminants, often characterized as a subclinical disease which can lead to reduced milk production and spontaneous abortions

(Cabezas-Cruz et al., 2019). Of note, all positive pools for *Anaplasma* spp. came from ticks removed from livestock, primarily sheep and goat, with zero detections occurring in environmental samples. A previous study of *Anaplasma* spp. and *Ehrlichia* spp. within ticks collected from central Mongolia reported a similar pattern of high MLE rates when ticks collected from animals were considered separately from ticks collected from the environment (Shao et al., 2020; von Fricken et al., 2020a). This pattern of pathogen distribution within ticks may result from ticks taking partial blood meals from infected livestock hosts and then detaching and reattaching to other livestock hosts, spreading the disease within herds in the process. We do not believe rodents or transovarial transmission plays a significant role in *A. ovis* transmission cycles given the absence of detection in such a large sample. These findings highlight an important component of *Anaplasma* spp. disease ecology where livestock act as amplifying hosts. Given these observed patterns of *Anaplasma* spp. transmission within Mongolia, the lack of *Anaplasma* spp. detection in many of the aimags in this study should not be interpreted as an actual absence of this microbe group, as many of the pools within these *Anaplasma*-negative aimags were primarily pools collected from the environment. Further investigation into co-feeding transmission between ticks and potential vectors is warranted, when paired with the high seroprevalence and pathogen detection found in previous studies screening livestock (Zhang et al., 2008; Jiao et al., 2021).

Bartonella spp. are typically transmitted by fleas and lice, however there is an ongoing larger discussion about what potential role ticks play within transmission cycles (Angelakis et al., 2010; Cheslock and Embers, 2019). Studies that have suggested the possible role of ticks as vectors of *Bartonella* spp. include the reported presence of *Bartonella* spp. DNA within various tick groups collected from around the world, including *Dermacentor* ticks, and epidemiological studies have often noted tick bites preceding *Bartonella* spp.-related illnesses (Wikswa et al., 2007; Angelakis et al., 2010; Zając et al., 2015). Experimental data supporting tick-mediated transmission of *Bartonella* spp. is limited, but includes the ability of *Bartonella* spp. to replicate within multiple tick species cell lines (Billeter et al., 2009), the ability of *Bartonella*-infected ticks to transmit *Bartonella* infection between animal models while feeding (Noguchi, 1926; Reis et al., 2011), and the detection of *Bartonella* spp. DNA or bacilli in infected tick midgut, salivary glands, and feces (Cotté et al., 2008; Reis et al., 2011; Wechtaison et al., 2021). Despite these findings, there is still a lack of consensus regarding the ability of ticks to vector *Bartonella* species. Regardless of whether ticks play a role in transmitting *Bartonella* in Mongolia, here we provide further evidence that *Bartonella melophagi* is present in Mongolia. Within this study, *Bartonella melophagi* was found in 11.1% of tick pools from the Arkhangai aimag, including a tick pool collected from sheep. Sheep are considered the reservoir host for *B. melophagi*, with sheep keds being the common insect vector of this pathogen (Maggi et al., 2009). Human infections with this microbe have been reported, resulting in flu-like symptoms, bite-site lesion,

neurological symptoms, and heart irregularities (Maggi et al., 2009). Although our group has previously reported *B. melophagi* within Mongolian sheep, to our knowledge, this is the first report of *B. melophagi* within ticks collected from Mongolia (Chaorattanakawee et al., 2022). Of note, this *Bartonella* species has been reported in *Dermacentor*, *Hyalomma*, and *Rhipicephalus* ticks collected from Xinjiang, China (Ni et al., 2021). The detection of *B. melophagi* in both domestic animals and ticks of Mongolia emphasizes a need for future studies to characterize the disease ecology of this pathogen, including determining the role of ticks in disease transmission and the possibility of transboundary disease movement between Mongolia and China.

All tick pools analyzed by qPCR and conventional PCR targeting the *Coxiella burnetii* transposase gene were negative, suggesting the absence of this pathogen from the ticks sampled. Of note, 35.8% of pools tested positive for a *Coxiella*-like bacteria, warranting further investigation.

This study illustrates the utility of NGS in characterizing the diversity of tick-borne pathogens found in *Dermacentor* ticks collected from geographically distinct locations. We applied NGS to receive a “snapshot” of the various bacterial groups present within tick pools, which then guided confirmatory assays to allow for accurate identification of tick-borne pathogen species. Given the large number of tick-borne pathogens present within Mongolia, the use of a nontargeting analytical method is appropriate to avoid unintentionally excluding the detection of certain microbial species which might be missed relying on other detection processes. Importantly, certain results of this study corroborate what has previously been reported concerning the epidemiology of tick-borne pathogens within Mongolia. This includes the high prevalence of *Rickettsia* spp., particularly *R. raoultii*, among *Dermacentor* ticks collected across the country (Fischer et al., 2020; von Fricken et al., 2020b). We also report the detection of *A. ovis* exclusively in ticks removed from livestock, which is in agreement with a previous study demonstrating higher levels of *Anaplasma* spp. in ticks collected from animals (Shao et al., 2020; von Fricken et al., 2020a). While this study expands on the knowledge concerning the geographical distribution of *Rickettsia* spp. and *Anaplasma* spp. within *Dermacentor* ticks, the observation that many of the results are in agreement with previous studies indicates that NGS offers a valid, novel approach for the characterization of tick-borne pathogens. Of note, the novel detection of *Bartonella* spp. DNA within *Dermacentor* spp. ticks collected within Mongolia also demonstrates the ability of NGS to discover new pathogen-vector relationships, which may be more difficult to detect using other molecular processes. Future use of NGS to describe the microbial diversity found in the various tick species from Mongolia will further contribute to a more complete characterization of the tick-borne pathogens in circulation within the country. The results from this study will contribute to more detailed risk mapping for tick-borne pathogens, which will help inform disease prevention interventions that benefit populations at increased risk of disease exposure. The abilities of NGS to identify novel vector-pathogen

associations will also prove to be vital for local health care practitioners by informing them on the various tick-borne diseases they should include as part of differential diagnoses strategies.

The use of NGS in epidemiological surveys of vector species such as ticks has many advantages over pathogen detection methods that are typically used in such studies, such as PCR and immunofluorescence. While this study focused primarily on clinically relevant pathogens, sequencing data resulting from NGS allows for the creation of a library of microbial sequences, which can promote the tracking of microbial evolution overtime (Deurenberg et al., 2017). Importantly, by surpassing the need for selecting pathogen-specific molecular probes, the use of NGS streamlines the ability of groups to rapidly identify uncommon or previously uncharacterized microbial agents, which may represent emergent diseases (Wu et al., 2021).

Limitations

In this study, ticks were only identified morphologically, which limits our ability to infer findings beyond the genus level. The reliance on morphological identification of ticks may have led to misidentification of the tick species analyzed in this study, hence the decision to keep most of our discussion at the genus level. The decision to pool ticks within this study also introduces some limitations, including difficulties in determining the true prevalence of the various microbial agents that were detected. For example, in instances in which 100% of tick pools are positive, it is not possible to calculate a maximum likelihood estimate, which is what occurred for the Tuv aimag pools for our *Rickettsia* results. Although there was one pool in which all three pathogens were detected and 12 pools in which two pathogens were detected, discussion of co-infection status of ticks is also complicated by pooling of ticks. While the use of NGS may prove useful for the characterization of pathogenic microbes within ticks and other vector insect species, it is important to note that the detection of microbial DNA does not necessarily indicate the presence of viable microbial organisms within the tick sample. The detection of microbial DNA may also represent remnant DNA from a recent bloodmeal (Tokarz et al., 2019). Therefore, caution must be taken when interpreting NGS results from blood-feeding arthropods.

Conclusion

Here we report the use of NGS to assess the diversity of pathogens within *Dermacentor* ticks collected from 15 different aimags of Mongolia. The results of this study highlight a high level of *Rickettsia* detected across all sampled aimags, including the presence of *R. sibirica*/*R. slovaca* in Govi-Altai, as well as detections of *A. ovis* in samples removed from livestock. These findings also

highlight the first reported detection of *B. melophagi* in ticks from this region. Future studies should make use of NGS analysis to further characterize the diversity of pathogens found in other medically relevant tick species within Mongolia, as this method allows for the detection of multiple pathogens simultaneously.

Author's note

The material in this manuscript has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author, and are not to be construed as official, or as reflecting true views of the Department of the Army, Department of the Navy, the Department of Defense, or the US Government. Multiple authors are military service members and/or federal/contracted employees of the United States government. The work prepared by DoD personnel in this manuscript was prepared as part of their official duties. Title 17 U.S.C. 105 provides that 'copyright protection under this title is not available for any work of the United States Government.' Title 17 U.S.C. 101 defines a United States Government work as work prepared by a military service member or employee of the United States Government as part of that person's official duties.

Data availability statement

The data presented in the study are deposited in the NCBI GenBank repository and can be found at: <https://www.ncbi.nlm.nih.gov/genbank/>, OM281134-OM281136; OM281112; OM281162; OM281137-OM281146; OM281148 -OM281155; OM281157; OM281160-OM281168; OM281170-OM281171; OM281173 -OM281177; OM281179-OM281185; OM281187 -OM281192; OM281156; OM281158; OM281159; OM281169; OM281178; OM281186; OM281147; OM281193-OM281217; OM320148 -OM320155; OM320157; OM281118-OM281120; OM281122 -OM281128; OM28121; OM281232; OM281229 -OM281231; OM333168-OM333184.

Author contributions

DA: methodology, formal analysis, sample collection, investigation, and writing—original draft. AL and GM: formal analysis, data visualization, and writing—original draft. RT: methodology, formal analysis, investigation, data curation, writing—review and editing, and visualization. JS and NC: methodology, formal analysis, and investigation. BP-S, JF, and EL: conceptualization, writing—review and editing, supervision, project administration, and funding acquisition. BBd: investigation, writing—review and editing, and project administration. SD and JH: sample collection, conceptualization, and funding acquisition. BBr: morphological identification and

sample collection. NT: writing—review and editing, supervision, and project administration. MF: conceptualization, formal analysis, investigation, data curation, writing—original draft, visualization, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This work was supported/funded by work unit number D0016 with funding provided by the Armed Forces Health Surveillance Division (AFHSD) Global Emerging Infections Surveillance (GEIS) Branch, ProMIS ID # PO133_19_AF_N2. Lab analysis was funded by the AFHSD-GEIS under study #P0128_20_AF_14.

Acknowledgments

We would like to thank Mongolian counterparts at the National Center for Zoonotic Disease satellite branches for

assisting with sample collection. We also acknowledge the staff from the Diagnostic and Reemerging Disease section of the Entomology Department, AFRIMS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Angelakis, E., Billeter, S. A., Breitschwerdt, E. B., Chomel, B. B., and Raoult, D. (2010). Potential for tick-borne Bartonellosis. *Emerg. Infect. Dis.* 16, 385–391. doi: 10.3201/eid1603.091685
- Aung, A. K., Spelman, D. W., Murray, R. J., and Graves, S. (2014). Rickettsial infections in Southeast Asia: implications for local populace and febrile returned travelers. *Am. J. Trop. Med. Hyg.* 91, 451–460. doi: 10.4269/ajtmh.14-0191
- Barnes, A. N., Baasandavga, U., Davaasuren, A., Gonchigoo, B., and Gray, G. C. (2020). Knowledge and practices surrounding zoonotic disease among Mongolian herding households. *Pastoralism* 10:8. doi: 10.1186/s13570-020-00162-5
- Battsetseg, B., Xuan, X., Ikadai, H., Bautista, J. L., Byambaa, B., Boldbaatar, D., et al. (2001). Detection of *Babesia caballi* and *Babesia equi* in *Dermacentor nuttalli* adult ticks. *Int. J. Parasitol.* 31, 384–386. doi: 10.1016/s0020-7519(01)00120-5
- Biggs, H. M. (2016). Diagnosis and management of tickborne rickettsial diseases: Rocky Mountain spotted fever and other spotted fever group Rickettsioses, Ehrlichioses, and Anaplasmosis—United States. *MMWR Recomm. Rep.* 65, 1–44. doi: 10.15585/mmwr.rr6502a1
- Billeter, S. A., Diniz, P. P. V. P., Battisti, J. M., Munderloh, U. G., Breitschwerdt, E. B., and Levy, M. G. (2009). Infection and replication of *Bartonella* species within a tick cell line. *Exp. Appl. Acarol.* 49, 193–208. doi: 10.1007/s10493-009-9255-1
- Boldbaatar, D., and Byambaa, B. (2015). *Bloodsucking Ticks of Mongolia*. Ulaanbaatar, Mongolia.
- Boldbaatar, B., Jiang, R.-R., von Fricken, M. E., Lkhagvatseren, S., Nymadawa, P., Baigalmaa, B., et al. (2017). Distribution and molecular characteristics of *rickettsiae* found in ticks across Central Mongolia. *Parasite Vector* 10:61. doi: 10.1186/s13071-017-1981-3
- Cabezas-Cruz, A., Gallois, M., Fontugne, M., Allain, E., Denoual, M., Moutailler, S., et al. (2019). Epidemiology and genetic diversity of *Anaplasma ovis* in goats in Corsica. *France. Parasite Vector* 12:3. doi: 10.1186/s13071-018-3269-7
- Černý, J., Buyannemekh, B., Needham, T., Gankhuyag, G., and Oyuntsetseg, D. (2019). Hard ticks and tick-borne pathogens in Mongolia—a review. *Ticks Tick-Borne Dis.* 10:101268. doi: 10.1016/j.ttbdis.2019.101268
- Chaorattanakawee, S., Wofford, R. N., Takhampunya, R., Katherine Poole-Smith, B., Boldbaatar, B., Lkhagvatseren, S., et al. (2022). Tracking tick-borne diseases in Mongolian livestock using next generation sequencing (NGS). *Ticks Tick-Borne Dis.* 13:101845. doi: 10.1016/j.ttbdis.2021.101845
- Cheslock, M. A., and Embers, M. E. (2019). Human bartonellosis: an underappreciated public health problem? *Trop. Med. Infect. Dis.* 4:69. doi: 10.3390/tropicalmed4020069
- Cotté, V., Bonnet, S., Le Rhun, D., Le Naour, E., Chauvin, A., Boulouis, H.-J., et al. (2008). Transmission of *Bartonella henselae* by *Ixodes ricinus*. *Emerg. Infect. Dis.* 14, 1074–1080. doi: 10.3201/eid1407.071110
- Deurenberg, R. H., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., García-Cobos, S., et al. (2017). Application of next generation sequencing in clinical microbiology and infection prevention. *J. Biotechnol.* 243, 16–24. doi: 10.1016/j.jbiotec.2016.12.022
- Enkhtaivan, B., Narantsatsral, S., Davaasuren, B., Otgonsuren, D., Amgalanbaatar, T., Uuganbayar, E., et al. (2019). Molecular detection of *Anaplasma ovis* in small ruminants and ixodid ticks from Mongolia. *Parasitol. Int.* 69, 47–53. doi: 10.1016/j.parint.2018.11.004
- Fischer, T., Myalkhaa, M., Krücken, J., Battsetseg, G., Batsukh, Z., Baumann, M. P. O., et al. (2020). Molecular detection of tick-borne pathogens in bovine blood and ticks from Khentii. *Mongolia. Transbound Emerg. Dis.* 67, 111–118. doi: 10.1111/tbed.13315
- Gao, Y., Lv, X.-L., Han, S.-Z., Wang, W., Liu, Q., and Song, M. (2021). First detection of *Borrelia miyamotoi* infections in ticks and humans from the northeast of Inner Mongolia. *China. Acta Trop.* 217:105857. doi: 10.1016/j.actatropica.2021.105857
- He, X., Zhao, J., Fu, S., Yao, L., Gao, X., Liu, Y., et al. (2018). Complete genomic characterization of three tick-borne encephalitis viruses detected along the China-North Korea border, 2011. *Vector-Borne Zoonot.* 18, 554–559. doi: 10.1089/vbz.2017.2173
- Huang, T., Zhang, J., Sun, C., Liu, Z., Haiyan, H., Wu, J., et al. (2020). A novel arthropod host of brucellosis in the arid steppe ecosystem. *Front. Vet. Sci.* 7:566253. doi: 10.3389/fvets.2020.566253
- Javkhan, G., Enkhtaivan, B., Baigal, B., Myagmarsuren, P., Battur, B., and Battsetseg, B. (2014). Natural *Anaplasma phagocytophilum* infection in ticks from a forest area of Selenge province. *Mongolia. Western Pacific Surv. Resp J.* 5, 21–24. doi: 10.5365/WPSAR.2013.4.3.001
- Jiao, J., Lu, Z., Yu, Y., Ou, Y., Fu, M., Zhao, Y., et al. (2021). Identification of tick-borne pathogens by metagenomic next-generation sequencing in *Dermacentor nuttalli* and *Ixodes persulcatus* in Inner Mongolia. *China. Parasite Vector* 14:287. doi: 10.1186/s13071-021-04740-3
- Li, H., Zhang, P.-H., Huang, Y., Du, J., Cui, N., Yang, Z.-D., et al. (2018). Isolation and identification of *Rickettsia raoultii* in human cases: a surveillance study in 3 medical centers in China. *Clin. Infect. Dis.* 66, 1109–1115. doi: 10.1093/cid/cix917
- Lkhagvatseren, S., Hogan, K. M., Boldbaatar, B., von Fricken, M. E., Anderson, B. D., Pulscher, L. A., et al. (2019). Discrepancies between self-reported tick bites and evidence of tick-borne disease exposure among nomadic Mongolian herders. *Zoonoses Public Hlth.* 66, 480–486. doi: 10.1111/zph.12579
- Maggi, R. G., Kosoy, M., Mintzer, M., and Breitschwerdt, E. B. (2009). Isolation of *Candidatus Bartonella melophagi* from human blood. *Emerg. Infect. Dis.* 15, 66–68. doi: 10.3201/eid1501.081080

