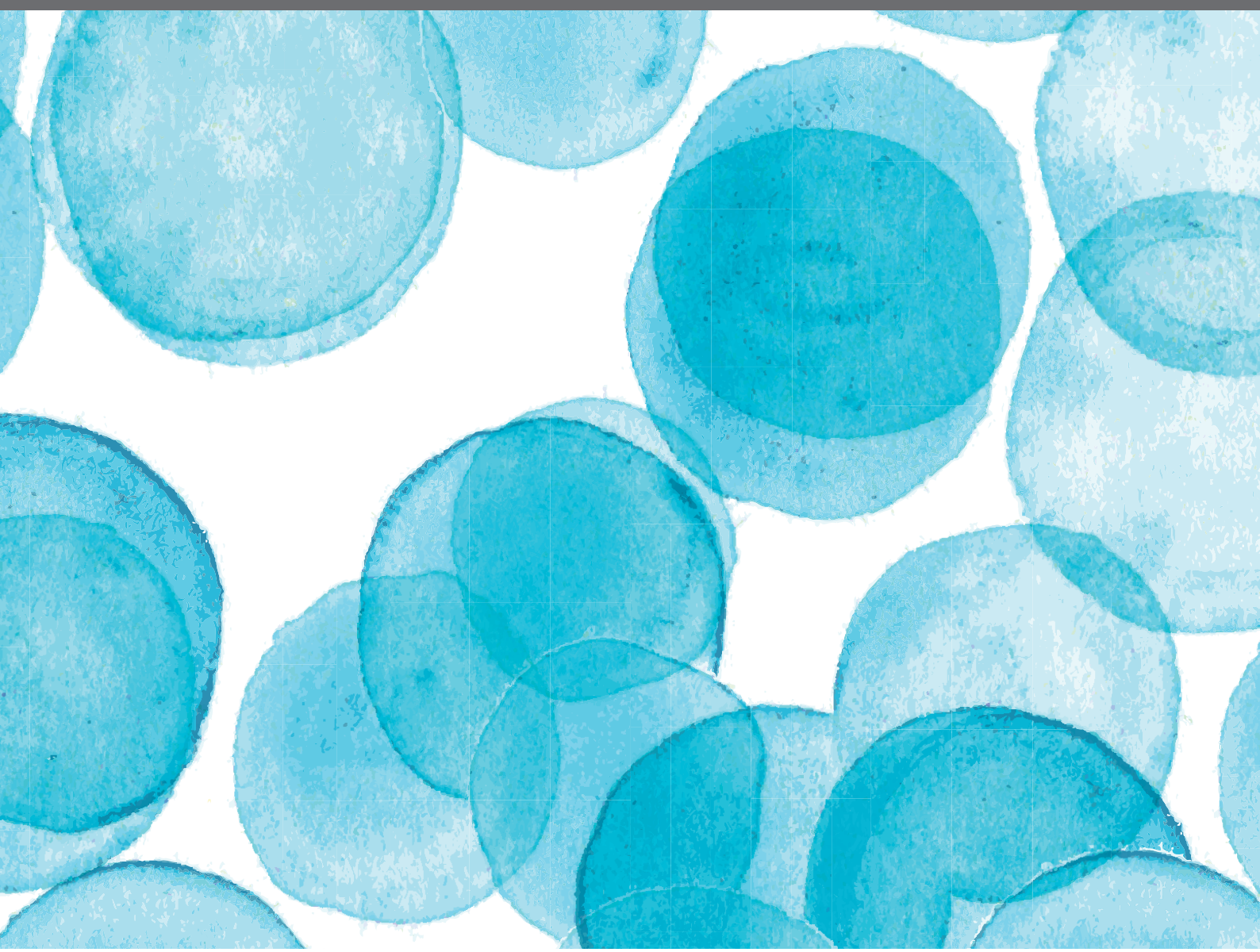
APPLICATION OF NEXT GENERATION SEQUENCING (NGS) IN INFECTION PREVENTION

EDITED BY: Michael Kemp and Martin Maiers
PUBLISHED IN: Frontiers in Microbiology, Frontiers in Public Health and Frontiers in Medicine



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-495-2
DOI 10.3389/978-2-88976-495-2

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

APPLICATION OF NEXT GENERATION SEQUENCING (NGS) IN INFECTION PREVENTION

Topic Editors:
Michael Kemp, University of Southern Denmark, Denmark
Martin Maiers, National Marrow Donor Program, United States

Citation: Kemp, M., Maiers, M., eds. (2022). Application of Next Generation Sequencing (NGS) in Infection Prevention. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-495-2

Table of Contents

frontiers | Frontiers in Public Health

EDITORIAL
published: 09 June 2022
doi: 10.3389/fpubh.2022.945595

Editorial: Application of Next Generation Sequencing (NGS) in Infection Prevention

Michael Kemp[1,2]* *and Martin Maiers*[2]

[1] Department of Regional Health Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark, [2] Marrow Donor Program, Minneapolis, MN, United States

Keywords: next generation sequencing, NGS, infection prevention, disease surveillance, antibiotic resistance, infection control

Editorial on the Research Topic

Application of Next Generation Sequencing (NGS) in Infection Prevention

In the past decade, Next Generation Sequencing (NGS) has become widely applied for diagnostics, surveillance, and research of infectious diseases. In 2017, Deurenberg et al. published a review on application of NGS in clinical microbiology and infection prevention (1). Since then, the field has moved fast. The continuous development of sequencing technology and data handling software has made NGS accessible even to smaller laboratories. NGS data are easily shared, allowing collaboration on *in silico* analyses between institutions. Both microorganisms and host responses may be analyzed by NGS. Thus, NGS offers new options for current and future prevention of infections. The aim of this Research Topic is to illustrate various uses of NGS in prevention of infections by virus, bacteria, and parasites.

The emergence of new variants during the SARS-CoV-2 pandemic has clearly demonstrated the importance of NGS based surveillance of viral diseases. However, NGS has also revealed other discoveries on the epidemiology of SARS-CoV-2. Wang et al. report co-existence of genetically distinct viruses within a host and a narrow transmission bottleneck between patients from the same households using viral whole genome sequencing (WGS). This finding is important in understanding the population dynamics of SARS-CoV-2.

For the study of bacteria, Rogers et al. designed a scoping review to determine the value of WGS in the surveillance of antibiotic resistance in enterococci. These are important bacteria, as *Enterococcus faecalis* and *E. faecium* were each responsible for between 100.000 and 250.000 deaths associated with antibiotic resistance globally in 2019 (2). Rogers et al. conclude that WGS has been used with success, especially for detection of new genes and typing of isolates of enterococci. WGS-based typing and resistance gene detection was used by Marbjerg et al. to characterize vancomycin-resistant *E. faecium* in a hospital over 5 years. Using an online software solution that does not require knowledge in bioinformatics, the authors detected accumulation of specific clusters over time. Mao et al. used pulsed field gel electrophoresis and WGS to describe the epidemiology of multidrug-resistant *Acinetobacter baumanii* in an intensive care unit (ITU). The high discriminatory power of WGS established transmission as the main mode of acquisition of multidrug-resistant *A. baumanii* by patients in the ITU. Surveillance at the local/regional level may benefit from WGS as discussed in the review by Asare et al.. The authors evaluate the relevance

OPEN ACCESS

Edited and reviewed by:
Marc Jean Struelens, Université libre de Bruxelles, Belgium

***Correspondence:**
Michael Kemp
mkemp@regionsjaelland.dk

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

Received: *16 May 2022*
Accepted: *27 May 2022*
Published: *09 June 2022*

Citation:
Kemp M and Maiers M (2022) Editorial: Application of Next Generation Sequencing (NGS) in Infection Prevention. Front. Public Health 10:945595. doi: 10.3389/fpubh.2022.945595

of genomic epidemiology in surveillance of tuberculosis in West Africa with special focus on drug resistant strains of *Mycobacterium tuberculosis*.

Although WGS shows superior performance compared to almost any other microbial typing system, replacement of previously used methods requires translation of WGS data into the previously used system, especially during a transition phase. Examples of established tools for predicting *Salmonella enterica* serotypes, *E. coli* serotypes, Fimtypes, and FumCtypes, *S. aureus* SPAtypes, and *Pseudomonas aeruginosa* serotypes from WGS data are found on http://www.genomicepidemiology.org/services/. *Brucella* sp. are traditionally typed by Multiple Locus Variable-number Tandem Repeat Analysis (MLVA), which is a laborious methology. The work of Pelerito et al. describes the successful prediction of MLVA types from WGS data, allowing comparison of present isolates to previously typed isolates of the bacterium. Qiu et al. used an established genome database to compare a patient isolate of a rare *Salmonella enterica* serovar (*S. enterica* serovar Telelkebir) to published sequences of this serovar from various parts of the world. Core genome Multi Locus Sequencing Typing and single nucleotide polymorphism analyses suggested international transmission.

Application of metagenomic NGS (mNGS) for rapid diagnostics of infectious diseases is demonstrated by a case story by Zeng et al., who detected *Klebsiella pneumoniae* DNA in culture-negative cerebrospinal fluid and blood from a patient with liver abscesses and meningitis. Also using mWGS, Michel et al. describes a case of lung infection by *Ureaplasma* spp. and *Mycoplasma hominis* in a patient after lung transplantation and establishing cure of infection. The diversity of bile bacteria in cholecystitis was analyzed in four patients by Yan et al. using a combination of culture, conventional 16S rRNA sequencing and WGS. This approach brings new knowledge on the role of bile bacteria in cholecystitis.

Application of NGS for research in parasitic infections is demonstrated by He et al., who used Single Molecule Real Time sequencing and Illumina sequencing for transcriptome analyses of different developmental stages of the mite *Otodectes cynotis*. The mite infests a variety of animals, causing allergic itching external otitis. Almost 2,700 genes that were differentially expressed by adult mites and nymph/larva were detected. The analyses allowed identification of almost 400 putative allergen genes with potential importance for pathogenesis. Furthermore, the analyses indicated regulation of KEGG pathways in the two stages. The study provides new knowledge of the biology of the parasite and the nature of the disease it causes and identifies targets for further studies.

Next Generation Sequencing provides new insights into the genetics of pathogens. The papers included in this Research Topic demonstrate applications of NGS on research in basic microbiology as well as diagnostics and surveillance of infectious diseases. Both the laboratory handling and subsequent data analyses of NGS are under constant development. As NGS becomes easier, cheaper, more rapid and more scalable, it will be more widely accessible and gain an even more important role in future infection prevention.

AUTHOR CONTRIBUTIONS

MK was a guest associate editor of the Research Topic and drafted the paper text. MM was a guest associate editor of the Research Topic and edited the text. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol.* (2017) 10:16-24. doi: 10.1016/j.jbiotec.2016.12.022
2. Antibiotic resistance collaboraters. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet.* (2022) 399:629-55. doi: 10.1016/S0140-6736(21)02724-0

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.

frontiers in Medicine

BRIEF RESEARCH REPORT
published: 15 February 2021
doi: 10.3389/fmed.2021.585358

Population Bottlenecks and Intra-host Evolution During Human-to-Human Transmission of SARS-CoV-2

Daxi Wang[1,2†], Yanqun Wang[3†], Wanying Sun[1,2,4†], Lu Zhang[3,5†], Jingkai Ji[1,2†], Zhaoyong Zhang[3†], Xinyi Cheng[1,2,6†], Yimin Li[3†], Fei Xiao[7], Airu Zhu[3], Bei Zhong[8], Shicong Ruan[9], Jiandong Li[1,2,4], Peidi Ren[1,2], Zhihua Ou[1,2], Minfeng Xiao[1,2], Min Li[1,2,4], Ziqing Deng[1,2], Huanzi Zhong[1,2,10], Fuqiang Li[1,2,11], Wen-jing Wang[1,11], Yongwei Zhang[1], Weijun Chen[4,12], Shida Zhu[1,13], Xun Xu[1,14], Xin Jin[1], Jingxian Zhao[3], Nanshan Zhong[3], Wenwei Zhang[1], Jincun Zhao[3,5]*, Junhua Li[1,2,6]* and Yonghao Xu[3]**

[1] BGI-Shenzhen, Shenzhen, China, [2] Shenzhen Key Laboratory of Unknown Pathogen Identification, BGI-Shenzhen, Shenzhen, China, [3] State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [4] BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China, [5] Institute of Infectious Disease, Guangzhou Eighth People's Hospital of Guangzhou Medical University, Guangzhou, China, [6] School of Biology and Biological Engineering, South China University of Technology, Guangzhou, China, [7] Guangdong Provincial Key Laboratory of Biomedical Imaging, Department of Infectious Diseases, Guangdong Provincial Engineering Research Center of Molecular Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China, [8] The Sixth Affiliated Hospital of Guangzhou Medical University, Qingyuan People's Hospital, Qingyuan, China, [9] Yangjiang People's Hospital, Yangjiang, China, [10] Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Copenhagen, Denmark, [11] Guangdong Provincial Key Laboratory of Human Disease Genomics, Shenzhen Key Laboratory of Genomics, BGI-Shenzhen, Shenzhen, China, [12] BGI PathoGenesis Pharmaceutical Technology, BGI-Shenzhen, Shenzhen, China, [13] Shenzhen Engineering Laboratory for Innovative Molecular Diagnostics, BGI-Shenzhen, Shenzhen, China, [14] Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen, China

The emergence of the novel human coronavirus, SARS-CoV-2, causes a global COVID-19 (coronavirus disease 2019) pandemic. Here, we have characterized and compared viral populations of SARS-CoV-2 among COVID-19 patients within and across households. Our work showed an active viral replication activity in the human respiratory tract and the co-existence of genetically distinct viruses within the same host. The inter-host comparison among viral populations further revealed a narrow transmission bottleneck between patients from the same households, suggesting a dominated role of stochastic dynamics in both inter-host and intra-host evolutions.

Keywords: SARS-CoV-2, population bottleneck, intra-host variation, human to human transmission, evolution

AUTHOR SUMMARY

In this study, we compared SARS-CoV-2 populations of 13 Chinese COVID-19 patients from three hospitals in different cities of Guangdong province. Those viral populations contained a considerable proportion of viral subgenomic messenger RNAs (sgmRNAs), reflecting an active viral replication activity in the respiratory tract tissues. The comparison of identified intra-host variants further showed a low viral genetic distance between intra-household patients and a narrow transmission bottleneck size. Despite the co-existence of genetically distinct viruses within the same host, most intra-host minor variants were not shared between transmission pairs, suggesting a

OPEN ACCESS

Edited by:
Michael Kemp, University of Southern Denmark, Denmark

Reviewed by:
Aine Niamh O'Toole, University of Edinburgh, United Kingdom
Roshan Kumar, Magadh University, India

***Correspondence:**
Wenwei Zhang
zhangww@genomics.cn
Jincun Zhao
zhaojincun@gird.cn
Junhua Li
lijunhua@genomics.cn
orcid.org/0000-0001-6784-1873
Yonghao Xu
dryonghao@163.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 Medicine

Received: *20 July 2020*
Accepted: *11 January 2021*
Published: *15 February 2021*

Citation:
Wang D, Wang Y, Sun W, Zhang L, Ji J, Zhang Z, Cheng X, Li Y, Xiao F, Zhu A, Zhong B, Ruan S, Li J, Ren P, Ou Z, Xiao M, Li M, Deng Z, Zhong H, Li F, Wang W-j, Zhang Y, Chen W, Zhu S, Xu X, Jin X, Zhao J, Zhong N, Zhang W, Zhao J, Li J and Xu Y (2021) Population Bottlenecks and Intra-host Evolution During Human-to-Human Transmission of SARS-CoV-2. Front. Med. 8:585358. doi: 10.3389/fmed.2021.585358

dominated role of stochastic dynamics in both inter-host and intra-host evolutions. Furthermore, the narrow bottleneck and active viral activity in the respiratory tract show that the passage of a small number of virions can cause infection. Our data have therefore delivered a key genomic resource for the SARS-CoV-2 transmission research and enhanced our understanding of the evolutionary dynamics of SARS-CoV-2.

INTRODUCTION

The rapid spread of the novel human coronavirus, SARS-CoV-2, has been causing millions of COVID-19 (coronavirus disease 2019) cases with high mortality rate worldwide (1, 2). As an RNA virus, SARS-CoV-2 mutates frequently (8.5×10^{-4} nucleotide substitutions per site per year) during genome replication (3–5), leading to the development of genetically different viruses within the same host. Several studies have reported intra-host single nucleotide variants (iSNVs) in SARS-CoV-2 (7, 8, 25). Recently, we investigated the intra-host evolution of SARS-CoV-2 and revealed genetic differentiation among tissue-specific populations (9). However, it is still not clear how the intra-host variants circulate among individuals. Here, we described and compared viral populations of SARS-CoV-2 among COVID-19 patients within and across households. Our work here demonstrated the utilization of viral genomic information to identify transmission linkage of this virus.

RESULTS AND DISCUSSION

Using both metatranscriptomic and hybrid-capture based techniques, we newly deep sequenced respiratory tract (RT) samples of seven COVID-19 patients in Guangdong, China, including two pairs of patients from the same households, respectively (P03 and P11; P23 and P24). Among the two intra-household pairs, patient P03, P23, and P24 had a travel history to Wuhan city during the early pandemic. The data were then combined with those of 23 RT samples used in our previous study (9), yielding a combined data set of 30 RT samples from 13 COVID-19 patients (**Supplementary Table 1**).

A sustained viral population should be supported by an active viral replication (10). We firstly estimated the viral transcription activity within RT samples using viral subgenomic messenger RNAs (**sgmRNAs**), which is only synthesized in infected host cells (11). The sgmRNA abundance was measured as the ratio of short reads spanning the transcription regulatory sequence (TRS) sites to the viral genomic reads (as demonstrated in **Supplementary Figure 1**). It should be noted that the sgmRNA abundance might be underestimated, given that only the short reads with sufficient length to simultaneously cover both leader and coding flanking regions of the TRS site, which might be improved with long read sequencing in future. Nonetheless, the sgmRNA abundance within nasal and throat swab samples was similar to that within sputum samples (**Figure 1A**), reflecting an active viral replication in the upper respiratory tract. Notably, the patient P01, who eventually passed away due to COVID-19, showed the highest level of sgmRNA abundance (**Supplementary Figure 2**). However, due to the limited samples of mild cases, we did not observe a significant difference of sgmRNA abundance between severe and mild cases. For the patients with chronological samples and improved clinical outcomes (P10 and P13), their viral load measured by real-time reverse transcription PCR (qRT-PCR) negatively correlate with the days post symptoms onset with marginal significance (**Figures 1B,C**). Interestingly, the sgmRNA abundance showed a similar trend across time (**Figures 1D,E**), reflecting a direct biological association between viral replication and viral shedding in the respiratory tract tissues.

Using the metatranscriptomic data, we identified 66 iSNVs in protein encoding regions with the alternative allele frequency (AAF) ranged from 5 to 95% (**Supplementary Tables 2, 3**; **Supplementary Figure 3**). Here, an alternative allele was defined as the allele that is different from the allele at the same position of the reference genome. The identified iSNVs showed a high concordance between the AAFs derived from metatranscriptomic and that from hybrid-capture sequences (Spearman's $\rho = 0.81$, $P < 2.2e\text{-}16$; **Supplementary Figure 4**). We firstly looked for signals of natural selection against intra-host variants. Using the Fisher's exact test, we compared the number of iSNV sites on each codon position against that of the other two positions and detected a marginal but significant difference among them (codon position 1 [$n = 10$, $P = 0.02$], 2 [$n = 21$; $P = 1$], and 3 [$n = 35$; $P = 0.03$]). In contrast to the numbers of iSNV sites, the alternative allele frequency of those iSNVs did not discriminate among the non-synonymous and synonymous categories (**Figure 2A**), suggesting that most non-synonymous intra-host variants were not under an effective purifying selection within the host. Among the 66 identified iSNVs, 30 were coincided with the consensus variants in the public database as of April 5, 2020 (**Supplementary Table 2**). Those iSNVs were categorized into common iSNVs, while the iSNVs presented in a single patient were categorized into rare iSNVs. Interestingly, the common iSNVs had a significant higher minor allele frequency compared to the rare iSNVs (**Supplementary Figure 5**; Wilcoxon rank sum test, $P = 2.7e\text{-}05$), suggesting that they may have been developed in earlier strains before the most recent infection.

We then estimated the viral genetic distance among samples in a pairwise manner based on their iSNVs and allele frequencies. The samples were firstly categorized into intra-host pairs (serial samples from the same host), intra-household pairs and inter-household pairs (**Figure 2B** and **Supplementary Table 4**). As expected, the intra-host pairs had the lowest genetic distance compared to either intra-household pairs (Wilcoxon rank sum test, $P = 0.018$) and inter-household pairs (Wilcoxon rank sum test, $P < 2.22e\text{-}16$). Interestingly, the genetic distance between intra-household pairs was significantly lower than that of inter-household pairs (**Figure 2B**; Wilcoxon rank sum test, $P = 0.03$), supporting a direct passage of virions among intra-household individuals. Nonetheless, we only observed a small proportion of (3/14 for P03 and P11; 1/20 for P23 and P24) minor intra-host variants shared among intra-household pairs, suggesting that the estimated genetic similarity was mostly determined by consensus nucleotide differences (**Figures 2C,D**). Based on the AAF of iSNVs in transmission pairs, it seems only the minor virion groups carrying three (from P03) and one variants (from

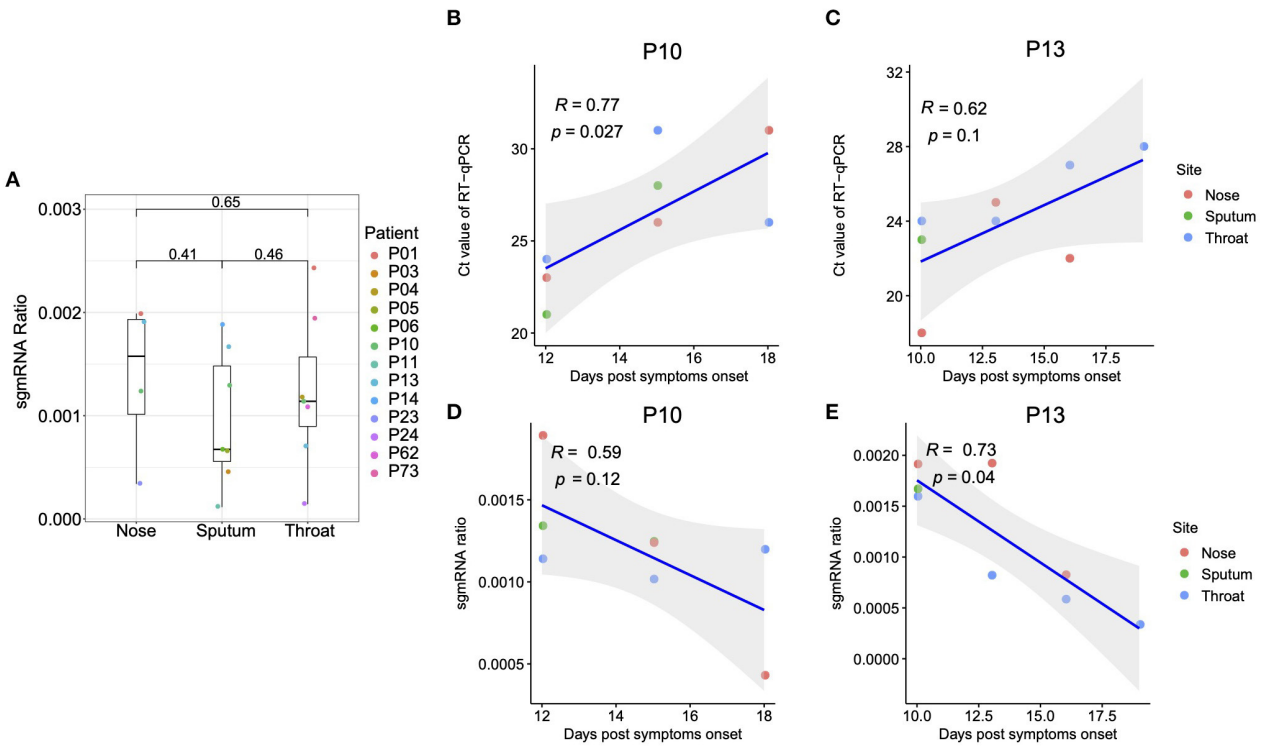


FIGURE 1 | Subgenomic messenger RNAs (sgmRNAs). **(A)** The ratio of sgmRNA of each respiratory sample type (nasal, throat swabs, and sputum). For the patients with multiple samples from the same sample type, only the samples with the median sgmRNA ratio were displayed. **(B,C)** Correlation between the cycle threshold and the days post symptoms onset for patient P10 and P13, respectively. **(D,E)** Correlation between the sgmRNA ratio and the days post symptoms onset for patient P10 and P13, respectively.

P23) were passed to the recipient, respectively. Specifically, in one intra-household pair (P23 and P24), one patient (P23) contained iSNVs that were coincided with the linked variants, C8782T and T28144C, suggesting that this patient may have been co-infected by genetically distinct viruses. However, the strain carrying 8782C and 28144T was not observed in the intra-household counterpart (P24). Given the small number of COVID-19 cases in Guangdong (about 2,000 total cases), secondary infection from other sources is not likely. Within this intra-household pair, it is likely that there is a narrow transmission bottleneck allowing only the major strain to be circulated, if P23 was infected by all the observed viral strains before the transmission.

The transmission bottlenecks among intra-household pairs were estimated using a beta binomial model, which was designed to allow some temporal stochastic dynamics of viral population in the recipient (12). Here, we defined the donor and recipient within the intra-household pairs according to their dates of the first symptom onset. The estimated bottleneck sizes were 6 (P03 and P11) and 8 (P23 and P24) for the two intra-household pairs (**Supplementary Table 5**). The observed narrow bottleneck is consistent with two recent studies of SARS-CoV-2 (13, 14). Nonetheless, a loose transmission bottleneck was also observed (8). Similarly, many animal viruses and human respiratory viruses showed a narrow transmission bottleneck (15, 16), while the only study reporting a loose bottleneck among human respiratory viral infections (17) was argued as the generic consequence of shared iSNVs caused by read mapping artifacts (18). The relatively narrow transmission bottleneck sizes are expected to increase the variance of viral variants being circulated between transmission pairs (19). However, given that we can only measure the viral population that were descendants of the founding population, the actual population could have been much larger. Even after successful transmission, virions carrying the minor variants are likely to be purged out due to the frequent stochastic dynamics within the respiratory tract (9), which is also consistent with the low diversity and instable iSNV observed among the RT samples.

The observed narrow transmission bottleneck suggests that, in general, only a few virions successfully enter host cells and eventually cause infection. Although the number of transmitted virions is sparse, they can easily replicate in the respiratory tract, given the observed viral replication activities in all the RT sample types and the high host-cell receptor binding affinity of SARS-CoV-2 (6). The narrow transmission bottleneck also indicate that instant hand hygiene and mask-wearing might be particular effective in blocking the transmission chain of SARS-CoV-2.

In summary, we have characterized and compared SARS-CoV-2 populations of patients within and across households using both metatranscriptomic and hybrid-capture based techniques. Our work showed an active viral replication activity

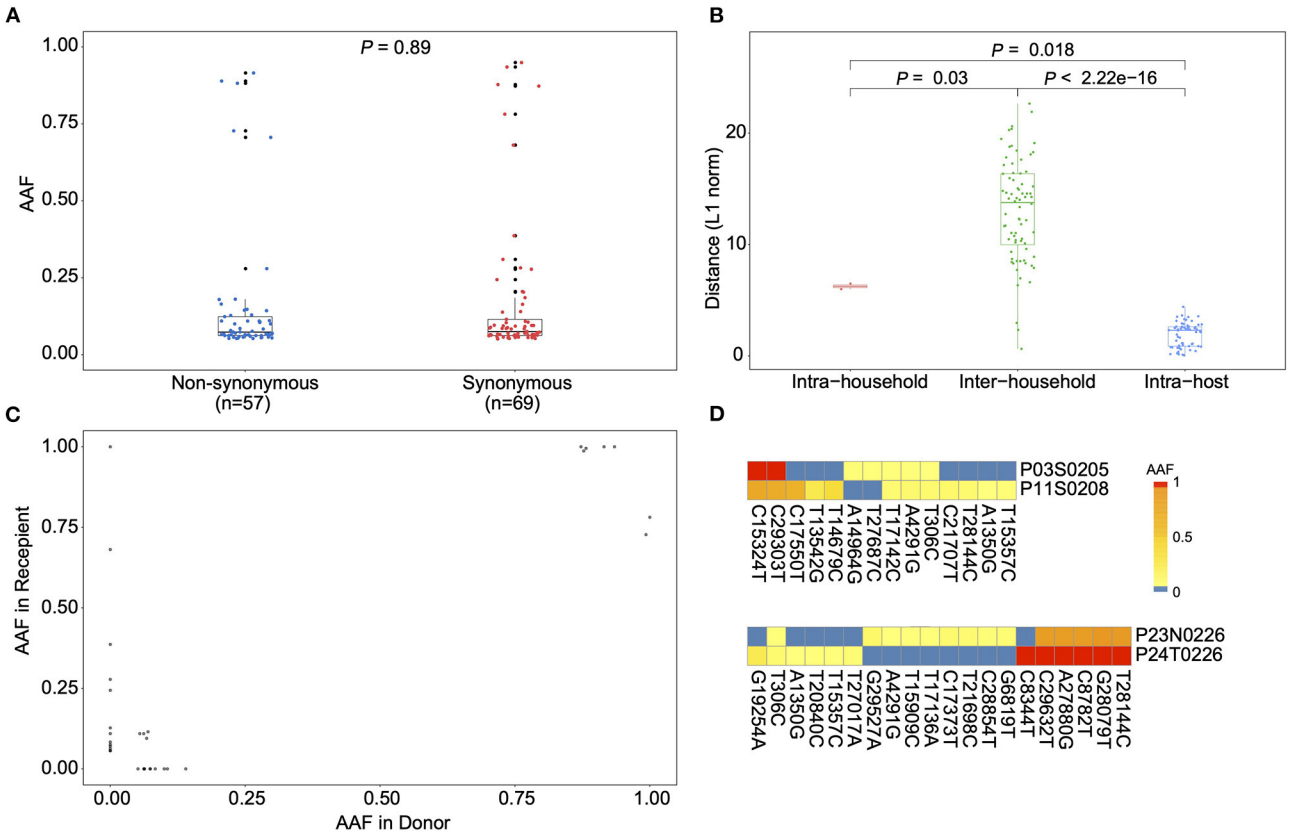


FIGURE 2 | Allele frequency changes of transmission pairs. **(A)** Box plots showing the alternative allele frequency (AAF) distribution of synonymous and non-synonymous intra-host variants. Each dot indicates the median AAF of each iSNV sites of samples from same patient. **(B)** Box plots representing the L1-norm distance distribution among sample pairs. Each dot represents the genetic distance between each sample pair. **(C)** The AAF of donor iSNVs in transmission pairs. Allele frequencies under 5% and over 95% were adjusted to 0% and 1, respectively. **(D)** Heatmap representing the alternative allele frequencies (AAFs) of consensus and intra-host single nucleotide variants (iSNVs) of the two transmission pairs.

in the human respiratory tract and the co-existence of genetically distinct viruses within the same host. The inter-host comparison among viral populations further revealed a narrow transmission bottleneck between patients from the same households, suggesting a dominated role of stochastic dynamics in both inter-host and intra-host evolution. The present work enhanced our understanding of SARS-CoV-2 virus transmission and shed light on the integration of genomic and epidemiological in the control of this virus.

MATERIALS AND METHODS

Patients

Respiratory tract (RT) samples, including nasal swabs, throat swabs, sputum, were collected from 13 COVID-19 patients during the early outbreak of the pandemic (from January 25 to February 10 of 2020). Those patients were hospitalized at the first affiliated hospital of Guangzhou Medical University (nine patients) in Guangzhou, the fifth affiliated hospital of Sun Yat-sen University (two patients), Qingyuan People's Hospital (1 patient) in Zhuhai and Yangjiang People's Hospital (one patient) Yangjiang. The research plan was assessed and approved by the Ethics Committee of each hospital. All the privacy information was anonymized.

Dataset Description

Public consensus sequences were downloaded from GISAID on April 5, 2020.

Sample Preparation and Sequencing

RNA was extracted from the clinical RT samples using QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), which was then tested for SARS-CoV-2 using qRT-PCR. Human DNA was removed using DNase I and RNA concentration was measured using Qubit RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). After human DNA-depletion, the samples were RNA purified and then subjected to double-stranded DNA library construction using the MGIEasy RNA Library preparation reagent set (MGI, Shenzhen, China) following the method used in the previous study (20). Possible environmental or cross contamination during library preparation was tracked using the control RNA samples from human breast cell lines

(Michigan Cancer Foundation-7). The constructed libraries were converted to DNA nanoballs (DNBs) and then sequenced on the DNBSEQ-T7 platform (MGI, Shenzhen, China), generating paired-end short reads with 100 bp in length. Most samples were also sequenced using hybrid capture-based enrichment approach that was described in previous study (20). Briefly, the SARS-CoV-2 genomic content was enriched from the double-stranded DNA libraries using the 2019-nCoVirus DNA/RNA Capture Panel (BOKE, Jiangsu, China). The enriched SARS-CoV-2 genomic contents were converted to DNBs and then sequenced on the MGISEQ-2000 platform, generating paired-end short reads with 100 bp in length.

Data Filtering

Read data from both metatranscriptomic and hybrid capture based sequencing were filtered following the steps described in the previous research (20). Short read data were mapped to a database that contains all the available reference genomes of coronaviridae (including SARS, SARS-CoV-2 and MERS genomes from GISAID, NCBI and CNGB) using Kraken v0.10.5 with default parameters. Low-quality, adaptor contaminations, duplications within the mapped reads were removed using fastp v0.19.5 and SOAPnuke v1.5.6. Low-complexity reads were then filtered using PRINSEQ v0.20.4.

Profiling of Subgenomic Messenger RNAs (SgmRNAs)

Coronaviridae-like short reads were mapped to the reference genome (EPI_ISL_402119) using the aligner HISAT2 (21). Reads spanning the transcription regulatory sequence (TRS) sites of both leader region and the coding genes (S gene, ORF3a, 6, 7a, 8, E, M, and N gene) were selected to represent the sgmRNAs. The junction sites were predicted using RegTools junctions extract (22). The ratio of sgmRNA reads to the viral genomic RNA reads (sgmRNA ratio) was used to estimate the relative transcription activity of SARS-CoV-2. The sgmRNA ratio and its correlation with days post the first symptom were plotted using the R package *ggplot* (v.3.3.0). To avoid oversampling, for the patients with more than one sample, only the median sgmRNA ratio from samples of that patient was used for comparison among patients.

Detection of Intra-Host Variants

We defined an intra-host single nucleotide variant (iSNV) as the co-existence of an alternative allele and the reference allele at the same genomic position within the same sample. To identify iSNV sites, paired-end metatranscriptomic coronaviridae-like short read data were mapped to the reference genome (EPI_ISL_402119) using BWA aln (v.0.7.16) with default parameters (23). The duplicated reads were detected and marked using Picard MarkDuplicates (v. 2.10.10) (http://broadinstitute.github.io/picard). Nucleotide composition of each genomic position was characterized from the read mapping results using pysamstats (v. 1.1.2) (https://github.com/alimanfoo/pysamstats). The variable sites of each sample were identified using the variant caller LoFreq with default filters and the cut-off of 5% minor allele frequency ($n = 89$). After removing variable sites at UTR regions ($n = 12$), the sites with more than one alternative allele ($n = 0$), and the sites with only fixed variants (AAF > 95%) were filtered out ($n = 9$). All the iSNVs with less than five metatranscriptomic reads were verified using the hybrid capture data (at least two reads), and thus removed two iSNV sites. The rest 66 sites were regarded as iSNV sites. The identified iSNVs were then annotated using the SnpEff (v.2.0.5) with default settings (24). Alternative allele frequencies between synonymous and non-synonymous iSNV sites were tested with Wilcoxon rank sum test. Each dot indicates the median AAF of the same iSNV site of samples from same patient. All the plots were visualized using the R package ggplot (v.3.3.0).