- Moore, T. C., Pulscher, L. A., Caddell, L., von Fricken, M. E., Anderson, B. D., Gonchigoo, B., et al. (2018). Evidence for transovarial transmission of tick-borne *rickettsiae* circulating in northern Mongolia. *PLOS Neglect. Trop. Dis.* 12:e0006696. doi: 10.1371/journal.pntd.0006696
- Ni, J., Ren, Q., Lin, H., Aizezi, M., Luo, J., Luo, Y., et al. (2021). Molecular evidence of *Bartonella melophagi* in ticks in border areas of Xinjiang. *China. Front. Vet. Sci.* 8:675457. doi: 10.3389/fvets.2021.675457
- Noguchi, H. (1926). Etiology of Oroya fever. *J. Exp. Med.* 44, 729–734. doi: 10.1084/jem.44.5.729
- Nouchi, A., Monsel, G., Jaspard, M., Jannic, A., Angelakis, E., and Caumes, E. (2018). *Rickettsia sibirica mongolitimonae* infection in a woman travelling from Cameroon: a case report and review of the literature. *J. Travel Med.* 25:tax074. doi: 10.1093/jtm/tax074
- Ochirkhuu, N., Konnai, S., Odbileg, R., Murata, S., and Ohashi, K. (2017). Molecular epidemiological survey and genetic characterization of *Anaplasma* species in Mongolian livestock. *Vector-Borne Zoonot.* 17, 539–549. doi: 10.1089/vbz.2017.2111
- Odontsetseg, N., Uuganbayar, D., Tserendorj, S., and Adiyasuren, Z. (2009). Animal and human rabies in Mongolia. *Rev Sci tech. OIE.* 28, 995–1003. doi: 10.20506/rst.28.3.1942
- Pulscher, L. A., Moore, T. C., Caddell, L., Sukhbaatar, L., von Fricken, M. E., Anderson, B. D., et al. (2018). A cross-sectional study of small mammals for tick-borne pathogen infection in northern Mongolia. *Infect. Ecol. Epidemiol.* 8:1450591. doi: 10.1080/2008686.2018.1450591
- Reis, C., Cote, M., Le Rhun, D., Lecuelle, B., Levin, M. L., Vayssier-Taussat, M., et al. (2011). Vector competence of the tick *Ixodes ricinus* for transmission of *Bartonella birtlesii*. *PLoS Negl. Trop. Dis.* 5:e1186. doi: 10.1371/journal.pntd.0001186
- Shao, J.-W., Zhang, X.-L., Li, W.-J., Huang, H.-L., and Yan, J. (2020). Distribution and molecular characterization of *rickettsiae* in ticks in Harbin area of northeastern China. *PLoS Negl. Trop. Dis.* 14:e0008342. doi: 10.1371/journal.pntd.0008342
- Takhampunya, R., Korkusol, A., Pongpichit, C., Yodin, K., Rungroj, A., Chanarat, N., et al. (2019). Metagenomic approach to characterizing disease epidemiology in a disease-endemic environment in northern Thailand. *Front. Microbiol.* 10:319. doi: 10.3389/fmicb.2019.00319
- Tokarz, R., Tagliafierro, T., Sameroff, S., Cucura, D. M., Oleynik, A., Che, X., et al. (2019). Microbiome analysis of *Ixodes scapularis* ticks from New York and Connecticut. *Ticks Tick Borne Dis.* 10:894–900. doi: 10.1016/j.ttbdis.2019.04.011
- von Fricken, M. E., Lkhagvatseren, S., Boldbaatar, B., Nymadawa, P., Weppelmann, T. A., Baigalmaa, B.-O., et al. (2018). Estimated seroprevalence of *Anaplasma* spp. and spotted fever group *Rickettsia* exposure among herders and livestock in Mongolia. *Acta Trop.* 177, 179–185. doi: 10.1016/j.actatropica.2017.10.015
- von Fricken, M. E., Qurollo, B. A., Boldbaatar, B., Wang, Y.-W., Jiang, R.-R., Lkhagvatseren, S., et al. (2020a). Genetic diversity of *Anaplasma* and *Ehrlichia* bacteria found in *Dermacentor* and *Ixodes* ticks in Mongolia. *Ticks Tick-Borne Dis.* 11:101316. doi: 10.1016/j.ttbdis.2019.101316
- von Fricken, M. E., Voorhees, M. A., Koehler, J. W., Asbun, C., Lam, B., Qurollo, B., et al. (2020b). Molecular characteristics of *Rickettsia* in ticks collected along the southern border of Mongolia. *Pathogens* 9:943. doi: 10.3390/pathogens9110943
- Voorhees, M. A., Padilla, S. L., Jamsransuren, D., Koehler, J. W., Delp, K. L., Adiyadorj, D., et al. (2018). Crimean-Congo hemorrhagic fever virus, Mongolia, 2013–2014. *Emerg. Infect. Dis.* 24, 2202–2209. doi: 10.3201/eid2412.180175
- Wechtaisong, W., Bonnet, S. I., Chomel, B. B., Lien, Y.-Y., Chuang, S.-T., and Tsai, Y.-L. (2021). Investigation of Transovarial transmission of *Bartonella henselae* in *Rhipicephalus sanguineus* sensu lato ticks using artificial feeding. *Microorganisms* 9:2501. doi: 10.3390/microorganisms9122501
- Wei, F., Song, M., Liu, H., Wang, B., Wang, S., Wang, Z., et al. (2016). Molecular detection and characterization of zoonotic and veterinary pathogens in ticks from northeastern China. *Front. Microbiol.* 7:1913. doi: 10.3389/fmicb.2016.01913
- Wiksw, M. E., Hu, R., Metzger, M. E., and Ereemeeva, M. E. (2007). Detection of *Rickettsia rickettsii* and *Bartonella henselae* in *Rhipicephalus sanguineus* ticks from California. *J. Med. Entomol.* 44, 158–162. doi: 10.1603/0022-2585(2007)44[158:do rrab]2.0.co;2
- Wu, Q., Li, J., Wang, W., Zhou, J., Wang, D., Fan, B., et al. (2021). Next-generation sequencing reveals four novel viruses associated with calf diarrhea. *Viruses* 13:1907. doi: 10.3390/v13101907
- Yin, X., Guo, S., Ding, C., Cao, M., Kawabata, H., Sato, K., et al. (2018). Spotted fever group *Rickettsiae* in Inner Mongolia, China, 2015–2016. *Emerg. Infect. Dis.* 24, 2105–2107. doi: 10.3201/eid2411.162094
- Yin, H., and Luo, J. (2007). Ticks of small ruminants in China. *Parasitol. Res.* 101, 187–189. doi: 10.1007/s00436-007-0688-3
- Zajac, V., Wójcik-Fatla, A., Dutkiewicz, J., and Szymańska, J. (2015). *Bartonella henselae* in eastern Poland: The relationship between tick infection rates and the serological response of individuals occupationally exposed to tick bites. *J. Vector Ecol.* 40, 75–82. doi: 10.1111/jvec.12135
- Zhang, F., Liu, W., Wu, X.-M., Xin, Z.-T., Zhao, Q.-M., Yang, H., et al. (2008). Detection of *Francisella tularensis* in ticks and identification of their genotypes using multiple-locus variable-number tandem repeat analysis. *BMC Microbiol.* 8:152. doi: 10.1186/1471-2180-8-152



OPEN ACCESS

EDITED BY

Mel C. Melendrez,
Anoka-Ramsey Community College,
United States

REVIEWED BY

Jun Hang,
Walter Reed Army Institute of Research,
United States
Kai Song,
Qingdao University,
China

*CORRESPONDENCE

Nicole L. Achee
nachee@nd.edu

SPECIALTY SECTION

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

RECEIVED 03 June 2022

ACCEPTED 12 August 2022

PUBLISHED 02 September 2022

CITATION

Achee NL and The Remote Emerging
Disease Intelligence—NETwork (REDI-NET)
Consortium (2022) The Remote Emerging
Disease Intelligence—NETwork.
Front. Microbiol. 13:961065.
doi: 10.3389/fmicb.2022.961065

COPYRIGHT

© 2022 Achee and The Remote Emerging
Disease Intelligence—NETwork (REDI-NET)
Consortium. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

The Remote Emerging Disease Intelligence—NETwork

Nicole L. Achee^{1,2*} and The Remote Emerging Disease
Intelligence—NETwork (REDI-NET) Consortium

¹Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, United States,

²Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN, United States

Accurate prediction of zoonotic spillover events requires a detailed understanding of baseline pathogens circulating in differing global environments. By characterizing the diversity and determining the natural baseline of pathogens in a given biological system, any perturbations to this balance can be detected, leading to estimates of risk for emerging diseases. As epidemics and probability for pandemics increase, there is a fundamental need for building global collaborations to fill gaps in the surveillance effort, especially to build remote in-county capacity and standardize timely sample processing and data analysis. To this point, a new consortium, the Remote Emerging Disease Intelligence-NETwork (*REDI-NET*) has been established to enhance surveillance approaches and characterize natural pathogens in temperate, tropical forest, and tropical grassland biomes. The *REDI-NET* is envisioned to be a long-term, phased initiative. All phases will integrate accompanying training resources such as videos reflecting SOPs and Quick Reference Guides. Routine bio- and xenosurveillance will facilitate the characterization of ecological parameters, enhance the accuracy of vector species identification using artificial intelligence technology, and guide the establishment of epidemiological risk thresholds critical for mitigating disease outbreaks in a timely manner. A key deliverable of the *REDI-NET* is a custom-designed electronically merged (e-MERGE) data pipeline and alert dashboard that integrates remotely captured data with state-of-the-art metagenomic next-generation sequencing technology. This pipeline incorporates data generated from field and laboratory best practices, to furnish health decision-makers with a centralized, timely, and rigorous database to efficiently search interdisciplinary and heterogeneous data sources necessary to alert, prepare and mitigate health threats. The e-MERGE pipeline, once fully established, will be a flexible, scalable, and expandable tool for varied health applications. Program success will result in an operational framework that addresses resource gaps in pathogen surveillance and enhances health protection with broad global applicability. The objective of this manuscript is to introduce the *REDI-NET* framework to anticipated stakeholders engaged in metagenomics, epidemiological surveillance, and One Health with a focus on Phase 1.

KEYWORDS

REDI-NET, bio-/xeno-surveillance, emerging infectious disease, zoonoses, metagenomics, next-generation sequencing, mobile data entry, capacity building

Introduction

An estimated 75% of emerging infectious diseases are zoonotic in nature, arising when pathogens in animals are passed to humans (Gebreyes et al., 2014), with vector-borne diseases accounting for more than 17% of all infectious diseases (World Health Organization, 2020). As the human population expands in number and into new geographical regions, the possibility that humans will come into close contact with animal species that are potential hosts of an infectious agent increases, including vector-borne diseases (VBD). Never before has the catastrophic impact of zoonotic spillover on humans needed less introduction. The World Health Organization (WHO) warned in its 2007 (World Health Organization, 2007) report that infectious diseases are emerging at a rate that has not been seen before. Since the 1970s, about 40 infectious diseases have been discovered, including SARS (Likhacheva, 2006), MERS (Lu and Liu, 2012), Ebola (Peterson et al., 2004), swine flu (Cohen and Enserink, 2009) and most recently SARS-CoV-2 (Sun et al., 2020). VBDs represent a major threat to public health worldwide. The re-emergence of dengue is recognized as a major concern by the WHO with more than 40% of the world's population at risk and estimates of more than 300 million infections every year (World Health Organization, 2021). Chikungunya virus (CHIKV) recently received considerable attention due to outbreaks in the Indian Ocean, India, Europe, and Americas, with over 1 million cases recorded to date. After limited early outbreaks in the Pacific in 2007 and 2013, the Zika virus has spread to more than 30 countries in the Americas and the Caribbean, infecting over 100 million people (Moore et al., 2020). The recent rise of microcephaly cases and other neurological disorders reported in Brazil prompted the WHO to declare Zika as a Public Health Emergency of International Concern (WHO, 2017). Beyond these diseases that receive global attention, other VBDs of human health importance occur in the United States, such as West Nile Virus (World Health Organization, 2016) and Eastern Equine Encephalitis (Lindsey et al., 2015). In addition, across North America, the incidence of tick-borne diseases is increasing. The Center for Disease Control and Prevention (CDC) estimates that there are approximately 300,000 cases of Lyme disease (LD) annually, making LD the most common vector-borne disease in the U.S. (Molaei et al., 2015).

Universities can serve as important resources for technical training and education related to estimating disease threats and are leaders in the development of risk models and forecasting. Local health jurisdictions represent the first line of defense for preventing and controlling disease outbreaks and are tasked with carrying out limited surveillance activities in response to emerging pathogens; however, surveillance efforts are often narrow in scope (targeting a predetermined cohort of biological samples and testing for known pathogens) and reports can be seriously delayed if samples are sent to overloaded reach-back laboratories, taking weeks or months to be processed with data release occurring much later. Despite this heavy reliance on local authorities for a coordinated response to emerging pathogens, critical gaps exist in

their ability to effectively do so. A recent evaluation found that local health personnel “expressed a lack of capacity to respond appropriately and effectively to emerging pathogens” (Center for Disease Control and Prevention, 2019). Respondents cited several areas of greatest weakness within their programs that included: (1) a lack of appropriate epidemiological surveillance tools (RESEARCH GAP), (2) a lack of adequately trained staff to carry out surveillance activities (TRAINING GAP), and (3) a lack of a uniform approach to data warehousing and analysis for risk assessments (COLLABORATION GAP).

The overarching aim of the *REDI-NET* is to develop a new U.S. and international laboratory consortium between academia and the Department of Defense (DoD) along with domestic and international partnering institutions to detect, predict, and mitigate emerging and re-emerging infectious disease threats and improve the accuracy and timeliness of the “data-to-decision” pipeline. Key expected deliverables include state-of-the-art standard operating procedures (SOPs), a functional e-MERGE data pipeline and *REDI-NET* database that includes custom-designed field data collection and web-based laboratory data collection applications, and a dashboard with built-in functionality to forecast emerging pathogens with user-friendly actionable outputs. Program success will result in an operational framework that addresses resource gaps in emerging disease surveillance and health protection with broad applicability.

Greater numbers of empowered remote laboratory and field personnel with a foundational background in monitoring and detecting pathogens will immediately benefit institutions and serve as local source guidance on appropriate surveillance methods, sample processing, and data usage to detect and mitigate potential zoonoses spill-over events. The computational infrastructure interface designed, built, and validated in this program will offer a broad range of tools for modeling data, simulations, data analysis, and the management of distributed data sources for health-decision making and future development of potential novel technologies such as vaccines and drugs.

Materials and methods

The *REDI-NET* is envisioned to be a long-term, phased initiative (Table 1): Phase 1 leverages existing partnerships to provide a solid foundation of excellence to construct robust Standard Operating Procedures (SOPs) for field sampling and pathogen assessment in ticks, leeches, water and sediment and validate genomic outputs across laboratories; Phase 2 will expand upon environmental and invertebrate sentinel sample types, integrate stakeholder training, and optimize the platform through user feedback; while Phase 3+ will advance toward a One Health Approach with the inclusion of active animal sampling in new ecologies related to DoD Force Readiness; while the goal of Phase 4 is to roll the platform out to DoD Commands. All phases will integrate accompanying training resources such as videos reflecting SOPs and Quick Reference Guides.

TABLE 1 REDI-NET program scope.

Phase I (completed)	<p>Aim 1: Establish robust SOPs for rigorous data capture and matched capabilities at <i>Gold</i> reach-back laboratories</p> <p>Aim 2: Active field surveillance across varied ecologies for broad-spectrum pathogen detection.</p> <p>Aim 3: Enable remote, verified <i>in-situ</i>, near real-time data acquisition for actionable reporting.</p> <p>Aim 4: Institute a data management pipeline for actionable reporting and threat forecasting.</p>
Phase II (current)	<p>Aim 1: Expand <i>REDI-NET</i> pathogen portfolio to include opportunistic vDNA outputs and an enhanced data warehouse.</p> <p>Aim 2: Develop an actionable workflow for early pathogen detection by remote forward-facing laboratories using the <i>REDI-NET</i> e-MERGE pipeline.</p> <p>Aim 3: Transfer knowledge of <i>REDI-NET</i> technologies and processes</p>
Phase III: (planned)	<p>Aim 1: Expand <i>REDI-NET</i> pathogen portfolio to include active vDNA sampling and field metadata collection for Force Health Protection / Force Readiness.</p> <p>Aim 2: Establish, maintain, and expand <i>REDI-NET</i> active surveillance across varied ecologies in existing and new Geographical Combatant Commands (GCCs) for broad-spectrum pathogen detection using established SOPs which include metagenomic approaches.</p> <p>Aim 3: Readiness for roll-out of the <i>REDI-NET</i> platform to Senior DoD Officials in six GCCs [NORTHCOM, SOUTHCOM, EUCOM, AFRICOM, CENTCOM, and INDOPACOM], and the AFHSC-GEIS partner network.</p>

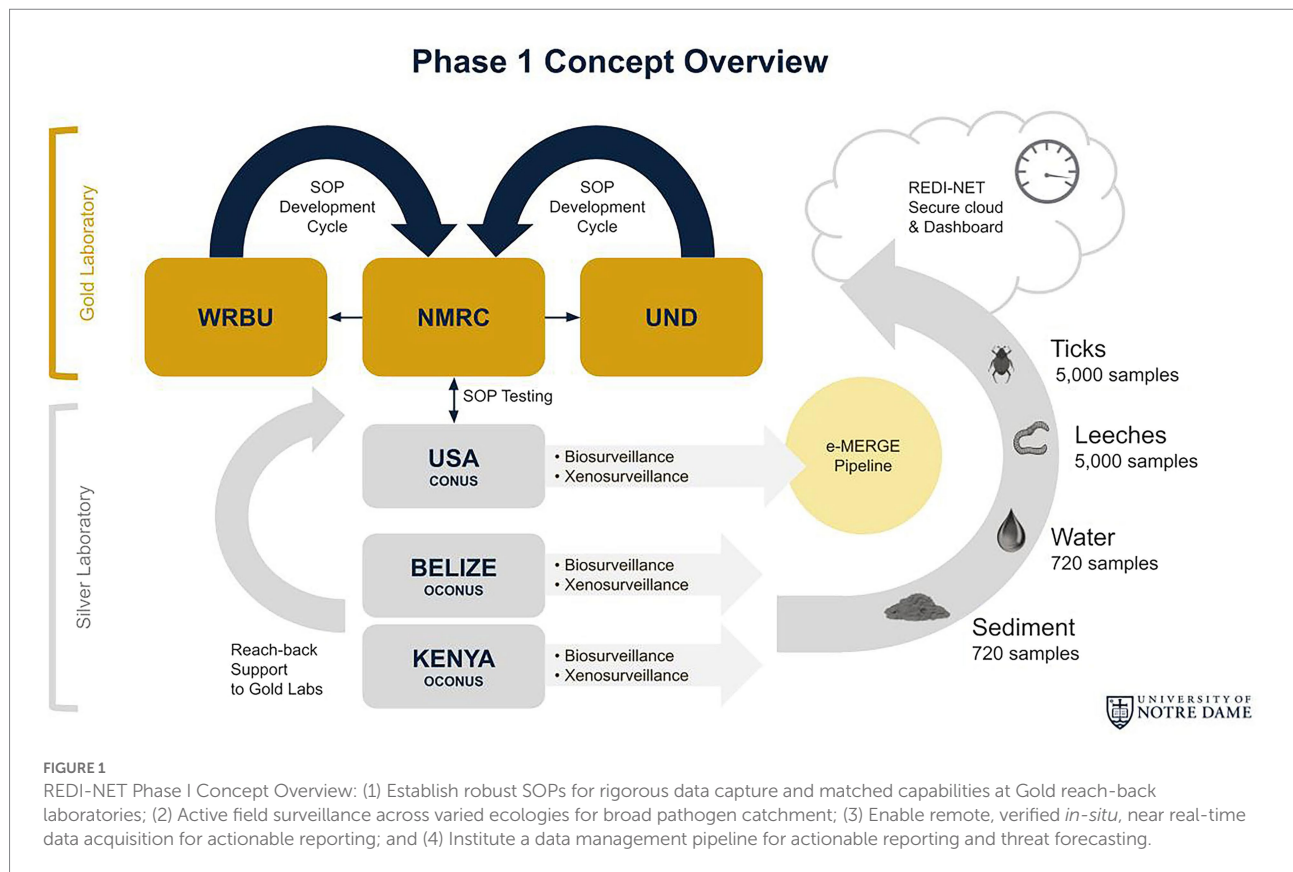
In each Phase, we will build new scientific collaborations and capacity by expanding consortium partnerships across institutions and bring the best HPDC practices to the utility of emerging infectious disease surveillance. A tiered laboratory schema will be applied throughout the initiative, whereby *Gold* Laboratories will serve as reach-back, verification centers and *Silver* Laboratories will serve as field-sample source locations. All laboratories will have matched capability for sample processing and testing.

Phase 1 centers on enhanced surveillance approaches to characterize natural pathogens in select temperate [CONUS—Navy Entomology Centers of Excellence (NECE)], tropical forest (OCONUS—Belize), and tropical grassland (OCONUS—Kenya) sites. Sentinel sample types will include ticks, leeches, water, and sediment (Figure 1). These invertebrate and environmental sample types were selected based on their high probability of containing pathogens that permit total nucleic acid (TNA) extraction for metagenomic analyses. Tick surveillance conducted in Belize has shown a presence of tick-borne rickettsioses relevant to humans and animals (ASTHO, 2007). Several studies have documented the recovery of pathogens (Polsomboon et al., 2017; Gogarten et al., 2019) in water bodies, including watering holes and wastewater (Huver et al., 2015; Mosher et al., 2017; Alfano et al., 2021), and in larger invertebrate “blood-bags,” such as leeches (Herder et al., 2014; Huver et al., 2015; Mosher et al., 2017; Alfano et al., 2021). These results suggest that iDNA/eDNA-based surveillance approaches may complement efforts to proactively identify pathogens that could potentially spill over to humans or livestock. In addition, high-throughput automated AI technology for morphological tick species identification will be developed thus facilitating distance and real-time tick identification. The remote Tick e-ID platform (IDX) will be verified through the molecular identification of tick species.

The four specific aims in Phase 1 (Table 1):

Aim 1: Establish Robust Sops for Rigorous Data Capture and Matched Laboratory Capabilities.

To ensure consistent and reproducible pathogen detection results, a set of robust SOPs encompassing field collection, sample storage, laboratory processing and testing, and sample shipping have been developed, tested, optimized, and validated by all consortium laboratories through an iterative process. Corresponding digital data collection sheets (DCS) have also been developed from each SOP, which reflects the essential field and laboratory data required to capture throughout the *REDI-NET* e-MERGE pipeline. Training, and enabling of *Gold* and *Silver* laboratories through capacity strengthening to proficiently perform SOPs, has occurred either in-person or through remote sessions due to COVID-19 restrictions to ensure reproducibility of electronically sourced results from the remote locations using the *REDI-NET* e-MERGE pipeline. To maintain data rigor and identify any initial capability deficiencies, aliquots of reference materials containing the same mixture of bacteria (3 Gram-negatives and 5 Gram-positives including *Pseudomonas aeruginosa*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Listeria monocytogenes*, and *Bacillus subtilis*), viruses (Epstein–Barr virus and Human Immunodeficiency Virus 1), and fungi (*Saccharomyces cerevisiae* and *Cryptococcus neoformans*) organisms have been distributed to all current Consortium laboratories for output verification. Proficiency testing based on nanopore sequencing read number, length, quality, and coverage depth of each microbe has been conducted for current *Gold/Silver* individual operators and will be performed for any future laboratories joining the Consortium. This proficiency testing will be performed on an annual basis across the entire Consortium to ensure optimal and consistent data output from each partner laboratory. A second layer of verification has been implemented to have periodical submission of subsets of field-collected samples from *Silver* laboratories to the *Gold* reach-back laboratories to allow verification on the sequencing output and resultant microbe classification and abundance. In the event that human pathogens are detected, an Illumina sequencer equipped in all *Gold* laboratories will be used to confirm the initial findings from nanopore data.



Aim 2: Active Field Surveillance Across Varied Ecologies for Broad-Spectrum Pathogen Detection.