Genetic Distance

The genetic distance between sample pairs was calculated using L1-norm distance, as defined by the following formula. To avoid oversampling, for the patients with more than one sample, only the median AAF among all samples of that patient was used for distance comparison. The L1-norm distance (D) between sample pairs is calculated by summing the distance of all the variable loci (N). The distance on each variable locus is calculated between vectors (p and q for each sample) of possible base frequencies ($n = 4$).

$$D = \sum_{k=1}^{N} \sum_{i=1}^{n} |p_i - q_i|$$

To verify the result, L2-norm distance (Euclidean distance) between sample pairs was calculated. The L2-norm distance $d(p, q)$ between two samples (p, q) is the square root of sum of distance across all the variable loci (N), as defined by the following formula.

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

The comparison of genetic distances among sample pair categories was performed using the Wilcoxon rank-sum test.

Beta Binomial Model of Bottleneck Size Estimation

A beta-binomial model was used to estimate bottleneck sizes between donor and recipient (12) (https://github.com/weissmanlab/BB_bottleneck). The beta-binomial model can estimate the probability of variant being detected in the recipient viral population under the prior condition of founding population, allowing variant frequency changes between founding time and sampling time. Here, the bottleneck size represents the number of virions that pass into the recipient and finally shape the sequenced viral population. The patient with the earlier symptom onset date was defined as the donor, while the other was defined as the recipient. For each variable site, variant frequencies within both donor and recipient, read depth and number of reads supporting the mutation within the recipient were used as input of the beta-binomial model. In this model, the virus transmission from donor to the recipient was regarded as a Bernoulli trial, and the probability of a given number of

virions carrying this mutation follows a binomial distribution. The maximum-likelihood estimates (MLE) of bottleneck sizes were estimated within 95% confidence intervals. In our data, we got 6 and 8 virions as the estimated transmission bottleneck size of the two donor-recipient pairs, as the probabilities of their beta-binomial distributions reached maximums, respectively.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study have been deposited into CNSA (CNGB Sequence Archive) of CNGBdb with the accession number CNP0001111 (https://db.cngb.org/cnsa/).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the first affiliated hospital of Guangzhou Medical University, the fifth affiliated hospital of Sun Yat-sen University, Qingyuan People's Hospital and Yangjiang People's Hospital, respectively. Informed consent was obtained from all participants enrolled in the study. All the privacy information was anonymized.

AUTHOR CONTRIBUTIONS

DW, YX, JL, WZ, and JZ conceived the study. YW, LZ, and YL collected clinical specimen and executed the experiments. DW, WS, XC, and JJ analyzed the data. DW, YW, and ZZ wrote the manuscript. All the authors participated in discussion and result interpretation and revised and approved the final version.

FUNDING

This study was funded by grants from The National Key Research and Development Program of China (2018YFC1200100, 2018ZX10301403, 2018YFC1311900), the emergency grants for prevention and control of SARS-CoV-2 of Ministry of Science and Technology (2020YFC0841400), Guangdong province (2020B111107001, 2020B111108001, 2020B111109001, 2018B020207013, 2020B111112003), the Guangdong Province Basic and Applied Basic Research Fund (2020A1515010911), Guangdong Science and Technology Foundation (2019B030316028), Shenzhen Peacock Team Plan grant (No. KQTD2015033117210153), Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011), and Guangzhou Medical University High-level University Innovation Team Training Program (Guangzhou Medical University released [2017] No.159). This work was supported by the Shenzhen Municipal Government of China Peacock Plan (KQTD2015033017150531). This work was supported by China National GeneBank (CNGB).

ACKNOWLEDGMENTS

This study was approved by the Health Commission of Guangdong Province to use patients' specimen for this study. We thank the patients who took part in this study. This manuscript has been released as a pre-print at the bioRxiv platform (Wang et al.).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.585358/full#supplementary-material

Supplementary Figure 1 | Structure of subgenomic messenger RNAs (sgmRNAs).

Supplementary Figure 2 | Transcription profile of subgenomic messenger RNAs (sgmRNAs) of each patient.

Supplementary Figure 3 | Heatmap representing the alternative allele frequencies (AAFs) of consensus and intra-host single nucleotide variants (iSNVs) of 30 patients.

Supplementary Figure 4 | Concordance between minor alternative allele frequencies (AAFs) derived from metagenomic and hybrid capture data.

Supplementary Figure 5 | Alternative allele frequency (AAF) distribution of rare and common iSNVs. Each dot indicates the median AAF of each iSNV sites of samples from same patient.

Supplementary Table 1 | Demography and clinical outcomes of COVID-19 patients.

Supplementary Table 2 | Summary of iSNVs.

Supplementary Table 3 | Frequency of iSNVs.

Supplementary Table 4 | Inter-host genetic distance (L1 and L2-norm).

Supplementary Table 5 | Bottleneck size of intra-household pairs.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. (2020) 579:265–9. doi: 10.1038/s41586-020-2008-3
2. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. (2020) 579:270–3. doi: 10.1038/s41586-020-2012-7
3. Smith EC, Sexton NR, Denison MR. Thinking outside the triangle: replication fidelity of the largest RNA viruses. *Annu Rev Virol.* (2014) 1:111–32. doi: 10.1146/annurev-virology-031413-085507
4. Day T, Gandon S, Lion S, Otto SP. On the evolutionary epidemiology of SARS-CoV-2. *Curr Biol.* (2020) 30:R849–57. doi: 10.1016/j.cub.2020.06.031
5. Nakagawa S, Miyazawa T. Genome evolution of SARS-CoV-2 and its virological characteristics. *Inflamm Regen.* (2020) 40:17. doi: 10.1186/s41232-020-00126-7
6. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature*. (2020) 581:221–4. doi: 10.1038/s41586-020-2179-y
7. Butler DJ, Mozsary C, Meydan C, Danko D, Foox J, Rosiene J, et al. Shotgun transcriptome and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *bioRxiv [Preprint]*. (2020). doi: 10.1101/2020.04.20.048066
8. Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, et al. Shared SARS-CoV-2 diversity suggests localised transmission of minority variants. *bioRxiv [Preprint]*. (2020). doi: 10.1101/2020.05.28.118992

9. Wang Y, Wang D, Zhang L, Sun W, Zhang Z, Chen W, et al. Intra-host variation and evolutionary dynamics of SARS-CoV-2 population in COVID-19 patients. *bioRxiv [Preprint]*. (2020). doi: 10.1101/2020.05.20.103549
10. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature*. (2020) 581:465–9. doi: 10.1038/s41586-020-2196-x
11. Sola I, Almazán F, Zúñiga S, Enjuanes L. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu Rev Virol.* (2015) 2:265–88. doi: 10.1146/annurev-virology-100114-055218
12. Sobel Leonard A, Weissman DB, Greenbaum B, Ghedin E, Koelle K. Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. *J Virol.* (2017) 91:0–19. doi: 10.1128/JVI.00171-17
13. Braun KM, Moreno GK, Halfmann PJ, Baker DA, Weiler AM, Haj AK, et al. Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. *bioRxiv [Preprint]*. (2020). doi: 10.1101/2020.11.16.384917
14. James SE, Ngcapu S, Kanzi AM, Tegally H, Fonseca V, Giandhari J, et al. High resolution analysis of transmission dynamics of SARS-CoV-2 in two major hospital outbreaks in South Africa leveraging intrahost diversity. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.11.15.20231993
15. Zwart MP, Elena SF. Matters of size: genetic bottlenecks in virus infection and their potential impact on evolution. *Annu Rev Virol.* (2015) 2:161–79. doi: 10.1146/annurev-virology-100114-055135
16. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes constrain the within and between host evolution of influenza virus. *Elife.* (2018) 7:e35962. doi: 10.7554/eLife.35962.036
17. Poon LLM, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, et al. Quantifying influenza virus diversity and transmission in humans. *Nat Genet.* (2016) 48:195–200. doi: 10.1038/ng.3479
18. Xue KS, Bloom JD. Reconciling disparate estimates of viral genetic diversity during human influenza infections. *Nat Genet.* (2019) 51:1298–301. doi: 10.1038/s41588-019-0349-3
19. Xue KS, Moncla LH, Bedford T, Bloom JD. Within-host evolution of human influenza virus. *Trends Microbiol.* (2018) 26:781–93. doi: 10.1016/j.tim.2018.02.007
20. Xiao M, Liu X, Ji J, Li M, Li J, Sun W, et al. Multiple approaches for massively parallel sequencing of HCoV-19 genomes directly from clinical samples. *Genome Med.* (2020) 12:57. doi: 10.1186/s13073-020-00751-4
21. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. (2015) 12:357–60. doi: 10.1038/nmeth.3317
22. Cotto KC, Feng Y-Y, Ramu A, Skidmore ZL, Kunisaki J, Conrad DF, et al. RegTools: Integrated analysis of genomic and transcriptomic data for the discovery of splicing variants in cancer. *bioRxiv [Preprint]*. (2020). doi: 10.1101/43663423
23. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. (2010) 26:589–95. doi: 10.1093/bioinformatics/btp698
24. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. (2012) 6:80–92. doi: 10.4161/fly.19695
25. Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, et al. Genomic diversity of severe acute respiratory syndrome-coronavirus 2 in patients with coronavirus disease 2019. *Clin Infect Dis*. (2020) 71:713–20. doi: 10.1093/cid/ciaa203

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.

frontiers in Public Health

SYSTEMATIC REVIEW
published: 10 June 2021
doi: 10.3389/fpubh.2021.599285

The Role of Whole Genome Sequencing in the Surveillance of Antimicrobial Resistant *Enterococcus* spp.: A Scoping Review

*Lindsay A. Rogers[1], Kayla Strong[1], Susan C. Cork[1], Tim A. McAllister[2], Karen Liljebjelke[1], Rahat Zaheer[2] and Sylvia L. Checkley[1]**

[1] Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada, [2] Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

OPEN ACCESS

Edited by:
Michael Kemp, University of Southern Denmark, Denmark

Reviewed by:
Claire Gorrie, The University of Melbourne, Australia
Henri Verbrugh, Erasmus University Medical Center, Netherlands

***Correspondence:**
Sylvia L. Checkley
slcheckl@ucalgary.ca

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

Received: *26 August 2020*
Accepted: *04 May 2021*
Published: *10 June 2021*

Citation:
Rogers LA, Strong K, Cork SC, McAllister TA, Liljebjelke K, Zaheer R and Checkley SL (2021) The Role of Whole Genome Sequencing in the Surveillance of Antimicrobial Resistant Enterococcus spp.: A Scoping Review. Front. Public Health 9:599285. doi: 10.3389/fpubh.2021.599285

Enterococcus spp. have arisen as important nosocomial pathogens and are ubiquitous in the gastrointestinal tracts of animals and the environment. They carry many intrinsic and acquired antimicrobial resistance genes. Because of this, surveillance of *Enterococcus* spp. has become important with whole genome sequencing emerging as the preferred method for the characterization of enterococci. A scoping review was designed to determine how the use of whole genome sequencing in the surveillance of *Enterococcus* spp. adds to our knowledge of antimicrobial resistance in *Enterococcus* spp. Scoping review design was guided by the PRISMA extension and checklist and JBI Reviewer's Guide for scoping reviews. A total of 72 articles were included in the review. Of the 72 articles included, 48.6% did not state an association with a surveillance program and 87.5% of articles identified *Enterococcus faecium*. The majority of articles included isolates from human clinical or screening samples. Significant findings from the articles included novel sequence types, the increasing prevalence of vancomycin-resistant enterococci in hospitals, and the importance of surveillance or screening for enterococci. The ability of enterococci to adapt and persist within a wide range of environments was also a key finding. These studies emphasize the importance of ongoing surveillance of enterococci from a One Health perspective. More studies are needed to compare the whole genome sequences of human enterococcal isolates to those from food animals, food products, the environment, and companion animals.

Keywords: *Enterococcus*, whole genome sequencing, scoping review, antimicrobial resistance, surveillance, One Health, nosocomial infection

INTRODUCTION

A variety of *Enterococcus* spp. are commensals within the gastrointestinal tract (GIT) of humans and animals, while others exist within the broader environment; some enterococcal species have also emerged as important human pathogens, especially in nosocomial infections (1). Two enterococcal species, *Enterococcus faecium* and *Enterococcus faecalis*, are most commonly implicated in human disease (2). These

species, in particular, can acquire antimicrobial resistance (AMR) and harbor virulence genes that give them advantages as opportunistic pathogens (3). Their acquisition of antimicrobial resistance genes (ARGs) can be chromosomally- and plasmid-mediated, arising from selection pressure through antimicrobial use and the transfer of ARGs on mobile genetic elements (MGEs) such as plasmids and transposons (4). Of particular concern is the rising emergence of vancomycin-resistant enterococci (VRE) (4). Acquired vancomycin resistance is mediated by various gene clusters termed VanA/B/D/E/G/L/M/N (4, 5). Each gene cluster codes for a different resistance mechanism. VanA and VanB are the most common clusters seen, are often hospital-acquired, and can be plasmid- or chromosomally- mediated (5–7). Two species, *Enterococcus gallinarum* and *Enterococcus casseliflavus*, have intrinsic vancomycin resistance that is chromosomally-mediated by the VanC gene cluster (4, 5).

These two species, along with others such as *Enterococcus villorum, Enterococcus thailandicus, Enterococcus durans,* and *Enterococcus hirae*, are more typical of animal- and environmentally- adapted species (8, 9). Their genomes reflect adaptations to specific niches; for example, clusters of orthologous groups (COGs) have been found for ethanolamine utilization as a carbon source in environmental species that are not found in *E. faecium* (8).

Other antimicrobials can co-select for vancomycin-resistance genes if there are multiple ARGs on a single mobile genetic element. This means that the use of other antimicrobials can lead to the acquisition of vancomycin-resistance genes, even if vancomycin is not used (10). Thus, the VRE that arise in this manner are also multi-drug resistant (MDR) or multi-class resistant (7) referring to organisms that have acquired resistance to two or more antimicrobials or those organisms that have acquired resistance to two or more classes of antimicrobials, respectively.

As AMR pathogens have emerged, the use of antimicrobials for prophylaxis and metaphylaxis in food producing animals has come under scrutiny for its potential to apply a selective pressure that contributes to the dissemination of AMR and MDR enterococci (7, 11). Perhaps the most classic example is in poultry and swine where the previous use of a vancomycin-related glycopeptide, avoparcin, as a growth promoter was associated with carriage of vancomycin-resistant *E. faecium* (VREfm) in treated herds or flocks through cross-resistance (7, 12). The occurrence of VREfm in food animals decreased after the avoparcin use in animals was banned (12); however, prolonged persistence of specific VREfm clusters in agricultural settings (e.g., VanA gene cluster on the Tn*1546* transposon) were observed possibly due to co-selection for VRE through continued use of other antimicrobials, such as macrolide use in swine (7, 12, 13). Retrospective molecular and genetic studies have demonstrated that the VREfm isolated from hospital and agricultural setting are usually separate sequence types (7, 12, 13). In contrast, the same sequence types of vancomycin-resistant *E. faecalis* can be isolated from hospital settings and from farm animals (7, 12, 14). In human medicine, enterococci are often opportunistic pathogens that acquire resistance and arise in immunocompromised individuals in hospital settings. Often these patients have been treated with multiple classes of antimicrobials in an effort to control difficult to treat infections (6). These antimicrobials may include those deemed critically important for use in human medicine by WHO (10), leading to enterococci resistant to these antimicrobials circulating in the human population (10).

Due to the importance of *Enterococcus* spp. as potential human pathogens, their ability to easily acquire ARGs, and their ubiquitous nature in the GIT and the broader environment, many countries have added enterococcal species to their list of pathogens under surveillance. Their role as a GIT commensal also make enterococci useful as fecal indicator bacteria (15). Surveillance for *Enterococcus* spp. and the other so-called ESKAPE pathogens (ESKAPE is an acronym for the following six nosocomial pathogens: *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter* spp.) (16) is done on a national level in several countries. These include programs such as the Danish Integrated Antimicrobial Resistance Monitoring and Research Program (DANMAP) (17), USA's National Antimicrobial Resistance Monitoring System (NARMS) (18), the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) (19), and the Colombian Integrated Program for Antimicrobial Resistance Surveillance (COIPARS) (15). Many hospitals and laboratories also have surveillance systems in place for specific organisms independent of national surveillance programs. Examples include JMI Laboratories' SENTRY Antimicrobial Surveillance Program and the University of Pittsburgh Medical Center-Presbyterian Hospital's Enhanced Detection System for Hospital-Acquired Transmission (UPMC EDS-HAT) (6, 20, 21). The European Antimicrobial Resistance Surveillance Network (EARS-Net) is a large surveillance program based on clinical antimicrobial resistance data from laboratories across Europe (22). The pharmaceutical industry also runs some important post-marketing surveillance programs to comply with licensing requirements of new antimicrobials, looking at potency and spectrum. Examples of these programs are the Zyvox Annual Appraisal of Potency and Spectrum (ZAAPS) and the Linezolid Experience and Accurate Determination of Resistance (LEADER) (23). Integrated surveillance programs survey and address AMR in humans, animals, and the environment from a One Health perspective, emphasizing the interfaces within the system. These programs collect samples and process isolates from various sources, including animal fecal samples, human screening samples, retail meats, wastewater, surface water, groundwater and soils (15). Human screening samples include those from hospital surveillance programs of in-patients and samples submitted to laboratories performing surveillance (6, 24–26). The data generated from the wide range of samples can be integrated with antimicrobial use data allowing for the monitoring of changes in antimicrobial resistance found in bacteria important to public and animal health (15). Information gleaned from surveillance can then inform policy and risk mitigation strategies to combat increasing AMR and protect antimicrobials important to human health (15). For example, surveillance of VRE through DANMAP allowed for the detection

TABLE 1 | Search results table from Web of Science database search with search terms and resulting number of articles.

Set*	Results¥	Search terms‡
4	283	#3 AND #2 AND #1 *Databases= WOS, BIOSIS, KJD, MEDLINE, RSCI, SCIELO Timespan=2002-2020* *Search language=English*
3	3,139,503	**TOPIC:** (survey) *OR* **TOPIC:** (surveillance) *OR* **TOPIC:** (epidemiolog*) *Databases= WOS, BIOSIS, KJD, MEDLINE, RSCI, SCIELO Timespan=2002-2020* *Search language=English*
2	485,212	**TOPIC:** (whole genome sequenc*) *OR* **TOPIC:** (WGS) *OR* **TOPIC:** (next generation sequenc*) *OR* **TOPIC:** (NGS) *OR* **TOPIC:** (genomic*) *Databases= WOS, BIOSIS, KJD, MEDLINE, RSCI, SCIELO Timespan=2002-2020* *Search language=English*
1	46,024	**TOPIC:** (enterococc*) *OR* **TOPIC:** (enterococcus) *Databases= WOS, BIOSIS, KJD, MEDLINE, RSCI, SCIELO Timespan=2002-2020* *Search language=English*

**Set: Number assigned by Web of Science database to search term.*

¥ Results: Number of articles in Web of Science database matching search terms for each set.

‡Search terms: Terms used in each set for the search. The numbers given in Set 4 represent the combination of other sets used for the search.

of VRE strains in broiler chickens and human isolates connected to the use of avoparcin for growth promotion in broilers. This detection led to the ban of avoparcin use in food animal production (7, 15).

Early surveillance was primarily based on traditional microbiology to determine phenotypic antimicrobial susceptibility profiles, molecular genomics methods such as polymerase chain reactions (PCR) to assess for the presence of resistance genes, and pulsed-field gel electrophoresis (PFGE) for DNA fingerprinting Multi-locus sequence typing (MLST) arose more recently to better assess genetic relationships among isolates (7, 27). The use of these technologies allowed for the phylogenetic study of sequence types, epidemiologic investigation and determination of the presence of specific ARGs. However, the development of whole genome sequencing (WGS) has provided a more in-depth and detailed analysis of enterococcal ARGs, phylogenetics, and virulence (7, 27). As whole genome sequencing has become more widely available and less expensive, many archived isolate collections are being reanalyzed and their genomes compared with new isolates (13, 28). Following WGS, it has become possible to utilize new sequence typing methods for enterococci, such as core-genome multi-locus sequence typing (cgMLST), allowing for better analysis of isolate relatedness across sample sources (28). WGS also allows for the identification of emerging strains, analysis of outbreaks, and the characterization of resistance and virulence genes and their locations and context in the bacterial genome. Due to these advantages many surveillance research groups have been transitioning to WGS-based approaches of isolate characterization. Sequencing compliments traditional microbiology approaches and offers a reliable method of characterizing ARGs and sequence types (27).

With the increasing popularity of WGS for bacterial pathogen surveillance, it is now imperative to review the progress that has been made toward surveillance methods and identify gaps in our surveillance knowledge. We have undertaken this scoping review to investigate and summarize the extent to which whole genome sequencing in surveillance studies has advanced our understanding of AMR in *Enterococcus* spp.

METHODS

To investigate our research question, a scoping review was designed following the PRISMA-ScR extension for scoping reviews (29) and the guidelines laid out by the JBI Reviewer's Manual (30). This protocol was not registered with an online registration platform.

Search Terms and Strategy

The Population, Concept, Context (PCC) framework (30) was used to develop the research question and search strategy. A search strategy was developed to return a broad range of studies that fit within the following population, concept, and context:

- Population: *Enterococcus* spp. that underwent whole genome sequencing
- Context: *Enterococcus* spp. isolates derived from surveillance-type studies (as described below)
- Concept: use of whole genome sequencing to better understand antimicrobial resistance.

After consultation with a librarian, three separate databases were selected for searching: PubMed (NCBI)[1], Web of Science[2], and CAB Abstracts[3] All databases within Web of Science were included in order to include both the Web of Science Core Collection and BIOSIS. Searches were performed by a single reviewer (LR) on November 17, 2019, and email search alerts

[1]https://www.ncbi.nlm.nih.gov/pubmed/

[2]https://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=5D5o7Naqz9xk9WXkIZ8&preferencesSaved=

[3]http://web.a.ebscohost.com.ezproxy.lib.ucalgary.ca/ehost/search/advanced?vid=0&sid=59def56a-2134-4464-bd34-368a168a49ab%40sessionmgr4008

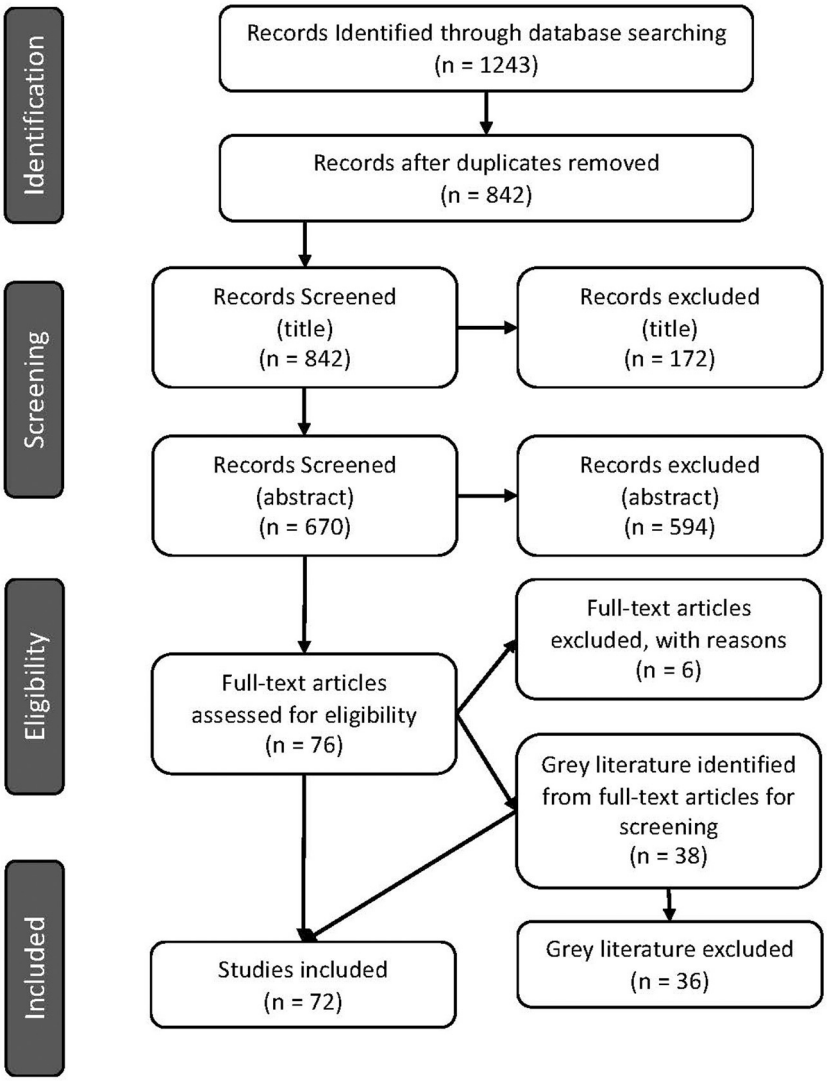


FIGURE 1 | PRISMA flow diagram indicating number of articles screened and identified for inclusion in the scoping review (29).

from each database were implemented to inform the reviewer of any new articles. New articles from alerts up to and including December 31, 2020 were included. The same search terms were used for all three databases, except for differences due to specific database formatting.The search terms and resulting number of articles for one database (Web of Science) are described in **Table 1**. To capture relevant gray literature, a manual search of the reference lists of included articles was completed during data extraction. Gray literature is information produced and distributed outside of academic publications such as government reports. The gray literature underwent the same screening process; however, the full text was read for screening if no abstract was available. The web application, Rayyan, was used for the organization of articles during the screening process (31).

Article Screening and Selection

Only journal articles and abstracts in the English language, published after 2002 were eligible for screening. The entire genome of *E. faecium* was first sequenced in 2000 (32); however, the assembly was not completed until 2012 (33, 34). The earliest available genome sequence of an *Enterococcus* sp. from the NCBI database is from 2002 (35). Given this information, 2002 was considered the earliest year that a publication would contain the information relevant to this scoping review. Any publication that was not an article or abstract (e.g., textbook, poster, or conference presentation) was excluded. Relevant gray literature articles were included and searched for, as described above. All screening was done independently by two reviewers (LR and KS). Any discrepancies were resolved in discussion with two other reviewers (SLC and SCC).

Title Screening

The article titles were screened initially and any article that was clearly about bacteria other than *Enterococcus* spp. was excluded. These articles needed to explicitly include the name of bacteria other than *Enterococcus* spp. in the title and not include terms relating to the taxonomy of enterococci. All articles that did not meet this exclusion criterion were included for the next screening step. The title screening was intentionally left broad to maximize the number of articles included.

Abstract Screening

Two screening steps were applied to the included abstracts. The first abstract screening step was performed to exclude any articles that did not include whole genome sequencing and antimicrobial resistance of *Enterococcus* spp. The abstract had to include all three pieces of information (i.e., WGS, AMR, and *Enterococcus*) in the abstract text. This step was also intentionally left broad and any articles that mentioned sequencing without providing information about whether or not the whole genome was sequenced were included to be screened based on methodology (described below). Following this, the abstracts were screened a second time using the following question, "did all or a portion of the *Enterococcus* spp. isolates in this study result from surveillance or screening for enterococci?" The following criteria for surveillance or screening were used:

- Isolates were from a collection maintained by a surveillance group (a surveillance group is defined as an organization collecting and analyzing bacterial isolates for surveillance of those particular bacteria such as CIPARS, SENTRY, or DANMAP). OR,
- A statement was included that the isolates were collected for screening or surveillance purposes. OR,
- Isolates were collected for the sole purpose of genomic comparison. OR,
- The article was published in a journal which included "Surveillance" in the journal name.

Articles needed to meet one or more criteria. Articles that did not meet these criteria, such as those with only clinical isolates, were excluded. Articles that were unclear if they met the surveillance inclusion criteria through their abstract were screened based on methodology as described below.

Methods Screening

As stated previously, some abstracts did not contain enough detail to determine if they met the inclusion/exclusion criteria. These articles were further screened through the reading of their methods sections, following the same criteria as for abstract screening. Articles that did not meet the inclusion criteria during the methods review were excluded.

The number of articles excluded at each screening step is displayed in **Figure 1**.

Data Extraction and Charting

Data was extracted independently by two reviewers (LR and KS) to answer the research question. The chart was trialed with 15 articles to ensure the reviewers were extracting comparable information. The completed tables from data extraction were compared by one reviewer (LR), and any discrepancies in information were resolved in discussion between the two original reviewers (LR and KS). No critical appraisal of articles was performed and all articles were included regardless of study quality. The methodology of data extraction is outlined in **Table 2**.

RESULTS

Article Characteristics

Seventy-two articles were included after the full-text review (**Figure 1**). Of these, 70 were primary research articles and two were gray literature (government reports) (25, 36). All articles were published in 2015 or later. The corresponding authors were from seventeen (17) different countries with most from Australia (16.7%), Denmark (13.9%), and Germany (13.9%), followed by the USA (12.5%), then the UK (6.9%) and 5.6% from each of Canada, China, the Netherlands and Portugal. Corresponding authors were also from Brazil (4.2%), Italy (1.4%), Japan (1.4%), Colombia (1.4%), Spain (1.4%), South Africa (1.4%), South Korea (1.4%), and Saudi Arabia (1.4%) (**Figure 2**). Additional information is available in **Supplementary Table 1**.

Just under half of the studies were not associated with a specified surveillance group (48.6%). The remaining articles were associated with government funded programs (e.g., DANMAP, NARMS), within hospital screening or surveillance programs, or private/industry funded surveillance programs (**Table 3** and **Supplementary Table 1**).

Objectives of Included Articles

The study objectives could be broadly divided into five categories, as shown in **Table 4**. These include (i) epidemiology or prevalence studies, (ii) genetic and/or molecular characterization, (iii) comparative genomics, (iv) the description of a novel finding, and (v) the comparison of techniques or development of a new technique. The categories were formed by assessing keywords within the stated objectives. For example, objectives to "determine the prevalence of…" or "perform comparative genomics of…" would fit into the categories of epidemiology or prevalence studies and comparative genomics, respectively. The categories were further broken down into the sample source location, and whether or not the article was studying a specific resistance gene (e.g., *vanA*), AMR profile (e.g., VRE), or another specific category (e.g., another specific genetic element such as a transposon). Articles could fall into more than one type of objective but generally stayed within the same category as the source location. The majority of articles were an epidemiological or prevalence study (54.2%), most commonly conducted within a defined region. Often these studies included human clinical data and hospital isolates but were not specific to just one hospital. Only three studies compared techniques or developed new ones, and this was in conjunction with other objectives of the study. Details of

TABLE 2 | Data extraction chart with description and example article.

Data	Description	Example
Article characteristics		
Citation	Article reference using APA 6th style	Leong, K. W., Cooley, L. A., Anderson, T. L., Gautam, S. S., McEwan, B., Wells, A., Wilson, F., Hughson, L., and O'Toole, R. F. (2018). Emergence of vancomycin-resistant *Enterococcus faecium* at an Australian Hospital: a whole genome sequencing analysis. *Scientific Reports* 8:6274.
Article type	List the study design or type	Cross-sectional
Research group country	Name the country from where the research group originates	Australia
Surveillance group	List the surveillance group(s) performing the study and/or with the isolate collection (if applicable)	Tasmania Infection Prevention and Control Healthcare Associated Infection Surveillance Program
Funders	List the funding groups for the study	- Royal Hobart Hospital Research Foundation grant - Tasmanian Infection Prevention and Control Unit
Introduction		
Study objectives	Bullet point summary of stated objectives	- To better define hospital spread of VREfm at the RHH - Will correlate genomic information with epidemiologic data
Methods		
Study location	List where the samples are from or where the study was performed	Royal Hobart Hospital (RHH) in Tasmania
Enterococcus spp.	List the enterococcal species isolated and studied	*E. faecium*
VREfm specific study (yes/no)	Yes: the study is specifically studying VREfm No: the study is not specifically looking for VREfm (although it may be included in the results)	Yes
Sample sources	List the sources from where the isolates were obtained	- Human screening - Human clinical
AMS testing	List the methods used for antimicrobial susceptibility testing	Done in previous study—disc diffusion (EUCAST)
WGS Platform	List the WGS platform(s) used	Illumina MiSeq
Archive accession numbers provided (yes/no)	Yes: accession numbers provided No: accession numbers not provided	Yes
Bioinformatic tools	List any bioinformatics tools used according to their purpose	Alignment: Snippy Phylogeny (SNPs): RaxML Assembly: Velvet AMR: Resfinder Typing: MLST Tool
Other genomics	List any other genomic analyses performed	None
Results		
AMR phenotypes	List the antimicrobials to which resistance was found (percent or ratio of isolates in parentheses) Make note of any particular phenotypic patterns	- Vancomycin (100%)
Sequence types and/or clonal complexes	List any sequence types and/or clonal complexes found	ST796 (47/80) ST80 (16/80) ST1421 (10/80) ST203 (4/80) ST78 (1/80) ST192 (1/80) ST555 (1/80)
AMR genes	List of AMR genes found through WGS, according to sample source If >10 AMR genes were found, reference where in the article they may be found, rather than providing an exhaustive list	*vanA* *vanB*

(Continued)

TABLE 2 | Continued

Data	Description	Example
Plasmids	List of any plasmids found	None described
Relatedness assessed	Brief description of the relatedness that was assessed	Study isolates to reference genome (SNPs) Comparison of genomic to epidemiologic data
Discussion		
Addition to AMR knowledge	Summary of addition to knowledge about AMR in *Enterococcus*	- VREfm profile at RHH has shifted to ST796 and ST80 - Screening is important to detect isolates that may be involved in transmission - WGS is helpful for more accurate typing of *E. faecium* for better investigations

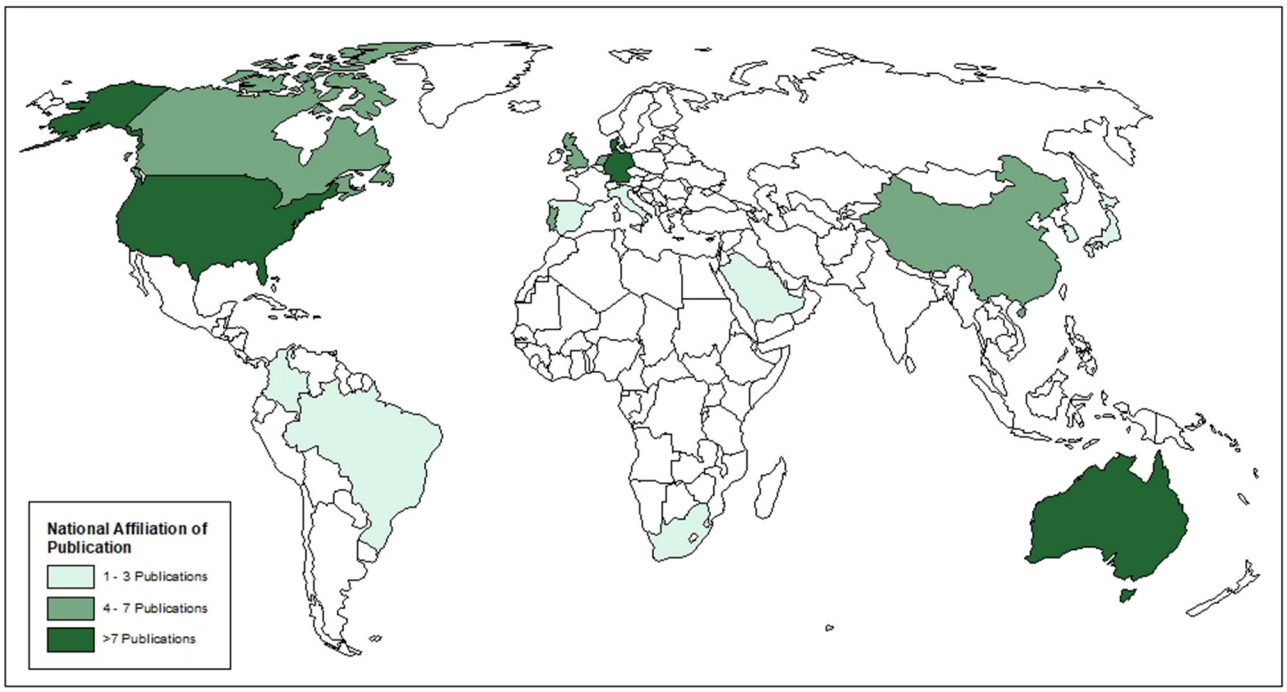


FIGURE 2 | Country of affiliation of corresponding author for included articles.

TABLE 3 | Proportion of articles funded and/or performed by a specific surveillance group.