Tick collections are conducted monthly using dragging methodology in Belize, Florida, and Kenya from ecologically diverse areas of high animal footfall (i.e., woodland, forest-field margin, and around watering holes). Tick dragging is standardized across *REDI-NET* Silver field locations using a fixed sampling scheme, allowing aggregated analysis of presence/absence data along with densities and species composition. Ticks from domestic and/or wild animal collection based on the local approved protocols with the assistance of local veterinarians are also being processed as opportunistic samples. Standardized tick species identification capability in *Gold* laboratories will be developed from computer-vision, deep-learning algorithms using the IDX device (Brey et al., 2022) originally developed for mosquito species identification (Goodwin et al., 2021). Free-living leeches are being collected from water bodies using traps baited with beef liver (Ratnasingham and Hebert, 2013). Environmental samples (i.e., water and sediment) are being collected from the same Belize, Florida, and Kenya surveillance sites. Water and sediment samples are being gathered in triplicate using standard dip cups at each edge and 1 m into the water body. Invertebrates are kept alive, and both invertebrate and environmental samples are being transported to *Silver* laboratories at 4°C. All field samples will be stored at -80°C where possible or placed in the best RNA

stabilizers (as determined by *REDI-NET* SOP development) or kept alive at 4°C (ticks, leeches; Herder et al., 2014; Mosher et al., 2017) until processing. Any samples shipped to the *Gold* labs for reach-back support and remote data validation will be shipped without breaking cold-chain and accompanied by the proper approved import/export permits according to the *REDI-NET* best practice shipping SOP. All specimens in each *Gold* and *Silver* laboratories are inventoried and entered into a custom-designed data management system.

Aim 3: Enable Remote, Verified *in-Situ*, Near Real-Time Data Acquisition for Actionable Reporting.

A newly developed *REDI-NET* data warehouse with seamless integration of environmental and bio-/xeno-surveillance pathogen data will facilitate the development of the comprehensive e-MERGE pipeline, accessed through the newly established *REDI-NET* website (REDI-NET Consortium, 2022). The UND CRC team has developed and maintains a secure data repository to meet the *REDI-NET* project needs. The repository has been developed using modern web and database technologies to support data ingestion, validation, storage, and export for analysis (Figure 2).

An important aspect of the study will be the ability for *REDI-NET* consortium members to be able to monitor, access, and interact with the data and metadata being collected. The

REDI-NET database allows dynamic data entry from each study site using a custom-designed mobile application for field data entry, hosted by CommCare, as well as a custom-designed web-based application for laboratory data entry providing the ability for near real-time data uploading and viewing. Critical functions of the *REDI-NET* data management system are provision of a digital library and data warehouse containing all available pathogen and vector-related data in a format that facilitates users to search and retrieve data and references pertaining to user-specified areas of interest. The data portal also hosts digital SOPs, Data Collection Sheets (DCSs), training videos, and reference materials. The digital library holds datasets of extracted parameter values for all primary vectors. Along with raw access to the data and metadata, the data portal will provide access to filterable visualizations of risk projections. Ultimately, the data portal provides decision-support tools and brings together the data, cutting-edge mathematical and statistical methods to improve emerging infectious disease modeling, validation, and prediction.

The *REDI-NET* mobile and web-based applications have been designed to assure data rigor based on SOPs. The DCSs were

developed from the SOPs to establish quality control rules (i.e., drop-down lists, user entry fields) and inform the logic of the field data entry mobile application and *REDI-NET* database. In addition, the DCSs can be utilized if the field mobile application is unavailable all the while ensuring seamless data ingestion into the e-MERGE pipeline. The result will be a repository of both raw data and de-identified data, with accompanying metadata, immediately available for modeling, statistical analyses, and actionable reporting requirements. The data warehouse and file repository are backed up daily and weekly at the operating system level to ease recovery as needed. All data are stripped of personal identifiers (GPS information blurred) at the central data warehouse prior to any subsequent analysis and/or data sharing restrictions pursuant to DoD data sharing policies. Descriptive visualizations of pathogen occurrence or predicted trends in risk will be available *via* end-user interactive sessions with the database dashboard.

Aim 4: Institute a Data Management Pipeline for Actionable Reporting and Threat Forecasting.

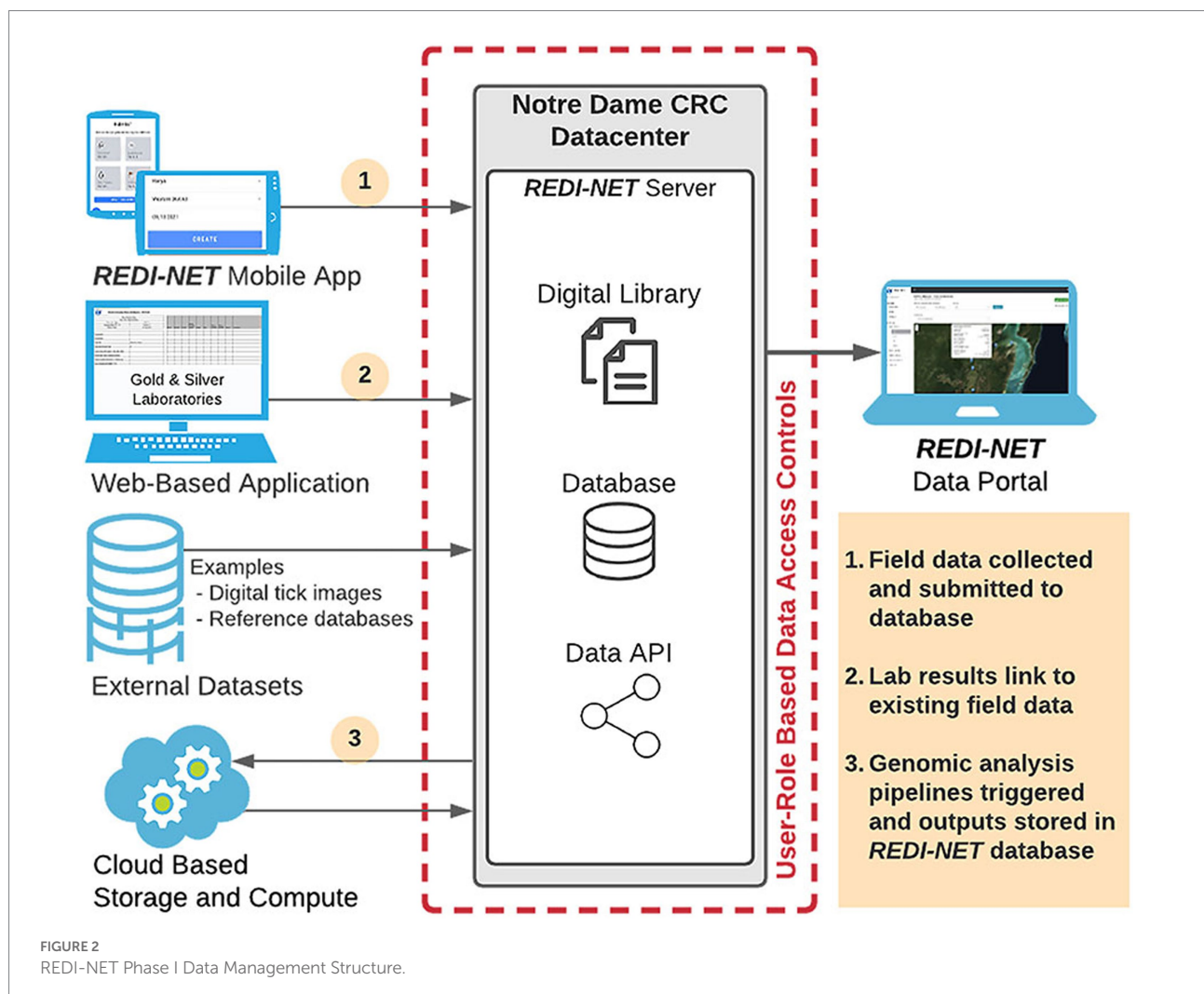


TABLE 2 Phase I REDI-NET standardized operating procedure (SOP) list.

No.	Title	Scope description
SOP T-1	Tick field sampling	To document the REDI-NET field processes for collecting samples
SOP L-1	Leech field sampling	
SOP W-1	Water field sampling	
SOP S-1	Sediment field sampling	
SOP T-2	Tick processing	To outline REDI-NET procedures to process samples for total nucleic acid extraction
SOP L-2	Leech processing	
SOP W-2	Water processing	
SOP S-2	Sediment processing	
SOP T-3	Tick storage	To outline steps for properly storing REDI-NET field-collected samples and nucleic acid samples purified from these samples
SOP L-3	Leech storage	
SOP W-3	Water storage	
SOP S-3	Sediment storage	
SOP T-4	Tick testing	To outline the REDI-NET procedures for properly using the Oxford Nanopore Sequencing platforms to sequence DNA and cDNA extracted from collected samples
SOP L-4	Leech testing	
SOP W-4	Water testing	
SOP S-4	Sediment testing	
SOP T-5	Tick shipping	To outline steps for proper packaging and shipping of preserved samples from a REDI-NET Silver lab to a REDI-NET Gold lab
SOP L-5	Leech shipping	
SOP W-5	Water shipping	
SOP S-5	Sediment shipping	
SOP DE	Data entry	To outline steps for properly recording and upload of data for the REDI-NET program

A central function of the *REDI-NET* e-MERGE analysis pipeline will be the development of novel risk maps and models. Outputs from these models will be accessible to the end user *via* the *REDI-NET* dashboard. Expected modeling outputs include pathogen presence/absence estimates, seasonal projections of vector abundance and pathogen prevalence, and predictive disease and pathogen spillover risk maps. Surveillance data collected through *REDI-NET* will be combined with environmental and ecological data to generate predictive disease risk maps *via* a hierarchical Bayesian modeling framework. Such a framework enables the incorporation of multiple data types into the modeling process, including both historical and prospective data, both epidemiological and environmental data, and data collected at different spatial and temporal scales. At the same time, this modeling framework allows for the incorporation of associations between outcomes of interest and numerous spatial covariates, such as satellite-derived measures of temperature, wetness, land cover, and elevation. While some data—e.g., satellite-derived covariates, data collected prospectively by the *REDI-NET*—are available globally, others are unique to a given setting—e.g., historical surveillance data. The modeling framework allows for all such data types to jointly inform model-based predictions of pathogen risk.

Several types of model predictions will be generated. We will use the Bayesian modeling framework to map infection thresholds in the environment, and in arthropod and vertebrate animal populations. Among other available data types, the e-MERGE data analysis pipeline will allow us to integrate novel covariate datasets into these models. Surveillance data will be divided into training

and testing sets, with the training data used to fit the models and the testing data used for model validation. Predicted infection thresholds from the model will then be resampled to generate predictive risk maps that can be used to estimate risk in unsampled locations. Where appropriate, we will also integrate point-based and polygon-based (areal) data into the predictive models using techniques to deal with the spatial misalignment of different datasets (Chilès and Delfiner, 2009; Rue et al., 2009). Last, our aim is to provide analytical tools that, once monitoring and surveillance become routine, can be used to investigate temporal trends to track changing pathogen patterns in the future. Spatial models will be extended to include a temporal component to capture seasonal patterns in pathogen prevalence and diversity. Spatiotemporal models can also be used to detect changes in pathogen distribution over time (seasonality) and to determine which ecological and environmental covariates are associated with any temporal trends in pathogen distribution, prevalence, or diversity (Christakos, 2000).

Results

A total of 21 SOPs spanning field sampling, laboratory specimen processing, pathogen testing, specimen storage, shipment, and data management have been developed (Table 2). Current findings indicate that active field surveillance in Florida, Belize, and Kenya using custom-optimized *REDI-NET* SOPs has been capable of generating sequencing data leading to the detection of human pathogens in ticks, leeches, water, and soil.

In addition, testing of reference material has generated consistent sequencing outputs indicating standardized capabilities throughout Consortium laboratories. Barcoding schemes have been developed to enable rigorous sample tracking and relating field metadata with laboratory pathogen testing outputs. The computer-vision algorithm for tick species identification has been developed from sample images taken with IDX submitted by all *Gold* laboratories for preliminary deployment in the IDX devices. The custom-designed *REDI-NET* database and both mobile field data entry and web-based lab data entry applications have been developed and verification of remotely captured data sync along with metadata relation functionality completed. The custom-developed *REDI-NET* technology (mobile/field, web-based/lab data entry applications) includes built-in functionality to ensure data integrity and data quality assurances. Additionally, quick reference guides and training videos have been developed for assuring standardization in methodology. A custom-designed *REDI-NET* sample storage web application has been integrated into the *REDI-NET* data portal to allow users to manage sample storage location for individual specimen records. A custom-designed *REDI-NET* dashboard and data portal have been designed as a resource repository to the database that hosts SOPs and DCSS, training videos, quick reference guides, equipment manufacturer data, and *REDI-NET* presentation templates with the ability to search/filter based on keywords. A NGS data upload mechanism has been established for the *REDI-NET* database and NGS software and data pipelines have been evaluated to identify those most appropriate for the *REDI-NET* pipeline. Risk models have been developed using publicly available databases (VectorMap and iNaturalist), published literature, and de-identified human case data from repositories to validate the mathematical model framework(s) for tick-borne disease risk estimation in Florida, Belize, and Kenya with model outputs available to visualize on the *REDI-NET* data portal. To ensure that the technical feasibility and viability of the *REDI-NET* framework is sustainable, and practical, as an early alarm of emerging infection threats, the Consortium will host end-user training sessions in each Phase to assess broad platform functionality and incorporate user feedback. Additionally, a digital *REDI-NET* resource package will be developed to facilitate ease of roll-out to global end users.

Discussion

Epidemics of emerging and exotic nature are becoming more frequent and diverse worldwide and these outbreaks will inevitably continue into the foreseeable future (Gebreyes et al., 2014). Recently, the Washington Post Editorial Board presented a roadmap for living with COVID (Editorial Board, 2022) which highlighted gaps in infrastructure for monitoring and informing public policies, including the need for systematic surveillance. Often there is no overarching database architecture and/or platform that supports scalable approaches for building the

necessary tools and services to facilitate the multi-user needs of emerging infectious disease surveillance data. This gap translates beyond COVID-19. Accurate prediction of zoonotic spillover events requires a detailed qualitative and quantitative understanding of the baseline environmental pathogens circulating in differing global land biomes, but detailed characterization regarding ecology, accurate species identification, and epidemiological thresholds, critical for actual success in mitigating disease outbreaks in a timely manner, has critical gaps. Any solution to alleviate these missing functional capabilities must also be cognizant of the underlying capacity, economic and sustainability issues at surveillance nodes.

The scientific premise of the *REDI-NET* was founded on the requirement of having tools and processes for near real-time, evidence-based, health decision-making with capacity strengthening for knowledgeable, competent health professionals and facilities for appropriate environmental pathogen surveillance, detection, and response. The *REDI-NET* has connected tangible, measurable partnerships amongst stakeholders to leverage expertise and resources for disease surveillance and management. The initial launch of the *REDI-NET* program has resulted in an operational framework that is meant to address some of the known resource gaps in infectious disease surveillance. The *REDI-NET* military/civilian collaborations have and will continue to broaden the reach of emerging disease surveillance activities for more accurate and timely predictive tools. We anticipate continued success in capacity-strengthening to occur as the *REDI-NET* initiative progresses. New consortium members will join the *REDI-NET*, remote data collection methodologies will be expanded using AI technologies, including drones for semi-autonomous and/or autonomous sample collection and we will consider adding machine learning algorithms to our spatial modeling platform, as these methods do not assume an underlying model and are robust to sampling biases (Elith et al., 2008; Alpaydin, 2020). Independent publications presenting from *REDI-NET* Phase 1 Aims are anticipated.

REDI-NET Consortium (by organization then in alphabetical order)

University of Notre Dame

Nicole L. Achee, Department of Biological Sciences, Eck Institute for Global Health, University of Notre Dame, 239 Galvin Life Science Center, Notre Dame, IN, United States.

Maria Dahn, Department of Biological Sciences, University of Notre Dame, 239 Galvin Life Science Center, Notre Dame, IN, United States.

Benedicte Fustec, Department of Biological Sciences, University of Notre Dame, 265 Galvin Life Science Center, Notre Dame, IN, United States.

Miriam Grady, Department of Biological Sciences, University of Notre Dame, 301 Galvin Life Science Center, Notre Dame, IN, United States.

John P. Grieco, Department of Biological Sciences, Eck Institute for Global Health, University of Notre Dame, 243 Galvin Life Science Center, Notre Dame, IN, United States.

Donovan Leiva, Department of Biological Sciences, University of Notre Dame, 301 Galvin Life Science Center, Notre Dame, IN, United States.

Kara Linder, Department of Biological Sciences, University of Notre Dame, 301 Galvin Life Science Center, Notre Dame, IN, United States.

Sean Moore, Department of Biological Sciences, University of Notre Dame, 345 Galvin Life Science Center, Notre Dame, IN, United States.

Stacy Mowry, Department of Biological Sciences, University of Notre Dame, 347 Galvin Life Science Center, Notre Dame, IN, United States.

Alex Perkins, Department of Biological Sciences, Eck Institute for Global Health, University of Notre Dame, 347 Galvin Life Science Center, Notre Dame, IN, United States.

Caroline Pitts, Department of Biological Sciences, University of Notre Dame, 301 Galvin Life Science Center, Notre Dame, IN, United States.

Brooke Rodriguez, Department of Biological Sciences, University of Notre Dame, 301 Galvin Life Science Center, Notre Dame, IN, United States.

UND Center for Research Computing

Jaroslav Nabrzyski, Center for Research Computing, University of Notre Dame, 829 Flanner Hall, Notre Dame, IN, United States.

Samuel Njoroge, Center for Research Computing, University of Notre Dame, 829 Flanner Hall, Notre Dame, IN, United States.

Matthew Noffsinger, Center for Research Computing, University of Notre Dame, 829 Flanner Hall, Notre Dame, IN, USA.

Kinsey Poland, Center for Research Computing, University of Notre Dame, 829 Flanner Hall, Notre Dame, IN, United States.

Caleb Reinking, Center for Research Computing, University of Notre Dame, 807 Flanner Hall, Notre Dame, IN, United States.

Bradley Sandberg, Center for Research Computing, University of Notre Dame, 829 Flanner Hall, Notre Dame, IN, United States.

Belize Vector and Ecology Center

Arlo Cansino, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize, Central America.

Jailene Castillo, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize, Central America.

Alvaro Cruz, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize, Central America.

Marie C. Pott, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize, Central America.

Uziel Romero, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize, Central America.

Naval Medical Research Center

Tatyana Belinskaya, Naval Medical Research Center (NMRC), 503 Robert Grant Avenue, Silver Spring, MD 20910, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD, United States.

Erica Cimo, Naval Medical Research Center (NMRC), 503 Robert Grant Avenue, Silver Spring, MD 20910, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD, United States.

Le Jiang, Naval Medical Research Center (NMRC), 503 Robert Grant Avenue, Silver Spring, MD 20910, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD, United States.

Hsiao-Mei Liao, Naval Medical Research Center (NMRC), 503 Robert Grant Avenue, Silver Spring, MD 20910, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD, United States.

Zhiwen Zhang, Naval Medical Research Center (NMRC), 503 Robert Grant Avenue, Silver Spring, MD 20910, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD, United States.

Navy Entomology Center of Excellence

Jason Fajardo, Navy Entomology Center of Excellence, 937 Child St, Jacksonville, FL 32212, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD.

Edward Traczyk, Navy Entomology Center of Excellence, 937 Child St, Jacksonville, FL 32212, United States.

Melissa Vizza, Navy Entomology Center of Excellence, 937 Child St, Jacksonville, FL 32212, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD.

Christy Waits, Navy Entomology Center of Excellence, 937 Child St, Jacksonville, FL 32212, United States; Henry M Jackson Foundation for the Advancement of Military Medicine, 6720A Rockledge Dr., Bethesda, MD.

Walter Reed Biosystematics Unit

Brian Bourke, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; One Health Branch, Center for Infectious Disease Research, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States.

Laura Caicedo-Quiroga, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; One Health Branch, Center for Infectious Disease Research, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States.

Koray Ergunay, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; One Health Branch, Center for Infectious Disease Research, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States; Virology Unit, Department of

Medical Microbiology, Faculty of Medicine, Hacettepe University Ankara, Turkey.

Yvonne-Marie Linton, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; One Health Branch, Center for Infectious Disease Research, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States; Department of Entomology, Smithsonian Institution National Museum of Natural History, Washington, DC, United States.

Suppaluck Nelson, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; One Health Branch, Center for Infectious Disease Research, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States.

David B. Pecor, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; One Health Branch, Center for Infectious Disease Research, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States.

Alexander Potter, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; One Health Branch, Center for Infectious Disease Research, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD, United States.

Dawn Zimmerman, Walter Reed Biosystematics Unit (WRBU), Smithsonian Institution Museum Support Center, Suitland, MD, United States; Department of Epidemiology of Microbial Disease, Yale School of Public Health, New Haven, CT, United States.

Smithsonian Institution

James Hassell, Global Health Program, Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC, United States; Department of Epidemiology of Microbial Disease, Yale School of Public Health, New Haven, CT, United States.

Mpala Research Centre

Maureen Kamau, Mpala Research Centre, Laikipia, Kenya; Global Health Program, Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC, United States.

Rashid Lebung, Mpala Research Centre, Laikipia, Kenya.

Janerose Mutura, Mpala Research Centre, Laikipia, Kenya.

George Mason University

Michael von Fricken, George Mason University, 4,400 University Drive, Fairfax, VA, United States.

Abigail A. Lilak, George Mason University, 4,400 University Drive, Fairfax, VA, United States.

Graham A. Matulis, George Mason University, 4,400 University Drive, Fairfax, VA, United States.

Vectech

Jewell Brey, Vectech, 3,600 Clipper Mill Rd., STE 205, Baltimore MD, United States.

Tristan Ford, Vectech, 3,600 Clipper Mill Rd., STE 205, Baltimore MD, United States.

Adam Goodwin, Vectech, 3,600 Clipper Mill Rd., STE 205, Baltimore MD, United States.

Ghnana Madineni, Vectech, 3,600 Clipper Mill Rd., STE 205, Baltimore MD, United States.

Bala Murali Manoghar Sai Sudhakar, Vectech, 3,600 Clipper Mill Rd., STE 205, Baltimore MD, United States.

Sanket Padmanabhan, Vectech, 3,600 Clipper Mill Rd., STE 205, Baltimore MD, United States.

Former consortium members (no longer actively affiliated)

Robert Ang'ila, Mpala Research Centre, Laikipia, Kenya.

Margaret Elliott, Department of Biological Sciences, University of Notre Dame, 301 Galvin Life Science Center, Notre Dame, IN, United States.

Joanna Gomez, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize.

Lauren Maestas, Navy Entomology Center of Excellence, 937 Child St, Jacksonville, FL, United States.

Marla S. Magaña, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize.

Mohamed Sallam, Navy Entomology Center of Excellence, 937 Child St, Jacksonville, FL, United States.

Alexia Thompson, Belize Vector and Ecology Center, Slaughterhouse St, Orange Walk Town, Orange Walk, Belize.

Data availability statement

The datasets presented in this article are not readily available because REDI-NET Phase 1 data generation and analysis is ongoing with associated datasets to become available at the time specific outcomes are published. Requests to access the datasets should be directed to NA, redinet@nd.edu.

Author contributions

The REDI-NET was conceived by JPG, LJ, Y-ML, and NA. LJ and H-ML led the development of laboratory sample processing and testing SOPs. NA, JPG, LJ, Y-ML, BF, MD, and CP contributed to the development of field sampling, storage, and shipment SOPs. Gold laboratory data are being generated by BF, DL, CP, H-ML, TB, ZZ, EC, BB, LC-Q, KE, SN, DP, and APo. Silver field and laboratory data are being generated in Belize by ACa, JC, ACr, MP, and UR, in Florida by ET, JF, MV, and CW, and in Kenya by DZ, JH, MK, RL, JM, MF, AL, and GM. Management of the central data repository is led by JN, SN, MN, KP, CR, and BS. Mathematical modeling is being conducted by APe, SMoo, and SMow. Literature searches were conducted by ME, BR, and MG. Human and animal use

applications are managed by NA. Tick e-ID platform development led by TF, AG, JB, GM, BS, and SP. The first draft of the manuscript was written by JG, LJ, Y-ML, MF, MD, BS, APe, SMoo, SMow, and NA. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the United States Army Medical Research and Development Command under contract no. W81XWH-21-C-0001.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Alfano, N., Dayaram, A., Axtner, J., Tsangaras, K., Kampmann, M., Mohamed, A., et al. (2021). Non-invasive surveys of mammalian viruses using environmental DNA. *Methods Ecol. Evol.* 12, 1941–1952.
- Alpaydin, E. (2020). *Introduction to Machine Learning, 4th Edn.*, MIT Press, Cambridge, MA.
- ASTHO (2007). *State Public Health Vector Control Conference: Workforce and Disease Priorities Needs Assessment Summary*. Arlington, VA: Association of State and Territorial Health Officials.
- Brey, J., Sudhakar, S., Manoghar, B. M., Gersch, K., Ford, T., Glancey, M., et al. (2022). Modified mosquito programs' surveillance needs and An image-based identification tool to address them. *Front. Trop. Dis.* 2:810062. doi: 10.3389/ftrd.2021.810062
- Center for Disease Control and Prevention (2019). Lyme disease data and surveillance. Available at: <https://www.cdc.gov/lyme/datasurveillance/index.html> (Accessed 2020).
- Chilès, J., and Delfiner, P. (2009). *Geostatistics: Modeling Spatial Uncertainty, 1st Edn.* John Wiley & Sons, Hoboken, NJ.
- Christakos, G. (2000). *Modern Spatiotemporal Geostatistics, 1st Edn.* Oxford University Press, Oxford, England.
- Cohen, J., and Enserink, M. (2009). Infectious diseases. As swine flu circles globe, scientists grapple With basic questions. *Science* 324, 572–573. doi: 10.1126/science.324.572
- Editorial Board (2022). *This is what 'living with covid' might look like*. Washington Post.
- Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. doi: 10.1111/j.1365-2656.2008.01390.x
- Gebreyes, W. A., Dupouy-Camet, J., Newport, M. J., Oliveira, C. J. B., Schlesinger, L. S., Saif, Y. M., et al. (2014). The global one health paradigm: challenges and opportunities for tackling infectious diseases at the human, animal, and environment Interface in low-resource settings. *PLoS Negl. Trop. Dis.* 8:e3257. doi: 10.1371/journal.pntd.0003257
- Gogarten, J. F., Düx, A., Mubemba, B., Pléh, K., Hoffmann, C., Mielke, A., et al. (2019). Tropical rainforest flies carrying pathogens form stable associations With social nonhuman primates. *Mol. Ecol.* 28, 4242–4258. doi: 10.1111/mec.15145
- Goodwin, A., Padmanabhan, S., Hira, S., Glancey, M., Slinowsky, M., Immidiseti, R., et al. (2021). Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection. *Sci. Rep.* 11, 1–15. doi: 10.1038/s41598-021-92891-9
- Herder, J., Valentini, A., Bellemain, E., Dejean, T., van Delft, J., Thomsen, P. F., et al. (2014). *Environmental DNA a Review of the Possible Applications for the Detection of (Invasive) Species*, Bureau Risicobeoordeling & Onderzoeksprogrammering (BuRO),

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other documentation. The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

part of the Netherlands Food and Consumer Product Safety Authority, Nijmegen, the Netherlands.

Huwer, J. R., Koprivnikar, J., Johnson, P. T. J., and Whyard, S. (2015). Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic Ecosystems. *Ecol. Appl.* 25, 991–1002.

Likhacheva, A. (2006). SARS revisited. *Virtual Mentor* 8, 219–222. doi: 10.1001/virtualmentor.2006.8.4.jdsc1-0604

Lindsey, N. P., Lehman, J. A., Staples, J. E., and Fischer, M. (2015). West Nile virus and other nationally Notifiable Arboviral diseases - United States, 2014. *MMWR Morb. Mortal. Wkly Rep.* 64, 929–934. doi: 10.15585/mmwr.mm6434a1

Lu, G., and Liu, D. (2012). SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases. *Protein Cell* 3, 803–805. doi: 10.1007/s13238-012-2811-1

Molaei, G., Armstrong, P. M., Graham, A. C., Kramer, L. D., and Andreadis, T. G. (2015). Insights Into the recent emergence and expansion of eastern equine encephalitis virus in a new focus in the northern New England USA. *Parasit. Vectors* 8:516. doi: 10.1186/s13071-015-1145-2

Moore, S. M., Oidtman, R. J., Soda, K. J., Siraj, A. S., Barker, C. M., and Perkins, T. A. (2020). Leveraging multiple data types to estimate the size of the Zika epidemic in the Americas. *PLoS Negl. Trop. Dis.* 14:e0008640. doi: 10.1371/journal.pntd.0008640

Mosher, B. A., Huyvaert, K. P., Chestnut, T., Kerby, J. L., Madison, J. D., and Bailey, L. L. (2017). Design- and model-based recommendations for detecting and quantifying an amphibian pathogen in environmental samples. *Ecol. Evol.* 7, 10952–10962. doi: 10.1002/ece3.3616

Peterson, A. T., Bauer, J. T., and Mills, J. N. (2004). Ecologic and geographic distribution of Filovirus disease. *Emerg. Infect. Dis.* 10, 40–47. doi: 10.3201/eid1001.030125

Polsomboon, S., Hoel, D. F., Murphy, J. R., Linton, Y., Motoki, M., Robbins, R. G., et al. (2017). Molecular detection and identification of rickettsia species in ticks (Acari: Ixodidae) collected from Belize, Central America. *J. Med. Entomol.* 54, 1718–1726. doi: 10.1093/jme/tjx141

Ratnasingham, S., and Hebert, P. D. N. (2013). A DNA-based registry for All animal species: The barcode index number (BIN) system. *PLoS One* 8:e66213. doi: 10.1371/journal.pone.0066213

REDI-NET Consortium (2022). Remote emerging disease intelligence - NETWORK (REDI-NET). Available at: <https://redi-net.nd.edu>

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. B Stat. Methodol.* 71, 319–392. doi: 10.1111/j.1467-9868.2008.00700.x

Sun, J., He, W., Wang, L., Lai, A., Ji, X., Zhai, X., et al. (2020). COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. *Trends Mol. Med.* 26, 483–495. doi: 10.1016/j.molmed.2020.02.008

WHO (2017). *Zika Situation Report*, WHO, Geneva, Switzerland.

World Health Organization (2007). *The World Health Report 2007: a Safer Future: Global public Health Security in the 21st Century: Overview*, World Health Organization, Geneva, Switzerland.

World Health Organization (2016). WHO Statement on the First Meeting of the International Health Regulations (2005) (IHR 2005) Emergency Committee on Zika Virus and Observed Increase in Neurological Disorders and Neonatal Malformations. Available at: [https://www.who.int/news-room/detail/01-02-2016-who-statement-on-the-first-meeting-of-the-international-health-regulations-\(2005\)-\(ihr-2005\)-emergency-committee-on-zika-virus-and-observed-increase-in-neurological-disorders-and-neonatal-malformations](https://www.who.int/news-room/detail/01-02-2016-who-statement-on-the-first-meeting-of-the-international-health-regulations-(2005)-(ihr-2005)-emergency-committee-on-zika-virus-and-observed-increase-in-neurological-disorders-and-neonatal-malformations)

(2005)-(ihr-2005)-emergency-committee-on-zika-virus-and-observed-increase-in-neurological-disorders-and-neonatal-malformations

World Health Organization (2020). Vector-borne diseases fact sheet. Available at: <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases> (Accessed 24 May 2022).

World Health Organization (2021). *WHO Coronavirus (COVID-19) Dashboard*. World Health Organization.



OPEN ACCESS

EDITED BY

Michael E. von Fricken,
George Mason University,
United States

REVIEWED BY

Alejandro Cabezas-Cruz,
Institut National de Recherche pour
l'Agriculture, l'Alimentation et
l'Environnement (INRAE), France
Ioana Adriana Matei,
University of Agricultural Sciences
and Veterinary Medicine
of Cluj-Napoca, Romania

*CORRESPONDENCE

Jun Hang
jun.hang.civ@health.mil

SPECIALTY SECTION

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

RECEIVED 03 June 2022

ACCEPTED 27 July 2022

PUBLISHED 08 September 2022

CITATION

Pollio AR, Jiang J, Lee SS, Gandhi JS,
Knott BD, Chunashvili T, Conte MA,
Walls SD, Hulseberg CE, Farris CM,
Reinbold-Wasson DD and Hang J
(2022) Discovery of *Rickettsia* spp.
in mosquitoes collected in Georgia by
metagenomics analysis and molecular
characterization.
Front. Microbiol. 13:961090.
doi: 10.3389/fmicb.2022.961090

COPYRIGHT

© 2022 Pollio, Jiang, Lee, Gandhi,
Knott, Chunashvili, Conte, Walls,
Hulseberg, Farris, Reinbold-Wasson
and Hang. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Discovery of *Rickettsia* spp. in mosquitoes collected in Georgia by metagenomics analysis and molecular characterization

Adam R. Pollio¹, Ju Jiang^{2,3}, Sam S. Lee¹,
Jaykumar S. Gandhi¹, Brian D. Knott⁴, Tamar Chunashvili⁴,
Matthew A. Conte¹, Shannon D. Walls⁴,
Christine E. Hulseberg⁴, Christina M. Farris²,
Drew D. Reinbold-Wasson⁴ and Jun Hang^{1*}

¹Walter Reed Army Institute of Research, Silver Spring, MD, United States, ²Naval Medical Research Center, Silver Spring, MD, United States, ³Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD, United States, ⁴U.S. Army Medical Research Directorate - Georgia (USAMRD-G), Walter Reed Army Institute of Research, Tbilisi, Georgia

Arthropods have a broad and expanding worldwide presence and can transmit a variety of viral, bacterial, and parasite pathogens. A number of *Rickettsia* and *Orientia* species associated with ticks, fleas, lice, and mites have been detected in, or isolated from, patients with febrile illness and/or animal reservoirs throughout the world. Mosquitoes are not currently considered vectors for *Rickettsia* spp. pathogens to humans or to animals. In this study, we conducted a random metagenome next-generation sequencing (NGS) of 475 pools of *Aedes*, *Culex*, and *Culiseta* species of mosquitoes collected in Georgia from 2018 to 2019, identifying rickettsial gene sequences in 33 pools of mosquitoes. We further confirmed the findings of the *Rickettsia* by genus-specific quantitative PCR (qPCR) and multi-locus sequence typing (MLST). The NGS and MLST results indicate that *Rickettsia* spp. are closely related to *Rickettsia bellii*, which is not known to be pathogenic in humans. The results, together with other reports of *Rickettsia* spp. in mosquitoes and the susceptibility and transmissibility experiments, suggest that mosquitoes may play a role in the transmission cycle of *Rickettsia* spp.

KEYWORDS

pathogen discovery, *Rickettsia*, NGS, vector-borne disease surveillance, country of Georgia, One Health, metagenomic, mosquito

Introduction

Arthropod vectors, such as mosquitoes, ticks, fleas, flies, lice, and midges, have a broad and expanding worldwide presence and can transmit a variety of viral, bacterial, and parasite pathogens between animals and humans. Vector-borne diseases (VBDs) result in millions of cases of illness and hundreds of thousands of deaths each year

(Molyneux, 1998; Schmidt et al., 2013; Huntington et al., 2016; Abdad et al., 2018). VBDs pose enormous social and economic burdens on the world, particularly on those living in low-income countries and rural areas, where the public health resources are insufficient for effective transmission prevention and medical treatment (Eisen et al., 2017; Wilcox et al., 2019). In addition, the impact of climate change is increasingly affecting developed countries due to the expanded mosquito and tick habitats in these warming areas (Semenza and Suk, 2018; Wilcox et al., 2019). These interrelated issues have increased VBDs to one of the top concerns of the One Health program, a global and multidisciplinary effort to confront health issues that arise from complex environmental, agricultural, and human factors (Schmidt et al., 2013; Sánchez-Montes et al., 2021). VBDs by different pathogens or strains of a pathogen can lead to highly variable clinical symptoms ranging from mild to severe to even fatal. The systematic and in-depth surveillance of vectors, vector-borne pathogens, and genetic evolution of pathogens has become essential for the prevention and control of VBDs and prompt outbreak response (Gillespie et al., 2008; Schmidt et al., 2013; Vayssier-Taussat et al., 2015; Chala and Hamde, 2021).

Rickettsial diseases are caused by infection with members of the genera *Rickettsia* and *Orientia*, the family *Rickettsiaceae*. The agents responsible for these infections include spotted fever group rickettsiae (SFGR), typhus group rickettsiae (TGR), scrub typhus group orientia, and transition group rickettsiae (Hechemy et al., 2003; Abdad et al., 2018). Additionally, there is an ancestral group of rickettsiae, which diverged genetically earlier than the aforementioned groups and includes agents, such as *Rickettsia bellii*, which are not known to cause disease in humans (Stothard et al., 1994; Weinert et al., 2009). A number of *Rickettsia* and *Orientia* species associated with ticks, fleas, lice, and mites have been detected in, or isolated from, patients with febrile illness and/or animal reservoirs throughout the world (Blair et al., 2004; Abdad et al., 2018). Mosquitoes are not currently considered vectors for *Rickettsia* spp. pathogens. However, there are an increasing number of reports of *Rickettsia* spp. in mosquitoes, and, importantly, a recent study by Dieme C. et al. demonstrated the ability of *Anopheles gambiae* mosquitoes to act as a vector and transmit *R. felis* to mice in a laboratory setting (Socolovschi et al., 2012a,b; Dieme et al., 2015; Guo et al., 2016; Maina et al., 2017; Barua et al., 2020). More studies are needed to identify *Rickettsia* spp. in mosquitoes, characterize mosquito-borne rickettsiae, assess the role of mosquitoes in the transmission cycle of rickettsiae to humans or animals, and ultimately find evidence for human infections.

In this study, we conducted a random metagenome next-generation sequencing (NGS) of *Aedes*, *Culex*, and *Culiseta* mosquitoes captured in the country of Georgia to identify rickettsial gene sequences in mosquito pools. The findings were confirmed by *Rickettsia* genus-specific quantitative PCR (qPCR) and multi-locus sequence typing (MLST). The results support our previous report on mosquito-borne rickettsia in

the Republic of Korea (ROK), suggesting a potential role for mosquitoes in the transmission cycle of *Rickettsia* spp.

Materials and methods

Mosquito collection

Mosquitoes were collected in Georgia throughout the 2018 and 2019 surveillance seasons (April–October). Multiple collection methods were utilized to survey a diversity of mosquito species including battery powered mosquito traps—BG Sentinel 2 (Biogents AG, Regensburg, Germany), a CDC light trap (Model 1012 and 1212, John W. Hock Company, Gainesville, FL, United States), a Stealth Trap (Model 214, John W. Hock Company, Gainesville, FL, United States), and a Fay-Prince Trap (Model 812, John W. Company, Gainesville, FL, United States); mosquito aspiration, mouth aspirators (Model 612, John W. Hock Company, Gainesville, FL, United States), and an Improved Prokopack aspirator (Model 1419, John W. Hock Company, Gainesville, FL, United States); mosquito larval collection—a mosquito dipper (Model 320, John W. Hock Company, Gainesville, FL, United States). Powered traps were equipped with mosquito attractants during surveillance: dry ice (1 KG/trap/24 h) in insulated dry ice containers (John W. Hock Company, Gainesville, FL, United States) utilized in a majority of the collections; BG-Lure (Biogents AG, Regensburg, Germany) with the BG Sentinel 2 traps; CDC-LT and Fay-Prince traps used either incandescent and ultraviolet (UV) light sources in addition to CO₂. Powered traps were set to work for 24 h collection periods and all trapped and aspirated specimens were transported frozen for storage and processing. Collected larval mosquitoes were placed in sealed containers until emergence, then frozen and processed with the other specimens. Mosquitoes were morphologically identified using a stereomicroscope (Leica S4E, Leica microsystems, Germany) and the ECDC MosKey Tool.¹ Female specimens were sorted by species into pools not greater than 30 individuals and stored at −80°C. A subset of the total collection was shipped frozen from Tbilisi, Georgia to Walter Reed Army Institute of Research (WRAIR), Silver Spring, Maryland, United States.

Mosquito extraction and nucleic acids purification for metagenome sequencing

Mosquito pools were homogenized in the cell culture medium by bead beating using a Mini-Beadbeater-16 (BioSpec Products, Inc., Bartlesville, OK, United States) and centrifuged

¹ <https://www.medilabsecure.com/moskeytool.html>

TABLE 1 Primers used for PCR, nested PCR, and Sanger sequencing.

Gene	Primer	Sequence (5'-3')	References
<i>rrs</i>	16SU17F	AGAGTTTGATCCTGGCTCAG	Jiang et al., 2012
	16SOR1198R	TTCCTATAGTTCCCGGCATT	Maina et al., 2017
	16SU547F	CAGCAGCCGCGGTAATAC	Maina et al., 2017
	16SU833R	CTACCAGGGTATCTAATCCTGTT	Maina et al., 2017
23S	23SU14F	AAGAGCATTTCGGTGGATG	This study
	23SR2753R	AATCAATCGAGCTATTAGTATC	This study
	23SOR655F	TGAATTAGACCCGAAACCG	This study
	23SU1240F	TCGGAAGTGAGAATGCT	This study
	23SOR1897F	GTGAAGATGCGGAGTTC	This study
	23SR722R	CCTTCAGCGGATTTTACTC	This study
	23SOR1371R	TACGCCTTTCAGCCTCA	This study
	23SU2054R	CAAAGGGTGGTATCTCAA	This study
	CS151F	CCGGGYTTTATGTCTACTGC	Guo et al., 2016
	CS1259R	AGCTGTCTWGGTCTGCTGATT	Guo et al., 2016
<i>ompA</i>	190-70F	ATGGCGAATATTTCTCCAAAA	Fournier et al., 1998
	RompA642R	ATTACCTATGTTCGGTTAATGGCA	Jiang et al., 2012; Minh et al., 2020
	190-701R	GTTCCGTTAATGGCAGCATCT	Fournier et al., 1998
	RompA58F	GGAGTAHKTTAGAKTTTAACGG	Jiang et al., 2005
<i>ompB</i>	RompA657R	TATTTGCATCAATCSYATAAGWA	Jiang et al., 2005
	120-M59F	CCGCAGGGTTGGTAACTGC	Roux and Raoult, 2000
	120-807R	CCTTTTAGATTACCGCCTAA	Roux and Raoult, 2000
	ompB1570R	TCGCCGGTAATTRTAGCACT	Jiang et al., 2013
	RBelB-1F	ATGATGATGAATGAAGCCTCTAAT	This study
	RBelB-28F	ATTTTAGGACCTAATGGTGT	This study
	RbelB2742R	CTAGAAGTTTAGCGGACT	This study
	RbelB1427R	TCACCTTGGATTAAAGTATAGG	This study
	RbelB1283F	CTTGACATCAGATGAAGTTATG	This study
	RrD749F	TGGTAGCATTAAAGCTGATGG	Jiang et al., 2005
<i>sca4</i>	RrD928F	ATTATACACTTGC GGTAACAC	Jiang et al., 2005
	RrD1826R	TCTAAATKCTGCTGMATCAAT	Jiang et al., 2005
	RrD2685R	TTCAGTAGAAGATTAGTACCAAAT	Jiang et al., 2005

to harvest clear supernatant, as described previously (Sanborn et al., 2019, 2021). After pre-treatment of incubation with nucleases to reduce host DNA and RNA contents, clear supernatant was extracted to purify nucleic acids using the MagMAX Pathogen RNA/DNA Kit and the KingFisher Flex Purification System (Thermo Fisher Scientific) by following the manufacturer's user guides.