Specified surveillance group	Proportion of articles (%)	Article citation
None listed	48.6	(8, 9, 11, 24, 37–67)
Government funded program	30.6	(25, 27, 28, 36, 68–85)
Private/Industry funded program	4.2	(13, 26, 86)
Within hospital program	16.7	(6, 20, 87–96)

the objectives for each study are presented in **Table 4** and **Supplementary Table 2**.

Enterococcal Source and Species

Nine different source types were sampled for surveillance of *Enterococcus* spp. isolates, with the majority of articles using isolates from human clinical infections (47 articles, 65.3%). Just over half of the articles (38 articles, 52.8%) used isolates from human screening samples. For animal samples, 12 articles (16.7%) pulled isolates directly from food animals (fecal, cecal, or other samples), 12 articles (16.7%) had retail meat samples, and three articles (4.2%) included isolates from milk products. Isolates were also from environmental samples with eight articles (11.1%) including samples from

TABLE 4 | Categorization of stated objectives for articles included in scoping review and proportion of included articles in each category.

Objective category and source	Article specific to a gene, AMR profile, or other genetic element?	Proportion of articles (%)	Article citation
Epidemiology and/or prevalence study (54.2%)			
Hospital or clinical setting only	Yes	16.7	(6, 26, 46, 49, 62, 72, 73, 90, 92, 93, 95, 96)
	No	1.4	(64)
Regionally specific study (may include clinical setting)	Yes	20.8	(24, 27, 28, 37, 45, 47, 50, 63, 74, 75, 78–80, 82, 85)
	No	4.2	(25, 36, 65)
Outbreak study	Yes	6.9	(20, 51, 52, 87, 94)
	No	–	
Food animal related (including retail meats and animal samples)	Yes	2.8	(55, 67)
	No	–	
Other	Yes	–	
	No	1.4	(41)
Comparative genomics (29.1%)			
Regionally specific study (may include clinical setting)	Yes	8.3	(28, 56, 63, 75, 77, 83)
	No	1.4	(53)
Food animal related (including retail meats and animal samples)	Yes	–	
	No	6.9	(8, 13, 38, 68, 84)
Other (multiple sources)	Yes	2.8	(58, 59)
	No	9.7	(9, 38–41, 54, 57)
Genetic and/or molecular characterization (23.5%)			
Hospital or clinical setting only	Yes	8.3	(26, 48, 49, 60, 89, 93)
	No	–	
Regionally specific study (may include clinical setting)	Yes	6.9	(44, 66, 75, 81, 91)
	No	–	
Food animal related (including retail meats and animal samples)	Yes	6.9	(55, 61, 67, 69, 70)
	No	1.4	(54)
Description of novel finding (14.0%)			
Hospital or clinical setting only	Yes	2.8	(60, 86)
	No	–	
Regionally specific study (may include clinical setting)	Yes	4.2	(37, 74, 81)
	No	–	
Outbreak study	Yes	1.4	(88)
	No	–	
Food animal related (including retail meats and animal samples)	Yes	2.8	(11, 71)
	No	–	
Other	Yes	1.4	(43)
	No	1.4	(76)
Comparison or development of technique(s) (5.6%)			
Hospital or clinical setting only	Yes	2.8	(6, 92)
	No	–	
Regionally specific study (may include clinical setting)	Yes	1.4	(27)
	No	–	
Outbreak study	Yes	–	
	No	1.4	(42)

TABLE 5 | Summary of enterococcal source and species from articles included in scoping review and proportion of included articles identifying each source or species.

Enterococcal characteristics	Proportion of articles (%)	Article citation
Sample sources		
Human clinical infection	65.3	(24–28, 36–39, 41–53, 56, 58, 60, 63, 64, 66, 72–75, 77, 78, 80–83, 86–92, 94, 96)
Human screening samples	52.8	(6, 20, 24, 25, 27, 42–48, 50, 52, 58, 60, 62–65, 72, 74–77, 81–83, 85, 87, 88, 90–96)
Retail meats	16.7	(11, 13, 25, 38, 39, 43, 55, 58, 63, 67, 69, 70)
Food animal samples	16.7	(8, 25, 38, 39, 43, 54, 58, 61, 63, 68, 71, 84)
Wastewater	11.1	(9, 38, 39, 41, 43, 58, 59, 63)
Hospital environment	11.1	(46, 51, 56, 60, 64, 87, 88, 94)
GenBank sequence	6.9	(11, 40, 57, 58, 90)
Milk products	4.2	(40, 43, 57)
Other	1.4	(65)
***Enterococcus* spp.**		
E. faecium	87.5	(6, 8, 9, 13, 20, 24–28, 36–39, 41–43, 45–52, 54–57, 59, 60, 62–67, 70–83, 85–96)
E. faecalis	34.7	(6, 8, 9, 13, 25, 26, 28, 36, 38, 40, 43, 44, 48, 53, 54, 58, 59, 61, 65, 66, 68, 69, 71, 78, 79)
E. gallinarum	13.9	(6, 8, 9, 13, 36, 38, 54, 59, 60, 67)
E. hirae	11.1	(8, 11, 13, 36, 38, 54, 59, 65)
E durans	9.7	(8, 13, 38, 43, 54, 59, 67)
E. casseliflavus	9.7	(6, 8, 9, 36, 38, 65, 81)
Other	13.9	(8, 11, 13, 36, 38, 43, 59, 65, 81, 84)
VREfm specific study?		
Yes	48.6	(20, 24, 27, 28, 37, 41, 42, 45–47, 50–52, 55, 62, 70, 72–77, 80, 82, 83, 85, 87, 88, 90–96)
No	51.4	(6, 8, 9, 11, 13, 25, 26, 36, 38–40, 43, 44, 48, 49, 53, 54, 56–61, 63–69, 71, 78, 79, 81, 84, 86, 89)

the hospital environment and eight articles (11.1%) using wastewater samples. Five articles (6.9%) included sequences from GenBank to supplement their isolates from sample sources (**Table 5**). No studies sampled from companion animals or equids.

At least 14 different species of *Enterococcus* were described in the 72 articles (one article did not identify all species). The majority of articles (87.5%) described *Enterococcus faecium* with 35 (48.6%) of these articles specifically studying VREfm. Twenty-five articles (34.7%) identified *E. faecalis*, 10 articles (13.9%) *E. gallinarum*, eight articles (11.1%) *E. hirae*, and seven studies (9.7%) each described *E. casseliflavus* and *E. durans*. Ten articles described other species, including *E. villorum*, *E. thailandicus*, *E. cecorum*, *E. mundtii*, *E. pseudoavium*, *E. ratti*, *E. avium*, and *E. raffinosus*.

All 72 articles provided some information on the AMR phenotypes of their isolates. The majority of articles (80.6%) describe isolates with resistance to glycopeptide antibiotics (vancomycin or teicoplanin). Twenty-one articles (29.2%) described isolates with resistance to oxazolidinones (linezolid or tedizolid). Other antimicrobial classes with identified phenotypic resistance included fluoroquinolones, macrolides, aminoglycosides, penicillins, and tetracyclines (**Figure 3** and **Supplementary Table 3**). Methods used to assess phenotypic resistance in each article are outlined in **Supplementary Table 4**. Nineteen articles (26.4%) provided no information on the methodology used to define phenotypic resistance.

WGS Platforms and Results

All articles selected performed whole genome sequencing of the isolates (as a part of the inclusion criteria) and Illumina was the most commonly used platform for sequencing. Sixty-four articles (88.9%) used a version of Illumina for sequencing, whereas four articles did not provide sequencing methodology. Illumina MiSeq was the most commonly used version of Illumina employed. Other platforms included PacBio, Ion Torrent PGM, or Illumina in combination with PacBio, Ion Torrent PGM, or MinIon platforms (**Table 6**). Sixty-two articles (86.1%) provided archive accession numbers for access to their resultant sequences.

Whole genome sequencing generated information about AMR genes that was described in almost all articles, with two articles (2.8%) failing to report AMR genes (**Figure 4** and **Supplementary Table 3**). Forty-one articles (56.9%) reported the vanA gene cluster, 26 articles (36.1%) reported vanB, and 15 articles (20.8%) reported optrA.

From the sequencing, 61 articles (84.7%) described the sequence type (ST) and/or clonal complex (CC) of isolates. Over 60 sequence types were reported, including

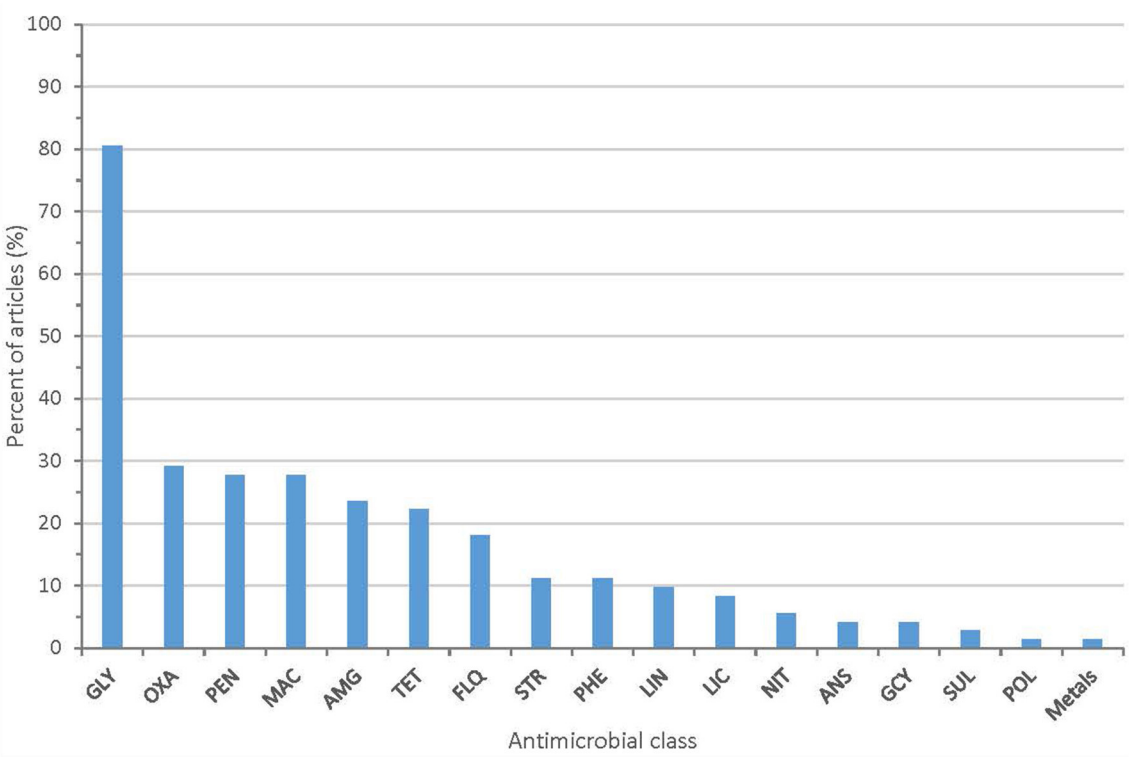


FIGURE 3 | Proportion of articles included in scoping review identifying resistance to each antimicrobial class (GLY, glycopeptides; PEN, penicillins; MAC, macrolides; AMG, aminoglycosides; OXA, oxazolidinones; TET, tetracyclines; FLQ, fluoroquinolones; PHE, phenicols; LIN, lincopeptides; STR, streptogramins; LIC, lincosamides; ANS, ansamycins; GCY, glycylcyclines; NIT, nitrofurans; SUL, sulfonamides; POL, polypeptides).

TABLE 6 | Summary of WGS platforms used in each article included in scoping review and whether archive accession numbers were provided in the article.

Sequencing platform		Proportion of articles (%)	Article citation
WGS platform			
Illumina	MiSeq	48.6	(6, 8, 9, 24–28, 38, 40, 44, 46, 53, 57, 60, 61, 66, 68, 72, 73, 78–80, 82–88, 90, 92, 94–96)
	NextSeq	9.7	(20, 50–52, 54, 55, 80)
	HiSeq	11.1	(41, 43, 59, 63–65, 69, 89)
	Version not specified	6.9	(13, 49, 56, 67, 76)
Illumina & ...	PacBio	6.9	(39, 47, 71, 74, 75)
	Ion Torrent	2.8	(42, 48)
	MinIon	4.2	(62, 70, 81)
PacBio		2.8	(11, 91)
IonTorrent PGM		2.8	(37, 45)
No information		5.6	(36, 58, 77, 93)
Archive accession numbers provided?			
Yes		86.1	(6, 8, 9, 11, 13, 24, 26, 27, 37–45, 47–66, 68–72, 74–86, 89–95)
No		13.9	(20, 25, 28, 36, 46, 67, 73, 87, 88, 96)

ST203 (25 articles), ST80 (27 articles), ST78 (19 articles), and ST18 (15 articles). Seven articles reported a pstS-null isolate. The summary of clonal complexes is in **Figure 5** and **Supplementary Table 3**.

Many articles (75.0%) reported other molecular techniques, including PCR, PFGE, and MALDI-TOF MS. These techniques were primarily used for speciation or screening for AMR genes (**Supplementary Table 4**).

Article Findings and Conclusions

The articles' findings or conclusions related to antimicrobial resistance were summarized into categories, and articles

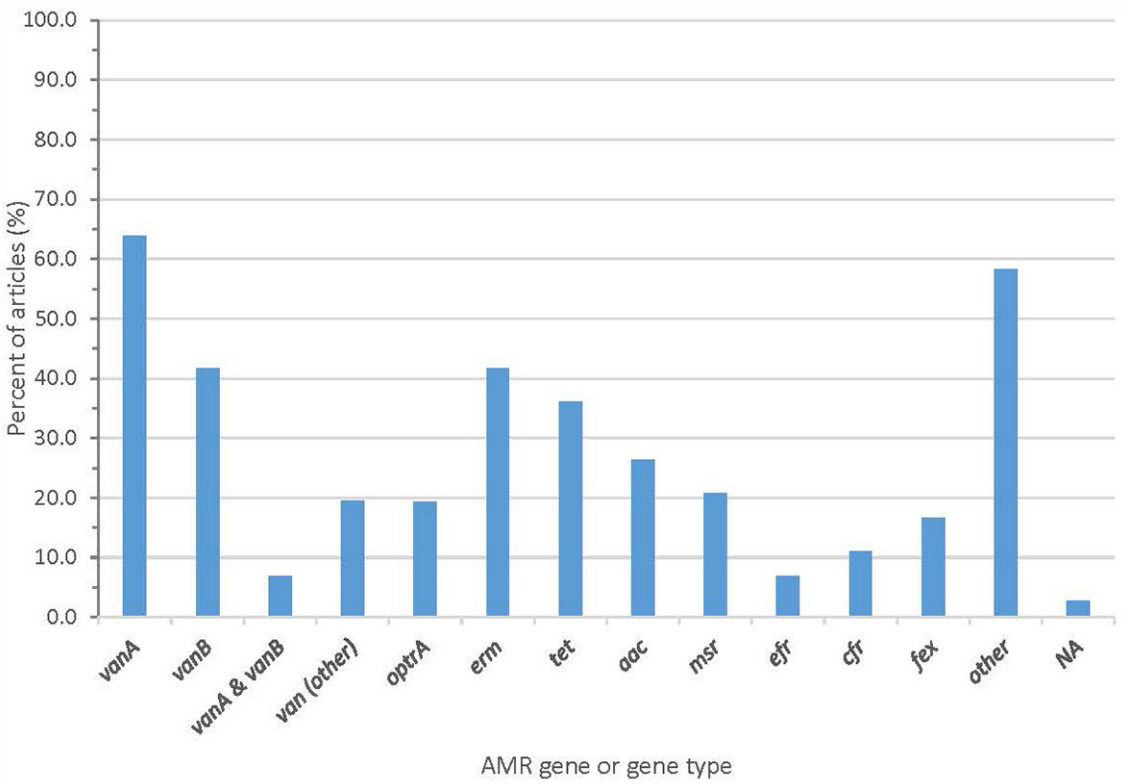


FIGURE 4 | Proportion of articles included in scoping review finding an AMR gene or gene type.

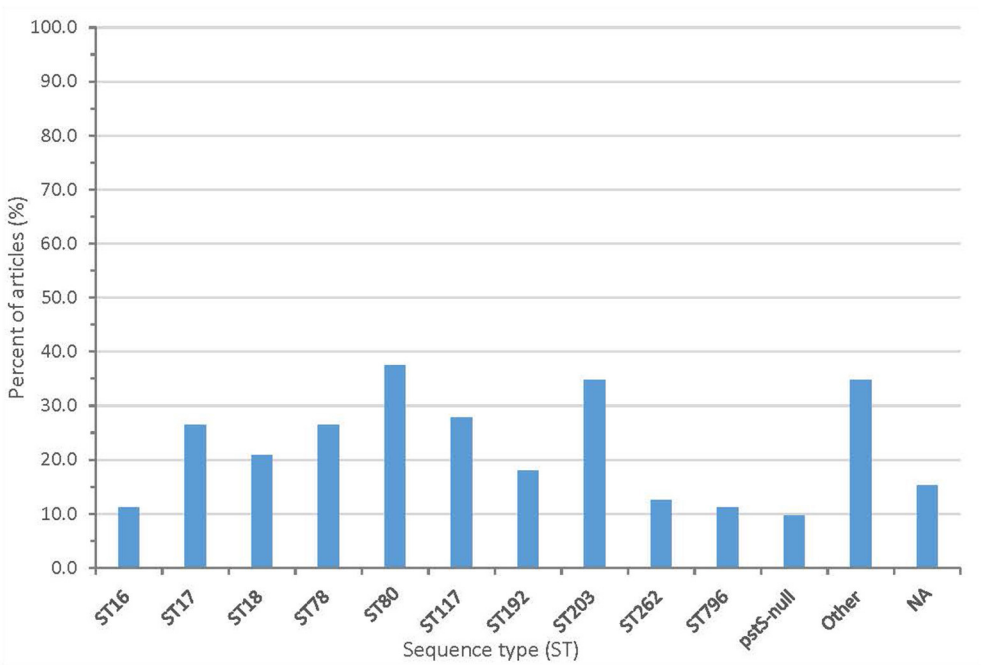


FIGURE 5 | Proportion of articles included in scoping review finding different sequence types (NA, no sequence type described).

TABLE 7 | Main categories of findings and conclusions and proportion of articles included in scoping review within each category.

Finding and/or conclusion	Proportion of articles (%)	Article citation
Report of a new or uncommon finding	31.9	(6, 28, 37, 43–45, 49, 52, 53, 59–63, 66, 69, 71, 74, 76, 77, 81, 86, 87)
Enterococcus is well optimized for adaptation and survival in their environment	20.8	(8, 9, 11, 26, 39, 46, 57, 62, 64, 68, 75, 78, 89, 94, 95)
WGS is the best means of detecting and differentiating *Enterococcus* spp.	16.7	(20, 27, 42, 50, 64, 72, 74, 80, 82, 88, 91, 92)
Surveillance of *Enterococcus* spp. is important	11.1	(28, 69, 73, 81, 82, 88–90)
Finding of suspected HGT or co-selection	9.7	(11, 49, 58, 59, 68, 70, 75)
Other*	16.7	(6, 38, 47, 48, 65, 78, 79, 83, 84, 91, 93, 95)
Importance of infection and control measures in hospitals	11.1	(20, 27, 41, 79, 87, 88, 94, 95)
VREfm prevalence is increasing in hospitals	11.1	(25, 28, 73, 74, 82, 85, 88, 96)
Human clinical strains have a higher number of ARGs than animal or human screening isolates or no relationship found between human/animal strains	9.7	(27, 38–40, 54, 57, 63)
Animal strains can carry ARGs and MGEs	8.3	(8, 13, 39, 41, 70, 84)
Information on antimicrobials and related resistance	8.3	(24, 44, 47, 53, 56, 67)
Finding of a shift to a new dominant strain	11.1	(50, 53, 72, 75, 80, 83, 85, 96)
Surveillance Program Report	5.6	(24, 25, 36, 51)
There is apparent transmission between human and animal enterococci (either human to animal or animal to human)	4.2	(43, 55, 58)
Importance of screening in hospitals	2.8	(72, 87)
Human wastewater contains human clinical strains	2.8	(9, 41)

**Other refers to article findings that did not fit into another defined category.*

could fit into more than one category. Of all articles, 31.9% reported a new or uncommon finding, which could include a novel strain or gene, or a previously described finding in a novel location. A fifth (20.8%) of articles reported how *Enterococcus* spp. were optimized for adaptation and survival in their environment. Twelve articles (16.7%) specifically stated that WGS was a better means of detecting or differentiating *Enterococcus* spp. than other genomic methods (such as PCR or PFGE). The other summarized findings are described in **Table 7** with more details are available in **Supplementary Table 2**.

DISCUSSION

In this scoping review, we aimed to assess the value added by the use of WGS in the surveillance of enterococcal AMR. The use of WGS for the surveillance of *Enterococcus* spp. has added to our knowledge about *Enterococcus* spp. through the detection of previously unidentified strains and finding that WGS was better at detecting and differentiating *Enterococcus* spp. than other genomic methods. European countries provided the most surveillance studies, with many of these coming from the Danish DANMAP program. This is perhaps unsurprising as DANMAP is an extensive and well-established program implemented in 1995 (25).

Importance of WGS to Detect AMR *Enterococcus* spp.

While the majority of *Enterococcus* spp. are adapted to the natural environment and animal GITs, rarely causing disease in humans, *E. faecium* and *E. faecalis* are the species most likely to cause human disease (2). Thus, it was not surprising that they were the most studied species in the included articles and the majority of articles were conducted in hospital settings. Nearly half of the included articles were studies specific to VREfm, showing the importance of vancomycin resistance in enterococci. WGS was important in studies to better understand VRE, especially in the assessment for vancomycin-variable enterococci (VVE). These are enterococcal isolates that are phenotypically sensitive to vancomycin but carry vancomycin-resistance genes. These isolates become phenotypically resistant to vancomycin when exposed to the antibiotic *in vivo* (25). Even though the resistance genes could be identified via PCR, the importance of WGS to further characterize VVE isolates was shown in the 2018 DANMAP report. The use of WGS and cgMLST allowed for the identification of new complexes and sequence types. DANMAP can now perform surveillance specific to these VVE strains. This should allow for earlier detection of VVE in patients and more appropriate treatment (25). WGS also allowed for a better understanding of *E. faecium* as it determined new sequence types, including *pstS*-null types, through cgMLST (37). The *pstS* gene locus is a housekeeping gene used for MLST but is missing in

some VREfm strains. The use of WGS and cgMLST allowed for more robust sequence typing and identification of these isolates (37). Surveillance of enterococci within a hospital using WGS also allowed for the identification of a VREfm outbreak. A combination of sequencing data and an epidemiological investigation allowed for the identification of transmission route and the implementation of measures to prevent further outbreaks (20).

A One-Health Approach

The ubiquitous nature of *Enterococcus* spp. naturally requires a One-Health approach to the surveillance of AMR in enterococci (38). This means a transdisciplinary approach across the human-animal-environment continuum in order to better understand the problem of AMR in enterococci (97). While enterococcal species other than *E. faecium* and *E. faecalis* were discussed in a few studies, in general, there were relatively few studies using a One-Health approach to compare animal, environmental (e.g., water and soil), and human samples (9, 11, 13, 39–41, 68–71). No studies sampled companion animals or equids, even though these animals live in close proximity to humans. This could be due to the complexity and cost of coordinating a study with so many sample sources or that human health studies are more easily funded. A collaborative transdisciplinary approach would bring a broad perspective to study design and allow for a better interpretation of the results. This would ease the complexity of designing and coordinating a One-Health surveillance study and create a more robust understanding of the issue of AMR in enterococci. Two studies included in this review did produce very informative results across multiple sample sources to address the One-Health continuum (38, 39). These studies pulled samples from livestock, retail meat, wastewater, and human bloodstream infections and showed limited sharing of genes between isolates from humans and animals (38, 39). Research using this One-Health approach will provide a means to assess the risk of AMR enterococci moving from food animals, through the food chain, into human populations as well as through the environment.

Importance of Surveillance of Enterococci

Many articles included in this study stressed the importance of the surveillance of *Enterococcus* spp., especially in hospital settings. This is because enterococci are optimized for adaptation and survival in their environment, whether the hospital environment or natural environment (42, 72). Both targeted surveillance of at-risk patients (e.g., immunocompromised) and passive surveillance of incoming hospital patients allowed for early recognition of outbreaks. Outbreaks could be controlled before becoming a significant problem and new hospital protocols surrounding cleaning and isolation could be developed (20, 72, 87). WGS allowed for more accurate sequence typing and identification of AMR genes (27, 73, 88).

Sequencing Platforms

From this review, Illumina sequencing platforms are currently the most popular for whole genome sequencing studies. They are historically reliable, with low error rates and have become accessible, abundant, and cost-effective (98). The combination of Illumina short-read with long-read sequencing (usually PacBio) was occasionally used to close a chromosome or a plasmid to accomplish genomic integrity as well as complete understanding of MGEs and their context. Unfortunately, the cost of running large numbers of isolates on a PacBio system is prohibitive (99). The use of long-read sequencing is likely to increase as inexpensive and portable bench-top platforms such as the Nanopore MinION become more reliable with lower error rates (100, 101). This will allow for the rapid identification of an isolate and its genetic composition, including MGEs (99). The majority of papers (86.1%) also provided archive accession numbers for their sequences, highlighting the importance of sharing raw genomic data with the scientific community and the requirements for publication in many journals.

Limitations

This scoping review held limitations similar to other review papers in the possible omission of relevant literature, such as gray literature or articles written in a language other than English. Findings from government surveillance programs may be published online or as peer-reviewed articles, but in order to maintain an efficient and reproducible search method, gray literature was only searched from the reference lists in the primary research articles included in the review. No separate gray literature search was performed, which may have resulted in the omission of relevant information. Two non-English articles were excluded in our search which otherwise might have been included.

In order to minimize the omission of articles, several databases were searched and inclusion criteria were intentionally left broad until the abstract screening steps. The authors did maintain a rigid definition of surveillance, which could have excluded epidemiological articles that did not fit the selection criteria. This largely eliminated studies on human clinical isolates of *Enterococcus* spp. as the isolates would have been selected for a study based on certain characteristics.

Another limitation of the study is that a critical appraisal of the included articles was not conducted. This was intentional as one of the objectives of the present study was to identify gaps in the literature, but it means that studies of lower and higher quality would carry equal weight. The findings of some studies may not share the same validity based on their study design, but this was not determined in this review.

CONCLUSION AND FUTURE DIRECTIONS

Whole genome sequencing has added value to the surveillance efforts of *Enterococcus* spp. by identifying new genes and strains, adding to the knowledge about its prevalence in various settings, and finding that WGS is a better means of detecting and differentiating *Enterococcus* spp. than other molecular methods. The ability of *Enterococcus* spp. to adapt and survive in its environment was frequently stated as a reason for the importance of using WGS for the surveillance of this bacterium. Future studies should focus on the state of *Enterococcus* spp. in companion animal veterinary medicine and determining the link between humans, animals, food products, and the

environment for a better One-Health approach to *Enterococcus* spp. surveillance.

AUTHOR CONTRIBUTIONS

LR, KS, SLC, and SCC developed the research question and scoping review protocol. LR and KS performed the literature search, article screening, and data extraction of included articles. LR drafted the complete manuscript. SCC and SLC were secondary reviewers of articles and primary reviewers of the manuscript. All authors assisted with editing and content review of the manuscript.

FUNDING

LR and KS stipend funding was through the University of Calgary, Faculty of Veterinary Medicine One Health Award, and the University of Calgary Master's Award. This research is part of the AMR – One Health Consortium, partially funded by the Major Innovation Fund program of the Ministry of Jobs, Economy and Innovation, Government of Alberta. The funders had no role in the study design, data interpretation, or writing of the manuscript.

ACKNOWLEDGMENTS

The authors would like to acknowledge Lorraine Toews for her advice on search databases and Dr. Sheryl Gow and Dr. Barb Wilhelm for their advice on review papers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.599285/full#supplementary-material

REFERENCES

1. Klein G. Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. *Int J Food Microbiol.* (2003) 88:123–31. doi: 10.1016/S0168-1605(03)00175-2
2. Strateva T, Atanasova D, Savov E, Petrova G, Mitov I. Incidence of virulence determinants in clinical *Enterococcus faecalis* and *Enterococcus faecium* isolates collected in Bulgaria. *Braz J Infect Dis.* (2016) 20:127–33. doi: 10.1016/j.bjid.2015.11.011
3. Arias CA, Murray BE. Emergence and management of drug-resistant enterococcal infections. *Expert Rev Anti Infect Ther.* (2008) 6:637–55. doi: 10.1586/14787210.6.5.637
4. Miller WR, Munita JM, Arias CA. Mechanisms of antibiotic resistance in enterococci. *Expert Rev Anti Infect Ther.* (2014) 12:1221–36. doi: 10.1586/14787210.2014.956092
5. Werner G, Coque TM, Hammerum AM, Hope R, Hryniewicz W, Johnson A, et al. Emergence and spread of vancomycin resistance among enterococci in Europe. *Eurosurveillance.* (2008) 13:19046–. doi: 10.2807/ese.13.47.19046-en
6. Flipse J, von Wintersdorf CJH, van Niekerk JM, Jamin C, van Tiel FH, Hasman H, et al. Appearance of vanD-positive *Enterococcus faecium* in a tertiary hospital in the Netherlands: prevalence of vanC and vanD in hospitalized patients. *Sci Rep.* (2019) 9:6949. doi: 10.1038/s41598-019-42824-4
7. Hammerum AM. Enterococci of animal origin and their significance for public health. *Clin Microbiol Infect.* (2012) 18:619–25. doi: 10.1111/j.1469-0691.2012.03829.x
8. Beukers AG, Zaheer R, Goji N, Amoako KK, Chaves AV, Ward MP, et al. Comparative genomics of *Enterococcus* spp. isolated from bovine feces. *BMC Microbiol.* (2017) 17:52. doi: 10.1186/s12866-017-0962-1
9. Sanderson H, Ortega-Polo R, Zaheer R, Goji N, Amoako KK, Brown RS, et al. Comparative genomics of multidrug-resistant *Enterococcus* spp. isolated from wastewater treatment plants. *BMC Microbiol.* (2020) 20:20. doi: 10.1186/s12866-019-1683-4
10. WHO. *Critically Important Antimicrobials for Human Medicine, 6th Revision*. Geneva: World Health Organization (2019).
11. Peng Z, Li M, Wang W, Liu H, Fanning S, Hu Y, et al. Genomic insights into the pathogenicity and environmental adaptability of *Enterococcus hirae* R17 isolated from pork offered for retail sale. *Microbiol Open.* (2017) 6:e514. doi: 10.1002/mbo3.514
12. Nilsson O. Vancomycin resistant enterococci in farm animals - occurrence and importance. *Infect Ecol Epidemiol.* (2012) 2:1. doi: 10.3402/iee.v2i0.16959
13. Manson AL, Van Tyne D, Straub TJ, Clock S, Crupain M, Rangan U, et al. Chicken meat-associated enterococci: influence of agricultural antibiotic use and connection to the clinic. *Appl Environ Microbiol.* (2019) 85:e01559–19. doi: 10.1128/AEM.01559-19
14. Rushton-Green R, Darnell RL, Taiaroa G, Carter GP, Cook GM, Morgan XC. Agricultural origins of a highly persistent lineage of vancomycin-resistant *Enterococcus faecalis* in New Zealand. *Appl Environ Microbiol.* (2019) 85:e00137-19. doi: 10.1128/AEM.00137-19
15. Donado-Godoy P, Castellanos R, Leon M, Arevalo A, Clavijo V, Bernal J, et al. The establishment of the colombian integrated program for antimicrobial resistance surveillance (COIPARS): a pilot project on poultry farms, slaughterhouses and retail market. *Zoonoses Public Health.* (2015) 62(Suppl. 1):58–69. doi: 10.1111/zph.12192
16. Mulani MS, Kamble EE, Kumkar SN, Tawre MS, Pardesi KR. Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: a review. *Front Microbiol.* (2019) 10:539. doi: 10.3389/fmicb.2019.00539
17. *About DANMAP*. (2018). Available online at: https://www.danmap.org/about-danmap (accessed February 17, 2020).
18. *About NARMS: US Centers for Disease Control.* (2019). Available online at: https://www.cdc.gov/narms/about/index.html (accessed February 17, 2020).
19. *About CIPARS Canada: Government of Canada*. (2007). Available online at: https://www.canada.ca/en/public-health/services/surveillance/canadian-integrated-program-antimicrobial-resistance-surveillance-cipars/background.html (accessed February 17, 2020).
20. Sundermann AJ, Babiker A, Marsh JW, Shutt KA, Mustapha MM, Pasculle AW, et al. Outbreak of vancomycin-resistant *Enterococcus faecium* in interventional radiology: detection through whole genome sequencing-based surveillance. *Clin Infect Dis.* (2019) doi: 10.1093/cid/ciz666
21. *SENTRY Antimicrobial Surveillance Program USA*. (2019). Available online at: https://www.jmilabs.com/sentry-surveillance-program/ (accessed February 17, 2020).
22. *European Antimicrobial Resistance Surveillance Network (EARS-Net): European Centre for Disease Prevention and Control.* (2021). Available online at: https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/ears-net (accessed April 15, 2020).
23. Sader HS, Rhomberg PR, Fuhrmeister AS, Mendes RE, Flamm RK, Jones RN. Antimicrobial resistance surveillance and new drug development. *Open Forum Infect Dis.* (2019) 6(Suppl. 1):S5–s13. doi: 10.1093/ofid/ofy345
24. Zhou X, Garcia-Cobos S, Ruijs G, Kampinga GA, Arends JP, Borst DM, et al. Epidemiology of extended-spectrum beta-lactamase-producing *E. coli* and vancomycin-resistant enterococci in the Northern Dutch-German cross-border region. *Front Microbiol.* (2017) 8:1914. doi: 10.3389/fmicb.2017.01914