Metagenome sequencing

Purified nucleic acid samples were subjected to unbiased random reverse transcription and PCR amplification (RT-PCR), as described previously (Maina et al., 2017; Sanborn et al., 2019, 2021). Separate reactions of negative control (molecular biology grade water) and positive control (MS2 bacteriophage RNA) were included as quality control to estimate background noise

in the lab and to track cross contaminations. Random RT-PCR amplicons were purified, examined using the TapeStation System and D5000 ScreenTape (Agilent Technologies, Inc., Santa Clara, CA, United States) or using the Quant-iT PicoGreen dsDNA Assay (Thermo Fisher Scientific), subjected to library preparation using the Illumina DNA Prep Kit, followed by NGS using the MiSeq System and Reagent Kit v3 (600-cycle) (Illumina, San Diego, CA, United States).

Metagenomics data analyses

Next-generation sequencing data were processed and analyzed using a metagenomics analysis pipeline, which includes sequence read quality processing, host sequence removal, *de novo* sequence data assembly and contig scaffolding, megablast, and discontinuous megablast of the

TABLE 2 Summary of mosquitoes collected across Georgia from 2018 to 2019 utilized in this study.

Mosquito species by site	Total number of mosquitoes (pools)	Number of pools <i>Rickettsia</i> positive
Training base Krtsanisi^a		
<i>Aedes albopictus</i>	1 (1)	0
<i>Aedes caspius</i>	930 (63)	2 ^a
<i>Aedes surcofi</i>	34 (5)	1 ^a
<i>Aedes vexans</i>	51 (6)	0
<i>Culex pipiens</i>	1037 (71)	25 ^a
<i>Culex pusillus</i>	156 (8)	3 ^a
<i>Culex theileri</i>	17 (4)	0
<i>Culex tritaeniorhynchus</i>	253 (20)	1 ^a
Training base Kutaisi^a		
<i>Aedes albopictus</i>	3 (2)	0
<i>Culex pipiens</i>	119 (13)	0
Training base Norio^{a,b}		
<i>Aedes albopictus</i>	5 (2)	1 ^b
<i>Aedes geniculatus</i>	45 (7)	0
<i>Aedes vexans</i>	119 (9)	0
<i>Culex pipiens</i>	145 (17)	0
<i>Culex pusillus</i>	1 (1)	0
Training base Senaki and Camp Eki^a		
<i>Aedes albopictus</i>	35 (6)	0
<i>Culex pipiens</i>	568 (51)	0
Training base Vaziani^a		
<i>Aedes albopictus</i>	5 (4)	0
<i>Aedes caspius</i>	55 (17)	0
<i>Aedes surcofi</i>	12 (1)	0
<i>Culex pipiens</i>	542 (60)	0
<i>Culex theileri</i>	43 (6)	0
Other sites within Georgia^{b,c}		
<i>Aedes albopictus</i>	182 (48)	0
<i>Aedes caspius</i>	2 (2)	0
<i>Aedes vexans</i>	40 (4)	0
<i>Culex pipiens</i>	919 (71)	0
<i>Culex pusillus</i>	28 (2)	0
<i>Culex theileri</i>	20 (2)	0
<i>Culiseta longiareolata</i>	51 (4)	0

Collection methods: ^a Powered traps.

^b Larval dipping.

^c Aspiration.

contigs and unassembled single reads, and the iteration of sequence-based taxonomic entities in the specimens (Kilianski et al., 2015). Assembled sequences from the pipeline were curated using bioinformatics tools, an NGS Mapper, Geneious R10 (Biomatters Ltd.),² an Integrative Genomics Viewer (IGV) (Broad Institute),³ for GenBank submission and phylogenetic analyses. The Multiple Sequence Comparison by Log-Expectation (MUSCLE) program was used for multiple sequence alignment. A maximum likelihood phylogenetic tree was generated using IQ-TREE version 1.6.12 using these alignments with 1,000 standard non-parametric bootstrap

replicates specified (-b 1,000) (Minh et al., 2020). Output trees were visualized and adapted for figures using FigTree version v1.4.4.⁴

De novo clustering of the metagenome sequences was performed using QIIME2 command line version 2022.2 (Bolyen et al., 2019). The “vsearch dereplicate-sequences” command was used first, followed by the “vsearch cluster-features-*de-novo*” command using a percentage identity (-p) of 0.99. Abundance counts are provided for each Operational Taxonomic Unit (OTU). Principal component analysis (PCA) was calculated on this OTU abundance output using the python scikit-learn decomposition package and plotted.

Mosquito DNA extraction, rickettsial real-time PCR assay, and amplicon sequencing

DNA was extracted from 100 µl of mosquito homogenates using the DNeasy blood & tissue kit (QIAGEN), along with one negative control, and the purified DNA was eluted in the 50 µl of elution buffer. Then, a 1:10 diluted stock was prepared for use in *Rickettsia* genus-specific qPCR assays.

Two qPCR assays were attempted: the Rick17b assay targets the 17 KDa antigen gene (Jiang et al., 2012) and the RickCS assay targets the citrate synthase gene (*gltA*) (Stenos et al., 2005). The reaction final conditions and the cycling parameters were the same as reported previously (Jiang et al., 2012).

PCR was performed using primers targeting the 16S, 23S, *gltA*, *ompA*, *ompB*, and *sca4* genes (Table 1) (Fournier et al., 1998; Roux and Raoult, 2000; Jiang et al., 2005, 2012, 2013; Guo et al., 2016; Maina et al., 2017). Additional primers for the 23S rRNA gene and *ompB* gene were designed targeting the conserved sequences of *Rickettsia*, or *Rickettsia* and *Orientia* genus based on the sequence alignment using MEGA 11 with 20 *Rickettsia* species and 5 *O. tsutsugamushi* strains. Primers for *ompB* in this study were designed specifically targeting *R. bellii* (GenBank CP000087). The 25 µl reaction mixture contained 2 µl of purified DNA, the Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific), and 0.3 µM of forward and reverse primers. PCR reactions were conducted on a T-Gradient Thermocycler (Biometra, Göttingen, Germany), incubated at 98°C for 10 s followed by 35 cycles of denaturation at 98°C for 1 s, annealing at 51–61°C (based on the T_m of the primers calculated on the website⁵) for 5 s, and elongation at 72°C for 30 s. Following the completion of the amplification steps, the reaction mixtures were exposed to a final elongation step at 72°C for 2 min. PCR products were visualized with GelRed on 1.0% agarose gels following electrophoresis. The

² <https://www.geneious.com>

³ <https://igv.org>

⁴ <http://tree.bio.ed.ac.uk/software/figtree/>

⁵ www.thermofisher.com/tmcalculator

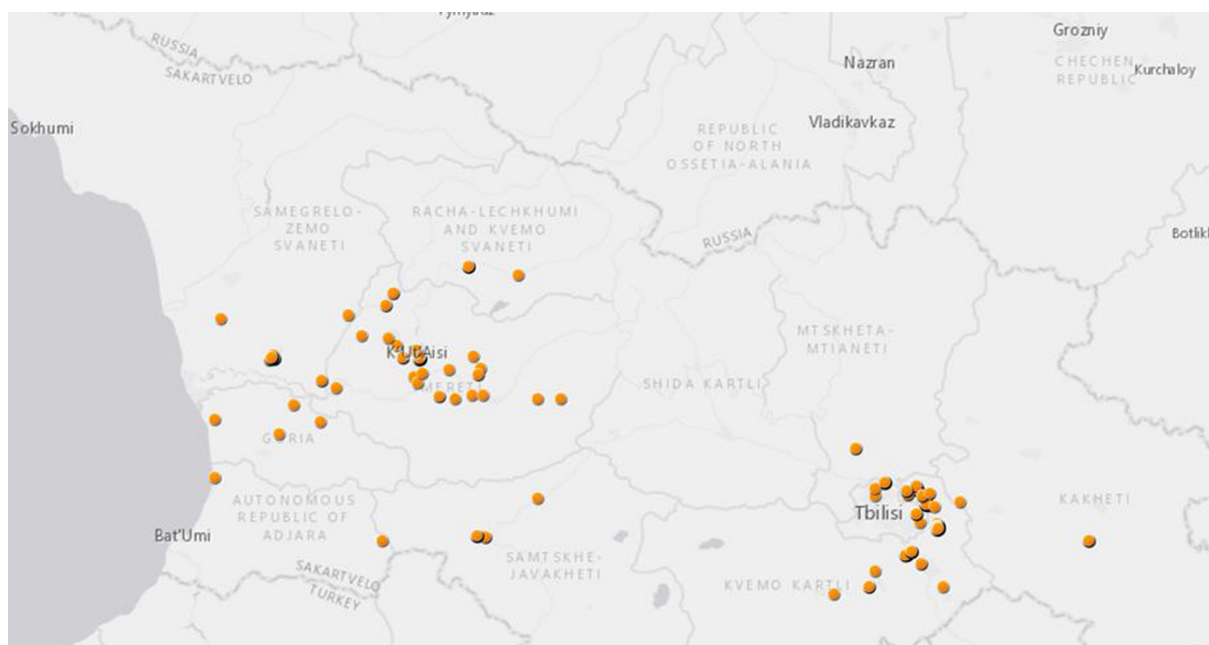


FIGURE 1

Mosquito collection in Georgia from 2018 to 2019 for this study. Each dot represents GPS coordinates of each mosquito collection site.

mastermix for PCR was prepared in a clean hood separated from where the DNA templates were added, and negative control (molecular biology grade water) was run at the same time under the same condition as the samples. The genomic DNA of *Rickettsia africae* was used as a positive control.

For Sanger sequencing, PCR amplicons were purified using the QIAquick PCR purification kit (QIAGEN). Sequencing reactions were performed for both DNA strands using the Big Dye Terminator v3.1 Ready Reaction Cycle Sequencing Kit (Thermo Fisher Scientific Applied Biosystems, Foster City, CA). After purification of sequenced products using Performa DTR Gel Filtration Cartridges (Calibre Scientific EdgeBio, Holland, OH, United States), Sanger sequencing was run on a 3500 Genetic Analyzer (Applied Biosystems). The primers used for PCR amplification were also used for the sequencing reactions. Sequences were assembled using the CodonCode aligner (CodonCode Corporation, Barnstable, MA, United States).

The nucleotide sequences from this study were deposited in GenBank with accession numbers ON960050-ON960051 for *gltA* genes, OP007139-OP007153 for 16S rRNA genes, and OP007301-OP007317 for 23S rRNA genes.

Results

In total, 475 pools (5,146 specimens) of *Aedes*, *Culex*, *Culiseta* mosquitoes were studied in this metagenomics analysis. The mosquitoes were collected at 112 Global Positioning System

(GPS) locations in 10 provinces across Georgia from 2018 to 2019 (Table 2 and Figure 1). Many mosquitoes were from two Georgian military training bases, Krtsanisi training area (KTA, 177 pools) and the Norio training area (NTA, 36 pools), both located in the southeastern province of Kvemo Kartli.

The unbiased metagenomics approach identified a large variety of viruses and other microbes associated with these pools of mosquitoes. *Rickettsia* spp. sequences were found in 33 mosquito pools, with each pool having over 500 rickettsial reads and/or assembled contig(s) of 500 bp or longer in sequence length (Table 3). As expected, only 23S and 16S ribosomal RNA genes were identified by metagenome sequencing, in which the abundant ribosome RNA transcripts were efficiently amplified and sequenced. Both Genbank nucleotide BLAST analysis (Table 3) and phylogenetic analysis (Figure 2) clearly indicated that Georgian mosquito-associated rickettsiae are closely related to *R. bellii*. The phylogeny also suggested that the rickettsiae found in Georgia clustered with *Rickettsia* spp. identified in mosquitoes collected in 2012 in the Republic of Korea (ROK) (Maina et al., 2017). Together *R. bellii* and the rickettsiae found in mosquitoes from ROK and Georgia form a clade with clear separation from SFGR and TGR.

Of the 475 mosquito pools, there are 302 pools (63.6%) of *Culex* and 173 pools (36.4%) of *Aedes*. Rickettsiae were found to be present in more *Culex* pools (29/33, 87.9%) than in *Aedes* pools (4/33, 12.1%). Additionally, 32/33 positive pools came from mosquitoes collected with powered traps. A single pool was positive from specimens collected *via* larval dipping and

TABLE 3 Summary of metagenome next-generation sequencing and comparison of assembled rickettsial gene sequences with *Rickettsia* sp. MEAM1 (*Bemisia tabaci*) strain MEAM1 (GenBank accession CP016305).

Mosquito pool ID	Mosquito species	Number of specimens	Total number of NGS reads	Total number of rickettsial reads	Rickettsial gene	Gene coverage	Percent nucleotide identity
18KTA68A-172.44	<i>Culex pipiens</i>	4	208636	594	23S rRNA	56.0%	99.87%
					16S rRNA	36.0%	99.59%
18KTA68B-172.09	<i>Culex pipiens</i>	15	880748	7398	23S rRNA	65.0%	99.78%
18KTA68B-172.19	<i>Culex tritaeniorhynchus</i>	6	119378	211	23S rRNA	50.0%	99.91%
18KTA70B-178.19	<i>Culex pipiens</i>	10	205590	4700	23S rRNA	42.0%	97.79%
					<i>gltA</i>	82.1%	98.23%
18KTA71-179.13	<i>Aedes caspius</i>	20	419430	13289	23S rRNA	88.0%	99.67%
					16S rRNA	85.0%	99.46%
					<i>gltA</i>	82.1%	98.23%
18KTA71-179.39	<i>Culex pipiens</i>	20	706380	991	23S rRNA	61.0%	99.54%
					16S rRNA	62.0%	99.26%
18KTA71-179.40	<i>Culex pipiens</i>	20	802878	2043	23S rRNA	25.0%	99.13%
					16S rRNA	43.0%	94.95%
18KTA71-179.45	<i>Culex pipiens</i>	20	517024	439	23S rRNA	78.0%	99.37%
					16S rRNA	44.0%	99.56%
18KTA71-179.46	<i>Culex pipiens</i>	20	695016	35361	23S rRNA	78.0%	99.33%
					16S rRNA	48.0%	99.59%
18KTA71-179.48	<i>Culex pipiens</i>	20	804150	5914	23S rRNA	77.0%	99.46%
					16S rRNA	73.0%	99.52%
18KTA71-179.51	<i>Culex pipiens</i>	20	767540	13438	23S rRNA	70.0%	99.44%
					16S rRNA	87.0%	99.47%
18KTA71-179.55	<i>Culex pipiens</i>	20	918674	10320	23S rRNA	95.0%	99.49%
					16S rRNA	71.0%	99.00%
18KTA71-179.57	<i>Culex pipiens</i>	20	915140	558	23S rRNA	78.0%	99.77%
					16S rRNA	62.0%	99.48%
18KTA71-179.60	<i>Culex pipiens</i>	20	332702	22833	23S rRNA	97.0%	99.67%
					16S rRNA	91.0%	99.41%
18KTA71A-180.16	<i>Culex pipiens</i>	20	811394	72731	23S rRNA	96.0%	99.29%
					16S rRNA	65.0%	98.65%
18KTA71A-180.18	<i>Culex pipiens</i>	20	735290	8188	23S rRNA	49.0%	99.64%
					16S rRNA	38.0%	98.99%
18KTA71A-180.22	<i>Culex pipiens</i>	20	854426	56583	23S rRNA	92.0%	99.65%
					16S rRNA	79.0%	95.00%
18KTA71A-180.23	<i>Culex pipiens</i>	15	897974	3046	23S rRNA	80.0%	99.64%
18KTA71A-180.25	<i>Culex pipiens</i>	20	996682	15129	23S rRNA	63.0%	99.86%
18KTA71A-180.26	<i>Culex pipiens</i>	20	748538	118	23S rRNA	15.0%	99.69%
18KTA71A-180.27	<i>Culex pipiens</i>	20	767576	81	23S rRNA	27.0%	99.13%
18KTA71A-180.29	<i>Culex pipiens</i>	20	784246	110639	23S rRNA	79.0%	99.40%
18KTA71A-180.32	<i>Culex pusillus</i>	20	1148410	330756	23S rRNA	55.0%	99.56%
					16S rRNA	67.0%	97.84%
18KTA71A-180.33	<i>Culex pusillus</i>	20	1095486	37082	23S rRNA	83.0%	99.15%
					16S rRNA	64.0%	99.18%
18KTA74A-189.01	<i>Aedes caspius</i>	20	697732	32	23S rRNA	20.0%	97.88%
18KTA74A-189.03	<i>Aedes surcoufi</i>	4	926422	3840	23S rRNA	56.0%	99.39%
18KTA74A-189.10	<i>Culex pusillus</i>	16	744216	75	23S rRNA	25.0%	99.73%
18Steg 85-113.16	<i>Aedes albopictus</i>	1	271998	3444	23S rRNA	61.0%	99.67%
					16S rRNA	34.0%	99.43%
18Steg 86-115.01	<i>Culex pipiens</i>	1	199090	122	23S rRNA	59.0%	99.92%

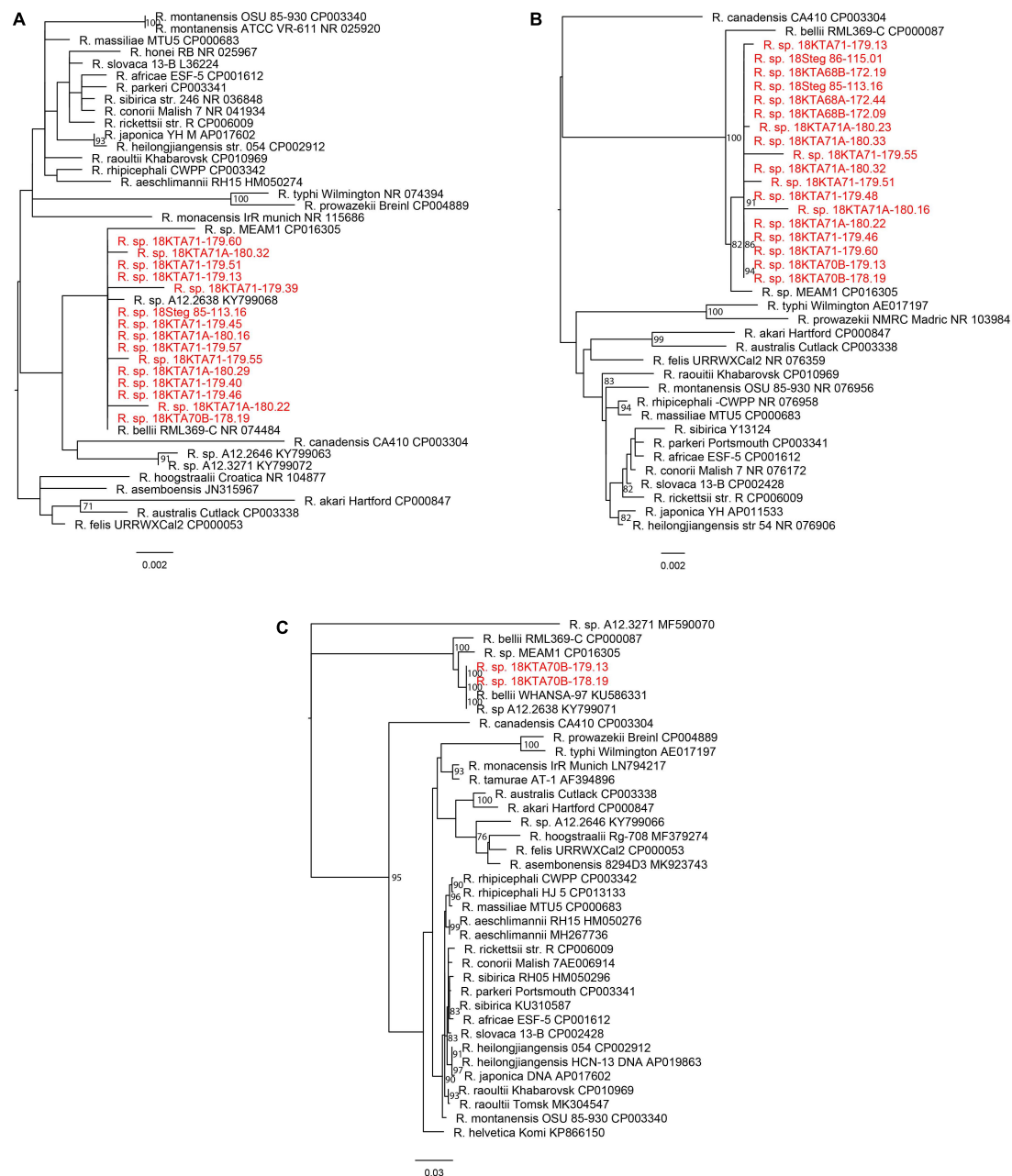


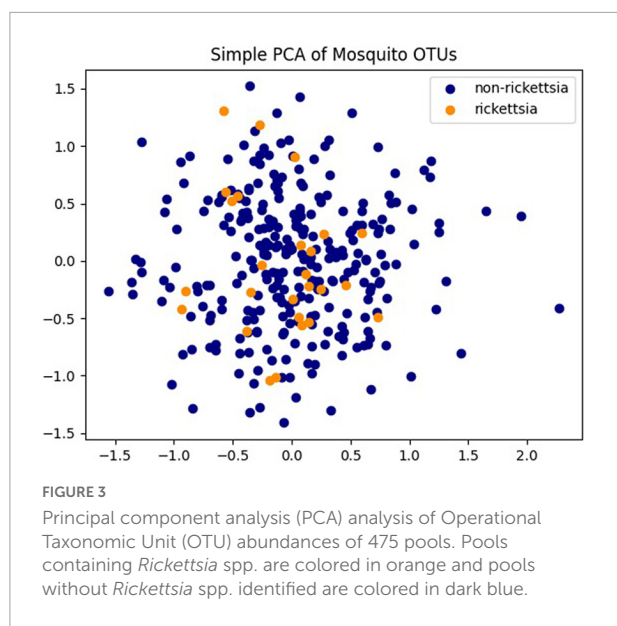
FIGURE 2

Phylogeny of *Rickettsia* sp. from this study and known rickettsial strains. (A) 16S rRNA gene; (B) 23S rRNA gene; and (C) *gltA* gene. The maximum likelihood phylogenetic tree with bootstrap support values from 1,000 replicates is shown at the branches. The scale bar represents estimated nucleotide substitutions per site. The sequences from this study are shown in red.

rearing them to adults prior to analysis. The 33 pools containing rickettsial sequences were from training areas KTA and NTA, with all but one found in KTA. The two training areas located in the cities of Kvemo Kartli and Norio, respectively, are both south of the capital city of Tbilisi and 30 km away from each other. Rickettsiae were not found in any other collection sites nearby. The PCA performed on all 475 pools revealed neither a distinct clustering of pools nor a distinct clustering of pools

containing rickettsia species or pools not containing *Rickettsia* species (Figure 3).