25. DANMAP. *DANMAP 2018 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark*. Denmark: Statens Serum Institut (2018).
26. Deshpande LM, Castanheira M, Flamm RK, Mendes RE. Evolving oxazolidinone resistance mechanisms in a worldwide collection of enterococcal clinical isolates: results from the SENTRY Antimicrobial Surveillance Program. *J Antimicrob Chemother.* (2018) 73:2314–22. doi: 10.1093/jac/dky188
27. Pinholt M, Larner-Svensson H, Littauer P, Moser CE, Pedersen M, Lemming LE, et al. Multiple hospital outbreaks of vanA *Enterococcus faecium* in Denmark, 2012-13, investigated by WGS, MLST and PFGE. *J Antimicrob Chemother.* (2015) 70:2474–82. doi: 10.1093/jac/dkv142
28. Hammerum AM, Baig S, Kamel Y, Roer L, Pinholt M, Gumpert H, et al. Emergence of vanA *Enterococcus faecium* in Denmark, 2005-15. *J Antimicrob Chemother.* (2017) 72:2184–90. doi: 10.1093/jac/dkx138
29. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* (2018) 169:467–73. doi: 10.7326/M18-0850
30. Aromataris E, Munn Z. *Joanna Briggs Institute Reviewer's Manual*. Adelaide, SA: Joanna Briggs Institute (2017).
31. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. *Syst Rev.* (2016) 5:210. doi: 10.1186/s13643-016-0384-4
32. *JGI Sequences "Supergerm" Genome in One Day*. Berkeley, CA: Joint Genome Institute (2000).
33. Qin X, Galloway-Pena JR, Sillanpaa J, Roh JH, Nallapareddy SR, Chowdhury S, et al. Complete genome sequence of *Enterococcus faecium* strain TX16 and comparative genomic analysis of *Enterococcus faecium* genomes. *BMC Microbiol.* (2012) 12: 135. doi: 10.1186/1471-2180-12-135
34. Lam MMC, Seemann T, Bulach DM, Gladman SL, Chen H, Haring V, et al. Comparative analysis of the first complete *Enterococcus faecium* genome. *J Bacteriol.* (2012) 194:2334–41. doi: 10.1128/JB.00259-12
35. *Genome*. Bethesda, MD: National Library of Medicine (US); National Center for Biotechnology Information. (2004).
36. Coombs G, Daley D. *Australian Enterococcal Sepsis Outcome Program (AESOP) 2016: Final Report*. Australian Group on Antimicrobial Resistance (2016).
37. Lemonidis K, Salih TS, Dancer SJ, Hunter IS, Tucker NP. Emergence of an Australian-like pstS-null vancomycin resistant *Enterococcus faecium* clone in Scotland. *PLoS ONE.* (2019) 14:e0218185. doi: 10.1371/journal.pone.0218185
38. Zaheer R, Cook SR, Barbieri R, Goji N, Cameron A, Petkau A, et al. Surveillance of *Enterococcus* spp. reveals distinct species and antimicrobial resistance diversity across a One-Health continuum. *Sci Rep.* (2020) 10:3937. doi: 10.1038/s41598-020-69044-5
39. Gouliouris T, Raven KE, Ludden C, Blane B, Corander J, Horner CS, et al. Genomic surveillance of *Enterococcus faecium* reveals limited sharing of strains and resistance genes between livestock and humans in the United Kingdom. *Mbio.* (2018) 9:e01780-18. doi: 10.1128/mBio.01780-18
40. He Q, Hou Q, Wang Y, Li J, Li W, Kwok L-Y, et al. Comparative genomic analysis of *Enterococcus faecalis*: insights into their environmental adaptations. *BMC Genomics.* (2018) 19:527. doi: 10.1186/s12864-018-4887-3
41. Gouliouris T, Raven KE, Moradigaravand D, Ludden C, Coll F, Blane B, et al. Detection of vancomycin-resistant *Enterococcus faecium* hospital-adapted lineages in municipal wastewater treatment plants indicates widespread distribution and release into the environment. *Genome Res.* (2019) 29:626–34. doi: 10.1101/gr.232629.117
42. Janes VA, Notermans DW, Spijkerman IJB, Visser CE, Jakobs ME, van Houdt R, et al. Amplified fragment length polymorphism and whole genome sequencing: a comparison of methods in the investigation of a nosocomial outbreak with vancomycin resistant enterococci. *Antimicrob Resist Infect Control.* (2019) 8:153. doi: 10.1186/s13756-019-0604-5
43. Freitas AR, Elghaieb H, Leon-Sampedro R, Abbassi MS, Novais C, Coque TM, et al. Detection of optrA in the African continent (Tunisia) within a mosaic *Enterococcus faecalis* plasmid from urban wastewaters. *J Antimicrob Chemother.* (2017) 72:3245–51. doi: 10.1093/jac/dkx321
44. Holzknecht BJ, Pinholt M, Gumpert H, Hammerum AM, Westh H. Genetic characterisation confirms sporadic occurrence of vancomycin-resistant *Enterococcus faecalis* in Copenhagen, Denmark. *Int J Antimicrob Agents.* (2017) 50:501–2. doi: 10.1016/j.ijantimicag.2017.07.001
45. van Hal SJ, Espedido BA, Coombs GW, Howden BP, Korman TM, Nimmo GR, et al. Polyclonal emergence of vanA vancomycin-resistant *Enterococcus faecium* in Australia. *J Antimicrob Chemother.* (2017) 72:998–1001. doi: 10.1093/jac/dkw539
46. Lee AS, White E, Monahan LG, Jensen SO, Chan R, van Hal SJ. Defining the role of the environment in the emergence and persistence of vana vancomycin-resistant *Enterococcus* (VRE) in an intensive care unit: a molecular epidemiological study. *Infect Control Hosp Epidemiol.* (2018) 39:668–75. doi: 10.1017/ice.2018.29
47. Lee RS, da Silva AG, Baines SL, Strachan J, Ballard S, Carter GP, et al. The changing landscape of vancomycin-resistant *Enterococcus faecium* in Australia: a population-level genomic study. *J Antimicrob Chemother.* (2018) 73:3268–78. doi: 10.1093/jac/dky331
48. Marchi AP, Perdigao Neto LV, Martins RCR, Rizek CF, Camargo CH, Moreno LZ, et al. Vancomycin-resistant enterococci isolates colonizing and infecting haematology patients: clonality, and virulence and resistance profile. *J Hosp Infect.* (2018) 99:346–55. doi: 10.1016/j.jhin.2017.10.010
49. Santona A, Taviani E, Hoang HM, Fiamma M, Deligios M, Ngo TVQ, et al. Emergence of unusual vanA/vanB2 genotype in a highly mutated vanB2-vancomycin-resistant hospital-associated *E. faecium* background in Vietnam. *Int J Antimicrob Agents.* (2018) 52:586–92. doi: 10.1016/j.ijantimicag.2018.07.006
50. van Hal SJ, Beukers AG, Timms VJ, Ellem JA, Taylor P, Maley MW, et al. Relentless spread and adaptation of non-typeable vanA vancomycin-resistant *Enterococcus faecium*: a genome-wide investigation. *J Antimicrob Chemother.* (2018) 73:1487–91. doi: 10.1093/jac/dky074
51. Andersson P, Beckingham W, Gorrie CL, Kennedy K, Daveson K, Ballard SA, et al. Vancomycin-resistant *Enterococcus* (VRE) outbreak in a neonatal intensive care unit and special care nursery at a tertiary-care hospital in Australia—A retrospective case-control study. *Infect Control Hosp Epidemiol.* (2019) 40:551–8. doi: 10.1017/ice.2019.41
52. Douglas AP, Marshall C, Baines SL, Ritchie D, Szer J, Madigan V, et al. Utilizing genomic analyses to investigate the first outbreak of vanA vancomycin-resistant *Enterococcus* in Australia with emergence of daptomycin non-susceptibility. *J Med Microbiol.* (2019) 68:303–8. doi: 10.1099/jmm.0.000916
53. Farman M, Yasir M, Al-Hindi RR, Farraj SA, Jiman-Fatani AA, Alawi M, et al. Genomic analysis of multidrug-resistant clinical *Enterococcus faecalis* isolates for antimicrobial resistance genes and virulence factors from the western region of Saudi Arabia. *Antimicrob Resist Infect Control.* (2019) 8:55. doi: 10.1186/s13756-019-0508-4
54. O'Dea M, Sahibzada S, Jordan D, Laird T, Lee T, Hewson K, et al. Genomic, antimicrobial resistance, and public health insights into *Enterococcus* spp. from Australian chickens. *J Clin Microbiol.* (2019) 57 e00319-19. doi: 10.1128/JCM.00319-19
55. Sacramento AG, Fernandes MR, Sellera FP, Dolabella SS, Zanella RC, Cerdeira L, et al. VanA-type vancomycin-resistant *Enterococcus faecium* ST1336 isolated from mussels in an anthropogenically impacted ecosystem. *Mar Pollut Bull.* (2019) 142:533–6. doi: 10.1016/j.marpolbul.2019.04.014
56. Wardenburg KE, Potter RF, D'Souza AW, Hussain T, Wallace MA, Andleeb S, et al. Phenotypic and genotypic characterization of linezolid-resistant *Enterococcus faecium* from the USA and Pakistan. *J Antimicrob Chemother.* (2019) 74:3445–52. doi: 10.1093/jac/dkz367
57. Zhong Z, Kwok L-Y, Hou Q, Sun Y, Li W, Zhang H, et al. Comparative genomic analysis revealed great plasticity and environmental adaptation of the genomes of *Enterococcus faecium*. *BMC Genomics.* (2019) 20:602. doi: 10.1186/s12864-019-5975-8
58. Freitas AR, Tedim AP, Novais C, Lanza VF, Peixe L. Comparative genomics of global optrA-carrying *Enterococcus faecalis* uncovers a common chromosomal hotspot for optrA acquisition within a diversity of core and accessory genomes. *Microb Genom.* (2020) 6:e000350. doi: 10.1099/mgen.0.000350
59. Ekwanzala MD, Dewar JB, Kamika I, Momba MNB. Comparative genomics of vancomycin-resistant *Enterococcus* spp. revealed common resistome

determinants from hospital wastewater to aquatic environments. *Sci Total Environ.* (2020) 719:137275. doi: 10.1016/j.scitotenv.2020.137275
60. Ahmed MO, Elramalli AK, Baptiste KE, Daw MA, Zorgani A, Brouwer E, et al. Whole genome sequence analysis of the first vancomycin-resistant *Enterococcus faecium* isolates from a Libyan Hospital in Tripoli. *Microb Drug Resist.* (2020) 26:1390–8. doi: 10.1089/mdr.2019.0095
61. Almeida LM, Gaca A, Bispo PM, Lebreton F, Saavedra JT, Silva RA, et al. Coexistence of the oxazolidinone resistance-associated genes *cfr* and *optrA* in *Enterococcus faecalis* from a healthy piglet in Brazil. *Front Public Health.* (2020) 8:518. doi: 10.3389/fpubh.2020.00518
62. Chilambi GS, Nordstrom HR, Evans DR, Ferrolino JA, Hayden RT, Maron GM, et al. Evolution of vancomycin-resistant *Enterococcus faecium* during colonization and infection in immunocompromised pediatric patients. *Proc Natl Acad Sci. USA*. (2020) 117:11703–14. doi: 10.1073/pnas.1917130117
63. Freitas AR, Tedim AP, Duarte B, Elghaieb H, Abbassi MS, Hassen A, et al. Linezolid-resistant (Tn6246::fexB-poxtA) *Enterococcus faecium* strains colonizing humans and bovines on different continents: similarity without epidemiological link. *J Antimicrob Chemother.* (2020) 75:2416–23. doi: 10.1093/jac/dkaa227
64. Gouliouris T, Coll F, Ludden C, Blane B, Raven KE, Naydenova P, et al. Quantifying acquisition and transmission of *Enterococcus faecium* using genomic surveillance. *Nat Microbiol.* (2020) 6:103–111. doi: 10.1038/s41564-020-00806-7
65. Gu J, Xie X-J, Liu J-X, Shui J-R, Zhang H-Y, Feng G-Y, et al. Prevalence and transmission of antimicrobial-resistant Staphylococci and Enterococci from shared bicycles in Chengdu, China. *Sci Total Environ.* (2020) 738:139735. doi: 10.1016/j.scitotenv.2020.139735
66. Ruiz-Ripa L, Fessler AT, Hanke D, Eichhorn I, Azcona-Gutierrez JM, Perez-Moreno MO, et al. Mechanisms of linezolid resistance among enterococci of clinical origin in spain-detection of optrA- and cfr(D)-carrying *E. faecalis*. *Microorganisms.* (2020) 8:1155. doi: 10.3390/microorganisms8081155
67. Sabenca C, de Sousa T, Oliveira S, Viala D, Theron L, Chambon C, et al. Next-generation sequencing and MALDI mass spectrometry in the study of multiresistant processed meat vancomycin-resistant enterococci (VRE). *Biol. Basel.* (2020) 9:89. doi: 10.3390/biology9050089
68. Zhang S, Wang D, Wang Y, Hasman H, Aarestrup FM, Alwathnani HA, et al. Genome sequences of copper resistant and sensitive *Enterococcus faecalis* strains isolated from copper-fed pigs in Denmark. *Stand Genomic Sci.* (2015) 10:35. doi: 10.1186/s40793-015-0021-1
69. Cavaco LM, Bernal JF, Zankari E, Leon M, Hendriksen RS, Perez-Gutierrez E, et al. Detection of linezolid resistance due to the optrA gene in *Enterococcus faecalis* from poultry meat from the American continent (Colombia). *J Antimicrob Chemother.* (2016) 72:678–83. doi: 10.1093/jac/dkw490
70. Leinweber H, Alotaibi SMI, Overballe-Petersen S, Hansen F, Hasman H, Bortolaia V, et al. Vancomycin resistance in *Enterococcus faecium* isolated from Danish chicken meat is located on a pVEF4-like plasmid persisting in poultry for 18 years. *Int J Antimicrob Agents.* (2018) 52:283–6. doi: 10.1016/j.ijantimicag.2018.03.019
71. Tyson GH, Sabo JL, Hoffmann M, Hsu C-H, Mukherjee S, Hernandez J, et al. Novel linezolid resistance plasmids in *Enterococcus* from food animals in the USA. *J Antimicrob Chemother.* (2018) 73:3254–8. doi: 10.1093/jac/dky369
72. Leong KWC, Cooley LA, Anderson TL, Gautam SS, McEwan B, Wells A, et al. Emergence of vancomycin-resistant *Enterococcus faecium* at an Australian hospital: a whole genome sequencing analysis. *Sci Rep.* (2018) 8:6274. doi: 10.1038/s41598-018-24614-6
73. Landerslev KG, Jakobsen L, Olsen SS, Pedersen MB, Kristensen B, Lemming LE, et al. Polyclonal spread of vanA *Enterococcus faecium* in Central Denmark Region, 2009-2013, investigated using PFGE, MLST and WGS. *Int J Antimicrob Agents.* (2016) 48:767–8. doi: 10.1016/j.ijantimicag.2016.09.001
74. Pinholt M, Gumpert H, Bayliss S, Nielsen JB, Vorobieva V, Pedersen M, et al. Genomic analysis of 495 vancomycin-resistant *Enterococcus faecium* reveals broad dissemination of a vanA plasmid in more than 19 clones from Copenhagen, Denmark. *J Antimicrob Chemother.* (2017) 72:40–7. doi: 10.1093/jac/dkw360
75. Pinholt M, Bayliss SC, Gumpert H, Worning P, Jensen VVS, Pedersen M, et al. WGS of 1058 *Enterococcus faecium* from Copenhagen, Denmark, reveals rapid clonal expansion of vancomycin-resistant clone ST80 combined with widespread dissemination of a vanA-containing plasmid and acquisition of a heterogeneous accessory genome. *J Antimicrob Chemother.* (2019) 74:1776–85. doi: 10.1093/jac/dkz118
76. Boyd DA, Lalancette C, Levesque S, Golding GR. Characterization of a genomic island harbouring a new vanD allele from *Enterococcus faecium* N15-508 isolated in Canada. *J Antimicrob Chemother.* (2016) 71:2052–4. doi: 10.1093/jac/dkw063
77. Falgenhauer L, Fritzenwanker M, Imirzalioglu C, Steul K, Scherer M, Rhine-Main VSG, et al. Near-ubiquitous presence of a vancomycin-resistant *Enterococcus faecium* ST117/CT71/vanB -clone in the Rhine-Main metropolitan area of Germany. *Antimicrob Resist Infect Control.* (2019) 8:128. doi: 10.1186/s13756-019-0573-8
78. Bender JK, Fleige C, Lange D, Klare I, Werner G. Rapid emergence of highly variable and transferable oxazolidinone and phenicol resistance gene optrA in German *Enterococcus* spp. clinical isolates. *Int J Antimicrob Agents.* (2018) 52:819–27. doi: 10.1016/j.ijantimicag.2018.09.009
79. Wang J, Foxman B, Pirani A, Lapp Z, Mody L, Snitkin E. Application of combined genomic and transfer analyses to identify factors mediating regional spread of antibiotic resistant bacterial lineages. *Clin Infect Dis.* (2020) 71: e642–9. doi: 10.1093/cid/ciaa364
80. Lee T, Pang S, Stegger M, Sahibzada S, Abraham S, Daley D, et al. A three-year whole genome sequencing perspective of *Enterococcus faecium* sepsis in Australia. *PLoS ONE.* (2020) 15:e0228781. doi: 10.1371/journal.pone.0228781
81. Hashimoto Y, Kita I, Suzuki M, Hirakawa H, Ohtaki H, Tomita H. First report of the local spread of vancomycin-resistant enterococci ascribed to the interspecies transmission of a vanA gene cluster-carrying linear plasmid. *mSphere.* (2020) 5:e00102-20. doi: 10.1128/mSphere.00102-20
82. Leong KWC, Kalukottege R, Cooley LA, Anderson TL, Wells A, Langford E, et al. State-wide genomic and epidemiological analyses of vancomycin-resistant *Enterococcus faecium* in Tasmania's Public Hospitals. *Front Microbiol.* (2019) 10:2940. doi: 10.3389/fmicb.2019.02940
83. Eisenberger D, Tuschak C, Werner M, Bogdan C, Bollinger T, Hossain H, et al. Whole-genome analysis of vancomycin-resistant *Enterococcus faecium* causing nosocomial outbreaks suggests the occurrence of few endemic clonal lineages in Bavaria, Germany. *J Antimicrob Chemother.* (2020) 75:1398–404. doi: 10.1093/jac/dkaa041
84. Sharma P, Gupta SK, Barrett JB, Hiott LM, Woodley TA, Kariyawasam S, et al. Comparison of antimicrobial resistance and pan-genome of clinical and non-clinical *Enterococcus cecorum* from poultry using whole-genome sequencing. *Foods.* (2020) 9:686. doi: 10.3390/foods9060686
85. Xanthopoulou K, Peter S, Tobys D, Behnke M, Dinkelacker AG, Eisenbeis S, et al. Vancomycin-resistant *Enterococcus faecium* colonizing patients on hospital admission in Germany: prevalence and molecular epidemiology. *J Antimicrob Chemother.* (2020) 75:2743–51. doi: 10.1093/jac/dkaa271
86. Deshpande LM, Ashcraft DS, Kahn HP, Pankey G, Jones RN, Farrell DJ, et al. Detection of a new cfr-Like Gene, cfr(B), in *Enterococcus faecium* isolates recovered from human specimens in the United States as Part of the SENTRY Antimicrobial Surveillance Program. *Antimicrob Agents Chemother.* (2015) 59:6256–61. doi: 10.1128/AAC.01473-15
87. Kampmeier S, Knaack D, Kossow A, Willems S, Schliemann C, Berdel WE, et al. Weekly screening supports terminating nosocomial transmissions of vancomycin-resistant enterococci on an oncologic ward - a retrospective analysis. *Antimicrob Resist Infect Control.* (2017) 6:48. doi: 10.1186/s13756-017-0206-z
88. Lister DM, Kotsanas D, Ballard SA, Howden BP, Carse E, Tan K, et al. Outbreak of vanB vancomycin-resistant *Enterococcus faecium* colonization in a neonatal service. *Am J Infect Control.* (2015) 43:1061–5. doi: 10.1016/j.ajic.2015.05.047
89. Moradigaravand D, Gouliouris T, Blane B, Naydenova P, Ludden C, Crawley C, et al. Within-host evolution of *Enterococcus faecium* during longitudinal carriage and transition to bloodstream infection in immunocompromised patients. *Genome Medicine.* (2017) 9:119. doi: 10.1186/s13073-017-0507-0
90. Rios R, Reyes J, Carvajal LP, Rincon S, Panesso D, Echeverri AM, et al. Genomic epidemiology of vancomycin-resistant *Enterococcus faecium* (VREfm) in latin america: revisiting the global VRE population structure. *Sci Rep.* (2020) 10:5636. doi: 10.1038/s41598-020-62371-7
91. Kim HM, Chung DR, Cho SY, Huh K, Kang CI, Peck KR. Emergence of vancomycin-resistant *Enterococcus faecium* ST1421 lacking the pstS

gene in Korea. *Eur J Clin Microbiol Infect Dis.* (2020) 39:1349–56. doi: 10.1007/s10096-020-03853-4
92. Eichel V, Klein S, Bootsveld C, Frank U, Heeg K, Boutin S, et al. Challenges in interpretation of WGS and epidemiological data to investigate nosocomial transmission of vancomycin-resistant *Enterococcus faecium* in an endemic region: incorporation of patient movement network and admission screening. *J. Antimicrob Chemother.* (2020) 75:1716–21. doi: 10.1093/jac/dkaa122
93. Heininger A, Zimmermann S, Bootsveld C, Boutin S, Nurjadi D. Low prevalence of combined linezolid- and vancomycin-resistant *Enterococcus faecium* from hospital admission screening in an endemic region in Germany. *J Glob Antimicrob Resist.* (2020) 22:646–50. doi: 10.1016/j.jgar.2020.05.003
94. Kampmeier S, Toennies H, Correa-Martinez CL, Mellmann A, Schwierzeck V. A nosocomial cluster of vancomycin resistant enterococci among COVID-19 patients in an intensive care unit. *Antimicrob Resist Infect Control.* (2020) 9:154. doi: 10.1186/s13756-020-00820-8
95. Neumann B, Bender JK, Maier BF, Wittig A, Fuchs S, Brockmann D, et al. Comprehensive integrated NGS-based surveillance and contact-network modeling unravels transmission dynamics of vancomycin-resistant enterococci in a high-risk population within a tertiary care hospital. *PLoS ONE.* (2020) 15:e0235160. doi: 10.1371/journal.pone.0235160
96. Weber A, Maechler F, Schwab F, Gastmeier P, Kola A. Increase of vancomycin-resistantEnterococcus faeciumstrain type ST117 CT71 at Charite-Universitatsmedizin Berlin, 2008 to 2018. *Antimicrob Resist Infect Control.* (2020) 9:109. doi: 10.1186/s13756-020-00754-1
97. Hall DC, Cork SC. Chapter 1: introduction to one health concepts. In: Cork S, Hall D, Liljebjelke K, editors. *One Health Case Studies, Addressing Complex Problems in a Changing World*. Sheffield: 5M Publishing, Ltd. (2016).
98. Quainoo SC, Jordy PM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, et al. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clin Microbiol Rev.* (2017) 30:1015–63. doi: 10.1128/CMR.00016-17
99. George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, et al. Resolving plasmid structures in Enterobacteriaceae using the MinION nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly approaches. *Microb Genom.* (2017) 3:e000118. doi: 10.1099/mgen.0.000118
100. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Sci Rep.* (2018) 8:10931. doi: 10.1038/s41598-018-29334-5
101. Lui C. A long road/read to rapid high-resolution HLA typing: the nanopore perspective. *Hum Immunol.* (2020). doi: 10.1016/j.humimm.2020.04.009. [Epub ahead of print].

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.

frontiers in Medicine

CASE REPORT
published: 06 July 2021
doi: 10.3389/fmed.2021.684040

Case Report: About a Case of Hyperammonemia Syndrome Following Lung Transplantation: Could Metagenomic Next-Generation Sequencing Improve the Clinical Management?

Charlotte Michel[1], Michela Raimo[2], Vladimir Lazarevic[3], Nadia Gaïa[3], Nina Leduc[4], Christiane Knoop[4], Marie Hallin[1], Olivier Vandenberg[1], Jacques Schrenzel[3], David Grimaldi[5] and Maya Hites[2]*

[1] Department of Microbiology, Laboratoire Hospitalier Universitaire de Bruxelles, Université Libre de Bruxelles, Brussels, Belgium, [2] Clinic of Infectious Diseases, Cliniques Universitaires de Bruxelles, Erasme Hospital, Brussels, Belgium, [3] Genomic Research Laboratory, Geneva University Hospitals, Geneva, Switzerland, [4] Department of Pneumology, Cliniques Universitaires de Bruxelles, Erasme Hospital, Brussels, Belgium, [5] Intensive Care Unit, Cliniques Universitaires de Bruxelles, Erasme Hospital, Brussels, Belgium

OPEN ACCESS

Edited by:
Martin Maiers,
National Marrow Donor Program,
United States

Reviewed by:
Joyce Wang,
University of Michigan, United States
Tracy H. Erkkila,
Los Alamos National Laboratory (DOE), United States

***Correspondence:**
Charlotte Michel
charlotte.michel@lhub-ulb.be

Specialty section:
This article was submitted to Infectious Diseases–Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: 22 March 2021
Accepted: 19 May 2021
Published: 06 July 2021

Citation:
Michel C, Raimo M, Lazarevic V, Gaïa N, Leduc N, Knoop C, Hallin M, Vandenberg O, Schrenzel J, Grimaldi D and Hites M (2021) Case Report: About a Case of Hyperammonemia Syndrome Following Lung Transplantation: Could Metagenomic Next-Generation Sequencing Improve the Clinical Management? Front. Med. 8:684040. doi: 10.3389/fmed.2021.684040

Background: *Mycoplasma hominis* and *Ureaplasma* spp. are responsible for opportunistic infections in transplant patients, sometimes causing a life-threatening hyperammonemia syndrome. Both pathogens are not identified with standard microbiology techniques, resulting in missed or delayed diagnosis. We present a clinical case that illustrates the added value that next-generation sequencing (NGS) may offer in the diagnosis of respiratory infections in immune-compromised patients.

Results: A 55 years-old man with idiopathic pulmonary fibrosis underwent double lung transplantation. He received antibiotic prophylaxis with piperacillin-tazobactam and azythromycin. At day 4 post-transplantation (PTx), the patient presented an acute respiratory distress. A broncho-alveolar lavage (BAL) was performed. At day 5 PTx, the patient presented a status epilepticus due to diffuse cerebral oedema. Serum ammonia concentration was 661 μg/dL. BAL bacterial culture was negative. Because of the clinical presentation, special cultures were performed and identified 100.000 CFU/mL of *M. hominis* and *Ureaplasma* spp. and specific PCRs were positive for *M. hominis* and *Ureaplasma parvum*. Antibiotic therapy was shifted to therapeutic dose of azithromycin and doxycycline; within 48 h ammonia serum concentrations returned to normal but the coma persisted several weeks, followed by a persistent frontal lobe syndrome. A follow-up BAL was performed on day 11 Ptx. The *Mycoplasma/Ureaplasma* culture was negative, yet the specific PCRs remained positive. Bacterial culture found 100 CFU/mL of *Staphylococcus aureus* and viral culture was positive for Herpes Simplex Virus-1. These results were confirmed by metagenomic next-generation sequencing (mNGS). In the bacterial fraction, the majority of reads belonged to *Corynebacterium propinquum* (34.7%), *S. aureus* (24.1%) and *Staphylococcus epidermidis* (17.1%). Reads assigned

to *M. hominis*, *Ureaplasma urealyticum* and *parvum* represented 0.71, 0.13, and 0.04% of the bacterial fraction and corresponded to 6.9×10^3, 9.7×10^2, and 3.7×10^2 genome equivalents per mL of BAL fluid, respectively. These results are in favor of a cure of the atypical infection.

Conclusions: mNGS offered added diagnostic and quantitative values compared to PCR tests, which can remain positive after resolved infections. The initiation of appropriate antibiotic therapy would have occurred earlier on, possibly resulting in a better clinical outcome if mNGS had been performed in a routine fashion.

Keywords: metagenomics, mollicutes, hyperammonemia, lung trasnplantation, next-generation sequencing

BACKGROUND

We report here a clinical case that illustrates the limitations of both culture-dependent and molecular routine microbiological methods and the potential added value of metagenomic next-generation sequencing (mNGS) in the diagnosis of respiratory infections in immunocompromised patients.

CASE REPORT

A 55-year-old man with idiopathic pulmonary fibrosis, and treated since 2016 with pirfenidone, underwent a double lung transplantation in 2019. The patient had no significant medical history other than dyslipidemia treated with statins. During surgery, the patient presented an episode of ventricular fibrillation with cardiac arrest due to the passage of air bubbles when the right pulmonary artery was unclamped. He was quickly resuscitated after cardioversion.

The patient received a piperacillin-tazobactam and azithromycine antibiotic prophylaxis and a classic immunosuppressive regimen (Thymoglobulin, mycophenolic acid, and methylprednisolone). On day 4 post-transplantation, he presented an acute respiratory distress that required reintubation. As the perioperative swab culture from the donor lungs revealed *Escherichia coli* of intermediate susceptibility to piperacillin-tazobactam, an *E. coli* pneumonia was suspected, and piperacillin-tazobactam was shifted to cefuroxime. A broncho-alveolar lavage (BAL) was performed before antibiotic shift and sent for bacterial cultures, galactomannan measurement, and detection of 34 pulmonary pathogens by multiplex polymerase chain reaction (PCR) (TaqMan Array Card, Thermo Fisher Scientific, Waltham, MA, USA), which only revealed the presence of Herpes simplex virus type 1 (HSV-1) with a coefficient threshold at 34, suggesting no acute infection. At day 5 post-transplantation, fasciculations of the facial muscles appeared. The patient's parameters did not reveal malignant hypertension. A continuous electroencephalogram (EEG) revealed a status epilepticus, and the patient was treated by levitiracetam and midazolam. Computed tomography revealed the presence of a diffuse brain edema, therefore excluding the possibility of performing a spinal tap. Blood tests revealed severe hyperammonemia at 661 μg/dL; an aggressive treatment was initiated with lactulose, L-carnitine, sodium benzoate, sodium-phenylbutyrate, and continuous veno-venous hemodiafiltration (CVVH), as well as restriction of protein intake. A treatment with doxycycline was also initiated, and the dosage regimen of azithromycin was increased from 250 mg, three times weekly to 500 mg daily, the following day.

The "standard" bacterial culture of the BAL fluid revealed negative results. Since clinical features suggested an infection by atypical pathogens, a specific culture for *Mycoplasma*/*Ureaplasma* spp. was performed, identifying 100,000 CFU/mL of *Mycoplasma hominis* and *Ureaplasma* spp. after 24 h of incubation. Because of the small size of colonies, it was impossible to obtain pure cultures, and their identification relied on microscopic examination. Antibiotic susceptibility testing of a mixed population was challenging, and an accurate minimal inhibitory concentration (MIC) could not be determined with the Mycoplasma IST2 susceptibility gallery (bioMérieux, Marcy l'Etoile, France). The mixed culture was interpreted as resistant to erythromycin, clindamycin, and azithromycin but susceptible to tetracycline, doxycycline, ofloxacin, and ciprofloxacin. Ammonia serum concentrations returned to normal values (<102 μg/dL) within 48 h of initiating doxycycline treatment. The coma persisted several weeks and was followed by a severe, irreversible frontal lobe syndrome and massive cognitive impairment. The patient is at present (24 months after transplantation) bedridden.

To confirm microbiological cure, a follow-up BAL was performed on day 12 post-transplantation, after 7 days of doxycycline and 5 days of azithromycin. Bacterial culture yielded 100 CFU/mL of *Staphylococcus aureus*, and both multiplex PCR for respiratory pathogens (Taqman Array Card) and viral culture were positive for HSV-1, all of which were not considered clinically significant. The specific culture for *Mycoplasma*/*Ureaplasma* spp. revealed negative results. The antibiotic therapy was stopped after a total of 10 days of doxycycline and 8 days of azithromycin at therapeutic doses.

A retrospective analysis was performed by specific qualitative PCR for species identification and revealed *M. hominis* and *Ureaplasma parvum* in both BAL samples. mNGS was performed for scientific documentation on the second BAL sample (unfortunately, no material was left from the first BAL sample). Of 2,275,532 quality filtered sequencing reads, 2,264,070 were human, 4,222 mapped to HSV-1, and 4,773 were assigned by

TABLE 1 | Microbial species detected by mNGS.

Species identified by CLARK	Read (pair) count	Genome equivalents per mL of BAL fluid	Genome coverage (%)
Human alphaherpesvirus 1 (HSV-1)	4,222	3.9E+06	91.46
C. propinquum	1,658	9.3E+04	11.9
S. aureus	1,151	5.7E+04	9.51
S. epidermidis	816	4.6E+04	7.21
Corallococcus aberystwythensis	180	2.5E+03	0.04
Streptomyces turgidiscabies	141	1.9E+03	0.01
Corynebacterium pseudodiphtheriticum	87	5.4E+03	0.89
Bifidobacterium tibiigranuli	77	4.0E+03	0.03
Staphylococcus schweitzeri	68	3.4E+03	0.79
Staphylococcus argenteus	50	2.6E+03	0.51
Mogibacterium diversum	45	3.5E+03	0.54
Enterococcus faecalis	42	2.1E+03	0.33
[Candida] glabrata	40	4.6E+02	0.07
Alcanivorax hongdengensis	35	1.3E+03	0.02
M. hominis	34	6.9E+03	1.08
Aequorivita lutea	27	1.1E+03	0.05
Sphingomonas paucimobilis	18	6.3E+02	0.09
Parvimonas micra	13	1.1E+03	0.17
Porphyromonas endodontalis	12	8.0E+02	0.12
Staphylococcus haemolyticus	11	6.1E+02	0.1
Lancefieldella parvula	10	9.1E+02	0.13
Peptostreptococcus stomatis	8	5.7E+02	0.07
Bifidobacterium dentium	8	4.3E+02	0.05
Staphylococcus warneri	7	4.1E+02	0.07
Cutibacterium acnes	7	3.9E+02	0.06
U. urealyticum	6	9.7E+02	0.15
Gemella morbillorum	5	4.0E+02	0.06
Mogibacterium timidum	5	3.8E+02	0.06
Tropheryma whipplei	3	4.6E+02	0.07
U. parvum	2	3.7E+02	0.05

Only microorganisms with >3E+02 genome equivalents per mL of BAL fluid are reported.

CLARK to bacteria. In the bacterial fraction, the majority of reads belonged to *Corynebacterium propinquum* (34.7%), *S. aureus* (24.1%), and *Staphylococcus epidermidis* (17.1%). Reads assigned to *M. hominis*, *Ureaplasma urealyticum*, and *U. parvum* represented 0.71, 0.13, and 0.04% of the bacterial fraction and corresponded to 6.9×10^3, 9.7×10^2, and 3.7×10^2 genome equivalents per mL of BAL fluid, respectively (**Table 1**). BLASTn analysis confirmed CLARK assignments, with the exception of one read pair that was classified by CLARK as *U. parvum* but had best BLASTn scores to *M. hominis*. These results are in favor of a cure of the atypical infection. The chronology of the case is represented in **Figure 1**.

DISCUSSION

Mycoplasma and *Ureaplasma* (class Mollicutes) have no cell wall and therefore are undetectable under light microscopy after Gram staining and are intrinsically resistant to antibiotics targeting peptidoglycans. They also cannot be grown on standard bacterial culture media. *M. hominis* and *Ureaplasma* spp. are urogenital commensals, which are long known to be responsible for infections of the urogenital tract. They also have been found in a wide panel of opportunistic infections, such as periaortic abscess, pneumonia, and pericarditis (1–3). Since 2015, *M. hominis*, *U. urealyticum,* and *U. parvum*, alone or in co-infection, have been increasingly documented as causes of hyperammonemia syndrome (HS) within the first 30 days after lung or hematopoietic cell transplantation (3–7). HS occurs rarely (1–4% of transplant patients) but has major impact in terms of neurological complications (3, 4) and mortality. The mortality rate in patients with increased ammonia levels is 67% compared to 17% in patients with normal ammonia levels (4).

Ureaplasma spp. have the ability to generate energy by cytosolic urease-catalyzed hydrolysis of urea to ammonia. This has no clinical consequences in urogenital infections, but it can be severe in disseminated infections: after the serum ammonia level increases, the liver converts it back to urea, which in turn provides more energy for the microorganism, resulting in sustained hyperammonemia (3, 8). To investigate the source of opportunistic infections caused by *Ureaplasma* in lung-transplant patients, Fernandez et al. screened all lung donors. The study found a significant correlation between *Ureaplasma*-positive lung donors and hyperammonemia ($p < 0.001$). Additionally, all recipients of *Ureaplasma*-positive lungs presented a systemic inflammatory response, with lung infiltrates and required vasopressors to maintain a normal blood pressure after transplantation (7).

Mycoplasma/Ureaplasma spp. are naturally susceptible to macrolides, tetracyclines, or fluoroquinolones (9), but these antibiotics are not part of the classic antibiotic prophylaxis regimen used for lung transplantation (10). Acquired resistance in these bacteria is, however, not unusual, even though resistance rates vary significantly between studies. A French study showed resistance rates to tetracyclines of 7.5% for *Ureaplasma* spp. and 14.8% for *M. hominis* isolates, while the levofloxacin and moxifloxacin resistance rates were, respectively, 1.2 and 0.1%, for *Ureaplasma* spp., and 2.7 and 1.6% for *M. hominis* (11). An Italian study, on the other hand, showed significant levels of quinolone resistance in *Ureaplasma* spp., particularly for ciprofloxacin (77% of cases). Furthermore, *M. hominis* strains were non-susceptible to azithromycin and roxithromycin in about 90% of cases (12).

The increasing number of immuno-suppressed patients is one of the reasons for the emergence of infections caused by new opportunistic, fastidious or yet-to-be-cultured pathogens. The development of culture-independent diagnostic microbiology tools is essential to improve the healthcare of these patients. Our patient, for instance, suffered a delayed microbiological diagnosis and late documentation of the susceptibility profile, resulting in irreversible neurological complications, despite appropriate empirical antibiotic therapy initiated at the presentation of symptoms.

The specific qualitative PCRs for *M. hominis* and *U. parvum* were positive on the first and the second BAL. The first one allowed an accurate diagnosis to the species level but the second

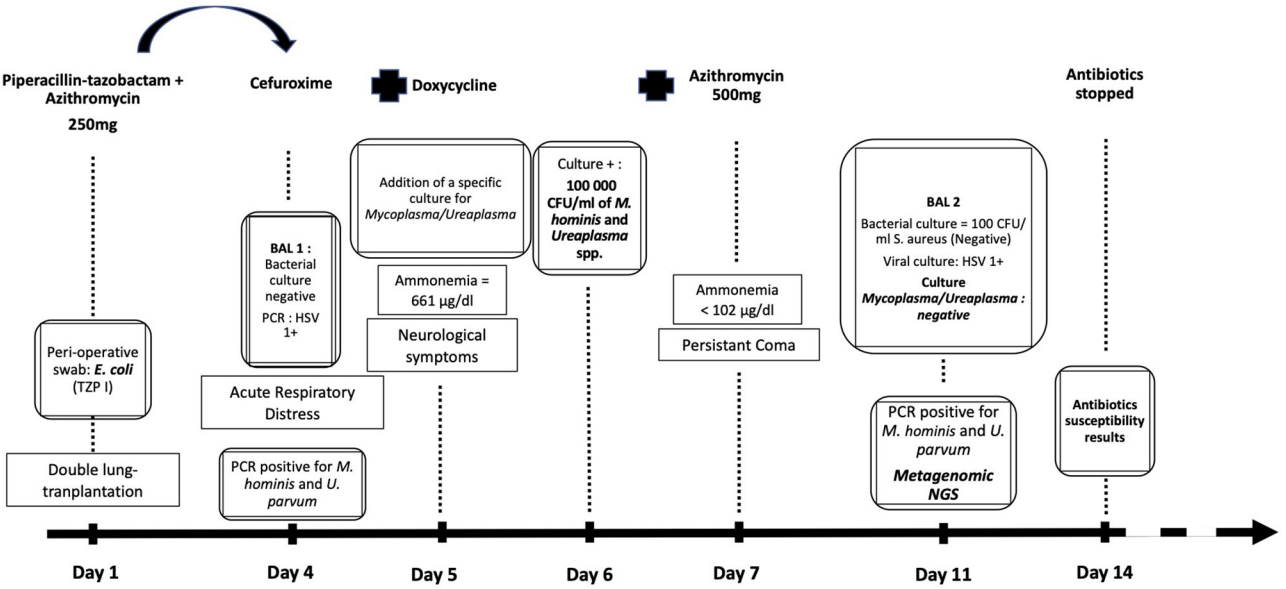


FIGURE 1 | Chronology of diagnostic and treatment of the patient according to the laboratory results. Apparition of the symptoms through time, Day 1 corresponding to the lung-transplantation. Antibiotic therapy shifts. Ammonemia levels. Results of microbiological cultures and biomolecular tests.

remained positive even after the microbiological cure of the patient. The positivity of a PCR can last after the patient is "cured" due to the presence of the pathogen's cell-free DNA, dead cells, and viable but not culturable cells. The procedure we used to extract DNA for mNGS included specific pretreatments aimed at removing free DNA and DNA from human and dead or damaged bacterial cells. Thus, the detected bacteria are expected to mostly correspond to viable populations, which, however, may contain non-culturable cells, as reported for *M. hominis* (13).

In our case, the results of mNGS performed on the follow-up BAL were in accordance with those found by bacterial and viral identifications by culture, PCR, and qPCR. mNGS also identified additional species, including *U. urealyticum*, and notably, *C. propinquum,* and *S. epidermidis*, both of which were considered as members of normal human microbiota in the culture report. However, *C. propinquum* is also an emerging cause of respiratory infections (14), and some strains are capable of producing urease (15).

Since the output of the iSeq 100 is lower than that of other Illumina (San Diego, CA, USA) instruments (MiniSeq, MiSeq or NextSeq), successful removal of host DNA is important for the detection of microbial DNA. However, the proportion of human reads in our iSeq NGS dataset remained high, in contrast to a previous report (16) in which the same method for microbial enrichment procedure had been used.

Most currently used culture-independent methods require a relatively precise suspicion of the pathogen causing the infection, while mNGS can "blindly" detect all microorganisms present within a clinical sample and assess their load (17). Provided that a sufficient number of bacterial reads are generated by mNGS (a situation which is more likely to occur before the initiation of the antibiotic treatment), the resistance profile (resistome) may be assessed (16–18) in a much quicker fashion than using culture-based methods (11, 12, 19). In some cases, mNGS may also be used for bacterial typing, bypassing the need for multiple PCR or qPCR tests (20). As a hypothesis-free approach, it is also well-suited for detecting new or unexpected pathogens (21). mNGS may provide an advantage over classical methods because it has the potential to predict infection and to help differentiate colonization from infection (22). The mNGS analysis of sequential BAL fluids of the same patient could allow follow-up of the infectious process and provide proof of a microbiological cure.

CONCLUSIONS

Life-threatening complications of *Mycoplasma*/*Ureaplasma* infections following lung or stem-cells transplantations have been reported, but the underlying mechanisms are not fully understood. Detection of *Mycoplasma*/*Ureaplasma* infections require specific molecular tests and growth media that many clinical microbiology laboratories do not perform routinely. It is likely that many infections caused by *Mycoplasma*/*Ureaplasma* remain undocumented.

mNGS as a diagnostic tool in clinical microbiology can help identify and quantify not only pathogens but also commensals, some of which have a pathogenic potential (pathobionts) in immunocompromised patients.

MATERIALS/METHODS

Standard BAL microbiological procedures: 10 µL of the BAL fluid was inoculated on a Columbia Agar with 5% sheep blood (BD,

Franklin Lakes, NJ, USA) and a HAEM2 Agar (bioMérieux) for 48 h under a 5% CO_2 atmosphere at 35 ± 2°C.

M. hominis/Ureaplasma spp. cultures and antibiogram: the BAL fluid of the patient was inoculated in the special liquid medium R2 for *Mycoplasma*/*Ureaplasma* and then plated on A7 agar plates (bioMérieux) for culture which were positive after 24 h. The diagnosis was made by optical microscopic observation. The antibiogram in liquid media was carried out by *Mycoplasma* IST2 gallery (bioMérieux), and the results were read at 48 h after the positive control turned positive.

Standard viral culture was performed on cells MRC5, LLC and VERO and incubated 2–3 weeks at 35 ± 2°C.

Multiplex Respiratory Pathogens Taqman Array Card (Thermo Fisher Scientific): customized multiplex PCR cards designed to detect the following pathogens (23): Adenovirus, Bocavirus, CMV, Coronavirus 229E/NL63/HKU1/OC43, HSV1/2, HHV6, VZV, RSV, Enterovirus, Influenza A, Influenza B, hMPV, Parainfluenza virus 1/2/3/4, Rhinovirus, *Bordetella pertussis, Chlamydia psittaci, Chlamydophila pneumoniae, Coxiella burnetii, Legionella pneumophila, Mycoplasma pneumoniae, Pneumocystis jirovecii, Aspergillus fumigatus, Aspergillus flavus, Aspergillus niger*, and *Aspergillus terreus.*

The specific PCRs for *M. hominis, U. urealyticum,* and *U. parvum* were performed by the National Reference Center for *Chlamydia trachomatis* and urogenital *Mycoplasma* species of the University Hospital of Bordeaux, France.

mNGS pipeline: The BAL fluid (400 μL, conserved at −80°C) was mixed with an equal volume of 10 mg/mL Liberase (TL Research grade, Sigma-Aldrich, St. Louis, MO, USA) in Dulbecco's phosphate-buffered saline with $MgCl_2$ and $CaCl_2$ (D8662, Sigma-Aldrich) and incubated with agitation (1,000 rpm) at 37°C for 1.5 h. The suspension was passed through a 100-μm nylon strainer (Corning), which was washed with 200 μL of the SU buffer from Ultra-Deep Microbiome Prep kit (Molzym, Bremen, Germany). For microbial DNA enrichment, we used the Ultra-Deep Microbiome Prep kit as per manufacturer's recommendations for liquid samples. Negative extraction was performed using 400 μL of the SU buffer (Molzym) instead of the clinical sample. Metagenomic libraries were constructed from extracted DNA using Nextera DNA Flex Library Prep kit (Illumina) and sequenced (2 × 151) on an Illumina iSeq 100 System. Our bioinformatics pipeline included (1) quality filtering (24); (2) elimination of replicate reads; (3) removal of reads matching human genome sequence; and (4) classification of reads using CLARK (25) against the Latest RefSeq/NCBI reference/representative (26) bacterial, archaeal, and fungal genomes, as well as genomes of DNA viruses infecting humans and DNA bacteriophages (selected from https://viralzone.expasy.org). To calculate genome coverage, metagenomic reads classified by CLARK were mapped to corresponding reference genomic sequence using USEARCH 11.0.667 (27) (-ublast -id 0.8 query_cov 0.5 -top_hit_only -strand both -evalue 0.00001). The reads assigned by CLARK to *Mycoplasma* and *Ureaplasma* were queried against the entire NCBI database using the BLASTn (28) tool on the NCBI website (29).

After filtering out the sequencing read pairs matching the human genome, sequencing data were submitted to European Nucleotide Archive (ENA) under study number PRJEB44898.

qPCR/mNGS based quantification of bacteria: The abundance of a given bacterial species was computed from relevant bacterial to human read counts ratio combined with a TaqMan-based quantification of human DNA (30). Obtained values were corrected for genome sizes of reference bacterial genomes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repository at: https://www.ebi.ac.uk/ena, under study number PRJEB44898.

ETHICS STATEMENT

The Ethics Committee waived the requirement for informed consent for this case report. Written informed consent was not obtained from the individual for the publication of any potentially identifiable data included in this article.

AUTHOR CONTRIBUTIONS

CM: microbiological diagnosis and writing. MR: clinical informations and writing. VL and JS: mNGS, correction of the manuscript, and writing. NG: bioinformatics. NL and CK: clinical care of the patient and follow-up. MHa and OV: microbiological diagnosis and correction of the manuscript. DG: clinical care of the patient. MHi: clinical care of the patient and correction of the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Sampath R, Patel R, Cunningham SA, Arif S, Daly RC, Badley AD, et al. Cardiothoracic transplant recipient mycoplasma hominis: an uncommon infection with probable donor transmission. *EBioMed.* (2017) 19:84–90. doi: 10.1016/j.ebiom.2017.04.026
2. Spiller OB. Emerging pathogenic respiratory mycoplasma hominis infections in lung transplant patients: time to reassesses it's role as a pathogen? *EBioMed.* (2017) 19:8–9. doi: 10.1016/j.ebiom.2017.05.002
3. Bharat A, Cunningham SA, Scott Budinger GR, Kreisel D, DeWet CJ, Gelman AE, et al. Disseminated ureaplasma infection as a cause of fatal hyperammonemia in humans. *Sci Transl Med.* (2015) 7:284re3. doi: 10.1126/scitranslmed.aaa8419
4. Nowbakht C, Edwards AR, Rodriguez-Buritica DF, Luce AM, Doshi PB, De Golovine A, et al. Two cases of fatal hyperammonemia syndrome due to mycoplasma hominis and ureaplasma urealyticum in immunocompromised patients outside lung transplant recipients. *Open Forum Infect Dis.* (2019) 6:ofz033. doi: 10.1093/ofid/ofz033
5. Smith M, Crews JD, Cheek N, Srivastava R, Appachi E. Hyperammonemic encephalopathy due to ureaplasma parvum infection in an immunocompromised child. *Pediatrics.* (2019) 144:e20190601. doi: 10.1542/peds.2019-0601

6. Matson KM, Sonetti DA. Successful treatment of ureaplasma-induced hyperammonemia syndrome post-lung transplant. *Transpl Infect Dis.* (2019) 21:e13022. doi: 10.1111/tid.13022
7. Fernandez R, Ratliff A, Crabb D, Waites KB, Bharat A. Ureaplasma transmitted from donor lungs is pathogenic after lung transplantation. *Ann Thorac Surg.* (2017) 103:670–1. doi: 10.1016/j.athoracsur.2016.09.026
8. Smith DG, Russell WC, Ingledew WJ, Thirkell D. Hydrolysis of urea by ureaplasma urealyticum generates a transmembrane potential with resultant aTP synthesis. *J Bacteriol.* (1993) 175:3253–8. doi: 10.1128/jb.175.11.3253-3258.1993
9. Lanao AE, Chakraborty RK, Pearson-Shaver AL. Mycoplasma infections. In: *StatPearls* [Internet]. Treasure Island, FL: StatPearls Publishing (2021). Available online at: https://www.ncbi.nlm.nih.gov/books/NBK536927/
10. Nosotti M, Tarsia P, Morlacchi LC. Infections after lung transplantation. *J Thorac Dis.* (2018) 10:3849–68. doi: 10.21037/jtd.2018.05.204
11. Meygret A, Le Roy C, Renaudin H, Bébéar C, Pereyre S. Tetracycline and fluoroquinolone resistance in clinical *Ureaplasma* spp. and *Mycoplasma hominis* isolates in France between 2010 and 2015. *J Antimicrob Chemother.* (2018) 73:2696–703. doi: 10.1093/jac/dky238
12. Foschi C, Salvo M, Galli S, Moroni A, Cevenini R, Marangoni A. Prevalence and antimicrobial resistance of genital mollicutes in italy over a two-year period. *New Microbiol.* (2018) 41:153–8.
13. Trushin MV, Chernov VM, Gorshkov OV, Baranova NB, Chernova OA. Atomic force microscopy analysis of DNA extracted from the vegetative cells and the viable, but nonculturable, cells of two mycoplasmas (*Acholeplasma laidlawii* PG8 and *Mycoplasma hominis* PG37). *Sci World J.* (2010) 10:894–900. doi: 10.1100/tsw.2010.78
14. Díez-Aguilar M, Ruiz-Garbajosa P, Fernández-Olmos A, Guisado P, Del Campo R, Quereda C, et al. Non-diphtheriae corynebacterium species: an emerging respiratory pathogen. *Eur J Clin Microbiol Infect Dis.* (2013) 32:769–72. doi: 10.1007/s10096-012-1805-5
15. Bernard K, Pacheco AL, Cunningham I, Gill N, Burdz T, Wiebe D. Emendation of the description of the species corynebacterium propinquum to include strains which produce urease. *Int J Syst Evol Microbiol.* (2013) 63:2146–54. doi: 10.1099/ijs.0.046979-0
16. Leo S, Gaïa N, Ruppé E, Emonet S, Girard M, Lazarevic V, et al. Detection of bacterial pathogens from broncho-Alveolar lavage by next-Generation sequencing. *Int J Mol Sci.* (2017) 18:2011. doi: 10.3390/ijms18092011
17. Rossen JWA, Friedrich AW, Moran-Gilad J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clin Microbiol Infect.* (2018) 24:355–60. doi: 10.1016/j.cmi.2017.11.001
18. Ruppé E, Cherkaoui A, Lazarevic V, Emonet S, Schrenzel J. Establishing genotype-to-phenotype relationships in bacteria causing hospital-acquired pneumonia: a prelude to the application of clinical metagenomics. *Antibiotics.* (2017) 6:30. doi: 10.3390/antibiotics6040030
19. Valentine-King MA, Brown MB. Antibacterial resistance in ureaplasma species and mycoplasma hominis isolates from urine cultures in college-Aged females. *Antimicrob Agents Chemother.* (2017) 22 61:e01104-17. doi: 10.1128/AAC.01104-17
20. Choutko V, Lazarevic V, Gaïa N, Girard M, Renzi G, Leo S, et al. Rare case of community-acquired endocarditis caused by neisseria meningitidis assessed by clinical metagenomics. *Front Cardiovasc Med.* (2019) 6:112. doi: 10.3389/fcvm.2019.00112
21. 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
22. Emonet S, Lazarevic V, Leemann Refondini C, Gaïa N, Leo S, Girard M, et al. Identification of respiratory microbiota markers in ventilator-associated pneumonia. *Intensive Care Med.* (2019) 45:1082–92. doi: 10.1007/s00134-019-05660-8
23. Steensels D, Reynders M, Descheemaeker P, Curran MD, Jacobs F, Denis O, et al. Clinical evaluation of a multi-parameter customized respiratory taqMan(®) array card compared to conventional methods in immunocompromised patients. *J Clin Virol.* (2015) 72:36–41. doi: 10.1016/j.jcv.2015.08.022
24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
25. Ounit R, Wanamaker S, Close TJ, Lonardi S. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics.* (2015) 16:236. doi: 10.1186/s12864-015-1419-2
26. Pruitt KD, Tatusova T, Maglott DR. (2007) NCBI reference sequences (RefSeq): a curated non- redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35:D61–5. doi: 10.1093/nar/gkl842
27. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* (2010) 26:2460–1. doi: 10.1093/bioinformatics/btq461
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2
29. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* (2008) 36:W5–9. doi: 10.1093/nar/gkn201
30. Lazarevic V, Gaïa N, Girard M, François P, Schrenzel J. Comparison of dNA extraction methods in analysis of salivary bacterial communities. *PLoS ONE.* (2013) 8:e67699. doi: 10.1371/journal.pone.0067699

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.

frontiers in Public Health

REVIEW
published: 23 July 2021
doi: 10.3389/fpubh.2021.706651

The Relevance of Genomic Epidemiology for Control of Tuberculosis in West Africa

Prince Asare, Adwoa Asante-Poku, Stephen Osei-Wusu, Isaac Darko Otchere and Dorothy Yeboah-Manu**

College of Health Sciences, Noguchi Memorial Institute for Medical Research, University of Ghana, Accra, Ghana

Tuberculosis (TB), an airborne infectious disease caused by *Mycobacterium tuberculosis* complex (MTBC), remains a global health problem. West Africa has a unique epidemiology of TB that is characterized by medium- to high-prevalence. Moreover, the geographical restriction of *M. africanum* to the sub-region makes West Africa have an extra burden to deal with a two-in-one pathogen. The region is also burdened with low case detection, late reporting, poor treatment adherence leading to development of drug resistance and relapse. Sporadic studies conducted within the subregion report higher burden of drug resistant TB (DRTB) than previously thought. The need for more sensitive and robust tools for routine surveillance as well as to understand the mechanisms of DRTB and transmission dynamics for the design of effective control tools, cannot be overemphasized. The advancement in molecular biology tools including traditional fingerprinting and next generation sequencing (NGS) technologies offer reliable tools for genomic epidemiology. Genomic epidemiology provides in-depth insight of the nature of pathogens, circulating strains and their spread as well as prompt detection of the emergence of new strains. It also offers the opportunity to monitor treatment and evaluate interventions. Furthermore, genomic epidemiology can be used to understand potential emergence and spread of drug resistant strains and resistance mechanisms allowing the design of simple but rapid tools. In this review, we will describe the local epidemiology of MTBC, highlight past and current investigations toward understanding their biology and spread as well as discuss the relevance of genomic epidemiology studies to TB control in West Africa.

Keywords: genomic epidemiology, tuberculosis, West Africa, *Mycobacterium tuberculosis* complex, *Mycobacterium africanum*, tuberculosis control

OPEN ACCESS

Edited by:
Martin Maiers, National Marrow Donor Program, United States

Reviewed by:
Joyce Wang, University of Michigan, United States
Sanjay Gautam, The University of Melbourne, Australia
Anou M. Somboro, University of KwaZulu-Natal, South Africa

****Correspondence:***
Prince Asare
pasare@noguchi.ug.edu.gh
Dorothy Yeboah-Manu
dyeboah-manu@noguchi.ug.edu.gh

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

Received: *07 May 2021*
Accepted: *29 June 2021*
Published: *23 July 2021*

Citation:
Asare P, Asante-Poku A, Osei-Wusu S, Otchere ID and Yeboah-Manu D (2021) The Relevance of Genomic Epidemiology for Control of Tuberculosis in West Africa. Front. Public Health 9:706651. doi: 10.3389/fpubh.2021.706651

INTRODUCTION AND BACKGROUND

Tuberculosis Historical Trends and Current Burden

Tuberculosis (TB) is a disease of antiquity and eradication of it has been man's dream throughout history. Before the 19th century, very little was known about the causative pathogen and disease mechanisms. During the 17th to 19th centuries, reports indicated that 1 in every 5 adults had TB and mortality was 900 deaths per 100,000 population in the western world. TB accounted for 20% of all human deaths at the time (1). The history and perspective of TB was changed dramatically on March 24, 1882, with presentation by Robert Koch titled, *Die Aetiologie der Tuberkulose*, to the

Berlin Physiological Society where Dr. Koch demonstrated the etiology of the disease and presented *Mycobacterium tuberculosis* as the causative agent (2).

The identification of the causative pathogen paved the way for several studies aimed at understanding the biology and the development of control tools including therapy. Antimycobacterial treatment began with the isolation of streptomycin (first isolated from *Streptomyces griseus* in 1944 by Albert Schatz, Elizabeth Bugie and Selman Waksman) followed in the 1950s and 60s by isoniazid and rifampicin (1). Nevertheless, TB persisted and does remain one of the leading causes of death among adults by a single infectious disease. TB affects millions of people annually so much so that in 1993 it became the first infectious disease to be declared a global health emergency by the World Health Organization (WHO). According to WHO estimates, about a quarter of the world's population are latently infected with the causative microorganism (3), thus, creating a pool of future active cases. Globally, in 2019 alone, an estimated 10 million new TB cases occurred, out of which 1.4 million died of TB making TB still the number one infectious disease killer by a single agent (4). Although the WHO African region is home to only 14% of the world's population, in 2019 it reported a quarter (25%, 2,460,000) of the global TB incidence, and currently has the highest HIV-associated TB cases and case fatality rates (4). This makes sub-Saharan Africa the most burdened region based on case to population ratio. Three of the 17 West African countries (Nigeria, Liberia, and Sierra Leone) are among WHO's list of 30 high TB burden countries globally. In addition, Nigeria, Liberia, Ghana, and Guinea Bissau also add up to WHO's list of the 30 high TB/HIV burden countries in the world. In 2019, 9 out of the 17 West African countries had TB incidence rate of >99 per 100,000 population per year compared to global incidence of 130 per 100,000 population per year.

To reduce this high TB burden, the WHO put in a strategy known as the "End TB Strategy" in 2014 with set targets to reduce the absolute number of TB deaths and TB incidence by 90 and 80% respectively by 2030 and 95 and 90% respectively by 2035 (5). The End TB strategy was unanimously endorsed in May 2014 by all members of the WHO and the United Nations (UN) who proceeded to adopting the UN Sustainable Development Goal (SDGs) in September 2015. The End TB strategy outlines three pillars including; (1) an integrated, patient-centered care and prevention, (2) bold policies and supportive systems, and (3) intensified research and innovation (5). Generally, the control strategy calls for improving diagnostic, intervention, and research tools to facilitate achieving the set targets. Currently, the annual rate of global TB incidence decline is about 2%, and this is far lower than the target of 10% set by the End-TB and SDG strategies. Also, per the End TB strategy, between 2015 and 2020, the total number of TB incidence rate and deaths were expected to have been reduced by 20 and 35%, respectively; however, only 9 and 14%, respectively were achieved.

The Causative Agent of Tuberculosis

Tuberculosis in mammals is mainly caused by 9 genetically related mycobacterial species comprising *Mycobacterium tuberculosis* sensu stricto (Mtb), *M. africanum* (Maf), *M. bovis, M. mungi, M. microti, M. caprae, M. pinnipedii, M. suricattae, and M. orygis* together referred to as the *M. tuberculosis* complex (MTBC) (6–9). The members of the MTBC are intracellular pathogens of mammals whose primary niche is the lungs (6, 10–15). Despite their close genetic relatedness, the MTBC differ in host specificity, although there are occasional cross-species infections (**Figure 1**). The main human pathogens are Mtb and Maf together referred to as human-adapted MTBC (hMTBC) (6, 12, 13). The animal adapted MTBC (aMTBC) comprising *M. bovis* mainly infects cattle and sheep, *M. caprae* infects goats, *M. microti* infects rodents, *M. pinnipedii* infects sea seals and sea lions, *M. mungi* infects Mangoose, Dassie bacillus infects Dassies, *M. suricattae* infects meerkats, and Chimp bacillus infects Chimpanzees whereas *M. orygis* infects antelopes (6–11, 16).

Maf is endemic in only West-African countries and is responsible for about 50% of TB cases in some of the countries (12, 17–19). Thus, in addition to dealing with the general burden of TB, West Africa has an extra burden to deal with a two-in-one pathogen.

Control of Tuberculosis and Its Challenges

The traditional methods for TB control depend on vaccination, early case detection of the affected using both clinical and laboratory-based tests followed by antimicrobial treatment of confirmed cases. TB vaccination has however, largely failed the fight against adult TB because the only WHO approved *M. bovis*—bacille Calmette-Guérin (BCG) vaccine administered to over 90% of newborns and in use since 1921 offers mainly protection against disseminated TB in children under 5 years as its efficacy wanes with time (20, 21). The current TB burden could be reduced considerably with a potent vaccine that is able to either induce clearance of latent infections or protect against new infection or both. Thus, early case detection followed by appropriate treatment remains the better option for TB control. Nevertheless, due to severe stigmatization, TB cases delay in reporting to the formal sector case management (22), which contributes to the high case fatality report by some of the countries including Ghana (23).

Laboratory methods used for diagnosis of TB include sputum smear microscopy, nucleic acid based assays, and culture. Smear microscopy prepared directly from sputum specimens is the most widely used test for diagnosing TB in West-Africa, though slowly being replaced by Gene Xpert. Molecular based tests are becoming the preferred test for diagnosing TB as most of them can simultaneously detected drug resistance. Two most widely used assays within the region are the Gene Xpert® MTB/RIF assay (Cepheid, USA) and the line probe assay (LPA) developed by Hain life sciences (GmbH, Nehren, Germany). The Gene Xpert which simultaneously detect TB as well as rifampicin (RIF) resistance directly from sputum has replaced direct sputum microscopy as the primary tool for diagnosing TB in countries like Ghana (3, 4, 24). The LPA which offers a wider spectrum of test including the ability to test for resistance to RIF, isoniazid (INH) fluoroquinolones (FQs), and injectable aminoglycosides (AMG) (GenoType MTBDR*sl*) (3, 25) is more used in second/third level laboratories.

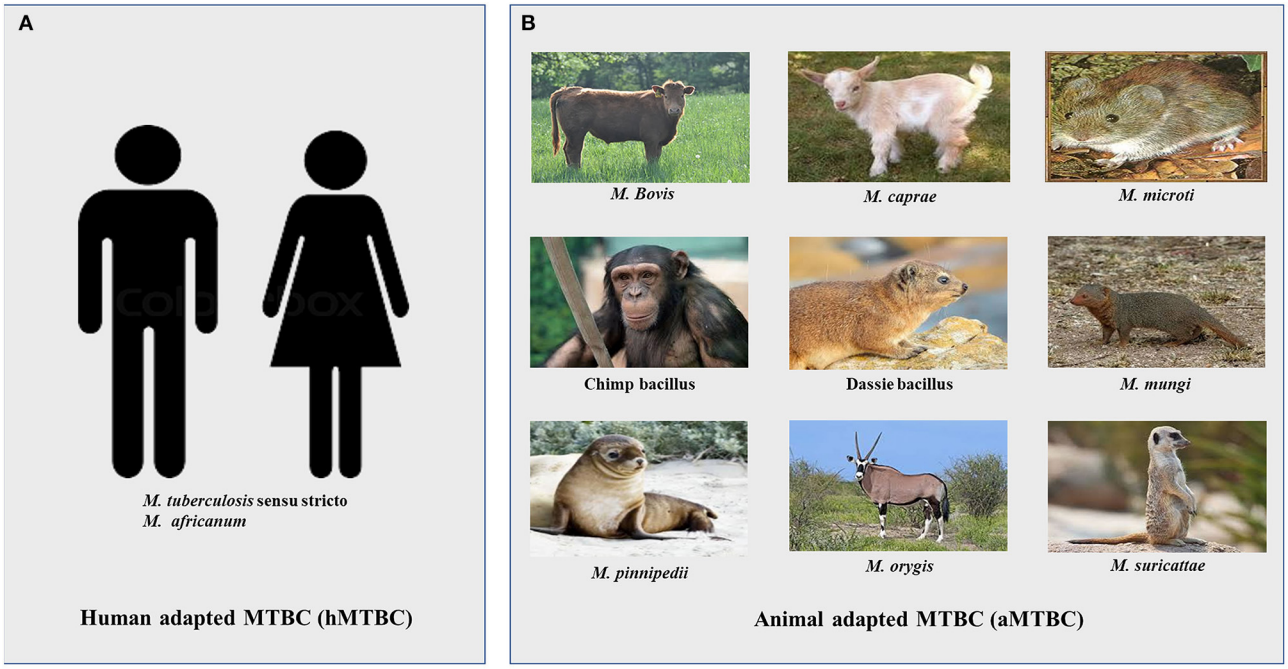


FIGURE 1 | Members of the *Mycobacterium tuberculosis* complex. **(A)** Human adapted *Mycobacterium tuberculosis* complex (hMTBC) **(B)** Animal adapted *Mycobacterium tuberculosis* complex (aMTBC).

The West-African regions follows the WHO approved case classification for anti-TB therapy. Drug sensitive active TB cases are treated with a 6-month multi-drug therapy which includes INH, RIF, pyrazinamide (PZA), and ethambutol (EMB) (26) which has ∼85% treatment success (4). However, there are recent reports of the emergence and high burden of drug resistant TB in the regions (27–30). Drug-resistant TB remains a public health threat. Treatment for drug resistant TB is quite cumbersome requiring at least 9 months (9–20 months) administration of relatively more toxic and expensive drugs such as FQs and AMGs sometimes in combination with linezolid, bedaquiline, and delamanid (4).

Challenges against TB control include socio-economic factors (such as weak health systems, increased urbanization and stigma leading to late reporting); pathogen related factors (such as emergence of drug resistant strains); lack of political will to commit funds and resources for control activities and the HIV epidemic. Lack of cheap but effective sensitive diagnostic tools and limited knowledge of the genome biology of the causative pathogen as well as the transmission dynamics of circulating strains are other equally important factors that hinder TB control. Nevertheless, traditional methods for evaluating TB control programs relies mainly on the number of cases detected and how many were cured, neglecting very crucial questions such as: the duration of infectivity, the frequency of reactivation, and the risk of progression among the infected contacts or the risk of transmission. Various molecular typing tools have been used in molecular epidemiological investigations for studying circulating MTBC strains to aid in TB control (31–33). However, whole genome sequencing (WGS), which has been made possible by the advent and increase in next generation sequencing (NGS) technologies offers the ability to study the genome of MTBC. WGS is crucial for genomic epidemiological investigations which is important for in-depth insight of the nature of pathogens, detecting circulating strains, monitoring resistance, evaluating interventions, and tracking the evolution of the pathogen hence providing a headway to achieve the End TB strategy (5). The SDG and End TB Strategy targets set for 2030 cannot be met without intensified research and innovation. In subsequent sections, we describe in detail some tools for probing MTBC genome, the local epidemiology of MTBC, highlight past and current investigations toward understanding their biology and spread as well as discuss the relevance of genomic studies to TB control in West Africa, the only sub-region that has to deal with a two-in-one pathogen.

GENOTYPING TECHNIQUES FOR EPIDEMIOLOGICAL STUDY OF *MYCOBACTERIUM TUBERCULOSIS* COMPLEX

Since the early 1990s, several genotyping tools have been proposed for the study of genetic diversity among the MTBC. These tools have been found to be discriminatory enough to distinguish unrelated strains as well as identify closely related strains. Genotyping of MTBC offers several advantages. In particular, it helps to distinguish between new infections and reactivated cases as well as identify predominant genotypes.

The classical genotyping methods that have been used to understand genetic diversity among MTBC include large sequence polymorphism (LSP) typing (34), single nucleotide polymorphism (SNP) typing (35), spacer oligonucleotide typing (spoligotyping) (36), insertion sequence 6110 (IS*6110)* restriction fragment length polymorphism (IS*6110* RFLP) typing (37) and mycobacteria interspersed repetitive unit—variable number of tandem repeats typing (MIRU-VNTR) (38) and currently, WGS (39, 40).

Large Sequence and Single Nucleotide Polymorphism Typing

Large sequence polymorphisms (LSPs) and single nucleotide polymorphisms (SNPs) are phylogenetically robust and stable molecular markers for strain identification. They are unique irreversible events and less prone to distortion by selective pressure due to lack of horizontal gene exchange in MTBC and this makes them less prone to convergent evolution (12, 41). Most importantly, LSPs also known as regions of differences (RDs) (34) have been used to define several discrete strain lineages within the hMTBC specific for different human populations and geographical regions and unravel the evolutionary scenario of ecotypes of MTBC (6, 42). SNPs have also been used to study the biology of the MTBC as a pathogen with very restricted genetic diversity (35). However, these typing tools do not allow the calculation of genetic distances and also cannot completely resolve all deep-rooting branches of the MTBC phylogeny (43).

Spoligotyping

This tool was developed in 1997 by Kamerbeek et al. (36) and is based on polymorphisms in the clustered regularly interspaced short palindromic repeats (CRISPRs) region of MTBC. Spoligotyping is the most frequently used PCR-based approach for studying the phylogeny of MTBC in high incidence areas. Spoligotyping is simple, cost-effective, and high-throughput with accurate and reproducible results within 2 days. However, it is less discriminatory; it targets only a single genetic locus, covering <0.1% of the MTBC genome. Its direct application on clinical samples without the need for prior culture and easy interpretation and computerized binary (present/absent) data format makes it suitable for molecular epidemiological studies. Direct (from sputum samples) and indirect (using cultured isolates) spoligotyping are both efficient in studying the phylogeny of MTBC, however, in regions with high prevalence of polyclonal infection such as sub-Saharan Africa where all MTBC lineages are present, it is recommended to rule out mixed infection by combining MIRU-VNTR and spoligotyping for more accurate results (44).

IS6110-RFLP

The first genotyping method developed in the early 1990s by van Embden et al., to be used for strain classification was RFLP based on IS*6110* insertion sequence *(*IS*6110*-RFLP) (37). Initially, considered as the gold standard for transmission studies, this method has been replaced by other methods for various reasons: it is labor intensive, requires high quality DNA, sophisticated and expensive computer software to analyse, experienced personnel of high technical expertise to interpret the results and most importantly, it is not discriminatory enough for strains with 6 or less IS*6110* copy numbers like some strains of *M. bovis*. Nonetheless, it paved the way for an in-depth understanding of the diversity among MTBC before the development of the more recent methods.

MIRU-VNTR Typing

This is one of the most widely used typing tool and is based on tandem repeat elements dispersed in intergenic regions of the MTBC genomes and copy number diversity (38). Currently, it has become the most reliable and efficient conventional genotyping system for TB transmission studies and has replaced *IS6110*-RFLP. This method has been widely adopted and successfully used in a variety of TB molecular epidemiological studies to trace on-going chains of TB transmission, differentiate relapse from re-infection cases and detect laboratory cross-contamination (43, 45) due to its reproducibility, portability, high discriminatory power, and standardization (33, 43, 45–49). However, it is labor-intensive due to a high number of individual PCRs required and less informative in areas with restricted MTBC lineages (43, 50–52).

Whole Genome Sequencing as a Typing Method

WGS is increasingly becoming the preferred technique for TB research. WGS determines the complete DNA sequence of an organism's genome at a single time and can provide several answers at a single time, making it the ideal tool for studying the pathogen. With WGS springing up, molecular epidemiology has gradually evolved to become genomic epidemiology. Several studies have applied large-scale WGS to different aspects of TB research; to accurately infer phylogeny (39, 40), to study the biology of the MTBC, and also to study chains of transmission (53, 54) and disease outbreaks (55). Furthermore, WGS has been used to identify drug-resistance associated mutations including; finding mutations compensating for the fitness defect associated with rifampicin resistance (56–58) and rapidly identify drug resistance mutations of an XDR-TB patient (59). These studies demonstrate the potential for future routine applications of WGS in research and genomic epidemiology. However, the use of WGS for large-scale applications especially in endemic areas is limited by its cost and the needed specialized expertise for analyses. A drawback of the current use of WGS in most TB research is that sequencing is mostly carried out on cultured isolates and analyzed using the dominant alleles present without considering within-host diversity. This can however be circumvented by investing in culture-free metagenomics-based approaches.

PHYLOGEOGRAPHY OF THE MTB AND MAF

Similar to other monomorphic bacterial pathogens such as *M. leprae* and *Bacillus anthracis*, MTBC exhibits low DNA sequence diversity and lack of horizontal gene transfer compared to other bacteria (35).