To confirm the detection of rickettsiae from metagenomic sequencing and to obtain sequences of additional genes for further identification, genomic DNA was extracted from total homogenate and subjected to rickettsial qPCR, PCR amplification, and Sanger sequencing. Five samples, including *Culex* pools 18KTA68B-172.09 and 18KTA70B-178.19 and



Aedes pools 18KTA70-176.01, 18KTA71-179.13, and 18Steg 85-113.16, were analyzed. Two samples, 18KTA70B-178.19 (*Culex pipiens* pool of 10 specimens) and 18KTA71-179.13 (*Aedes caspius* pool of 20 specimens), were both positive in RickCS qPCR with Ct values of 29.72 and 31.86, respectively. PCR amplification and Sanger sequencing for 16S, 23S, and *gltA* genes were also successful for these two samples. The genus specific Rick17b qPCR assay and PCR for *ompA*, *ompB*, and *sca4* genes were negative for all samples. 16S, 23S, and *gltA* sequences from Sanger sequencing were identical between the two mosquito pools and closely related to *R. bellii* (Figure 2).

Discussion

Mosquitoes are the most abundant and most widely distributed arthropod disease vectors, transmitting more VBDs than any other insects. Although rickettsial infections occur worldwide, “hot spot” focal areas of endemicity can present a risk to travelers as well. Rickettsial diseases are difficult to diagnose and can be severe or even fatal when treatment is delayed (Stewart and Stewart, 2021). Therefore, characterizing rickettsiae in mosquitoes and other arthropods from different regions in the world offers significant medical relevance. In this study, we sequenced several conserved and variable rickettsial genes with high sequence identity to *R. bellii* from three *Culex* species and three *Aedes* species. Other reports showed *Rickettsia* spp. of different genotypes in a variety of mosquito species, including *R. felis*, which is a tick-borne human pathogen, in multiple countries (Socolovschi et al., 2012a,b; Maina et al., 2017; Zhang et al., 2019; Li et al., 2022). Thus far, there has been no clear evidence that rickettsiae are transmitted from

mosquitoes to humans or animals outside of the laboratory setting, however, close attention and more investigation are needed to continue exploring the role of mosquitoes in the transmission cycle of rickettsiae and the potentiality of mosquito-borne rickettsioses.

This study detected rickettsial bacteria within multiple species of mosquito from the *Aedes* and *Culex* genera. We found one *Rickettsia* spp. positive pool from a larval dipped specimen, *Aedes albopictus* from NTA, which after the collection was allowed to emerge as an adult for identification and analysis. This result indicates that the vertical transmission of rickettsial pathogens is possible (trans-ovarial and trans-stadial). Our study only utilized female mosquitoes; however, previous studies have found male mosquitoes positive for *R. felis* (Zhang et al., 2016). These provide evidence that not only can rickettsia bacteria be acquired in the environment by mosquitoes but they can also be transmitted vertically.

This study showed the value of metagenomics analysis as an effective tool to support One Health efforts. The unbiased approach provides data for a comprehensive understanding of genetic contents in the study subject and possibly shed light on associations and interactions among host, viruses, other microbes, parasites, etc. Metagenomics analysis also reveals focus topics for further study. In our study, metagenome NGS detected the presence of rickettsial sequence reads in a significant number of Georgian mosquito pools, but, due to the relative abundance of rRNA (*rrs*) genes, these reads were limited to 16S and 23S rRNA genes. The results led to the application of established molecular characterization methods on selected samples to probe relatedness. It is intriguing that, in spite of the large number of NGS reads of rickettsial sequences, we were only able to succeed in relatively few rickettsial qPCR and MLST experiments. Further study is warranted to test more pools and obtain deeper sequencing data on a panel of rickettsial genes, including *rrs*, *gltA*, 17 kDa antigen, *ompB*, *ompA*, *sca4*, and *groEL* gene, to achieve taxonomic classification on the species level. We plan to use the hybridization NGS targeted enrichment method to obtain full gene sequences to complete MLST rickettsial genotyping.

We will continue the study to obtain solid evidence to further investigate the observations made in this work. Rickettsial DNA was found in both *Culex* and *Aedes* mosquitoes with *Culex* pools positive at a much higher rate than *Aedes* pools. It is noteworthy that all but one rickettsia-containing mosquito pool were from KTA, given the close proximity of KTA and NTA to each other. Future work will survey a broader diversity of mosquito species by surveying from more collection sites and conducting a statistical analysis on rickettsial carriage rates for mosquito species and the geographic distribution of mosquito-associated *Rickettsia* spp. in Georgia.

In addition to more robust and comprehensive geosurveillance and characterization studies for mosquito-

associated rickettsiae, further work on the transmission and pathogenic potential is also needed. With increasing evidence on the presence of mosquito species associated with diverse rickettsial genotypes in different parts of the world, these studies are important for addressing fundamental questions on the life cycle, transmissibility, and pathogenicity of rickettsiae to humans or animals.

Data availability statement

The original contributions presented in this study are publicly available. This data can be found here: GenBank, OP007139-OP007153, OP007301-OP007317, ON960050, and ON960051.

Author contributions

BK, CH, SW, and DR-W: sample acquisition. AP, JJ, SL, JG, BK, TC, MC, SW, CF, DR-W, and JH: methodology and data analysis. MC, SW, CH, CF, DR-W, and JH: project administration. AP, JJ, SL, JG, BK, TC, MC, SW, CH, CF, DR-W, and JH: writing and editing. All authors contributed to the article and approved the submitted version.

Funding

Funding was provided by the Global Emerging Infections Surveillance and Response System (GEIS), Division of the Armed Forces Health Surveillance Branch, projects P0020_21_NM, P0171_21_WR, P0175_22_WR. NMRC work unit number is A0047.

References

- Abdad, M. Y., Abou Abdallah, R., Fournier, P. E., Stenos, J., and Vasoo, S. (2018). A concise review of the epidemiology and diagnostics of Rickettsioses: *Rickettsia* and *Orientia* spp. *J. Clin. Microbiol.* 56, e1728–e1717. doi: 10.1128/JCM.01728-17
- Barua, S., Hoque, M. M., Kelly, P. J., Poudel, A., Adekanmbi, F., Kalalah, A., et al. (2020). First report of *Rickettsia felis* in mosquitoes, USA. *Emerg. Microbes Infect.* 9, 1008–1010.
- Blair, P. J., Jiang, J., Schoeler, G. B., Moron, C., Anaya, E., Cespedes, M., et al. (2004). Characterization of spotted fever group rickettsiae in flea and tick specimens from northern Peru. *J. Clin. Microbiol.* 42, 4961–4967. doi: 10.1128/JCM.42.11.4961-4967.2004
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Author correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37:1091. doi: 10.1038/s41587-019-0252-6
- Chala, B., and Hamde, F. (2021). Emerging and re-emerging vector-borne infectious diseases and the challenges for control: A review. *Front. Public Health* 9:715759. doi: 10.3389/fpubh.2021.715759
- Dieme, C., Bechah, Y., Socolovschi, C., Audoly, G., Berenger, J. M., Faye, O., et al. (2015). Transmission potential of *Rickettsia felis* infection by *Anopheles gambiae* mosquitoes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8088–8093.
- Eisen, R. J., Kugeler, K. J., Eisen, L., Beard, C. B., and Paddock, C. D. (2017). Tick-Borne Zoonoses in the United States: Persistent and Emerging Threats to Human Health. *ILAR J.* 58, 319–335. doi: 10.1093/ilar/ilx005
- Fournier, P. E., Roux, V., and Raoult, D. (1998). Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA. *Int. J. Syst. Bacteriol.* 48, 839–849.

Acknowledgments

We thank James S. Hilaire, Nicole R. Nicholas, Tuan K. Nguyen, and April N. Griggs for their assistance in project management and sample tracking, storage, and retrieval.

Conflict of interest

JJ was employed by Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author Disclaimer

The material has been reviewed by the authors' respective institutions. There is no objection to its presentation and/or publication. The views expressed here are those of the authors and do not reflect the official policy of the Department of the Army, Department of the Navy, Department of Defense, or the U.S. Government. This is the work of U.S. government employees and may not be copyrighted (17 USC 105).

- Gillespie, J. J., Williams, K., Shukla, M., Snyder, E. E., Nordberg, E. K., Ceraul, S. M., et al. (2008). *Rickettsia* phylogenomics: Unwinding the intricacies of obligate intracellular life. *PLoS One* 3:e2018. doi: 10.1371/journal.pone.0002018
- Guo, W. P., Tian, J. H., Lin, X. D., Ni, X. B., Chen, X. P., Liao, Y., et al. (2016). Extensive genetic diversity of Rickettsiales bacteria in multiple mosquito species. *Sci. Rep.* 6:38770. doi: 10.1038/srep38770
- Hechemy, K. E., Avsic-Zupanc, T., Childs, J. E., and Raoult, D. A. (2003). Rickettsiology: Present and future directions: Preface. *Ann. N. Y. Acad. Sci.* 990, xvii–xx. doi: 10.1111/j.1749-6632.2003.tb07330.x
- Huntington, M. K., Allison, J., and Nair, D. (2016). Emerging vector-borne diseases. *Am. Fam. Physician* 94, 551–557.
- Jiang, J., Blair, P. J., Felices, V., Moron, C., Cespedes, M., and Anaya, E. (2005). Phylogenetic analysis of a novel molecular isolate of spotted fever group *Rickettsiae* from northern Peru: *Candidatus Rickettsia andeanae*. *Ann. N. Y. Acad. Sci.* 1063, 337–342. doi: 10.1196/annals.1355.054
- Jiang, J., Maina, A. N., Knobel, D. L., Cleaveland, S., Laudoit, A., Wamburu, K., et al. (2013). Molecular detection of *Rickettsia felis* and *Candidatus Rickettsia asemboensis* in fleas from human habitats, Asembo, Kenya. *Vector Borne Zoonotic Dis.* 13, 550–558. doi: 10.1089/vbz.2012.1123
- Jiang, J., Stromdahl, E. Y., and Richards, A. L. (2012). Detection of *Rickettsia parkeri* and *Candidatus Rickettsia andeanae* in *Amblyomma maculatum* gulf coast ticks collected from humans in the United States. *Vector Borne Zoonotic Dis.* 12, 175–182.
- Kilianski, A., Carcel, P., Yao, S., Roth, P., Schulte, J., Donarum, G. B., et al. (2015). Pathosphere.org: Pathogen detection and characterization through a web-based, open source informatics platform. *BMC Bioinformatics* 16:416. doi: 10.1186/s12859-015-0840-5
- Li, F., Tian, J., Wang, L., Yang, Z., Lu, M., Qin, X., et al. (2022). High Prevalence of *Rickettsia bellii* in mosquitoes from Eastern China. *J. Med. Entomol.* 59, 390–393. doi: 10.1093/jme/tjab177
- Maina, A. N., Klein, T. A., Kim, H. C., Chong, S. T., Yang, Y., Mullins, K., et al. (2017). Molecular characterization of novel mosquito-borne *Rickettsia* spp. from mosquitoes collected at the demilitarized zone of the republic of Korea. *PLoS One* 12:e0188327. doi: 10.1371/journal.pone.0188327
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). Corrigendum to: IQ-TREE 2: New models and efficient methods for phylogenetic inference in the eonomic era. *Mol. Biol. Evol.* 37:2461. doi: 10.1093/molbev/msaa131
- Molyneux, D. H. (1998). Vector-borne parasitic diseases—an overview of recent changes. *Int. J. Parasitol.* 28, 927–934. doi: 10.1016/s0020-7519(98)00067-8
- Roux, V., and Raoult, D. (2000). Phylogenetic analysis of members of the genus *Rickettsia* using the gene encoding the outer-membrane protein rOmpB (ompB). *Int. J. Syst. Evol. Microbiol.* 50, 1449–1455. doi: 10.1099/00207713-50-4-1449
- Sanborn, M. A., Klein, T. A., Kim, H. C., Fung, C. K., Figueroa, K. L., Yang, Y., et al. (2019). Metagenomic analysis reveals three novel and prevalent mosquito viruses from a single pool of *Aedes vexans nipponii* collected in the republic of Korea. *Viruses* 11:222. doi: 10.3390/v11030222
- Sanborn, M. A., Wuertz, K. M., Kim, H. C., Yang, Y., Li, T., Pollett, S. D., et al. (2021). Metagenomic analysis reveals *Culex* mosquito virome diversity and Japanese encephalitis genotype V in the republic of Korea. *Mol. Ecol.* 30, 5470–5487. doi: 10.1111/mec.16133
- Sánchez-Montes, S., Colunga-Salas, P., Lozano-Sardaneta, Y. N., Zazueta-Islas, H. M., Ballados-González, G. G., Salceda-Sánchez, B., et al. (2021). The genus *Rickettsia* in Mexico: Current knowledge and perspectives. *Ticks Tick Borne Dis.* 12:101633. doi: 10.1016/j.ttbdis.2020.101633
- Schmidt, K., Dressel, K. M., Niedrig, M., Mertens, M., Schüle, S. A., and Groschup, M. H. (2013). Public health and vector-borne diseases - a new concept for risk governance. *Zoonoses Public Health* 60, 528–538.
- Semenza, J. C., and Suk, J. E. (2018). Vector-borne diseases and climate change: A European perspective. *FEMS Microbiol. Lett.* 365:fnx244.
- Socolovschi, C., Pages, F., Ndiath, M. O., Ratmanov, P., and Raoult, D. (2012a). *Rickettsia* species in African *Anopheles* mosquitoes. *PLoS One* 7:e48254. doi: 10.1371/journal.pone.0048254
- Socolovschi, C., Pagés, F., and Raoult, D. (2012b). *Rickettsia felis* in *Aedes albopictus* mosquitoes, Libreville, Gabon. *Emerg. Infect. Dis.* 18, 1687–1689. doi: 10.3201/eid1810.120178
- Stenos, J., Graves, S. R., and Unsworth, N. B. (2005). A highly sensitive and specific real-time PCR assay for the detection of spotted fever and typhus group *Rickettsiae*. *Am. J. Trop. Med. Hyg.* 73, 1083–1085.
- Stewart, A. G., and Stewart, A. G. A. (2021). An update on the laboratory diagnosis of *Rickettsia* spp. infection. *Pathogens* 10:1319.
- Stothard, D. R., Clark, J. B., and Fuerst, P. A. (1994). Ancestral divergence of *Rickettsia bellii* from the spotted fever and typhus groups of *Rickettsia* and antiquity of the genus *Rickettsia*. *Int. J. Syst. Bacteriol.* 44, 798–804. doi: 10.1099/00207713-44-4-798
- Vayssier-Taussat, M., Kazimirova, M., Hubalek, Z., Hornok, S., Farkas, R., Cosson, J. F., et al. (2015). Emerging horizons for tick-borne pathogens: From the ‘one pathogen-one disease’ vision to the pathobiome paradigm. *Future Microbiol.* 10, 2033–2043. doi: 10.2217/fmb.15.114
- Weinert, L. A., Werren, J. H., Aebi, A., Stone, G. N., and Jiggins, F. M. (2009). Evolution and diversity of *Rickettsia* bacteria. *BMC Biol.* 7:6. doi: 10.1186/1741-7007-7-6
- Wilcox, B. A., Echaubard, P., de Garine-Wichatitsky, M., and Ramirez, B. (2019). Vector-borne disease and climate change adaptation in African dryland social-ecological systems. *Infect. Dis. Poverty* 8:36. doi: 10.1186/s40249-019-0539-3
- Zhang, J., John Kelly, P., Lu, G., Cruz-Martinez, L., and Wang, C. (2016). *Rickettsia* in mosquitoes, Yangzhou, China. *Emerg. Microbes Infect.* 5:e108. doi: 10.1038/emi.2016.107
- Zhang, J., Lu, G., Li, J., Kelly, P., Li, M., Wang, J., et al. (2019). Molecular detection of *Rickettsia felis* and *Rickettsia bellii* in mosquitoes. *Vector Borne Zoonotic Dis.* 19, 802–809. doi: 10.1089/vbz.2019.2456



OPEN ACCESS

EDITED BY

Ratree Takhampunya,
Armed Forces Research Institute of Medical
Science, Thailand

REVIEWED BY

Junping Peng,
Institute of Pathogen Biology (CAMS),
China
Andrew Whitelaw,
Stellenbosch University,
South Africa
Susanna J Sabin,
Centers for Disease Control and Prevention
(CDC), United States

*CORRESPONDENCE

Feng-sheng Liu
liu_fs_doctor@163.com
Shi-bing Qin
qinshibing17@163.com

[†]These authors have contributed equally to
this work and share first authorship

SPECIALTY SECTION

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

RECEIVED 15 August 2022

ACCEPTED 17 November 2022

PUBLISHED 07 December 2022

CITATION

Li Y, Yao X-w, Tang L, Dong W-j, Lan T-l,
Fan J, Liu F-s and Qin S-b (2022)
Diagnostic efficiency of metagenomic
next-generation sequencing for suspected
spinal tuberculosis in China: A multicenter
prospective study.
Front. Microbiol. 13:1018938.
doi: 10.3389/fmicb.2022.1018938

COPYRIGHT

© 2022 Li, Yao, Tang, Dong, Lan, Fan, Liu
and Qin. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Diagnostic efficiency of metagenomic next-generation sequencing for suspected spinal tuberculosis in China: A multicenter prospective study

Yuan Li^{1†}, Xiao-wei Yao^{2†}, Liang Tang^{3†}, Wei-jie Dong²,
Ting-long Lan², Jun Fan², Feng-sheng Liu^{3*} and Shi-bing Qin^{1*}

¹Department of Orthopedics, Beijing Chest Hospital, Capital Medical University, Beijing, China,

²Department of Orthopedics, Hebei Chest Hospital, Shijiazhuang, China, ³Department of
Orthopedics, Tianjin Haihe Hospital, Tianjin, China

Background: The pathogens of suspected spinal tuberculosis (TB) include TB and non-TB bacteria. A rapid and effective diagnostic method that can detect TB and non-TB pathogens simultaneously remains lacking. Here, we used metagenomic next-generation sequencing (mNGS) to detect the pathogens in patients with suspected spinal TB.

Methods: The enrolled patients with suspected spinal TB were regrouped three times into patients with spinal infection and controls, patients with spinal TB and controls, and patients with non-TB spinal infection and controls. We tested the three groups separately by using mNGS and conventional detection methods.

Results: Ultimately, 100 patients were included in this study. Pathogens were detected in 82 patients. Among the 82 patients, 37 had TB and 45 were infected with other bacteria. In patients with spinal infection, the sensitivity of the mNGS assay was higher than that of culture and pathological examination ($p < 0.001$, $p < 0.001$). The specificity of the mNGS assay was not statistically different from that of culture and pathological examination ($p = 1.000$, $p = 1.000$). In patients with spinal TB, no statistical difference was found between the sensitivity of the mNGS assay and that of Xpert and T-SPOT.TB ($p = 1.000$, $p = 0.430$). The sensitivity of the mNGS assay was higher than that of MGIT 960 culture and pathological examination ($p < 0.001$, $p = 0.006$). The specificities of the mNGS assay, Xpert, MGIT 960 culture, and pathological examination were all 100%. The specificity of T-SPOT.TB (78.3%) was lower than that of the mNGS assay (100%; $p < 0.001$). In patients with non-TB spinal infection, the sensitivity of the mNGS assay was higher than that of bacterial culture and pathological examination ($p < 0.001$, $p < 0.001$). The specificity of the mNGS assay was not statistically different from that of bacterial culture and pathological examination ($p = 1.000$, $p = 1.000$).