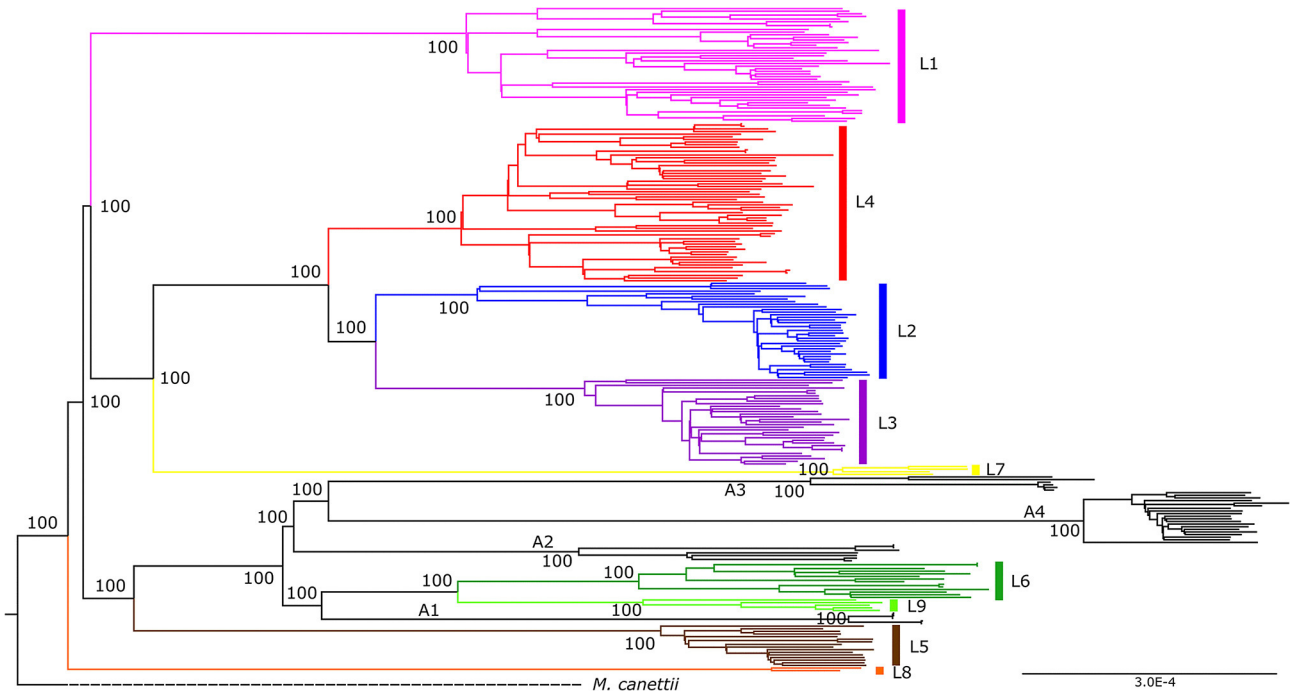


FIGURE 2 | Maximum-likelihood phylogeny of the *Mycobacterium tuberculosis* complex using whole genome sequencing. The phylogeny is rooted on *M. canettii* and bootstrap values shown for nodes. The traditional hMTBC color coding of the first 7 lineages (L1-L7) were used whereas animal adapted members are given the color black. The newly described lineages, L8 and L9 are colored orange and light green respectively [adapted from Coscolla et al. (40)].

The hMTBC were split into 7 main phylogenetic lineages (L) using specific LSPs and SNP markers (42, 60). Lineage 1, L2, L3, L4, and L7 are classified under Mtb whereas L5 and L6 are under Maf. Two additional lineages namely L8 (61) and L9 (40) have been recently identified resulting in 9 phylogenetic lineages of the hMTBC (**Figure 2**). Nevertheless, L1 to L6 are the main phylogenetic lineages based on the number and proportion of characterized isolates. Using the currently available highly discriminatory genotyping tools, several sub-lineages, and genotypes have been identified among these main phylogenetic lineages of the hMTBC (40, 62–66).

Analyzing the diversity of the hMTBC in conjunction with the country of isolation led to the discovery of different distribution patterns of the various lineages across the globe (12). Whereas, Mtb lineages with the exception of L7 (found at the Horn of Africa) are globally ubiquitous (prominent among them being L4), those of Maf (L5 and L6) are found in West Africa (67). Thus, in addition to dealing with the general burden of TB, West Africa has an extra burden to deal with a two-in-one pathogen. Although Maf is unique to West Africa, its prevalence varies by country. Using molecular genotyping results, the prevalence of L5 increases from West to East and appears highest in Benin (39%) and Ghana (21%), while that of L6 increases from East to West, highest in Guinea Bissau with 51% of smear-positive TB caused by L6 (67).

The observed associations between hMTBC lineage and country of origin under cosmopolitan clinical setting as well as between pathogen lineage and ethnicity within country point to a potential host-pathogen coevolution of the hMTBC and humans (42, 68, 69). Based on this geographical distribution pattern of hMTBC lineages, two groups namely *specialists* (limited to specified geographical locations) and *generalists* (found everywhere) have been proposed (70). L4 is generally described as a generalist lineage whereas L5 and L6 are specialists. Nevertheless, some distinct L4 sub-lineages, are restricted to specified geographical settings including the Uganda, Cameroun and Ghana genotypes (70). It is also possible that the disparate geographical spreading of hMTBC genotypes maybe explained by historical eventualities such as emerging of specific genotypes in regions that later championed colonization and globalization (39, 70, 71). Nevertheless, the potential contribution of biological traits that promote coevolution of specific hMTBC genotypes and certain human populations cannot be understated as evidenced by the San Francisco study which found associations between hMTBC lineage and country of origin of the affected TB patients (42).

GENOME BIOLOGY OF THE MTBC AND ITS SIGNIFICANCE TOWARD THE CONTROL OF TB IN WEST AFRICA

The annotated H37Rv genome revealed a genome size of 4.4 Mbp containing ∼4,000 genes. The annotated genes include those encoding proteins involved in intermediate metabolism and respiration (877; 22.0%), lipid metabolism

(225; 5.7%), information pathways (207; 5.2%), virulence, detoxification and adaptation (91; 2.3%), cell wall and cell processes (517; 13.0%), regulation (188; 4.7%), supposedly conserved hypothetical functions (911; 22.9%) and unknown functions (607; 15.3%). Non-protein-coding regions including, genes encoding stable RNAs (50; 1.3%), insertion elements and remnants of bacteriophages (137; 3.4%) and those rich with PE (Pro-Glu) and PPE (Pro-Pro-Glu) were also identified. The MTBC has extremely high number of genes involved with fatty acid metabolism, especially those associated with β-oxidation of fatty acids which are over 100 (2.5%) compared to 50 (1.2%) genes found to perform the same function in *E. coli* K-12 (72, 73). This large number of the MTBC enzymes dedicated to fatty acid catabolism enhances its ability to thrive in tissues of the infected host, where fatty acids are the major source of carbon (72). Additionally, the presence of glycine rich proteins of the PE (74) and PPE (71) families is unique to the members of the MTBC. Actual functions of these PE/PPE genes are not clearly deciphered. However, similar genes in *M. marinum* have been associated with virulence. Furthermore, antigenicity of some localized PE subfamily of proteins called PGRS (polymorphic GC-rich repetitive sequence) with conserved PE domain followed by Gly-Gly-Ala or Gly-Gly-Asn have been found which underscores the potential implication of these genes in the ability of the MTBC to cause disease.

Comparative genomics analysis of MTBC strains shows little to no evidence of horizontal gene transfer in either hMTBC or aMTBC strains (43, 75–78). This has been attributed to the evolved intracellular adaptation leading to the typical clonal nature of members of this complex compared to most non-mycobacteria and some atypical mycobacteria including *M. abscesses* and *M. avium* (39, 43, 79). Genome diversity within the MTBC thence arise mainly from SNPs comprising insertions, deletions, and substitutions and LSPs including duplication and/or transposition of mobile genetic elements as well as deletion of genetic elements accounting for the different host adaptations, disease phenotypes, and response to interventions (76, 80).

Despite lack of evidence of horizontal gene transfer within the MTBC, recent comparative genomics of the MTBC showed substantial strain diversity among the different members which could have functional implications especially in West Africa where the highest diversity of the MTBC is found (40, 81–87). Comparing the first Maf whole genome sequence (L6 strain GM041182) to H37Rv revealed the presence of a unique sequence RD900 encoding a protein involved with trans-membrane transportation of macromolecules. This sequence is also present in all so-called "ancient" lineages of the MTBC including L1, L5, L6, and L7 but independently lost in all "modern" lineages including L2, L3, L4, and *M. bovis*. Conversely, the Maf genome shares a number of uniquely lost genes with *M. bovis* but not Mtb including genes for biosynthesis of some vitamins arising from pseudogenization. In addition, the L6 genome has an intact copy of the gene (*iniA*) capable of increasing its susceptibility to antibiotics that are not active against Mtb (88, 89). It has also been shown that, all classical hMTBC strains have a conserved *mpt40 gene* which is missing from the genomes of all classical aMTBC an indication of specific host adaptation (90) a discovery that has been incorporated into a rapid nested assay for differential diagnosis of TB (91).

The effects of genomic diversity among bacterial pathogens such as *Escherichia coli*, *Neisseria menigitidis*, *Haemophylus influenzae*, B*ordetella*, and *Streptococcus* species are well-documented (92–97). In these species, some strains are more likely to cause invasive disease than others though expression of distinct virulent toxins (93). No such genetic marker (s) has been identified for the MTBC albeit infection by different genotypes results in a range of clinical phenotypes ranging from asymptomatic infection through localized diseased lungs to different forms of disseminated disease (98, 99). However, it is likely that specific MTBC genotypes have distinct genetic traits which influence the immune response elicited by the host, and subsequently the outcome of the host-pathogen interaction (100, 101). This assertion is supported by findings from *in vivo* and *ex vivo* model studies involving different MTBC lineages, indicating that Mtb lineages such as the L2 are more virulent compared to the West-African specific lineages, L5 and L6 (11, 45, 102, 103).

Clinical studies comparing Mtb and Maf (mostly L6), found statistical association between Mtb infection and early progression to active pulmonary TB disease relative to Maf, suggesting that Mtb is more virulent compared to Maf (67, 98, 104–106). Numerous genotype-specific mutations in genes of functional categories such as intermediate metabolism and respiration, cell wall and cell processes, lipid metabolism, regulatory proteins, information pathways and virulence, detoxification, and adaptation could be responsible for the differential presentation of infection by different lineages. Comparative target gene sequence analysis of MTBC lineages found that approximately two-thirds of all SNPs in coding genes are non-synonymous SNPs (nsSNPs) thus underscoring the potential implications of the limited genomic diversity within the MTBC compared to other bacteria (13, 74, 79, 107). Other than SNPs, the reductive evolution of the MTBC involving the genomic deletion of blocks of specific genomic regions including phylogenetic markers can involve blocks of diverse functional genes including but not limited to those associated with immunogenicity and/or host evasion (12, 108, 109). This observation is attributed to the deletion of RD1 a genetic evolutionary event similar to what transpired leading to the generation of the only WHO approved TB vaccine *M. bovis* BCG vaccine from continuous *in vitro* passage of *M. bovis*. The RD1 encodes the *esx1* locus which is responsible for the type VII secretory system driving the secretion of germane T cell antigens ESAT-6 and CFP-10 associated with pathogenicity of the MTBC (110–112).

Recent comparative genomics analysis of MTBC including large number of Maf genomes have revealed many insightful lineage-specific genomic events. For instance, different pairwise SNP distances within lineages (76, 84), different average pairwise nucleotide diversity of annotated genes (84), lineage-specific, and within-lineage accumulation of amino acid mutations (84, 85, 87, 105) some of which could potentially affect the applicability of some control tools (83, 85). Additionally, *in sillico* pan-genome analysis of hMTBC from West Africa identified impaired

expression of *mpt64* and *mlaD* genes specifically among L6. Whereas, the *mpt64* encodes the immunogenic protein which is the basis of the *mpt64* rapid TB diagnostics currently in use, the *mlaD* encodes a mammalian cell entry protein. Albeit there was no evidence of impaired expression of *mpt64* among L5, an amino acid mutation I43N within the gene was found specifically among L5. These findings may potentially explain the low sensitivity of *mpt64* -based diagnostics in West Africa where L5 an L6 cause up to 50% of TB (83, 85, 113) as well as the reported slow progression of Maf infection to disease relative to that of Mtb (98). Also, a couple of Maf specific amino acid mutations were found within the *esx-1* secretory system which drives secretion of T cell antigens ESAT-6 and CFP-10 associated with virulence in the MTBC (114). Interestingly, the ESAT-6 and CFP-10 are the backbone of the many interferon gamma release assays used for diagnosis of TB (115) and some potential TB vaccines in different phases of development (116, 117) which could potentially affect the applicability of these interventions in West Africa. Furthermore, essential genes of the MTBC irrespective of lineage were found to be highly conserved and under purifying selection. However, when comparing T cell epitopes, genotypes that are widely distributed such as L4 sub-lineage LAM were significantly diverse compared to T cell epitopes of specialist genotypes such as L5 and L4 Uganda genotypes (40, 70, 84). Unexpectedly, T cell epitopes of L6 which is described as a *specialist* pathogen due to its restriction to West Africa were under positive selection contrary to those of L5 which is also restricted to West Africa. This observation coupled with the statistical association of L5 with ethnicity in West Africa and association of L6 with HIV/AIDS suggest that L5 and L6 may be restricted to West Africa by different biological processes (18, 68). Thus, L6 could potentially be an opportunistic pathogen with an unknown environmental reservoir specific to West Africa (84).

In spite of these indications, exactly how MTBC genomic diversity influences disease progression and presentation as well as the distribution of various lineages and their potential impact on control of TB remain poorly understood. This calls for additional research that seeks to comparatively assess the clinical and ecological implications of MTBC genomic diversity using strain collection that encompass MTBC isolates from every part of the globe toward efficient control of TB.

MTBC SURVEILLANCE AND TRANSMISSION IN WEST AFRICA

Surveillance activities geared toward understanding MTBC transmission are necessary to complement conventional control efforts to allow the establishment of good preventive strategies, appropriate therapy, and a better understanding of the pathogen biology thereby contributing to the development of future control tools and ultimately helping eliminate TB. These surveillance activities are specifically needed (1) To correctly identify, characterize and track MTBC lineages/strains; (2) To detect risk factors associated with the disease; (3) To understand MTBC person-to-person transmission dynamics, which has been studied extensively in developed countries of North America and Europe as well as other parts of the world and has been useful for identification of outbreaks as well as most at risk groups (33, 48, 118–120) for targeted control activities and; (4) To track TB strains among recurring TB patients and provide indications of the cause of secondary case source (121–123), for appropriate treatment, evaluation of performance and epidemiology (32, 124). To effectively control TB in West Africa, it is therefore paramount to undertake such investigations in a population-based scale which will contribute to knowledge on factors that enhance spread of the disease in the sub-region.

The molecular typing tools discussed earlier have not only been used to study MTBC biology but also for surveillance purposes through molecular epidemiological investigations. Although the typing tools possess varying discriminatory power (32, 125–129), they have been used widely in advanced countries to help monitor MTBC spread especially among prisoners and cross country travelers (130–133). In West Africa, findings from surveillance studies have revealed in-depth knowledge of the varying distribution of Mtb and Maf (134, 135) and has called for further studies to investigate their transmission dynamics within respective geographical areas. This has a great public health value considering that members of the MTBC do not all have the same disease phenotype hence the need to survey to obtain knowledge of circulating genotypes. Some interesting observations has been made over the years from surveillance activities conducted in West Africa. For instance, as identified elsewhere (136–141), an association between Beijing strains of L2 and drug resistance has been reported in Benin including the identification of a possible streptomycin-resistant Beijing outbreak (142, 143) and similarly, through surveillance activities conducted in Ghana, the Ghana genotype of L4 has been linked to drug resistance (19, 144). The spread of difficult-to-treat drug resistant strains are also monitored through surveillance activities (31, 132, 145, 146). Until recently, drug resistant clones were thought to be less fit and less likely to transmit from person to person; however recent surveillance studies in Ghana, Nigeria, and other parts of the world have documented evidence of transmission of both INH resistant strains and MDR even though not involving large clusters (53, 118, 147, 148). There is therefore the need to identify and control such difficult-to-treat drug resistant clones to stop their spread through surveillance activities. WHO currently supports the inclusion of NGS to provide detailed information on drug resistance across multiple gene regions directly from sputum specimens, however the WHO is yet to review and approve current protocols (4). Also, through surveillance activities, zoonotic TB spillover has been observed in West Africa with indications that individuals who are in direct constant contact with livestock and/or dairy products are at major risk of contracting zoonotic TB (19, 149, 150). Finally, among others, surveillance activities in the sub-region has led to the observation that the West African restricted Maf specie has reduced transmissibility (45), has decreased sensitivity to some available diagnostics (83), has poor progression to active disease (98, 151), has poor treatment outcome (152, 153) and has been found to be associated with some endemic ethnic groups (18, 68).

With the advent of new genotyping techniques, surveillance activities are made more meaningful through MTBC

transmission studies. The MIRU-VNTR typing tool has over the years been widely adopted and together with epidemiological data used in a variety of TB transmission studies for the detection of recent TB transmission and outbreaks due to its portability, standardization, reproducibility, and high discriminatory power (31, 33, 45–49, 154, 155). However, only few West African studies have employed this tool for transmission studies to identify genetic and geographic TB clusters (45, 105, 135). The discriminatory power of this largely adopted MIRU-VNTR typing tool may not be sufficient to distinguish unrelated strains for some geographical settings (126) especially in West Africa with the most diverse MTBC lineages, hence for TB transmission studies it is recommended that WGS, which is by far the ultimate tool for strain differentiation, be used to make decisive conclusions. Data generated from WGS offers the ability to accurately identify recent TB transmission and also to trace the direction of transmission between epidemiologically linked cases (156). However, this tool has not been readily utilized for MTBC transmission studies in West Africa probably because of the huge cost and expertise needed to analyze the generated data. This may not be a problem in the near future as WGS is gradually becoming less expensive. The recent increase in NGS technologies coupled with the competition among available sequencing platforms and the availability of simpler data analysis tools have made WGS an attractive molecular tool used in many surveillance and transmission investigations. Globally, the first report of an effective use of WGS for MTBC transmission investigation was in 2011 from Vancouver which involved the delineation of two unrelated transmission events among a cohort of drug users having identical MIRU-VNTR profiles following which it has been used in a large array of studies (148, 157–161). WGS has been useful in detecting unsuspected outbreaks hence it should be used not only as a research tool but as a surveillance tool to aid in providing the necessary guided steps to track, monitor, and control MTBC strains.

There exist varying reports on the transmissibility of members of the MTBC (45, 98). However, limited surveillance activities have suggested that Mtb transmits better than Maf (45, 162). This is probably because Maf is thought to be attenuated compared to Mtb (67, 113, 163). Although it is argued that Mtb is fitter than Maf, and with time it will outcompete the Maf population, this may not be true as two studies from the Gambia and Ghana have proven otherwise by observing a constant prevalence of Maf over a period of 7 and 8 years, respectively (19, 152). However, decline in Maf prevalence have been reported in a number of studies from Guinea-Bissau, Côte d'Ivoire, and Cameroon (164–168). Some of these studies employed biochemical means for classification and perhaps might have misclassified the strains. There is therefore the need to invest more resources in using higher resolution tools such as the WGS for MTBC surveillance and hence transmission. Understanding MTBC transmission will contribute to knowledge on factors that enhance the spread of the disease, which is useful for developing preventive interventions and may have implications for the development and deployment of new TB vaccines and diagnostics. For instance, hotspots of TB transmission identified through a recent study in Ghana (53) offered the TB control program to direct some of their limited resources to targeted population groups for increased awareness and enhance screening activities. Also, using WGS, there is reported evidence of person-to-person transmission of Maf lineage 6 strains from Mali (87) confirming the propensity of the Maf species to also transmit (45).

Finally, through molecular surveillance activities, TB relapse are now correctly defined. Formally, all individuals who get a secondary episode of TB are referred to as having relapse. By definition, a disease is said to have relapse if the old infection bounces back. This is however not true for all secondary episodes of TB knowing that some individuals do come back with totally different strains compared to the previously infecting strain; such cases are technically referred to as exogenous re-infections rather than relapse. A secondary TB episode can be referred to as relapse only if the previous infecting strain is the same as the current one. Thankfully, current MTBC genotyping tools make it possible in most instances to distinguish between relapse and re-infection. This is possible because of the assumption that if strains from both episodes are genotypically/gnomically indistinguishable, it suggests relapse whereas distinguishable strains suggest re-infection. Predominance of relapse over re-infection indicates high-quality public health practices and a low risk of local transmission. Many studies have employed genotyping tools like MIRU-VNTR, IS*6110* RFLP, and WGS technique to explore relapse and re-infection among TB patients (121–123, 169–174). However, very few of these studies originated from West Africa. Majority of the few previous studies conducted in Africa have only employed large DNA-sequence based typing assays (ie. MIRU-VNTR and IS*6110* RFLP) which can potentially be confounded by convergence evolution. This means that established relapse cases may not actually be relapse events hence we advocate for the use of WGS which is more robust and relatively free from convergence evolution. Using WGS, it was possible to accurately detect relapse from a Ghanaian cohort; this study identified a couple of individuals who were previously infected with drug sensitive strains but later had TB recurrence harboring drug resistant strains (121). This shows the possibility to track such recurring cases and highlights the need to foster genomic epidemiology to aid early detection of drug resistance emergence to provide an effective TB control. Such surveillance activities carried out using WGS data are currently not absolute and has a few limitations as the common practice has been to make judgements based on sequencing one isolate per individual at each time point neglecting the possibility of within-host bacteria diversity. However, it is possible to detect mixed infections (175). It is worth noting that the confidence in differentiating relapse and re-infection can be reduced without considering the various bacterial populations that may exist at a given time point. We acknowledge that such within-host diversities do exist, and current and future studies should consider it in their investigations.

MECHANISM OF DRUG RESISTANCE EMERGENCE AND EVOLUTION

The control of tuberculosis has been hampered by the emergence of drug resistance globally which threatens to make TB

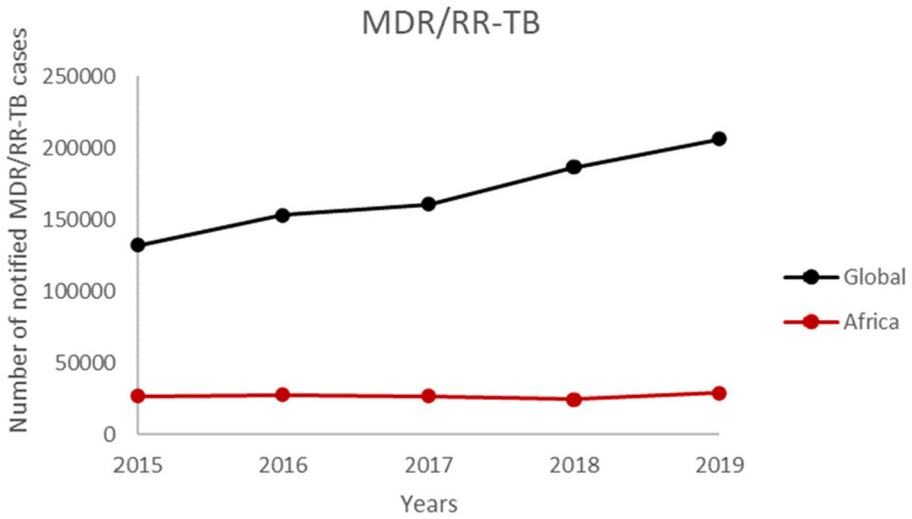


FIGURE 3 | The number of notified MDR/RR-TB cases by WHO member states from 2015 to 2019.

untreatable (4, 176). Drug resistant TB in the Africa region has been a major threat to the achievement of the goals of WHO's End TB Strategy and the SDGs in the region (4). The spread of these resistant strains mimics the pre-antibiotic era. According to the 2020 Global TB Report, about 0.5 million individuals developed rifampicin-resistant TB (RR-TB) and 78% of this number were confirmed as multi-drug resistant TB (MDR-TB) cases (4). The drug resistance TB cases contributed 3.3% of all new TB cases and 17.7% of previously treated people (4).

The threat posed by drug resistant TB strains can be mitigated by gaining a better understanding of the mechanism of the emergence of these strains and their evolution. With the recent advances in bacterial genomics using NGS, the molecular mechanisms of emergence, fixation, and transmission of drug resistant TB are being unraveled (177). However, there is the need to further assess the complexity of the emergence of drug resistant strains that have become a major challenge to the control of the disease. Comparative genomic studies have also shown that strain diversity could also be a major factor heightening the threat of TB drug resistance (40, 178, 179). This section seeks to throw more light on the advances made toward deciphering the mechanisms of drug resistance especially among the West African genotypes of MTBC and to provide new directions for future studies.

There is paucity of data on TB drug resistance in sub-Saharan Africa which has always called for active surveillance to determine the true burden of DR in West Africa and Africa as a whole (178). The WHO initially reported the burden of drug resistance in West Africa based mainly on projected estimates. An active surveillance conducted by the West African Network of Excellence for Tuberculosis, AIDS and Malaria (WANETAM), on isolates from 2009 to 2013 showed an unexpectedly higher MDR-TB of 6% for new TB cases as compared to the 3.5% WHO estimate in 2013 (180). Although the WHO estimated 20.5% MDR-TB for retreatment cases (classified as patients who have been previously treated for TB and have reported again with the disease), the network reported 35% for the eight West African countries (28, 180). In another study that screened isolates from the eight West African countries from 2012 to 2014, the authors also reported the same 6% MDR-TB for new cases and 34% for retreatment cases (30). This shows the prevalence of drug resistance in West Africa has been fairly constant and multi-drug resistance has become an emerging health challenge in West Africa. The same trend is seen in the WHO Africa region where the number of notified MDR/RR-TB cases have been constant over a 5-year period (**Figure 3**).

Several studies have linked specific lineages/sub-lineages with drug resistance (179, 181). A typical example is the Beijing sub-lineage of MTBC which is associated with MDR-TB in Asia (182, 183). The propensity for some of MTBC genotypes to harbor resistance toward anti-TB drugs have been reported in the sub-region. For instance, the Ghana sub-lineage of L4 which is mainly restricted to West Africa has been found to be associated with drug resistance in Ghana (144). Comparative genomics analysis of hMTBC from Ghana revealed that INH resistant Mtb and Maf were significantly associated with *katG* and *inhApro* mutations, respectively (144). In a separate study, a univariate analysis revealed that L6 was less likely to be associated with INH resistance (18). In support of this, a most recent comparative genomics analysis of Maf has shown that L5 has a high genetic inclination toward development of drug resistance compared to L6 (40). West Africa is unique in its MTBC genotypes and evolution of drug resistance which calls for further molecular investigation. Although a number of the resistant isolates reported in the sub-region were sequenced and analyzed for the presence of mutations associated with drug resistance, there is still a lot that were only characterized by phenotypic methods (18, 28, 30, 144).

The number of WHO States with at least one reported case of XDR-TB has been increasing over the years (**Figure 4**).

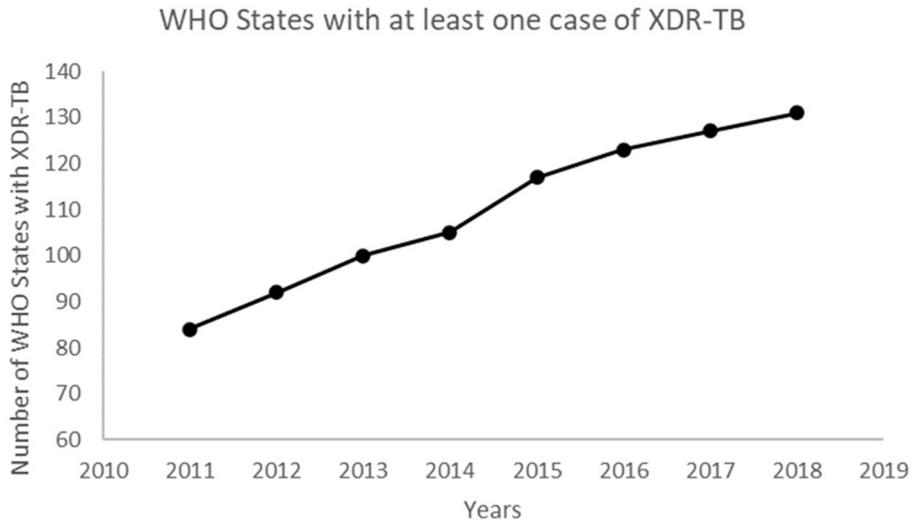


FIGURE 4 | The number of WHO member States with at least one confirmed case of XDR-TB.

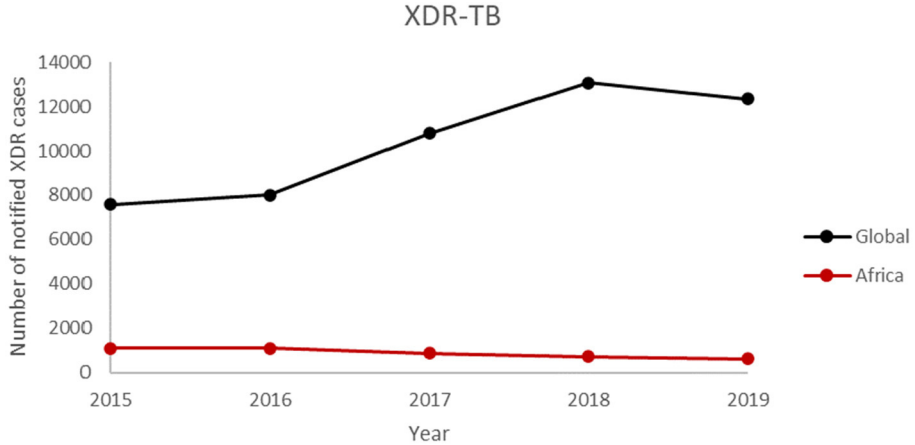


FIGURE 5 | The number of notified XDR-TB case from 2015 to 2019.

This confirms the spread of XDR-TB globally. The number of reported XDR-TB cases has been on the rise with that of Africa on a slight decrease (**Figure 5**). The DR picture for Africa might look different for different parts of the region.

The WANETAM reported the circulation of pre-XDR-TB, which is MDR-TB with additional resistance to either a fluoroquinolone or an aminoglycoside, among retreatment cases in all the eight countries in 2016 without any record of XDR-TB. Interestingly, only Ghana and Togo had reported pre-XDR TB among new cases. It was not surprising when Ghana reported its first XDR-TB case in 2018 (29). Togo, together with Burkina-Faso and Niger had already reported at least one XDR-TB case in 2011 (184). This calls for more genomic studies in West-Africa to understand the evolution and spread of DR-TB in West-Africa.

FUTURE PERSPECTIVE

Despite these evidence of genomic diversity among the MTBC supported by phenotypic data including variable outcome of TB infections, differences in macroscopic morphology, niacin production, and inhibition by pyrazinamide (185, 186), most research supporting global interventions are largely based on limited and biased collection of isolates. The presence of a unique TB causing bacteria restricted to West-Africa makes it imperative for more genomic studies in the sub-region to improve the understanding of the biology of Maf and the functional implications of genomic diversity between lineages of Maf and that of Mtb.

As has been detailed above, genomics and bioinformatics, though relatively new in biomedical disciplines, they are very useful tools for diseases surveillance and have a role to play in

the control of infectious diseases including tuberculosis. WGS of MTBC can be used in routine care settings for species identification, determination of drug resistance profiles and to complement epidemiological source investigation. Although few research works have used these tools in West-Africa to date, they are increasingly being used especially in universities and research institutions. Nevertheless, due to cost and infrastructural as well as expertise demands, its usefulness has not been felt in public health intervention programs. Some implementation challenges include funding for infrastructure as well as expertise for WGS in many research settings and even greater challenges in the context of the clinical settings. Major efforts need to be made in building human capacity as well as infrastructure investment in national public health reference laboratories to improve health care in high burdened countries. Moreover, the cost of WGS is coming down drastically and there are available simplified/portable platforms such as the minion that is field friendly and less expensive. The establishment of WGS in routine settings such as regional and national public health reference laboratories of the national health system will need to be done in a way that is relevant to the local health priorities. It is anticipated that current capacity building being championed by programs such as the African Centers of Excellence being financed through World Bank Loans, DELTAS financed by Wellcome Trust through the African Academy of Sciences will enhance implementation of WGS as part of routine health service surveillance activities, which improve health delivery in West African countries such as Ghana. This is evidenced in the contribution of some of the centers in the sequencing efforts of SARS-CoV 2 within the region (187).

The highest incidence and impact of antimicrobial resistance (AMR) is experienced in resource-poor settings (188). Underlying factors promoting AMR include misuse of antimicrobials, lower access to alternative antibiotics and the prevalence of multidrug-resistant bacterial strains (189). The slow growth nature of the MTBC negatively impacts the use of culture for routine surveillance of DR, hence, making molecular detection the better option. Routine and systematic surveillance of AMR infections is key to inform policy decisions and public health interventions to counter AMR. WGS and targeted sequencing offer the opportunity for the identification of the causative pathogen, to understand the genetic basis of resistance, as well as pathogen evolution and population dynamics at different spatial and temporal scales. It also offers the opportunity to probe the whole genome to detect not only susceptibility to a single antibiotic but multiple antibiotics at the same time. It is envisaged that potable sequencing platforms will be established at least in national and supra national reference laboratories within the region to support the DR surveillance efforts.

The use of WGS has a great prospect toward improved individual case management of infectious and non-communicable diseases. Advancement in simplified, high-throughput genomic technologies will in future assist West-Africa to sequence the whole exome or genome of a person at a price that is affordable for some health-care systems. More services based on these technologies will enhance host-directed therapies appropriate to individual patients with probing limited to analysis of specific (sets of) genes of clinical value (190, 191). WGS in future will be useful in West-Africa for evaluation of interventions such as vaccination for preventive policies through enhanced assessment of disease and drug resistance transmission dynamics. Furthermore, it will improve pathogen detection, especially with the emergence of new and un-culturable infections as well as biological risk prediction.

AUTHOR CONTRIBUTIONS

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

FUNDING

DY-M is a GSK-EDCTP Senior Fellow (TMA.2017.GSF.1942-TB-DM).

ACKNOWLEDGMENTS

We are grateful to Coscolla et al., for granting us the permission to reuse their figures in this manuscript.

REFERENCES

1. Daniel TM. The history of tuberculosis. *Respir Med.* (2006) 100:1862–70. doi: 10.1016/j.rmed.2006.08.006
2. Daniel TM. Robert Koch and the pathogenesis of tuberculosis. *Int J Tuberc Lung Dis.* (2005) 9:1181–2.
3. WHO. *Global Tuberculosis Report.* (2018). Available online at: http://www.who.int/tb/publications/global_report/en/ (accessed September 26, 2019).
4. WHO. *Global Tuberculosis Report.* (2020). Available online at: https://www.who.int/teams/global-tuberculosis-programme/data
5. WHO. *Gear Up to End TB: Introducing the End TB Strategy.* (2015). Available online at: https://www.who.int/tb/End_TB_brochure.pdf (accessed February 2, 2019).
6. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis.* complex. *Proc Natl Acad Sci USA.* (2002) 99:3684–9. doi: 10.1073/pnas.052548299
7. Huard RC, Lazzarini LC, Butler WR, van Soolingen D, Ho JL. PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. *J Clin Microbiol.* (2003) 41:1637–50. doi: 10.1128/JCM.41.4.1637-1650.2003
8. Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr MA. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J Infect Dis.* (2002) 186:74–80. doi: 10.1086/341068
9. van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R, Brosch R, et al. Characterization of Mycobacterium orygis as *M. tuberculosis* complex subspecies. *Emerg Infect Dis.* (2012) 18:653–5. doi: 10.3201/eid1804.110888
10. Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, et al. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int J Syst Evol Microbiol.* (2003) 53:1305–1314. doi: 10.1099/ijs.0.02401-0
11. Frota CC, Hunt DM, Buxton RS, Rickman L, Hinds J, Kremer K, et al. Genome structure in the vole bacillus, Mycobacterium microti, a member

of the *Mycobacterium tuberculosis* complex with a low virulence for humans. *Microbiology*. (2004) 150:1519–27. doi: 10.1099/mic.0.26660-0
12. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis.* (2007) 7:328–37. doi: 10.1016/S1473-3099(07)70108-1
13. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci USA*. (2003) 100:7877–82. doi: 10.1073/pnas.1130426100
14. Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, et al. (2013, Jun). Novel mycobacterium tuberculosis complex isolate from a wild chimpanzee. *Emerg. Infect. Dis*. 19:969–76. doi: 10.3201/eid1906.121012
15. Smith NH, Gordon SV, de la Rua-Domenech R, Clifton-Hadley RS, Hewinson RG. (2006, Sep). Bottlenecks and broomsticks: the molecular evolution of Mycobacterium bovis. *Nat Rev Microbiol*. 4:670–81. doi: 10.1038/nrmicro1472
16. Wells AQ, Oxon DM. *Tuberculosis in wild voles. Lancet*. (1937) 139:.
17. Addo K, Owusu-Darko K, Yeboah-Manu D, Caulley P, Minamikawa M, Bonsu F, et al. Mycobacterial species causing pulmonary tuberculosis at the korle bu teaching hospital, accra, ghana. *Ghana Med J.* (2007) 41:52–7. doi: 10.4314/gmj.v41i2.55293
18. Asante-Poku A, Otchere ID, Osei-Wusu S, Sarpong E, Baddoo A, Forson A, et al. Molecular epidemiology of *Mycobacterium africanum* in Ghana. *BMC Infect Dis.* (2016) 16:385. doi: 10.1186/s12879-016-1725-6
19. Yeboah-Manu D, Asare P, Asante-Poku A, Otchere ID, Osei-Wusu S, Danso E, et al. Spatio-temporal distribution of *Mycobacterium tuberculosis* complex strains in Ghana. *PLoS ONE*. (2016) 11:e0161892. doi: 10.1371/journal.pone.0161892
20. Davenne T, McShane H. Why don't we have an effective tuberculosis vaccine yet? *Expert RevVaccines*. (2016) 15:1009–13. doi: 10.1586/14760584.2016.1170599
21. Zwerling A, Behr MA, Verma A, Brewer TF, Menzies D, Pai M. The BCG World Atlas: a database of global BCG vaccination policies and practices. *PLoS Med.* (2011) 8:e1001012. doi: 10.1371/journal.pmed.1001012
22. Owolabi OA, Jallow AO, Jallow M, Sowe G, Jallow R, Genekah MD, et al. Delay in the diagnosis of pulmonary tuberculosis in the Gambia, West Africa: a cross-sectional study. *Int J Infect Dis.* (2020) 101:102–6. doi: 10.1016/j.ijid.2020.09.029
23. NTP. *World TB Day Press Release*. Ghana: National Tuberculosis Control Program (NTP) - (2019).
24. Togun TO, Egere U, Sillah AK, Ayorinde A, Mendy F, Tientcheu L, et al. Contribution of Xpert(R) MTB/RIF to the diagnosis of pulmonary tuberculosis among TB-exposed children in the Gambia. *Int J Tuberc Lung Dis.* (2015) 19:1091–7, i-ii. doi: 10.5588/ijtld.15.0228
25. Zumla A, Abubakar I, Raviglione M, Hoelscher M, Ditiu L, McHugh TD, et al. Drug-resistant tuberculosis–current dilemmas, unanswered questions, challenges, and priority needs. *J Infect Dis.* (2012) 205(Suppl. 2):S228–40. doi: 10.1093/infdis/jir858
26. WHO. *Guidelines on the Management of Latent Tuberculosis Infection*. (2015). Available online at: https://www.who.int/tb/publications/ltbi_document_page/en/ (accessed December 28, 2018).
27. Ekaza E, N'Guessan RK, Kacou-N'Douba A, Aka N, Kouakou J, Le Vacon F, et al. Emergence in Western African countries of MDR-TB, focus on Côte d'Ivoire. *Biomed Res Int*. (2013) 2013:426709. doi: 10.1155/2013/426709
28. Gehre F, Otu J, Kendall L, Forson A, Kwara A, Kudzawu S, et al. The emerging threat of pre-extensively drug-resistant tuberculosis in West Africa: preparing for large-scale tuberculosis research and drug resistance surveillance. *BMC Med.* (2016) 14:160. doi: 10.1186/s12916-016-0704-5
29. Osei-Wusu S, Amo Omari M, Asante-Poku A, Darko Otchere I, Asare P, Forson A, et al. Second-line anti-tuberculosis drug resistance testing in Ghana identifies the first extensively drug-resistant tuberculosis case. *Infect Drug Resist.* (2018) 11:239–46. doi: 10.2147/IDR.S152720
30. Otu J, Gehre F, Zingue D, Kudzawu S, Forson A, Mane M, et al. Multidrug-Resistant Tuberculosis (Mdr-Tb): an emerging problem in West Africa. *BMJ Global Health*. (2017) 2:A32–3. doi: 10.1136/bmjgh-2016-000260.85
31. Anderson LF, Tamne S, Brown T, Watson JP, Mullarkey C, Zenner D, et al. Transmission of multidrug-resistant tuberculosis in the UK: a cross-sectional molecular and epidemiological study of clustering and contact tracing. *Lancet Infect Dis.* (2014) 14:406–15. doi: 10.1016/S1473-3099(14)70022-2
32. Barnes PF, Cave MD. Molecular epidemiology of tuberculosis. *N Engl J Med*. (2003) 349:1149–56. doi: 10.1056/NEJMra021964
33. Hamblion EL, Le Menach A, Anderson LF, Lalor MK, Brown T, Abubakar I, et al. Recent TB transmission, clustering and predictors of large clusters in London, 2010-2012: results from first 3 years of universal MIRU-VNTR strain typing. *Thorax*. (2016) 71:749–56. doi: 10.1136/thoraxjnl-2014-206608
34. Huard RC, Fabre M, de Haas P, Lazzarini LC, van Soolingen D, Cousins D, et al. Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. *J Bacteriol*. (2006) 188:4271–87. doi: 10.1128/JB.01783-05
35. Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol*. (2008) 62:53–70. doi: 10.1146/annurev.micro.62.081307.162832
36. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*. (1997) 35:907–14. doi: 10.1128/jcm.35.4.907-914.1997
37. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol*. (1993) 31:406–9. doi: 10.1128/jcm.31.2.406-409.1993
38. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol*. (2001) 39:3563–71. doi: 10.1128/JCM.39.10.3563-3571.2001
39. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet*. (2013) 45:1176–82. doi: 10.1038/ng.2744
40. Coscolla M, Gagneux S, Menardo F, Loiseau C, Ruiz-Rodriguez P, Borrell S, et al. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microb Genom*. (2021) 7:000477. doi: 10.1099/mgen.0.000477
41. Schork NJ, Fallin D, Lanchbury JS. Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet*. (2000) 58:250–64. doi: 10.1034/j.1399-0004.2000.580402.x
42. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA*. (2006) 103:2869–73. doi: 10.1073/pnas.0511240103
43. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE*. (2009) 4:e7815. doi: 10.1371/journal.pone.0007815
44. Kargarpour Kamakoli M, Khanipour S, Hadifar S, Ghajavand H, Farmanfarmaei G, Fateh A, et al. Challenge in direct Spoligotyping of *Mycobacterium tuberculosis*: a problematic issue in the region with high prevalence of polyclonal infections. *BMC Res Notes*. (2018) 11:486. doi: 10.1186/s13104-018-3579-z
45. Asare P, Asante-Poku A, Prah DA, Borrell S, Osei-Wusu S, Otchere ID, et al. Reduced transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban West Africa. *Int J Infect Dis.* (2018) 73:30–42. doi: 10.1016/j.ijid.2018.05.014
46. Norheim G, Seterelv S, Arnesen TM, Mengshoel AT, Tonjum T, Ronning JO, et al. Tuberculosis outbreak in an educational institution in Norway. *J Clin Microbiol*. (2017) 55:1327–33. doi: 10.1128/JCM.01152-16
47. Streit E, Baboolal S, Akpaka PE, Millet J, Rastogi N. Finer characterization of *Mycobacterium tuberculosis* using spoligotyping and 15-loci MIRU-VNTRs reveals phylogeographical specificities of isolates circulating in Guyana and Suriname. *Infect Genetic Evolut.* (2015) 30:114–9. doi: 10.1016/j.meegid.2014.12.015
48. Vluggen C, Soetaert K, Groenen G, Wanlin M, Spitaels M, Arrazola de Onate W, et al. Molecular epidemiology of *Mycobacterium tuberculosis* complex in Brussels, 2010-2013. *PLoS ONE*. (2017) 12:e0172554. doi: 10.1371/journal.pone.0172554

49. Zamani S, Haeili M, Nasiri MJ, Imani Fooladi AA, Javadpour S, Feizabadi MM. Genotyping of *Mycobacterium tuberculosis* Isolates from hormozgan province of Iran based on 15-Locus MIRU-VNTR and spoligotyping. *Int J Bacteriol.* (2016) 2016:7146470. doi: 10.1155/2016/7146470
50. Dong H, Shi L, Zhao X, Sang B, Lv B, Liu Z, et al. Genetic diversity of *Mycobacterium tuberculosis* isolates from Tibetans in Tibet, China. *PLoS ONE.* (2012) 7:e33904. doi: 10.1371/journal.pone.0033904
51. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem repeats for typing Beijing family *Mycobacterium tuberculosis*. *J Med Microbiol.* (2008) 57:873–80. doi: 10.1099/jmm.0.47564-0
52. Shamputa IC, Lee J, Allix-Beguec C, Cho EJ, Lee JI, Rajan V, et al. Genetic diversity of *Mycobacterium tuberculosis* isolates from a tertiary care tuberculosis hospital in South Korea. *J Clin Microbiol.* (2010) 48:387–94. doi: 10.1128/JCM.02167-09
53. Asare P, Otchere ID, Bedeley E, Brites D, Loiseau C, Baddoo NA, et al. Whole genome sequencing and spatial analysis identifies recent tuberculosis transmission hotspots in Ghana. *Front Med.* (2020) 7:161. doi: 10.3389/fmed.2020.00161
54. Schurch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, Siezen RJ, et al. High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *J Clin Microbiol.* (2010) 48:3403–6. doi: 10.1128/JCM.00370-10
55. Bryant JM, Schurch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, et al. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect Dis.* (2013) 13:110. doi: 10.1186/1471-2334-13-110
56. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet.* (2014) 46:279–86. doi: 10.1038/ng.2878
57. Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, et al. Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res.* (2012) 22:735–45. doi: 10.1101/gr.128678.111
58. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet.* (2011) 44:106–10. doi: 10.1038/ng.1038
59. Koser CU, Bryant JM, Parkhill J, Peacock SJ. Consequences of whiB7 (Rv3197A) mutations in Beijing genotype isolates of the *Mycobacterium tuberculosis* complex. *Antimicrob Agents Chemother.* (2013) 57:3461. doi: 10.1128/AAC.00626-13
60. Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, et al. Significance of the identification in the horn of africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. *PLoS ONE.* (2012) 7:e52841. doi: 10.1371/journal.pone.0052841
61. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat Commun.* (2020) 11:2917. doi: 10.1038/s41467-020-16626-6
62. Goig GA, Torres-Puente M, Mariner-Llicer C, Villamayor LM, Chiner-Oms Á, Gil-Brusola A, et al. Towards next-generation diagnostics for tuberculosis: identification of novel molecular targets by large-scale comparative genomics. *Bioinformatics.* (2020) 36:985–9. doi: 10.1101/569384
63. Hijikata M, Keicho N, Duc LV, Maeda S, Hang NTL, Matsushita I, et al. Spoligotyping and whole-genome sequencing analysis of lineage 1 strains of *Mycobacterium tuberculosis* in Da Nang, Vietnam. *PLoS ONE.* (2017) 12:e0186800. doi: 10.1371/journal.pone.0186800
64. Malm S, Linguissi LS, Tekwu EM, Vouvoungui JC, Kohl TA, Beckert P, et al. New *Mycobacterium tuberculosis* complex sublineage, Brazzaville, Congo. *Emerging Infect Dis.* (2017) 23:423–9. doi: 10.3201/eid2303.160679
65. Rutaihwa LK, Sasamalo M, Jaleco A, Hella J, Kingazi A, Kamwela L, et al. Insights into the genetic diversity of *Mycobacterium tuberculosis* in Tanzania. *PLoS ONE.* (2019) 14:e0206334. doi: 10.1371/journal.pone.0206334
66. Shuaib YA, Khalil EAG, Wieler LH, Schaible UE, Bakheit MA, Mohamed-Noor SE, et al. *Mycobacterium tuberculosis* complex lineage 3 as causative agent of pulmonary tuberculosis, Eastern Sudan(1). *Emerg Infect Dis.* (2020) 26:427–36. doi: 10.3201/eid2603.191145
67. de Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*–review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis.* (2010) 4:e744. doi: 10.1371/journal.pntd.0000744
68. Asante-Poku A, Yeboah-Manu D, Otchere ID, Aboagye SY, Stucki D, Hattendorf J, et al. *Mycobacterium africanum* is associated with patient ethnicity in Ghana. *PLoS Negl Trop Dis.* (2015) 9:e3370. doi: 10.1371/journal.pntd.0003370
69. Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, et al. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *J Clin Microbiol.* (2009) 47:1119–28. doi: 10.1128/JCM.02142-08
70. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet.* (2016) 48:1535–43. doi: 10.1038/ng.3704
71. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, et al. (2018). Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Sci Adv.* 4:eaat. doi: 10.1126/sciadv.aat5869
72. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* (1998) 393:537–44. doi: 10.1038/31159
73. Serres MH, Gopal S, Nahum LA, Liang P, Gaasterland T, Riley M. A functional update of the *Escherichia coli* K-12 genome. *Genome Biol.* (2001) 2:Research0035. doi: 10.1186/gb-2001-2-9-research0035
74. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* (2008) 6:e311. doi: 10.1371/journal.pbio.0060311
75. Blouin Y, Cazajous G, Dehan C, Soler C, Vong R, Hassan MO, et al. Progenitor "*Mycobacterium canettii*" clone responsible for lymph node tuberculosis epidemic, Djibouti. *Emerg Infect Dis.* (2014) 20:21–8. doi: 10.3201/eid2001.130652
76. Coscolla M, Gagneux S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin Immunol.* (2014) 26:431–44. doi: 10.1016/j.smim.2014.09.012
77. Ioerger TR, Feng Y, Ganesula K, Chen X, Dobos KM, Fortune S, et al. Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. *J Bacteriol.* (2010) 192:3645–53. doi: 10.1128/JB.00166-10
78. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nat Genet.* (2013) 45:172–9. doi: 10.1038/ng.2517
79. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA.* (1997) 94:9869–74. doi: 10.1073/pnas.94.18.9869
80. Coscolla M, Gagneux S. Does *M. Tuberculosis* genomic diversity explain disease diversity? Drug discovery today. *Dis Mechan.* (2010) 7:e43–59. doi: 10.1016/j.ddmec.2010.09.004
81. Gehre F, Otu J, DeRiemer K, de Sessions PF, Hibberd ML, Mulders W, et al. Deciphering the growth behaviour of *Mycobacterium africanum*. *PLoS Negl Trop Dis.* (2013) 7:e2220. doi: 10.1371/journal.pntd.0002220
82. Krishnan N, Malaga W, Constant P, Caws M, Tran TH, Salmons J, et al. *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PLoS ONE.* (2011) 6:e23870. doi: 10.1371/journal.pone.0023870
83. Ofori-Anyinam B, Kanuteh F, Agbla SC, Adetifa I, Okoi C, Dolganov G, et al. Impact of the Mycobaterium africanum West Africa 2 Lineage on TB diagnostics in West Africa: decreased sensitivity of rapid identification tests in the Gambia. *PLoS Negl Trop Dis.* (2016) 10:e0004801. doi: 10.1371/journal.pntd.0004801
84. Otchere ID, Coscollá M, Sánchez-Busó L, Asante-Poku A, Brites D, Loiseau C, et al. Comparative genomics of *Mycobacterium africanum* Lineage 5 and Lineage 6 from Ghana suggests distinct ecological niches. *Sci Rep.* (2018) 8:11269. doi: 10.1038/s41598-018-29620-2
85. Sanoussi CN, de Jong BC, Odoun M, Arekpa K, Ali Ligali M, Bodi O, et al. Low sensitivity of the MPT64 identification test to detect lineage 5 of

the *Mycobacterium tuberculosis* complex. *J Med Microbiol.* (2018) 67:1718–27. doi: 10.1099/jmm.0.000846
86. Sarkar R, Lenders L, Wilkinson KA, Wilkinson RJ, Nicol MP. Modern lineages of *Mycobacterium tuberculosis* exhibit lineage-specific patterns of growth and cytokine induction in human monocyte-derived macrophages. *PLoS ONE.* (2012) 7:e43170. doi: 10.1371/journal.pone.0043170
87. Winglee K, Manson McGuire A, Maiga M, Abeel T, Shea T, Desjardins CA, et al. Whole genome sequencing of *Mycobacterium africanum* Strains from mali provides insights into the mechanisms of geographic restriction. *PLoS Negl Trop Dis.* (2016) 10:e0004332. doi: 10.1371/journal.pntd.0004332
88. Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, Harris SR, et al. The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. *PLoS Neglected Trop Dis.* (2012) 6:e1552. doi: 10.1371/journal.pntd.0001552
89. Colangeli R, Helb D, Sridharan S, Sun J, Varma-Basil M, Hazbón MH, et al. The *Mycobacterium tuberculosis* iniA gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. *Mol Microbiol.* (2005) 55:1829–40. doi: 10.1111/j.1365-2958.2005.04510.x
90. Koivula T, Svenson SB, Källenius G. The mtp40 gene is not present in *Mycobacterium bovis*. *Tuberculosis.* (2002) 82:183–5. doi: 10.1054/tube.2002.0338
91. Sinha P, Prakash P, Patne SC, Anupurba S, Gupta S, Srivastava GN. Performance of nested multiplex PCR assay targeting MTP40 and IS6110 gene sequences for the diagnosis of tubercular lymphadenitis. *J Microbiol.* (2017) 55:63–7. doi: 10.1007/s12275-017-6127-y
92. Bogaert D, van der Valk P, Ramdin R, Sluijter M, Monninkhof E, Hendrix R, et al. Host-pathogen interaction during pneumococcal infection in patients with chronic obstructive pulmonary disease. *Infect Immun.* (2004) 72:818–23. doi: 10.1128/IAI.72.2.818-823.2004
93. Buboltz AM, Nicholson TL, Karanikas AT, Preston A, Harvill ET. Evidence for horizontal gene transfer of two antigenically distinct O antigens in Bordetella bronchiseptica. *Infect Immun.* (2009) 77:3249–57. doi: 10.1128/IAI.01448-08
94. Lipsitch M. Bacterial vaccines and serotype replacement: lessons from Haemophilus influenzae and prospects for Streptococcus pneumoniae. *Emerging Infect Dis.* (1999) 5:336–45. doi: 10.3201/eid0503.990304
95. Madoff LC, Michel JL, Gong EW, Kling DE, Kasper DL. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proc Natl Acad Sci USA.* (1996) 93:4131–6. doi: 10.1073/pnas.93.9.4131
96. Schouls L, van der Heide H, Witteveen S, Zomer B, van der Ende A, Burger M, et al. Two variants among Haemophilus influenzae serotype b strains with distinct bcs4, hcsA and hcsB genes display differences in expression of the polysaccharide capsule. *BMC Microbiol.* (2008) 8:35. doi: 10.1186/1471-2180-8-35
97. Spratt BG, Greenwood BM. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet.* (2000) 356:1210–1. doi: 10.1016/S0140-6736(00)02779-3
98. de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, et al. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in the Gambia. *J Infect Dis.* (2008) 198:1037–43. doi: 10.1086/591504
99. Huang CC, Chu AL, Becerra MC, Galea JT, Calderon R, Contreras C, et al. *Mycobacterium tuberculosis* Beijing lineage and risk for tuberculosis in child household contacts, Peru. *Emerging Infect Dis.* (2020) 26:568–78. doi: 10.3201/eid2603.191314
100. Forrellad MA, Klepp LI, Gioffre A, Sabioy Garcia J, Morbidoni HR, de la Paz Santangelo M, et al. Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence.* (2013) 4:3–66. doi: 10.4161/viru.22329
101. Nunes-Alves C, Booty MG, Carpenter SM, Jayaraman P, Rothchild AC, Behar SM. In search of a new paradigm for protective immunity to TB. *Nat Rev Microbiol.* (2014) 12:289–99. doi: 10.1038/nrmicro3230
102. López B, Aguilar D, Orozco H, Burger M, Espitia C, Ritacco V, et al. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. *Clin Exp Immunol.* (2003) 133:30–7. doi: 10.1046/j.1365-2249.2003.02171.x
103. Portevin D, Gagneux S, Comas I, Young D. Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog.* (2011) 7:e1001307. doi: 10.1371/journal.ppat.1001307
104. Bold TD, Davis DC, Penberthy KK, Cox LM, Ernst JD, de Jong BC. Impaired fitness of *Mycobacterium africanum* despite secretion of ESAT-6. *J Infect Dis.* (2012) 205:984–90. doi: 10.1093/infdis/jir883
105. Gehre F, Antonio M, Faihun F, Odoun M, Uwizeye C, de Rijk P, et al. The first phylogeographic population structure and analysis of transmission dynamics of *M. africanum* West African 1–combining molecular data from Benin, Nigeria and Sierra Leone. *PLoS ONE.* (2013) 8:e77000. doi: 10.1371/journal.pone.0077000
106. Meyer CG, Scarisbrick G, Niemann S, Browne ENL, Chinbuah MA, Gyapong J, et al. Pulmonary tuberculosis: Virulence of *Mycobacterium africanum* and relevance in HIV co-infection. *Tuberculosis.* (2008) 88:482–9. doi: 10.1016/j.tube.2008.05.004
107. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol.* (2002) 184:5479–90. doi: 10.1128/JB.184.19.5479-5490.2002
108. Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, et al. Genomic analysis distinguishes *Mycobacterium africanum*. *J Clin Microbiol.* (2004) 42:3594–9. doi: 10.1128/JCM.42.8.3594-3599.2004
109. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, et al. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci USA.* (2004) 101:4865–70. doi: 10.1073/pnas.0305634101
110. Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, et al. Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol.* (2004) 51:359–70. doi: 10.1046/j.1365-2958.2003.03844.x
111. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and Mycobacterium microti. *Mol Microbiol.* (2002) 46:709–17. doi: 10.1046/j.1365-2958.2002.03237.x
112. Stanley SA, Raghavan S, Hwang WW, Cox JS. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci USA.* (2003) 100:13001–6. doi: 10.1073/pnas.2235593100
113. Otchere ID, Harris SR, Busso SL, Asante-Poku A, Osei-Wusu S, Koram K, et al. The First population structure and comparative genomics analysis of *Mycobacterium africanum* strains from Ghana reveals higher diversity of Lineage. *Int J Mycobacteriol.* (2016) 5:S80–1. doi: 10.1016/j.ijmyco.2016.09.051
114. Otchere ID. Comparative genomic analysis of *Mycobacterium tuberculosis* complex strains from Ghana [thesis]. PhD. Thesis published in University of Ghana digital collection (2017). Available online at: http://ugspace.ug.edu.gh/handle/123456789/24762 (accessed May 5, 2021).
115. Horvat RT. Gamma interferon assays used in the diagnosis of tuberculosis. *Clin Vaccine Immunol.* (2015) 22:845. doi: 10.1128/CVI.00199-15
116. Brandt L, Elhay M, Rosenkrands I, Lindblad EB, Andersen P. ESAT-6 subunit vaccination against *Mycobacterium tuberculosis*. *Infect Immun.* (2000) 68:791–5. doi: 10.1128/IAI.68.2.791-795.2000
117. Kim WS, Kim H, Kwon KW, Cho SN, Shin SJ. Immunogenicity and vaccine potential of InsB, an ESAT-6-Like antigen identified in the highly virulent *Mycobacterium tuberculosis* Beijing K strain. *Front Microbiol.* (2019) 10:220. doi: 10.3389/fmicb.2019.00220
118. Gautam SS, Haug G, Cooley LA, Mac Aogain M and O'Toole RF. Intercontinental translocation of latent multidrug-resistant tuberculosis to Australia demonstrated by whole genome sequencing. *Med J Australia.* (2019) 210:236–231. doi: 10.5694/mja2.50044
119. Gautam SS, Mac Aogain M, Cooley LA, Haug G, Fyfe JA, Globan M, et al. Molecular epidemiology of tuberculosis in Tasmania and genomic characterisation of its first known multi-drug resistant case. *PLoS ONE.* (2018) 13:e0192351. doi: 10.1371/journal.pone.0192351
120. Seraphin MN, Lauzardo M, Doggett RT, Zabala J, Morris JG Jr, Blackburn JK. Spatiotemporal clustering of *Mycobacterium tuberculosis* complex genotypes in florida: genetic diversity segregated by country of birth. *PLoS ONE.* (2016) 11:e0153575. doi: 10.1371/journal.pone.0153575

121. Asare P, Osei-Wusu S, Baddoo NA, Bedeley E, Otchere ID, Brites D, et al. Genomic epidemiological analysis identifies high relapse among individuals with recurring tuberculosis and provides evidence of recent household-related transmission of tuberculosis in Ghana. *Int J Infect Dis.* (2021) 106:13–22. doi: 10.1016/j.ijid.2021.02.110
122. Shanmugam S, Bachmann NL, Martinez E, Menon R, Narendran G, Narayanan S, et al. Whole genome sequencing based differentiation between re-infection and relapse in Indian patients with tuberculosis recurrence, with and without HIV co-infection. *Int J Infect Dis.* (2021). doi: 10.1016/j.ijid.2021.03.020. [Epub ahead of print].
123. Varghese B, al-Omari R, Grimshaw C, Al-Hajoj S. Endogenous reactivation followed by exogenous re-infection with drug resistant strains, a new challenge for tuberculosis control in Saudi Arabia. Tuberculosis. (2013) 93:246–249. doi: 10.1016/j.tube.2012.12.001
124. van Soolingen D, Kremer K, Vynycky E. New perspectives in the molecular epidemiology of tuberculosis. In: *Mycobacteria and TB*. Vol. 2. Karger Publishers (2003). p. 17–45. doi: 10.1159/000066887
125. Jagielski T, van Ingen J, Rastogi N, Dziadek J, Mazur PK, Bielecki J. Current methods in the molecular typing of *Mycobacterium tuberculosis* and other mycobacteria. *Biomed Res Int.* (2014) 2014:645802. doi: 10.1155/2014/645802
126. Jamieson FB, Teatero S, Guthrie JL, Neemuchwala A, Fittipaldi N, Mehaffy C. Whole-genome sequencing of the *Mycobacterium tuberculosis* Manila sublineage results in less clustering and better resolution than mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) typing and spoligotyping. *J Clin Microbiol.* (2014) 52:3795–8. doi: 10.1128/JCM.01726-14
127. Kanduma E, McHugh TD, Gillespie SH. Molecular methods for *Mycobacterium tuberculosis* strain typing: a users guide. *J Appl Microbiol.* (2003) 94:781–91. doi: 10.1046/j.1365-2672.2003.01918.x
128. Mendez MV, Leon C, Escalona A, Abadia E, Da Mata O, de Waard J, et al. [Evaluation of discriminatory power of molecular epidemiology techniques in *Mycobacterium tuberculosis* Venezuelan isolates]. *Invest Clin.* (2016) 57:25–37.
129. Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. *Am J Epidemiol.* (2002) 155:565–71. doi: 10.1093/aje/155.6.565
130. Abascal E, Perez-Lago L, Martinez-LirolaM, Chiner-Oms A, Herranz M, Chaoui I, et al. Whole genome sequencing-based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations. *Euro Surveill.* (2019) 24:1800005. doi: 10.2807/1560-7917.ES.2019.24.4.1800005
131. Martinez-Lirola M, Jajou R, Mathys V, Martin A, Cabibbe AM, Valera A, et al. Integrative transnational analysis to dissect tuberculosis transmission events along the migratory route from Africa to Europe. *J Travel Med.* (2021) 28:taab054. doi: 10.1093/jtm/taab054
132. Tagliani E, Anthony R, Kohl TA, de Neeling A, Nikolayevskyy V, Kodmon C, et al. Use of a whole genome sequencing-based approach for *Mycobacterium tuberculosis* surveillance in Europe in 2017-2019: an ECDC pilot study. *EuropRespir J.* (2021) 57:2002272. doi: 10.1183/13993003.02272-2020
133. Walter KS, Tatara MB, Esther da Silva K, Moreira FMF, Dos Santos PCP, de Melo Ferrari DD, et al. Local and travel-associated transmission of tuberculosis at central western border of Brazil, 2014-2017. *Emerging Infect Dis.* (2021) 27:905–14. doi: 10.3201/eid2703.203839
134. Gehre F, Kumar S, Kendall L, Ejo M, Secka O, Ofori-Anyinam B, et al. A mycobacterial perspective on tuberculosis in west africa: significant geographical variation of *M. africanum* and other m tuberculosis complex lineages. *PLoS Negl Trop Dis.* (2016) 10:e0004408. doi: 10.1371/journal.pntd.0004408
135. Mekonnen D, Derbie A, Abeje A, Shumet A, Kassahun Y, Nibret E, et al. Genomic diversity and transmission dynamics of *M. tuberculosis* in Africa: a systematic review and meta-analysis. *Int J Tuberc Lung Dis.* (2019) 23:1314–26. doi: 10.5588/ijtld.19.0127
136. Almeida D, Rodrigues C, Ashavaid TF, Lalvani A, Udwadia ZF, Mehta A. High incidence of the Beijing genotype among multidrug-resistant isolates of *Mycobacterium tuberculosis* in a tertiary care center in Mumbai, India. *Clin Infect Dis.* (2005) 40:881–6. doi: 10.1086/427940
137. Arora J, Bhalla M, Sidiq Z, Lal P, Behera D, Rastogi N, et al. Predominance of Beijing genotype in extensively drug resistant *Mycobacterium tuberculosis* isolates from a tertiary care hospital in New Delhi, India. *Int J Mycobacteriol.* (2013) 2:109–13. doi: 10.1016/j.ijmyco.2013.03.001
138. Buu TN, van Soolingen D, Huyen MN, Lan NT, Quy HT, Tiemersma EW, et al. Increased transmission of *Mycobacterium tuberculosis* Beijing genotype strains associated with resistance to streptomycin: a population-based study. *PLoS ONE.* (2012) 7:e42323. doi: 10.1371/journal.pone.0042323
139. de Steenwinkel JE, ten Kate MT, de Knegt GJ, Kremer K, Aarnoutse RE, Boeree MJ, et al. Drug susceptibility of *Mycobacterium tuberculosis* Beijing genotype and association with MDR TB. *Emerging Infect Dis.* (2012) 18:660–663. doi: 10.3201/eid1804.110912
140. Liang C, Zhang XX, Xing Q, Yi JL, Zhang YQ, Li CY, et al. [Study on the prevalence of Beijing genotype *Mycobacterium tuberculosis* and its relationship with second-line anti-tuberculosis drug resistance]. *Zhonghua Jie He He Hu Xi Za Zhi.* (2020) 43:356–61. doi: 10.3760/cma.j.cn112147-20191215-00827
141. Nieto Ramirez LM, Ferro BE, Diaz G, Anthony RM, de Beer J, van Soolingen D. Genetic profiling of *Mycobacterium tuberculosis* revealed "modern" Beijing strains linked to MDR-TB from Southwestern Colombia. *PLoS ONE.* (2020) 15:e0224908. doi: 10.1371/journal.pone.0224908
142. Affolabi D, Anyo G, Faihun F, Sanoussi N, Shamputa IC, Rigouts L, et al. First molecular epidemiological study of tuberculosis in Benin. *Int J Tuberc Lung Dis.* (2009) 13:317–22.
143. Affolabi D, Faihun F, Sanoussi N, Anyo G, Shamputa IC, Rigouts L, et al. Possible outbreak of streptomycin-resistant *Mycobacterium tuberculosis* Beijing in Benin. *Emerging Infect Dis.* (2009) 15:1123–25. doi: 10.3201/eid1507.080697
144. Otchere ID, Asante-Poku A, Osei-Wusu S, Baddoo A, Sarpong E, Ganiyu AH, et al. Detection and characterization of drug-resistant conferring genes in *Mycobacterium tuberculosis* complex strains: a prospective study in two distant regions of Ghana. *Tuberculosis.* (2016) 99:147–54. doi: 10.1016/j.tube.2016.05.014
145. Sanchini A, Jandrasits C, Tembrockhaus J, Kohl TA, Utpatel C, Maurer FP, et al. Improving tuberculosis surveillance by detecting international transmission using publicly available whole genome sequencing data. *Euro Surveill.* (2021) 26:1900677. doi: 10.2807/1560-7917.ES.2021.26.2.1900677
146. Wang F, Shao L, Fan X, Shen Y, Diao N, Jin J, et al. Evolution and transmission patterns of extensively drug-resistant tuberculosis in China. *Antimicrob Agents Chemother.* (2015) 59:818–25. doi: 10.1128/AAC.03504-14
147. Senghore M, Otu J, Witney A, Gehre F, Doughty EL, Kay GL, et al. Whole-genome sequencing illuminates the evolution and spread of multidrug-resistant tuberculosis in Southwest Nigeria. *PLoS ONE.* (2017) 12:e0184510. doi: 10.1371/journal.pone.0184510
148. Walker TM, Merker M, Knoblauch AM, Helbling P, Schoch OD, van der Werf MJ, et al. A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *Lancet Infect Dis.* (2018) 18:431–40. doi: 10.1016/S1473-3099(18)30004-5
149. Cadmus S, Akinseye VO, van Soolingen D. *Mycobacterium bovis* in humans and *M.* tuberculosis in animals in Nigeria: an overview from 1975-2014. *Int J Tuberc Lung Dis.* (2019) 23:1162–70. doi: 10.5588/ijtld.18.0641
150. Otchere ID, van Tonder AJ, Asante-Poku A, Sanchez-Buso L, Coscolla M, Osei-Wusu S, et al. Molecular epidemiology and whole genome sequencing analysis of clinical *Mycobacterium bovis* from Ghana. *PLoS ONE.* (2019) 14:e0209395. doi: 10.1371/journal.pone.0209395
151. Baya B, Diarra B, Diabate S, Kone B, Goita D, Sarro YDS, et al. Association of *Mycobacterium africanum* infection with slower disease progression compared with *Mycobacterium tuberculosis* in malian patients with tuberculosis. *Am J Trop Med Hyg.* (2020) 102:36–41. doi: 10.4269/ajtmh.19-0264
152. de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M, et al. Differences between tuberculosis cases infected with *Mycobacterium africanum*, West African type 2, relative to Euro-American *Mycobacterium tuberculosis*: an update. *FEMS Immunol Med Microbiol.* (2010) 58:102–5. doi: 10.1111/j.1574-695X.2009.00628.x