Conclusion: Data presented here demonstrated that mNGS can detect TB and non-TB bacteria simultaneously, with high sensitivity, specificity and short detection time. Compared with conventional detection methods, mNGS is a more rapid and effective diagnostic tool for suspected spinal TB.

KEYWORDS

diagnosis, spine, tuberculosis, metagenomic next-generation sequencing, infection

Introduction

Tuberculosis (TB) is the leading cause of death from a single infectious agent worldwide and remains one of the top 10 causes of death. According to the 2021 global TB report, approximately 5.8 million people fell ill with TB in 2020, and 1.5 million people died of TB globally (World Health Organization, 2021). Osteoarticular TB is one of the common types of extrapulmonary TB, which currently accounts for 15~20% of TB case in Asia, and spinal TB accounts for approximately half of bone TB cases (Pigrau-Serrallach and Rodríguez-Pardo, 2013). In recent years, the proportion of non-TB infections in patients with suspected spinal TB has increased gradually (Tsantes et al., 2020). Although non-TB spinal infection and spinal TB have similar clinical symptoms and imaging findings, their treatment is contradictory (Kafle et al., 2022). Thus, their differential diagnosis is required. The misdiagnosis and delayed treatment of suspected spinal TB often lead to death or disability. Therefore, making a rapid and accurate diagnosis becomes the key to controlling these two diseases.

Conventional detection methods, such as culture, targeted nucleic acid amplification tests, and immunological assays, can be challenging due to the wide variety of pathogens that cause clinically indistinguishable diseases. Bacterial culture has been considered as the gold standard for the diagnosis of infectious diseases. However, bacterial culture cannot detect many different types of bacteria simultaneously; for example, BACTEC MGIT 960 (MGIT 960) culture can only detect TB and not other bacteria (Ma et al., 2020; Tsang et al., 2020). Meanwhile, the administration of antimicrobial drugs before culture reduces organism recovery rates. In addition, bacterial culture is a time-consuming method, i.e., the general bacterial culture takes a few days, and TB bacterial culture takes 2 months (Stangenberg et al., 2021; Mingora et al., 2022). Conventional PCR-based tests are targeted methods, and thus cannot be used effectively without some prior knowledge regarding the identity of the pathogen in question. Although Xpert MTB/RIF (Xpert) has good TB detection ability, it cannot detect other bacteria (Ma J. et al., 2022; Wilson et al., 2014). Immunological tests, such as T-SPOT.TB, have poor specificity for TB infections and cannot be used as the main basis for diagnosis; moreover, they reflect previous TB infections (Ryang and Akbar, 2020; Berrocal-Almanza et al., 2022). Therefore, a rapid, accurate and extensive method for the detection of pathogenic microorganisms is the key to solving the above problem.

Metagenomic next-generation sequencing (mNGS) is an unbiased approach to the detection of pathogens. It can overcome the limitations of current diagnostic tests, thus allowing for hypothesis-free, culture-independent pathogen detection directly from clinical specimens. It can cover almost all clinical pathogens that range from viruses to bacteria, fungi, and parasites, and has a short detection time (Gu et al., 2019). In particular, when conventional testing methods cannot provide information about pathogens, mNGS can provide a timely and valuable reference for clinicians in most cases. Previous studies have shown the promises of mNGS as a diagnostic tool for infectious diseases (Simmer et al., 2018).

However, only a few reports on the diagnostic efficiency of mNGS for suspected spinal TB exist; furthermore, existing studies have shortcomings, such as small sample sizes and the lack of controls, and are mostly retrospective (Ruppé et al., 2017; Zhao et al., 2020; Ma C. et al., 2022). Therefore, the diagnostic ability of mNGS for suspected spinal TB has not been accurately evaluated. This situation limits the application of mNGS in the clinical diagnosis of suspected spinal TB. Thus, a multicenter prospective study with a large sample size and control group was designed to evaluate systematically the efficacy of mNGS in the diagnosis of suspected spinal TB and to guide its clinical application.

Materials and methods

Patient enrollment

From January 2021 to December 2021, patients with suspected spinal TB were prospectively enrolled in the Orthopedics Department of three TB-specialized hospitals: Beijing Chest Hospital, Hebei Chest Hospital, and Tianjin Haihe Hospital. The patients enrolled in the study had clinical manifestations suggestive of suspected spinal TB. These manifestations included (i) persistent back pain lasted for at least 3 weeks, (ii) low fever (<38°C), (iii) elevated erythrocyte sedimentation rate (male >15 mm/h, female >20 mm/h), and (iv) spinal magnetic resonance imaging abnormalities.

Sample collection and processing

Specimens were collected *via* surgery or CT-guided puncture. Specimens included granulation tissue and pus. Blood specimens were obtained when all patients were enrolled in this study. The samples then were sent to the laboratory for further processing. Granulation tissue samples were cut into small pieces on a disposable Petri dish support by using a scalpel. Each granulation tissue sample was weighed, and then added with phosphate-buffered saline at the rate of 1 g/ml. The mixture was homogenized with a FastPrep-24 instrument (MP Biomedicals Europe) for 100 s at 6 m/s by using MP Bio FASTPREP-24. During homogenization, the tube containing the mixture was removed from the FastPrep-24 instrument every 20 s and cooled on ice for 30 s. Pus samples or the mixtures of homogenized granulation tissue (a 600 µl volume of each) from all patients were each mixed with 1 g of 0.5 mm diameter glass beads and then placed on a vortex mixer for 30 min at 3,000 rpm.

Conventional testing

Conventional tests included Xpert assay, pathological examination, MGIT 960 culture, and T-SPOT.TB test. The experimental procedure was consistent with our previous studies (Li et al., 2018). Bacterial culture examination: Specimens were tested by using aerobic and anaerobic bacterial cultures. Briefly,

abscess specimens were plated and incubated for up to 5 days on 5% sheep blood and MacConkey agar for aerobic culture and on 5% sheep blood agar for anaerobic culture. Bacterial identification was performed with VITEK 2 Compact system (Bio Mérieux, France). Operation was conducted in accordance with the manufacturer's instructions.

mNGS

DNA was extracted from 300 µl of each pretreated sample by using a TIANamp Micro DNA Kit (Tiangen Biotech, Beijing, China) in accordance with the manufacturer's instructions. Purified DNA was fragmented into 200–300 bp segments by using ultrasound followed by end-repair, ligation with multiplex barcode adapters, and PCR amplification to complete the construction of DNA libraries. After the molarities of DNA libraries were estimated by using indexing PCR, the DNA concentrations were determined *via* the DNA Qubit Assay (Thermo Fisher). Meanwhile, DNA quality was evaluated electrophoretically by using an Agilent 2,100 system (Agilent Technologies, Santa Clara, CA, United States). Up to 20 qualified DNA libraries were pooled, and then pooled libraries were subjected to DNA sequencing analysis by using the MGISEQ-2000 platform (MGI Tech Co., Shenzhen, China).

Bioinformatics analysis

Low-quality sequences and adaptor sequences were first removed to generate clean reads. Subsequently, sequences mapped to the human reference genome (hg19) were subtracted from the clean reads by using Burrows–Wheeler Alignment software (version 0.7.10). Nonhuman sequence reads from each sample were submitted to the Genome Sequence Archive of the Beijing Institute of Genomics, Chinese Academy of Sciences under the accession number PRJCA000880. Additionally, the remaining data were further mapped against the RefSeq Microbial Genome Database of viruses, bacteria, fungi, and parasites by using Burrows–Wheeler Alignment software (version 0.7.10). RefSeq Microbial Genome Database was created and maintained by the National Center of Biotechnology Information. RefSeq analysis yielded 1,798 whole-genome sequences matching the DNA of viral taxa, 6,350 bacterial genomes or scaffolds, 1,064 pathogenic fungi of human infections, and 234 parasites associated with human diseases (Wang et al., 2019). Reporting criteria for infectious pathogens identified by using mNGS included: (i) >30% relative abundance at the genus level in bacteria or fungi; (ii) at least three unique reads from a single viral, bacterial, or fungal species; and (iii) at least one unique read matching *M. TB* complex species (Wang et al., 2019). If more than one pathogen was detected, the species present with the greatest relative abundance yielding the highest number of unique reads was deemed as the probable species associated with osteoarticular infection in that patient.

Patient categories

On the basis of the composite reference standard (CRS), the patients were categorized into three groups: (1) cases with spinal TB infection (including A: cases positive for mycobacterial culture, B: cases with the pathological result of TB and good response to anti-TB therapy, C: cases with the Xpert result of TB and good response to anti-TB therapy); (2) non-TB spinal infection cases (including A: cases with positive bacterial culture, B: cases with the pathological result of infection and good response to anti-infection therapy, and C: cases with the pathological result of inflammation and good response to anti-infection therapy); and (3) non-infection spinal diseases cases (negative results for spinal TB infection and non-TB spinal infection test, and patient improved without receiving anti-TB and anti-infection therapy). Pathological result of TB includes typical tuberculous granuloma or positive acid-fast staining. Pathological result of infection includes presence of white blood cells and pus cells. A good outcome was defined as follows: (1) resolution of clinical symptoms due to infection, (2) improvement of osteoarticular function, and (3) improvement of inflammation, as indicated by inflammatory biomarkers and radiological features (Wieland et al., 2012).

Statistical analysis

The demographic and clinical data of the study subjects were collected by using case report forms. The data included gender, age, comorbidities, clinical symptoms, laboratory results, radiological features, and treatment regimens. All data were entered through double manual data entry with the EpiData Entry program, version 3.1 (EpiData Association, Odense, Denmark). Statistical analysis was performed by using SPSS software, version 20.0 (IBM SPSS, Chicago, IL, United States). Chi-square test and Fisher's exact test were used for categorical variables, whereas *t*-test or Mann–Whitney *U*-test was used for continuous variables, as appropriate. A two-sided value of $p < 0.05$ was considered statistically significant.

Results

Study patients

A total of 114 consecutive patients with suspected spinal TB were prospectively enrolled. Subsequently, specimens were obtained from 100 of the 114 patients through surgery or CT-guided puncture. Ultimately, 100 patients were included in this study. In accordance with the CRS, 38 patients were diagnosed with spinal TB, 53 patients were diagnosed with non-TB spinal infection, and nine patients were diagnosed with non-infectious spinal diseases. Table 1 shows the demographic characteristics of the studied patients in different categories.

TABLE 1 Clinical characteristics of the studied patients.

Clinical characteristics	Spinal TB infection (n = 38)	Non-TB spinal infection (n = 53)	Non-infection spinal diseases (n = 9)
Age, years (mean \pm SD)	44.1 \pm 15.9	51.4 \pm 16.5	49.00 \pm 17.2
Female, n (%)	20 (52.6%)	23 (43.4%)	5 (55.6%)
Site of spine lesion, n (%)			
Lumbar vertebra	19 (50.0%)	36 (67.9%)	4 (44.4%)
Thoracic vertebra	16 (42.1%)	13 (24.5%)	5 (55.6%)
Cervical vertebra	3 (7.9%)	4 (7.6%)	0 (0.0%)
Clinical symptoms			
Fever	13 (34.2%)	31 (58.5%)	0 (0.0%)
Pain	28 (73.7%)	43 (81.1%)	7 (77.8%)
Complications			
Pulmonary tuberculosis	7 (18.4%)	0 (0.0%)	0 (0.0%)
Diabetes	8 (21.1%)	17 (32.1%)	1 (11.1%)
Autoimmune disease	5 (13.2%)	9 (17.0%)	0 (0.0%)

N: number of patients, TB: tuberculosis

mNGS assay

The detection time of mNGS ranged from 15.0 to 20.5 h, with an average of 17.7 ± 1.7 h. The numbers of sequence reads ranged from 3.2×10^6 to 6.3×10^7 reads, with an average of $(2.7 \pm 2.0) \times 10^7$ reads per specimen. The sequencing depth of mNGS for pathogens ranges from $1.0 \times$ to $6.5 \times$, with an average of $(3.4 \pm 1.7) \times$ per specimen (Supplementary Table S1). The bacteria detected by mNGS and culture in this study were consistent.

Pathogen composition

In this study, pathogens were detected in 82 patients with spinal infection by using mNGS, MGIT 960 culture, bacterial culture, and Xpert. Among the 82 patients, 37 had TB and 45 were infected with other bacteria. *Brucella*, *Staphylococcus aureus*, *Escherichia coli*, fungi, *Streptococcus anginosus*, and *Klebsiella pneumoniae* accounted for 28.9% (13/45), 22.2% (10/45), 8.9% (4/45), 6.7% (3/45), 6.7% (3/45), and 4.4% (2/45) of the 45 non-TB bacteria. In addition, 10 of the 45 non-tuberculous bacteria were detected only once. The 3 fungi included 1 *Candida albicans*, 1 *Candida glabrata* and 1 *Candida parapsilosis*. Figure 1 shows the composition of pathogens detected in 82 patients with spinal infection.

Performance of mNGS and conventional detection methods in patients with spinal infection

In accordance with the CRS, patients with spinal infection included patients with spinal TB and patients with non-TB spinal

infection, and patients with non-infection spinal diseases were used as the control group. In all patients with spinal infection, the sensitivities of the mNGS assay, culture (including MGIT 960 and bacterial cultures), and pathological examination were 89.0% (81/91), 28.1% (25/89), 42.9% (30/70), respectively. The sensitivity of the mNGS assay was higher than that of culture and pathological examination. Moreover, the sensitivity of the mNGS assay was statistically different from that of culture and pathological examination ($p < 0.001$, $p < 0.001$). In all patients with spinal infection, the specificities of the mNGS assay, culture (including MGIT 960 and bacterial cultures), and pathological examination were 88.9% (8/9), 100.0% (9/9), and 100.0% (9/9), respectively. The specificity of the mNGS assay was not statistically different from that of culture and pathological examination ($p = 1.000$, $p = 1.000$). Table 2 shows the performances of mNGS and conventional detection methods in patients with spinal infection.

Performance of mNGS and conventional detection methods in patients with spinal TB

The enrolled patients were divided into two groups in accordance with the CRS. One group comprised patients with spinal TB, and the other group, which included patients with non-TB spinal infection and non-infection spinal diseases, served as the control. In patients with spinal TB, the sensitivities of the mNGS assay, Xpert, MGIT 960 culture, pathological examination, and T-SPOT.TB. were 94.7% (36/38), 94.6% (35/37), 45.9% (17/37), 67.9% (19/28), and 89.2% (33/37) respectively. No statistical difference was found between the sensitivity of the mNGS assay and that of Xpert and T-SPOT.TB ($p = 1.000$, $p = 0.430$). Compared with that of MGIT 960 culture and pathological examination, the sensitivity of the mNGS assay was higher and was statistically different ($p < 0.001$, $p = 0.006$). In patients with spinal TB, the specificities of the mNGS assay, Xpert, MGIT 960 culture, and pathological examination were all 100%. In patients with spinal TB, the specificity of T-SPOT.TB (78.3%) was lower than that of the mNGS assay (100.0%), and the difference between the specificities of these assays were statistically significant ($p < 0.001$). Table 3 shows the performances of mNGS and conventional detection methods in patients with spinal TB.

Performance of mNGS and conventional detection methods in patients with non-TB spinal infection

The enrolled patients were divided into two groups on the basis of the CRS. One group comprised patients with non-TB spinal infections, and the other group, which served as the control, constituted patients with spinal TB and non-infection spinal diseases. The sensitivities of the mNGS assay, bacterial culture, and

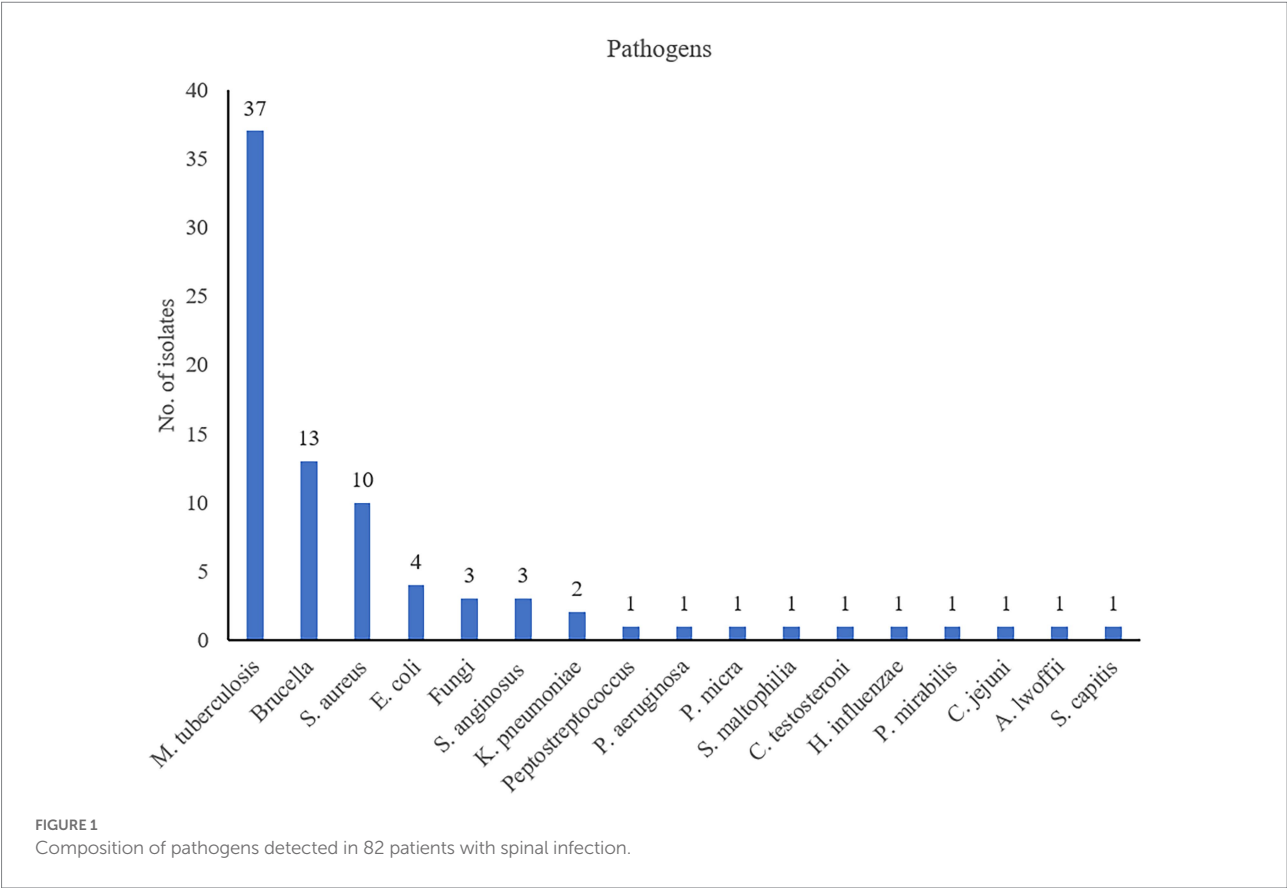


TABLE 2 Performance of mNGS and conventional detection methods in patients with spinal infection.

Methods	Sensitivity (%, N, 95% CI)	Specificity (%, N, 95% CI)	PPV (%, N, 95% CI)	NPV (%, N, 95% CI)	p value (sensitivity)	p value (specificity)
mNGS	89.0% (81/91) (83–96)	88.9% (8/9) (52–100)	98.8% (81/82) (96–101)	44.4% (8/18) (22–69)	–	–
Culture	28.1% (25/89) (19–38)	100.0% (6/6) (54–100)	100.0% (25/25) (86–100)	8.6% (6/70) (2–15)	$\chi^2 = 68.997$ $p < 0.001^a$	$p = 1.000^a$
Pathological examination	42.9% (30/70) (31–55)	100.0% (9/9) (66–100)	100.0% (30/30) (88–100)	18.4% (9/49) (9–32)	$\chi^2 = 39.636$ $p < 0.001^b$	$p = 1.000^b$

N: number of patients, PPV: positive predictive value, NPV: negative predictive value, CI: confidence interval, a: mNGS vs. culture. b: mNGS vs. Pathological examination.

pathological examination in patients with non-TB spinal infection were 84.9% (45/53), 15.4% (8/52), 26.2% (11/42), respectively. Compared with that of bacterial culture and pathological examination, the sensitivity of the mNGS assay was higher and statistically different ($p < 0.001$, $p < 0.001$). The specificities of the mNGS assay, bacterial culture, and pathological examination in patients with non-TB spinal infection were 97.9% (46/47), 100.0% (43/43), and 100.0% (37/37), respectively. The specificity of the mNGS assay was not statistically different from that of bacterial culture and pathological examination ($p = 1.000$, $p = 1.000$). Table 4 shows the performances of the mNGS and conventional detection methods in patients with non-TB spinal infection.

Discussion

Infectious diseases remain the leading causes of morbidity and mortality in all patient populations worldwide. They are accompanied by a mortality rate of approximately 15% (Michiels and Jäger, 2017). Accurate diagnosis can be challenging due to the wide variety of pathogens that cause clinically indistinguishable diseases. The global TB epidemic remains serious. Spinal TB is a special spinal infectious disease that accounts for < 25% of spinal infectious diseases in Southern China (Yee et al., 2010). In recent years, the proportion of non-TB spinal infections in patients with suspected spinal TB has gradually increased in TB-specialized

TABLE 3 Performance of mNGS and conventional detection methods in patients with spinal TB.

Methods	Sensitivity (%, N, 95% CI)	Specificity (%, N, 95% CI)	PPV (%, N, 95% CI)	NPV (%, N, 95% CI)	p value (sensitivity)	p value (specificity)
mNGS	94.7% (36/38) (82–99)	100.0% (62/62) (94–100)	100.0% (36/36) (90–100)	96.9% (62/64) (89–100)	–	–
Xpert	94.6% (35/37) (82–99)	100.0% (56/56) (94–100)	100.0% (35/35) (90–100)	96.6% (56/58) (88–100)	$p = 1.000^a$	–
MGIT 960 culture	45.9% (17/37) (30–63)	100.0% (41/41) (91–100)	100.0% (17/17) (80–100)	67.2% (41/61) (55–79)	$\chi^2 = 68.997$ $p < 0.001^b$	–
Pathological examination	67.9% (19/28) (48–84)	100.0% (51/51) (93–100)	100.0% (19/19) (82–100)	85.0% (51/60) (76–94)	$p = 0.006^c$	–
T-SPOT.TB	89.2% (33/37) (75–97)	78.3% (47/60) (68–89)	71.7% (33/46) (57–84)	92.2% (47/51) (85–100)	$p = 0.430^d$	$\chi^2 = 15.035$ $p < 0.001^d$

N: number of patients, PPV: positive predictive value, NPV: negative predictive value, CI: confidence interval, ^amNGS vs. Xpert, ^bmNGS vs. MGIT 960 culture, ^cmNGS vs. pathological examination, ^dmNGS vs. T-SPOT.TB.

TABLE 4 Performance of mNGS and conventional detection methods in patients with non-TB spinal infection.

Methods	Sensitivity (%, N, 95% CI)	Specificity (%, N, 95% CI)	PPV (%, N, 95% CI)	NPV (%, N, 95% CI)	p value (sensitivity)	p value (specificity)
mNGS	84.9% (45/53) (75–95)	97.9% (46/47) (88–100)	97.8% (45/46) (88–100)	85.2% (46/54) (75–95)	–	–
Bacterial culture	15.4% (8/52) (5–26)	100.0% (43/43) (91–100)	100.0% (8/8) (63–100)	49.4% (43/87) (39–60)	$\chi^2 = 50.748$ $p < 0.001^a$	$p = 1.000^a$
Pathological examination	26.2% (11/42) (14–42)	100.0% (37/37) (90–100)	100.0% (11/11) (72–100)	54.4% (37/68) (42–67)	$\chi^2 = 33.381$ $p < 0.001^b$	–

N: number of patients, PPV: positive predictive value, NPV: negative predictive value, CI: confidence interval, ^amNGS vs. bacterial culture, ^bmNGS vs. pathological examination

hospitals in China. In this study, non-TB spinal infection accounted for 53% of the patients with suspected spinal TB, whereas spinal TB accounted for only 38% of the patients. The data showed that the proportion of patients with spinal TB in patients with suspected spinal TB is significantly lower than before, whereas the proportion of spinal infection is significantly higher. This trend is similar to the results reported in the literature (Tsantes et al., 2020; World Health Organization, 2021).

In this study, *Brucella* and *S. aureus* accounted for a high proportion of the non-TB spinal infection bacteria and were present at considerably higher proportions than other bacteria. *S. aureus* is a common bone infection pathogen. The high proportion of *Brucella* is mainly due to the close location of the three hospitals to pastoral areas in northern China. At the same time, given that *Brucella* infection has similar symptoms as TB infection, the patients went to TB-specialized hospitals for treatment. Opportunistic pathogens also account for a certain proportion of cases with non-TB bacterial spinal infections (Gupta et al., 2022; Mehkri et al., 2022).

The accurate and rapid differential diagnosis of the two diseases poses a new challenge to the clinicians. Given that conventional detection methods cannot meet the needs of clinical diagnosis, so new technologies need to be introduced to solve the existing problems. mNGS is an unbiased approach for pathogen detection that allows for universal pathogen detection regardless

of the type of pathogen (viruses, bacteria, fungi, and parasites), and mNGS can even be applied for novel organism discovery (Gu et al., 2019). Therefore, mNGS is a powerful tool for differential diagnosis in patients with suspected spinal TB. Some studies have shown that mNGS performs well in the diagnosis of orthopedic infectious diseases (Ruppé et al., 2017; Ma C. et al., 2022). However, only a few reports on the ability of mNGS for the differential diagnosis ability of suspected spinal TB exist (Huang et al., 2019, 2020; Zhao et al., 2020). Therefore, a multicenter prospective study was conducted to evaluate the diagnostic performance of mNGS in patients with suspected spinal TB.

In this study, the sensitivity of mNGS (89.0%) was significantly higher than that of culture (28.1%) and pathological examination (42.9%) in patients with spinal infection. These results demonstrated the higher sensitivity of mNGS than that of conventional detection methods. However, mNGS had a lower specificity (88.9%) than culture (100.0%) and pathology (100.0%) because it detected bacteria in one uninfected patient. This result is similar to previously reported findings (Zhou et al., 2019; Ma C. et al., 2022). In this study, mNGS detected *Porphyromonas gingivalis* in a specimen from an uninfected patient. This situation indicated the possibility of false-positives in mNGS results. *P. gingivalis* was likely detected in the uninfected patient because it is a bacterium that commonly colonizes the oral cavity. Therefore, mNGS should be combined with other clinical tests to

enable comprehensive judgment. False-positive results in mNGS testing may be related to contamination, unbiased nucleic acid amplification, and human-colonizing bacteria.

In this study, mNGS and Xpert showed similar sensitivity (94.7% vs. 94.6%) and the same specificity (100.0%) without statistically different sensitivities ($p = 1.000$) for spinal TB. This result suggested that mNGS and Xpert have the comparable abilities for the diagnosis of spinal TB. The sensitivity of T-SPOT.TB was not statistically significantly different from that of mNGS and Xpert (89.2% vs. 94.7%, 94.6%). However, T-SPOT.TB had significantly lower specificity than mNGS and Xpert (78.3% vs. 100.0%, 100.0%). Therefore, the diagnostic capability of T-SPOT.TB was lower than that of mNGS and Xpert. This finding is similar to the reported results (Zhao et al., 2020; Chen et al., 2022). The poor specificity of T-SPOT.TB is mainly due to the high latent infection rate of mycobacterium TB in China (World Health Organization, 2021). MGIT 960 culture and pathological examination had significantly lower sensitivity than mNGS (45.9%, 67.9% vs. 94.7%). However, the specificity of these methods was consistent with that of mNGS (100.0%). This result suggested that the ability of these methods to diagnose spinal TB is lower than that of mNGS.

In patients with non-TB spinal infection, the sensitivity of mNGS was significantly higher than that of bacterial culture and pathological examination (84.9% vs. 15.4%, 26.2%). The specificity of mNGS was 97.9%, and although mNGS had a false-positive result for one case, no statistical difference was found among the specificities of mNGS, bacterial culture, and pathological examination. This result indicated that mNGS has a good diagnostic ability in the detection of non-TB spinal infection and is similar to previously reported findings (Miao et al., 2018; Huang et al., 2020). Given the existence of false-positive test results, the results of mNGS junction tests should be comprehensively analyzed in combination with clinical conditions.

The results of this study show that mNGS has higher sensitivity and specificity than conventional detection methods in the diagnosis of spinal TB and non-TB spinal disease infection, except for Xpert. Therefore, mNGS is a powerful diagnostic tool for patients with suspected spinal TB and avoids missed diagnosis and misdiagnosis. However, conventional methods also have their own advantages over mNGS; for example, although Xpert has the same ability as mNGS for the diagnosis of spinal TB, it can detect rifampicin-resistant gene mutations and provide guidance for TB treatment plans (Yu et al., 2021). In addition, MGIT 960 and bacterial cultures can be used to conduct drug sensitivity tests on bacteria (Zhou et al., 2019; Chen et al., 2022), and the results of drug sensitivity test have important guiding significance for the treatment of clinical spinal infections. Currently, mNGS provides limited information on the drug sensitivity of the detected bacteria, and its ability to detect drug-resistant mutations needs further study.

In this study, mNGS did not detect bacteria in ten patients with spinal infection and detected bacteria in an uninfected patient. Currently, mNGS also has some problems in the diagnosis

of infectious diseases. These problems are related to the sequencing principle of mNGS. Given that mNGS indiscriminately detects all nucleic acid molecules in specimens, including pathogenic bacteria, colonized bacteria, and exogenous nucleic acid molecules previously integrated into the human body, pathogens need to be distinguished from other bacteria. At the same time, the possibility of contamination, including contamination from specimens, reagents, and operating procedures, exists. mNGS provides a massive amount of data, 90% of which is on human nucleic acids, and requires information analysis to exclude interfering factors and identify pathogenic bacteria (Simner et al., 2018; Xiao et al., 2022). For the samples with low pathogen load, the sequencing depth of mNGS should be increased to improve the detection rate of bacteria, but this will lead to the increase of sequencing cost and sequencing time. Therefore, it is necessary to balance the relationship between the sequencing cost, sequencing depth, and sequencing time of mNGS from the aspects of sample processing, detection process, and bioinformatics analysis.

This study had several limitations. First, three TB-specialized hospitals were selected as research units. Therefore, the representation of pathogen composition is limited to a certain proportion. Second, RNA sequencing was not carried out because RNA is unstable and easily degraded. Therefore, pathogenic microorganism with RNA genomes could not be detected. Third, the sensitivity of nonhuman DNA detection was removed because human DNA contamination was not depleted during sample DNA purification.

In summary, mNGS is a rapid and effective diagnostic tool for patients with suspected spinal TB. In contrast to conventional detection method, mNGS can detect tuberculous and non-TB bacteria infection simultaneously, thus avoiding missed diagnosis and misdiagnosis. mNGS also has high sensitivity, specificity and short detection time. Nevertheless, mNGS also has some shortcomings. Thus, further research is needed.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by ethics committee of Hebei Chest Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

S-bQ, F-sL, and YL designed the study. YL, LT, W-jD, JF, T-IL, and X-wY participated in data collection. YL, X-wY, S-bQ, and

F-sL participated in data analysis. YL, X-wY, LT, W-jD, JF, and T-IL wrote the manuscript. Author ranking based on contributions to articles. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the National Science and Technology Major Project (grant number 2017ZX09304009-004; <http://www.nmp.gov.cn/>). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We thank the enrolled patients for cooperating with our study and acknowledge the professionalism and compassion demonstrated by all the workers involved in the three hospitals.

References

- Berrocal-Almanza, L. C., Harris, R. J., Collin, S. M., Muzyamba, M. C., Conroy, O. D., Mirza, A., et al. (2022). Effectiveness of nationwide programmatic testing and treatment for latent tuberculosis infection in migrants in England: a retrospective, population-based cohort study. *Lancet Public Health* 7, e305–e315. doi: 10.1016/S2468-2667(22)00031-7
- Chen, Y., Wang, Y., Liu, X., Li, W., Fu, H., Liu, X., et al. (2022). Comparative diagnostic utility of metagenomic next-generation sequencing, GeneXpert, modified Ziehl-Neelsen staining, and culture using cerebrospinal fluid for tuberculous meningitis: a multi-center, retrospective study in China. *J. Clin. Lab. Anal.* 36:e24307. doi: 10.1002/jcla.24307
- Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical metagenomic next generation sequencing for pathogen detection. *Annu. Rev. Pathol.* 14, 319–338. doi: 10.1146/annurev-pathmechdis-012418-012751
- Gupta, N., Bhat, S. N., Reddysetti, S., Afees Ahamed, M. A., Jose, D., Sarvepalli, A. S., et al. (2022). Clinical profile, diagnosis, treatment, and outcome of patients with tubercular versus nontubercular causes of spine involvement: a retrospective cohort study from India. *Int. J. Mycobacteriol.* 11, 75–82. doi: 10.4103/ijmy.ijmy_243_21
- Huang, Z., Zhang, C., Hu, D., Shi, K., Li, W., Zhang, C., et al. (2019). Diagnosis of osteoarticular tuberculosis via metagenomics next-generation sequencing: a case report. *Exp. Ther. Med.* 18, 1184–1188. doi: 10.3892/etm.2019.7655
- Huang, Z. D., Zhang, Z. J., Yang, B., Li, W. B., Zhang, C. J., Fang, X. Y., et al. (2020). Pathogenic detection by metagenomic next-generation sequencing in Osteoarticular infections. *Front. Cell. Infect. Microbiol.* 10:471. doi: 10.3389/fcimb.2020.00471
- Kafle, G., Garg, B., Mehta, N., Sharma, R., Singh, U., Kandasamy, D., et al. (2022). Diagnostic yield of image-guided biopsy in patients with suspected infectious spondylodiscitis: a prospective study from a tuberculosis-endemic country. *Bone Joint J.* 104-B(1), 120–126. doi: 10.1302/0301-620X.104B1.BJJ-2021-0848.R2
- Li, Y., Jia, W., Lei, G., Zhao, D., Wang, G., and Qin, S. (2018). Diagnostic efficiency of Xpert MTB/RIF assay for osteoarticular tuberculosis in patients with inflammatory arthritis in China. *PLoS One* 13:e0198600. doi: 10.1371/journal.pone.0209939
- Ma, Y., Fan, J., Li, S., Dong, L., Li, Y., Wang, F., et al. (2020). Comparison of Lowenstein-Jensen medium and MGIT culture system for recovery of mycobacterium tuberculosis from abscess samples. *Diagn. Microbiol. Infect. Dis.* 96:114969. doi: 10.1016/j.diagmicrobio.2019.114969
- Ma, J., Liu, H., Wang, J., Li, W., Fan, L., and Sun, W. (2022). HIV-negative rifampicin resistance /multidrug-resistance Extrapulmonary tuberculosis in China from 2015 to 2019: a clinical retrospective investigation study from a National Tuberculosis Clinical Research Center. *Infect Drug Resist.* 15, 1155–1165. doi: 10.2147/IDR.S342744
- Ma, C., Wu, H., Chen, G., Liang, C., Wu, L., and Xiao, Y. (2022). The potential of metagenomic next-generation sequencing in diagnosis of spinal infection: a retrospective study. *Eur. Spine J.* 31, 442–447. doi: 10.1007/s00586-021-07026-5
- Mehkri, Y., Felisma, P., Panther, E., and Lucke-Wold, B. (2022). Osteomyelitis of the spine: treatments and future directions. *Infect. Dis. Res.* 3:3. doi: 10.53388/IDR20220117003
- Miao, Q., Ma, Y., Wang, Q., Pan, J., Zhang, Y., Jin, W., et al. (2018). Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clin. Infect. Dis.* 67, S231–S240. doi: 10.1093/cid/ciy693
- Michiels, I., and Jäger, M. (2017). Spondylodiscitis: current strategies for diagnosis and treatment. *Orthopade* 46, 785–804. doi: 10.1007/s00132-017-3436-0
- Mingora, C. M., Garcia, B. A., Mange, K. C., Yuen, D. W., Ciesielska, M., van Ingen, J., et al. (2022). Time-to-positivity of Mycobacterium avium complex in broth culture associates with culture conversion. *BMC Infect. Dis.* 22:246.
- Pigrau-Serrallach, C., and Rodríguez-Pardo, D. (2013). Bone and joint tuberculosis. *Eur. Spine J. Suppl.* 4, 556–566. doi: 10.1007/s00586-012-2331-y
- Ruppé, E., Lazarevic, V., Girard, M., Mouton, W., Ferry, T., Laurent, F., et al. (2017). Clinical metagenomics of bone and joint infections: a proof of concept study. *Sci. Rep.* 7:7718. doi: 10.1038/s41598-017-07546-5
- Ryang, Y. M., and Akbar, M. (2020). Pyogenic spondylodiscitis: symptoms, diagnostics and therapeutic strategies. *Orthopade* 49, 691–701. doi: 10.1007/s00132-020-03945-1
- Simner, P. J., Miller, S., and Carroll, K. C. (2018). Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin. Infect. Dis.* 66, 778–788. doi: 10.1093/cid/cix881
- Stangenberg, M., Mende, K. C., Mohme, M., Krätzig, T., Viezens, L., Both, A., et al. (2021). Influence of microbiological diagnosis on the clinical course of spondylodiscitis. *Infection* 49, 1017–1027. doi: 10.1007/s15010-021-01642-5
- Tsang, S. J., Eyre, D. W., Atkins, B. L., and Simpson, A. H. R. W. (2020). Should modern molecular testing be routinely available for the diagnosis of musculoskeletal infection? *Bone Joint J.* 102-B, 1274–1276. doi: 10.1302/0301-620X.102B10.BJJ-2020-1496
- Tsantes, A. G., Papadopoulos, D. V., Vroni, G., Sioutis, S., Sapkas, G., Benzakour, A., et al. (2020). Spinal infections: an update. *Microorganisms* 8:476. doi: 10.3390/microorganisms8040476
- Wang, S., Chen, Y., Wang, D., Wu, Y., Zhao, D., Zhang, J., et al. (2019). The feasibility of metagenomic next-generation sequencing to identify pathogens causing tuberculous meningitis in cerebrospinal fluid. *Front. Microbiol.* 10:1993. doi: 10.3389/fmicb.2019.01993
- Wieland, B. W., Marcantoni, J. R., Bommarito, K. M., Warren, D. K., and Marschall, J. (2012). A retrospective comparison of ceftriaxone versus oxacillin for osteoarticular infections due to methicillin-susceptible *Staphylococcus aureus*. *Clin. Infect. Dis.* 54, 585–590. doi: 10.1093/cid/cir857
- Wilson, M. R., Naccache, S. N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., et al. (2014). Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* 370, 2408–2417. doi: 10.1056/NEJMoa1401268

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1018938/full#supplementary-material>

World Health Organization, (2021). *Global Tuberculosis Report 2021*. World Health Organization, Geneva.

Xiao, G., Cai, Z., Guo, Q., Ye, T., Tang, Y., Guan, P., et al. (2022). Insights into the unique lung microbiota profile of pulmonary tuberculosis patients using metagenomic next-generation sequencing. *Microbiol Spectr.* 10:e0190121. doi: 10.1128/spectrum.01901-21

Yee, D. K., Samartzis, D., Wong, Y. W., Luk, K. D., and Cheung, K. M. (2010). Infective spondylitis in southern Chinese: a descriptive and comparative study of ninety-one cases. *Spine (Phila Pa 1976)* 35, 635–641. doi: 10.1097/BRS.0b013e3181c4ff46

Yu, G., Wang, X., Zhu, P., Shen, Y., Zhao, W., and Zhou, L. (2021). Comparison of the efficacy of metagenomic next-generation sequencing and Xpert MTB/RIF in the

diagnosis of tuberculous meningitis. *J. Microbiol. Methods* 180:106124. doi: 10.1016/j.mimet.2020.106124

Zhao, M., Tang, K., Liu, F., Zhou, W., Fan, J., Yan, G., et al. (2020). Metagenomic next-generation sequencing improves diagnosis of Osteoarticular infections from abscess specimens: a multicenter retrospective study. *Front. Microbiol.* 11:2034. doi: 10.3389/fmicb.2020.02034

Zhou, X., Wu, H., Ruan, Q., Jiang, N., Chen, X., Shen, Y., et al. (2019). Clinical evaluation of diagnosis efficacy of active mycobacterium tuberculosis complex infection via metagenomic next-generation sequencing of direct clinical samples. *Front. Cell. Infect. Microbiol.* 9:351. doi: 10.3389/fcimb.2019.00351



OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE

Feng-sheng Liu
✉ liu_fs_doctor@163.com
Shi-bing Qin
✉ qinshibing17@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 02 April 2023

ACCEPTED 03 April 2023

PUBLISHED 18 April 2023

CITATION

Li Y, Yao X-w, Tang L, Dong W-j, Lan T-l, Fan J, Liu F-s and Qin S-b (2023) Corrigendum: Diagnostic efficiency of metagenomic next-generation sequencing for suspected spinal tuberculosis in China: a multicenter prospective study.
Front. Microbiol. 14:1198931.
doi: 10.3389/fmicb.2023.1198931

COPYRIGHT

© 2023 Li, Yao, Tang, Dong, Lan, Fan, Liu and Qin. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Corrigendum: Diagnostic efficiency of metagenomic next-generation sequencing for suspected spinal tuberculosis in China: a multicenter prospective study

Yuan Li^{1†}, Xiao-wei Yao^{2†}, Liang Tang^{3†}, Wei-jie Dong¹, Ting-long Lan¹, Jun Fan¹, Feng-sheng Liu^{2*} and Shi-bing Qin^{1*}

¹Department of Orthopedics, Beijing Chest Hospital, Capital Medical University, Beijing, China,

²Department of Orthopedics, Hebei Chest Hospital, Shijiazhuang, China, ³Department of Orthopedics, Tianjin Haihe Hospital, Tianjin, China

KEYWORDS

diagnosis, spine, tuberculosis, metagenomic next-generation sequencing, infection

A corrigendum on

Diagnostic efficiency of metagenomic next-generation sequencing for suspected spinal tuberculosis in China: a multicenter prospective study

by Li, Y., Yao, X.-w., Tang, L., Dong, W.-j., Lan, T.-l., Fan, J., Liu, F.-s., and Qin, S.-b. (2022). *Front. Microbiol.* 13:1018938. doi: 10.3389/fmicb.2022.1018938

In the published article, there was an error in the numbering of affiliations. Instead of “Wei-jie Dong², Ting-long Lan², Jun Fan², Feng-sheng Liu^{3*}”, it should be “Wei-jie Dong¹, Ting-long Lan¹, Jun Fan¹, Feng-sheng Liu^{2*}”.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