153. Tientcheu LD, Bell A, Secka O, Ayorinde A, Otu J, Garton NJ, et al. Association of slow recovery of *Mycobacterium africanum*-infected patients posttreatment with high content of Persister-Like bacilli in pretreatment sputum. *Int J Mycobacteriol.* (2016) 5(Suppl. 1):S99–100. doi: 10.1016/j.ijmyco.2016.09.033
154. Varghese B, Supply P, Shoukri M, Allix-Beguec C, Memish Z, Abuljadayel N, et al. Tuberculosis transmission among immigrants and autochthonous populations of the eastern province of Saudi Arabia. *PLoS ONE.* (2013) 8:e77635. doi: 10.1371/journal.pone.0077635
155. Zetola NM, Moonan PK, Click E, Oeltmann JE, Basotli J, Wen XJ, et al. Population-based geospatial and molecular epidemiologic study of tuberculosis transmission dynamics, Botswana, 2012-2016. *Emerging Infect Dis.* (2021) 27:835–44. doi: 10.3201/eid2703.203840
156. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis.* (2013) 13:137–46. doi: 10.1016/S1473-3099(12)70277-3
157. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med.* (2011) 364:730–9. doi: 10.1056/NEJMoa1003176
158. Lalor MK, Casali N, Walker TM, Anderson LF, Davidson JA, Ratna N, et al. The use of whole-genome sequencing in cluster investigation of a multidrug-resistant tuberculosis outbreak. *Europ Respir J.* (2018) 51:1702313. doi: 10.1183/13993003.02313-2017
159. Walker TM, Lalor MK, Broda A, Saldana Ortega L, Morgan M, Parker L, et al. Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med.* (2014) 2:285–92. doi: 10.1016/S2213-2600(14)70027-X
160. Walker TM, Merker M, Kohl TA, Crook DW, Niemann S, Peto TE. Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing. *Clin Microbiol Infect.* (2017) 23:161–6. doi: 10.1016/j.cmi.2016.10.014
161. Walker TM, Monk P, Smith EG, Peto TEA. Contact investigations for outbreaks of *Mycobacterium tuberculosis* advances through whole genome sequencing. *Clin Microbiol Infect.* (2013) 19:796–802. doi: 10.1111/1469-0691.12183
162. Nebenzahl-Guimaraes H, Verhagen LM, Borgdorff MW, van Soolingen D. Transmission and progression to disease of *Mycobacterium tuberculosis* phylogenetic lineages in the Netherlands. *J Clin Microbiol.* (2015) 53:3264–71. doi: 10.1128/JCM.01370-15
163. de Jong BC, Hill PC, Brookes RH, Gagneux S, Jeffries DJ, Otu JK, et al. *Mycobacterium africanum* elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. *J Infect Dis.* (2006) 193:1279–86. doi: 10.1086/502977
164. Dosso M, Bonard D, Msellati P, Bamba A, Doulhourou C, Vincent V, et al. Primary resistance to antituberculosis drugs: a national survey conducted in Cote d'Ivoire in 1995-1996. Ivoirian Study Group on Tuberculosis Resistance. *Int J Tuberc Lung Dis.* (1999) 3:805–9.
165. Huet M, Rist N, Boube G, Potier D. [Bacteriological study of tuberculosis in Cameroon]. *Revue de Tuberculose et de Pneumol.* (1971 35:413–26.
166. Kallenius G, Koivula T, Ghebremichael S, Hoffner SE, Norberg R, Svensson E, et al. Evolution and clonal traits of *Mycobacterium tuberculosis* complex in Guinea-Bissau. *J Clin Microbiol.* (1999) 37:3872–8. doi: 10.1128/JCM.37.12.3872-3878.1999
167. Koro Koro F, Kamdem Simo Y, Piam FF, Noeske J, Gutierrez C, Kuaban C, et al. Population dynamics of tuberculous Bacilli in Cameroon as assessed by spoligotyping. *J Clin Microbiol.* (2013) 51:299–302. doi: 10.1128/JCM.01196-12
168. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, Sola C, et al. Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. *J Clin Microbiol.* (2003) 41:2547–53. doi: 10.1128/JCM.41.6.2547-2553.2003
169. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P, et al. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir Med.* (2013) 1:786–92. doi: 10.1016/S2213-2600(13)70231-5
170. Guerra-Assuncao JA, Houben RM, Crampin AC, Mzembe T, Mallard K, Coll F, et al. Recurrence due to relapse or reinfection with *Mycobacterium tuberculosis*: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. *J Infect Dis.* (2015) 211:1154–63. doi: 10.1093/infdis/jiu574
171. McIvor A, Koornhof H, Kana BD. Relapse, re-infection and mixed infections in tuberculosis disease. *Pathog Dis.* (2017) 75. doi: 10.1093/femspd/ftx020
172. Unis G, Ribeiro AW, Esteves LS, Spies FS, Picon PD, Dalla Costa ER, et al. Tuberculosis recurrence in a high incidence setting for HIV and tuberculosis in Brazil. *BMC Infect Dis.* (2014) 14:548. doi: 10.1186/s12879-014-0548-6
173. Velayutham B, Chadha VK, Singla N, Narang P, Gangadhar Rao V, Nair S, et al. Recurrence of tuberculosis among newly diagnosed sputum positive pulmonary tuberculosis patients treated under the Revised National Tuberculosis Control Programme, India: a multi-centric prospective study. *PLoS ONE.* (2018) 13:e0200150. doi: 10.1371/journal.pone.0200150
174. Zong Z, Huo F, Shi J, Jing W, Ma Y, Liang Q, et al. Relapse versus reinfection of recurrent tuberculosis patients in a national tuberculosis specialized hospital in Beijing, China. *Front Microbiol.* (2018) 9:1858. doi: 10.3389/fmicb.2018.01858
175. Sobkowiak B, Glynn JR, Houben R, Mallard K, Phelan JE, Guerra-Assuncao JA, et al. Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data. *BMC Genom.* (2018) 19:613. doi: 10.1186/s12864-018-4988-z
176. WHO. *Global Antimicrobial Resistance Surveillance System (GLASS).* (2020). Available online at: https://www.who.int/glass/en/ (accessed May 4, 2021).
177. Fonseca JD, Knight GM, McHugh TD. The complex evolution of antibiotic resistance in *Mycobacterium tuberculosis*. *Int J Infect Dis.* (2015) 32:94–100. doi: 10.1016/j.ijid.2015.01.014
178. Chisompola NK, Streicher EM, Muchemwa CMK, Warren RM, Sampson SL. Molecular epidemiology of drug resistant *Mycobacterium tuberculosis* in Africa: a systematic review. *BMC Infect Dis.* (2020) 20:344. doi: 10.1186/s12879-020-05031-5
179. Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, et al. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. *JAMA.* (2005) 293:2726–31. doi: 10.1001/jama.293.22.2726
180. WHO. *Global Tuberculosis Report.* (2014). Available online at: http://www.who.int/tb/publications/global_report/en/ (accessed June 6, 2015).
181. Perdigão J, Silva C, Maltez F, Machado D, Miranda A, Couto I, et al. Emergence of multidrug-resistant *Mycobacterium tuberculosis* of the Beijing lineage in Portugal and Guinea-Bissau: a snapshot of moving clones by whole-genome sequencing. *Emerg Microbes Infect.* (2020) 9:1342–53. doi: 10.1080/22221751.2020.1774425
182. Hanekom M, van der Spuy GD, Streicher E, Ndabambi SL, McEvoy CRE, Kidd M, et al. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. *J Clin Microbiol.* (2007) 45:1483–90. doi: 10.1128/JCM.02191-06
183. San LL, Aye KS, Oo NAT, Shwe MM, Fukushima Y, Gordon SV, et al. Insight into multidrug-resistant Beijing genotype *Mycobacterium tuberculosis* isolates in Myanmar. *Int J Infect Dis.* (2018) 76:109–19. doi: 10.1016/j.ijid.2018.06.009
184. WHO. *Global Tuberculosis Report*. Geneva: World Health Organization (2012).
185. Collins CH, Yates MD, Grange JM. Subdivision of *Mycobacterium tuberculosis* into five variants for epidemiological purposes: methods and nomenclature. *J Hygiene.* (1982) 89:235–42. doi: 10.1017/S0022172400070765
186. Niemann S, Kubica T, Bange FC, Adjei O, Browne EN, Chinbuah MA, et al. The species *Mycobacterium africanum* in the light of new molecular markers. *J Clin Microbiol.* (2004) 42:3958. doi: 10.1128/JCM.42.9.3958-3962.2004
187. Ngoi JM, Quashie PK, Morang'a CM, Bonney JH, Amuzu DS, Kumordjie S, et al. Genomic analysis of SARS-CoV-2 reveals local viral evolution

in Ghana. *Exp Biol Med.* (2021) 246:960–70. doi: 10.1177/15353702209 75351
188. NIHR Global Health Research Unit on Genomic Surveillance of AMR. Whole-genome sequencing as part of national and international surveillance programmes for antimicrobial resistance: a roadmap. *BMJ Glob Health.* (2020) 5:e002244. doi: 10.1136/bmjgh-2019-00 2244
189. Prestinaci F, Pezzotti P, Pantosti A. Antimicrobial resistance: a global multifaceted phenomenon. *Pathog Glob Health.* (2015) 109:309–18. doi: 10.1179/2047773215Y.0000000030
190. Tobin DM. Host-directed therapies for tuberculosis. *Cold Spring Harb Perspect Med.* (2015) 5:a021196. doi: 10.1101/cshperspect. a021196
191. Zumla A, Rao M, Wallis RS, Kaufmann SH, Rustomjee R, Mwaba P, et al. Host-directed therapies for infectious diseases: current status, recent progress, and future prospects. *Lancet Infect Dis.* (2016) 16:e47–63. doi: 10.1016/S1473-3099(16) 00078-5

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 29 July 2021
doi: 10.3389/fmicb.2021.704152

Global Genomic Characterization of *Salmonella enterica* Serovar Telelkebir

Yu-feng Qiu[1,2†], Reshma B. Nambiar[3†], Xue-bin Xu[4†], Shun-tai Weng[1,2], Hang Pan[3], Kui-cheng Zheng[1,2,5] and Min Yue[3,6,7,8*]*

[1] Department of Bacterialogy, Fujian Provincial Center for Disease Control & Prevention, Fuzhou, China, [2] Department of Bacterialogy, Fujian Provincial Key Laboratory of Zoonosis Research, Fuzhou, China, [3] Department of Veterinary Medicine & Institute of Preventive Veterinary Science, Zhejiang University College of Animal Sciences, Hangzhou, China, [4] Department of Microbiology, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China, [5] School of Public Health, Fujian Medical University, Fuzhou, China, [6] Zhejiang Provincial Key Laboratory of Preventive Veterinary Medicine, Hangzhou, China, [7] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, National Medical Center for Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China, [8] Hainan Institute of Zhejiang University, Sanya, China

OPEN ACCESS

Edited by:
Michael Kemp,
University of Southern Denmark, Denmark

Reviewed by:
Sandeep Tamber,
Health Canada, Canada
Valentine Usongo,
Health Canada, Canada

***Correspondence:**
Min Yue
myue@zju.edu.cn
Kui-cheng Zheng
kingdadi9909@126.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 01 May 2021
Accepted: 25 June 2021
Published: 29 July 2021

Citation:
Qiu Y-f, Nambiar RB, Xu X-b, Weng S-t, Pan H, Zheng K-c and Yue M (2021) Global Genomic Characterization of Salmonella enterica Serovar Telelkebir. Front. Microbiol. 12:704152. doi: 10.3389/fmicb.2021.704152

Non-typhoidal *Salmonella* (NTS) is a common cause for self-limiting gastroenteritis, representing a public health concern globally. NTS is one of the leading causes of foodborne illnesses in China; however, the invasive infection caused by NTS is largely underappreciated. Here, we reported an NTS invasive infection caused by an infrequently reported serovar Telelkebir (13,23:d:e,n,z15) strain FJ001 in China, which carries antimicrobial-resistant genes [*fosA7* and *aac(6′)-Iaa*] and typhoid-toxin genes (*cdtB*, *pltA*, and *pltB*). By conducting the whole genomic sequencing, we also investigated the relatedness of this strain with an additional 120 global contextual *Salmonella enterica* serovar Telelkebir (*S.* Telelkebir) isolates, and assessed the antimicrobial-resistant determinants and key virulence factors using the available genomic dataset. Notably, all 121 (100%) of the *S.* Telelkebir strains possessed the typhoid toxin genes *cdtB*, *pltA*, and *pltB*, and 58.67% (71/121) of *S*. Telelkebir harbored antimicrobial-resistant gene *fosaA7*. The study by core genome multilocus sequence typing (cgMLST) and core single-nucleotide polymorphism (SNP)-based phylogenomic analysis demonstrated that the *S*. Telelkebir isolates from different sources and locations clustered together. This suggests that regular international travels might increase the likelihood of rapid and extensive transmissions of potentially pathogenic bacteria. For the first time, our study revealed the antimicrobial resistance, virulence patterns, and genetic diversity of the serovar *S.* Telelkebir isolate in humans and similar isolates over the world. The present study also suggests that genomic investigation can facilitate surveillance and could offer added knowledge of a previously unknown threat with the unique combination of virulent and antimicrobial-resistant determinants.

Keywords: *Salmonella* Telelkebir, invasive infection, antimicrobial resistance gene, typhoid toxin, virulence gene

INTRODUCTION

Salmonella enterica is a major global foodborne pathogen (Pan et al., 2018; World Health Organization (WHO), 2018; Jiang et al., 2021). *S. enterica* is divided into six distinct subspecies: *enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae*, and *indica* (Chand et al., 2020). The *S. enterica* subsp. *enterica* consists of more than 1,500 serotypes. A small number of *Salmonella* serovars (*S.* Typhi and *S.* Paratyphi A, B, or C) are human restricted and evoke an invasive, life-threatening systemic disease (Parry et al., 2002; Liu et al., 2021). In contrast, the non-typhoidal *Salmonella* (NTS) serovars generally cause self-limiting diarrheal illnesses with low case mortality. According to the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD 2017 Causes of Death Collaborators, 2018), it was estimated that NTS accounts for 95 million cases of enterocolitis and 50,771 related deaths.

In sub-Saharan Africa, NTS is the most common cause of bloodstream infection in immunocompromised adults and children with a fatality rate of 20–25% (Feasey et al., 2012). In developed countries, 5% of NTS cases are invasive, focal systemic infections, or extra-intestinal disease leading to bacteremia (Suez et al., 2013). The NTS infections in humans arise through the food chain, but infection can also be contracted through contact with the infected animals, person-to-person transmission, or contaminated water (World Health Organization (WHO), 2018). Various wildlife animals, reptiles, and their contaminated environment could act as a reservoir for various rare and infrequently reported serovars that are normally dissimilar from those strains isolated from the commercial food chain. Currently, most of the research is concentrated on *S.* Typhimurium and *S.* Enteritidis, and comparatively lesser consideration has been given to uncommon emerging serovars (Tankson et al., 2006; Mansour et al., 2020). There is an evolving need to investigate the incidence of these uncommon *S. enterica* serovars as a variety of these have been linked to numerous foodborne outbreaks, acute gastroenteritis, splenic abscesses, etc. (Rodriguez et al., 2006; Berendes et al., 2007; Li et al., 2017), which are largely underappreciated. In China, most of the foodborne bacterial outbreaks (70–80%) are ascribed to *Salmonella* infections (Wang et al., 2019; Paudyal et al., 2018; Li et al., 2020; Zhou et al., 2020). However, the study on invasive NTS infection remains incomplete in China.

Recently, various reports have indicated the efficiency of whole-genome sequencing (WGS) in the epidemiological investigation of transmittable disease at different geographical locations (Gardy and Loman, 2018). The WGS methods involve either the characterization of core single-nucleotide polymorphisms (SNP) or core-genome multilocus sequence typing (cgMLST) analysis (Mellmann et al., 2016). The cgMLST-based approach involves the direct assessment and comparison of newly determined genotypes with historically available data. However, the SNP-based analysis requires recalculation once there is a change in the data set unless a preliminarily established reference genome is provided (Ruan et al., 2020).

In this study, we have investigated a bloodstream infection case caused by *Salmonella* Telelkebir in China. To improve our knowledge of the genomic epidemiology of *S.* Telelkebir, phylogenetic relationship and comparative genomic analysis of additional 120 global *S.* Telelkebir isolates from various countries were also carried out.

MATERIALS AND METHODS

Ethics Statements

The FJ001 strain was collected and approved by the Shanghai and Fujian Center for Disease Control and Prevention. Written consent was obtained prior to the study for collecting the samples for surveillance purposes.

Bacterial Isolates

The publicly available isolates with their metadata were collected from Enterobase[1] or the sequence read archive[2]. The *S.* Telelkebir isolates were used in this study to embody various geographical locations, including the UK (n = 61), United States (n = 23), Germany (n = 18), Ireland (n = 4), France (n = 4), China (n = 1), Mali (n = 1), Netherlands (n = 1), and Turkey (n = 1). The locations of the remaining strains were not available (NA) (**Figure 1**).

Clinical Study

A patient with no significant medical history was admitted to a tertiary hospital for cardiovascular medicine in Fujian, with fever, abdominal pain, diarrhea, and syncope on December 12, 2018. The patient had no history of travel or contact with any wildlife or domestic animals. The patient had a history of eating outside (roasted meat, the type of meat was unclear) 1 week prior to the onset of the disease. However, other food sources may have been the vehicle of infection. During admission, the patient had a body temperature of 36.6°C, and the blood pressure was 124/60 mmHg. Blood tests showed a normal white blood count (9.4×10^9/L) and a monocyte count (1.12×10^9/L). The stool sample was normal and was negative for fungi and parasites *via* microscopic examination. The stool sample was negative for *Salmonella* and *Shigella*. On the second day of hospital admission, the patient developed diarrhea, abdominal pain, and fever. The injection treatments with Sulperazon (cefoperazone sulbactam) and heparin were ineffective. During the treatment, the patient had intermittent fever and chills. On December 16, a combination of treatments with Sulperazon (cefoperazone sulbactam) and Cravit (levofloxacin) injection also failed. On December 17, the bacterial blood culture indicated the presence of *Salmonella* Group G, and the patient was diagnosed for sepsis and multiple organ dysfunction syndromes. On December 18, the patient was transferred to the intensive care unit with combined treatment of imipenem–cilastatin and vancomycin, and the body temperature of the patient and chills decreased. After a successive 4-day treatment, the blood culture was negative for *Salmonella*. The patient completely recovered on December 26 and was discharged from the hospital (**Figure 2**).

[1] http://enterobase.warwick.ac.uk

[2] https://www.ncbi.nlm.nih.gov/sra

FIGURE 1 | Source location of *S.* Telelkebir strain genomes used in this study. The red color indicates the presence of this serovar. This map was created using an online service (https://mapchart.net/).

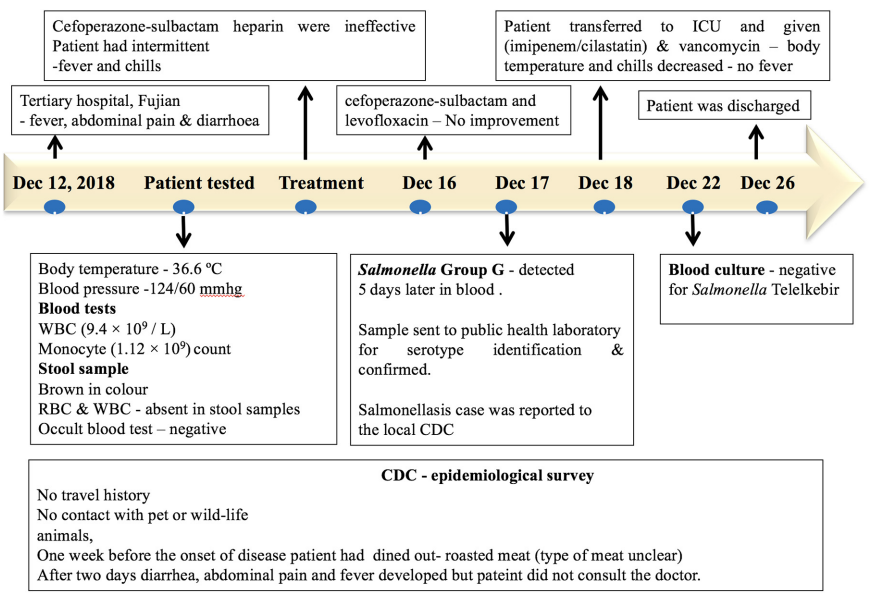


FIGURE 2 | Timeline of the salmonellosis diagnosis, treatment, and investigation since December 12, 2018.

Characterization of *Salmonella* Telelkebir

For screening *Salmonella* and *Shigella*, blood and stool samples were spread onto xylose lysine deoxycholate agar (XLD) and MacConkey agar, respectively, and were incubated at 36°C for 24 h. The stool cultures were negative for *Salmonella*, while the blood culture displayed positive colonies on XLD after 24 h. The plate was purple-red in color, and colonies with the typical characteristic of *Salmonella*, i.e., round, moist, smooth, colorless, translucent, and black in the center were selected and inoculated in triple sugar iron (TSI) plate for 24 h at 36°C. The *Salmonella* isolate was further subjected to biochemical analysis (VITEK2 COMPACT; bioMérieux, France) and identified using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). The serotype agglutination

assay was conducted using the anti-serum purchased from Denmark (SSI Diagnostica, Denmark), and was confirmed as a newly reported serovar (13,23:d:e,n,z15 or Telelkebir) in mainland China.

Antimicrobial Susceptibility Testing

The bacteria isolate was used for antimicrobial susceptibility testing (AST) against 17 antimicrobial agents. Broth microdilution minimum inhibitory concentration (MIC) determination was performed according to the Clinical and Laboratory Standards Institute (CLSI) guidelines and interpretation (CLSI, 2018), the European Committee for Antimicrobial Susceptibility Testing (EUCAST, 2018), and where no EUCAST or CLSI interpretative criteria were available, breakpoints were harmonized with those of the National Antimicrobial Resistance Monitoring System (NARMS), United States (FDA, 2017). The following antimicrobials were tested: fosfomycin, levofloxacin, cefoperazone sulbactam, streptomycin, ampicillin, amoxicillin–clavulanic acid, cefoxitin, imipenem, nalidixic acid, ciprofloxacin, chloramphenicol, tetracycline, kanamycin, gentamicin, trimethoprim–sulfamethoxazole, ceftiofur, and azithromycin (Sangon Biotech, China). *Escherichia coli* ATCC 25922 were used as a control strain. The AST against all antibiotics were carried out in Muller–Hinton broth (MHB) medium in both aerobic and anaerobic conditions and incubated at 37°C. The AST was also performed in Dulbecco's modified Eagle's medium (DMEM) (Dulbecco and Freeman, 1959) and incubated in a 5% CO_2 incubator.

DNA Extraction, Genomic Sequencing, and Data Analysis

The genomic DNA of FJ001 isolate was extracted from overnight cultures grown at 37°C in Luria–Bertani broth under 180 rpm shaking conditions by using a TIANamp bacteria DNA kit (Tiangen Biotech, China) according to the manufacturer's protocol. The Qubit Broad Range assay kit (Invitrogen, United States) was used for quantification. The Genomic DNA library was constructed using Nextera XT DNA library construction kit (No. FC-131-1024; Illumina, United States). High-throughput genome sequencing was accomplished by the Illumina NovaSeq 6000 platform, using paired-end sequencing of 2 × 150-bp reads as previously described (Biswas et al., 2019; Paudyal et al., 2019; Yu et al., 2020). For the comparative genomic analysis, an additional 119 isolates (**Supplementary Table 1**), with assembled contigs in FASTA format, were obtained from the Enterobase (see text footnote 1). One additional strain, with raw reads in SRA format was obtained from the NCBI SRA dataset (see text footnote 2). The quality of sequencing was checked with FastQC toolkit (Rouzeau-Szynalski et al., 2019). The raw reads of the FJ001 genome were trimmed with trimmomatic software prior to genome assembly. The raw reads for each strain were assembled by using SPAdes 4.0.1 (Bolger et al., 2014). QUAST (Gurevich et al., 2013) was used to assess the assembled genomes through basic statistics generation, including the total number of contig, the length of contig, and N50 (**Supplementary Table 2**). Prokka v.1.14 with default settings under the in-house galaxy platform and Rapid Annotation Subsystem Technology (RAST) server[3] were used for annotation of the assembled genomes. The NCBI Basic Local Alignment Search Tool (BLAST) BLASTp[4] program was used for similarity alignment. The plasmid types and antimicrobial-resistant genes were determined using the PlasmidFinder 2.0[5] and ResFinder 3.2,[6] respectively (Zankari et al., 2012). The virulence factors in the genome were examined using the Virulence Factors Database (VFDB) (Liu et al., 2019). *Salmonella* Telelkebir *in silico* serotyping was conducted by the *Salmonella In Silico* Typing Resource (SISTR) platform[7] and SeqSero2.[8] The multilocus sequence typing of the isolates were carried out using MLST.[9] The contigs were used for variant calling against reference strain 98-12414 and outgroup control strain Poona ATCC® BAA-1673 by software Snippy v4.4.4 to obtain core single-nucleotide polymorphism (SNP) for determining the population structure of 121 available *S.* Telelkebir isolates. After being filtered by 95% gap parameter to get the core SNPs, a total of 85,694 SNPs were used to build a maximum-likelihood phylogenetic tree with 1,000 bootstraps using IQ-TREE v.1.6.12 with the best model TVM + F + ASC (Letunic and Bork, 2007).

Core Genome Multilocus Sequence Typing Analysis of *Salmonella* Telelkebir Isolates

The cgMLST analysis was carried out using the Ridom SeqSphere+ software v6.0.2 (Ridom). An *ad hoc* core genome MLST (cgMLST) scheme was created for the gene-by-gene analysis with SeqSphere+ (Ridom® GmbH, Münster, Germany). Hence, the *S.* Typhimurium LT2 (NC_003197.1) genome comprising 4,451 genes was used as annotated reference. The cgMLST target definer tool was applied to 121 *Salmonella* Telelkebir genomes with the default settings of the software to define the core genome loci (Simon et al., 2018). A cgMLST tree was built using the neighbor-joining method. The cgMLST distance matrix showing pairwise comparison of allelic differences between 121 isolates is given in **Supplementary Table 3**.

Data Availability

The genome for the strain FJ001 was deposited in the NCBI (BioProject PRJNA666303). Associated metadata and virulence genes can be found in **Supplementary Table 1**.

[3]http://rast.nmpdr.org/

[4]https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

[5]https://cge.cbs.dtu.dk/services/PlasmidFinder/

[6]https://cge.cbs.dtu.dk/services/ResFinder/

[7]https://lfz.corefacility.ca/sistr-app/

[8]https://cge.cbs.dtu.dk/services/SeqSero/

[9]https://cge.cbs.dtu.dk/services/MLST/

RESULTS

Salmonella Telelkebir Clinical Isolate FJ001

In this study, we reported a bloodstream infection caused by *Salmonella* Telelkebir (FJ001) isolate in mainland China. The *Salmonella* isolate was serotyped by agglutination assay at both the Fujian and Shanghai Center for Disease Prevention and Control, and two independent groups confirmed this causative isolate as an uncommon serovar *S. enterica* subsp. *enterica* Telelkebir (13,23:d:e,n,z15). Here, a retrospective investigation confirmed that the patient had no history of travel and pet contact for 1 month prior to the onset of the disease. Exposure factors are mainly related to the history of unclean diet, i.e., roasted meat, but this cannot make accurate traceability judgments for the source of infection. Also, other food source may have been the vehicle of infection. However, Fujian province, located in the sub-tropic region of Eastern China, has abundant natural species resources, local processing, and consumption habits of wild and farmed snakes. Therefore, we cannot rule out the possibility of infection sources as snakes and other reptiles (Yu-feng et al., 2020), for this particular case.

Antimicrobial Resistance Profile for FJ001

The MIC analysis showed that the strain FJ001 was susceptible against all examined antimicrobial agents in both the tested medium under the tested conditions (**Supplementary Table 4**). It is interesting to note that earlier treatment with cefoperazone sulbactam alone or combined with levofloxacin failed to inhibit the bacterial replication in the patient.

Antimicrobial Resistance Genes and Plasmids in *Salmonella* Telelkebir Population

WGS analysis indicated that the FJ001 strain carries two antimicrobial-resistant genes, *fosA7* and *aac(6′)-Iaa*. To understand the global resistance patterns, plasmid profile, and virulence patterns of the *S.* Telelkebir strains, an additional 120 strains were retrieved from Enterobase and NCBI. The WGS analysis revealed that the global strains contains resistance genes against eight antibiotic families: aminoglycosides including *aac(3)-Vla* (1/121), *aac(6′)-Iaa_1* (121/121), *ant(3″)-Ia_1* (1/121), *aph(3)-Ib_5* (2/121), and *aph(6)-Id_1* (3/121); beta-lactamases, including $bla_{\mathrm{TEM-1B}}$ (3/121), $bla_{\mathrm{CTX-M-15}}$ (1/121), and $bla_{\mathrm{OXA-1}}$ (1/121); phenicols, including *catB3_1* (1/121) and *catA1_1* (1/121); trimethoprims including *dfrA14_5* (1/121); quinolones including *qnrB19_1* (10/121), *qnrB1_1* (2/121), and *qnrS1_1* (1/121); sulfonamides, including *sul1_5* (12/121) and *sul2_2* (2/121); tetracyclines, including *tet(A)* (3/121), and fosfomycins including *fosA7_1* (71/121) (**Figure 3**). All the *S.* Telelkebir genomes harbored at least one antimicrobial-resistant gene *aac(6′)-Iaa_1* (**Supplementary Table 1**). A total of 58.67% (71/121) isolates harbored resistance genes for fosfomycin, and 11.57% (14/121) isolates possessed resistance gene for quinolones. Three isolates recovered from human infections from different countries harbored antimicrobial resistance gene $bla_{\mathrm{TEM-1B}}$. Two human isolates from Mali (07-1331) and UK (56980) harbored resistance genes to more than three antibiotic families (**Figure 3**), which indicates that these strains could be multiple drug-resistant (Elbediwi et al., 2019; Mansour et al., 2020). A total of 35.2% isolates harbored plasmids, mainly IncI1α, IncFII, Col440I, IncHI2, IncN, and IncI2 (Delta) (**Figure 3**).

Virulence Genes in *Salmonella* Telelkebir Population

The virulence gene profile for the 121 strains was carried out using the virulence factor database (VFDB) (**Supplementary Table 1**). The FJ001 strain harbored cytolethal distending toxin (*cdtB*) gene and carried fimbrial adherence genes and secretion system genes. A total of 55.81% of the genes (72/129) were conserved among all strains. Twenty-one genes were not present in any of the strains, and the rest of the genes were variable. All the strains carried fimbrial adherence genes *fim*, *csg*, and *inv*, the secretion system genes, *sopA*, *sopB*, *sopE2*, *spaOPQRS*, *spiC, sscAB, sifA,* and *pipB,* and the operons *sse* and *ssa*. The secretion system genes *pipB2, sopD2*, and *slrP*, and the effector genes, *sseK1, and sseK2*, and bacteriophage-related genes *sodCI, sspH2*, and *sspH1* genes were variable among the strains. Importantly, all strains (100%) were positive for the cytolethal distending toxin (CDT) genes (*cdtB, pltA*, and *pltB*). The *cdtB* gene is considered as one of the typhoid toxins of *Salmonella* Typhi, which causes cell arrest due to DNA damage (Chang et al., 2019). The BLASTp analysis of the proteins CdtB, PltA, and PltB of *Salmonella* Telelkebir showed that all the proteins were highly similar (100–97%) to the CdtB, PltA, and PltB proteins of the available *S. enterica* strains in NCBI.

Phylogenetic Analysis of *Salmonella* Telelkebir Population

To determine the phylogenetic relationship, Snippy v.4.4.4 was used to obtain SNP alignment. A total of 18 ST types were identified in 121 isolates from 10 countries across four continents (**Figure 3** and **Supplementary Table 1**). The present study showed that ST450 is the predominant sequence type (33%, 40/121), followed by ST2222 (19%, 23/121), and ST2651 (9%, 11/121) (**Figure 3**). The FJ001 strain belongs to ST5494, and it clustered with a human isolate isolated from France. The R17.4518 obtained from Taiwan, China, had the ST2651 and clustered with the UK isolates. Interestingly, the two Chinese strains did not cluster together. Among the ST450 isolates, 29 isolates (72.5%) were obtained from humans; four isolates were recovered from animals (10%), whereas the source of the remaining seven isolates could not be obtained. The ST2222 and ST2651 strains were mainly isolated from humans (23 and 11 strains, respectively). The ST450 and ST2249 were associated with strains of animal origin. The phylogenomic analysis reveals that the closely

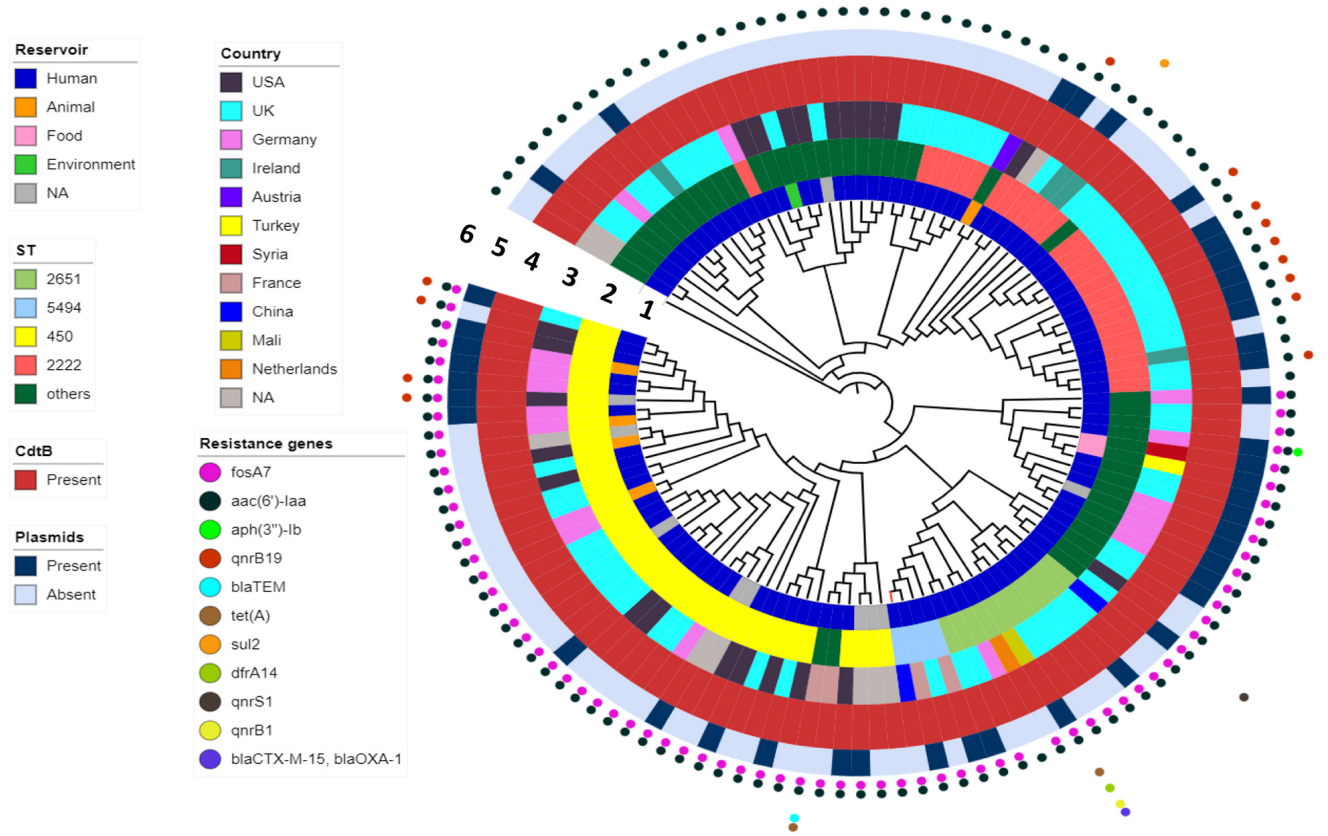


FIGURE 3 | The phylogenomic relationship among 121 *Salmonella* Telelkebir strains. The tree has been rooted using the serovar Poona ATCC® BAA-1673 as the outgroup. The numbers 1–6 represents the inner to outer rings corresponding to reservoir, ST, country, *cdtB*, plasmids, and resistance genes, respectively. The strain FJ001 containing leaf node is represented in red. The following plasmids were present in the *Telelkebir* strains: IncI1α, IncFII, Col440I, IncHI2, IncN, and IncI2 (Delta).

related strains are from very diverse geographical origins. In the ST profiles, the UK region presented a substantial higher diversity among the examined isolates (**Figure 3** and **Supplementary Table 1**). ST2155 was associated with the food products (baklava from Turkey).

Core Genome Multilocus Sequence Typing Analysis of *Salmonella* Telelkebir Population

The cgMLST based approach was employed to evaluate the genomic epidemiological features of the global *S.* Telelkebir strains. The cgMLST study was performed on a collection of loci that were shared by all *Salmonella* isolates, which were then used for gene by gene comparison. The difference in the nucleotide sequences of these loci determines the clustering of isolates (Tang et al., 2017; Pearce et al., 2018). According to cgMLST results, these samples were diverse as depicted by the different number of alleles between the isolates (**Figure 4**). Nineteen clusters and 68 singletons were identified in the global phylogenetic tree. The FJ001 strain did not form any cluster. The nearest neighbor for the FJ001 strain was 201702574, which were obtained from France with an allelic difference of 53 alleles. The clusters were perfectly coherent with the STs assignment. The ST450 group was subdivided into seven different clusters, with ST450-cluster (C1) being the dominant type. The ST450 strains were isolated from various geographical locations and different years and the reservoirs for these strains where humans (n = 4) and animals (n = 4) (**Supplementary Table 1**), whereas the cluster C1 mainly included the strains recovered from humans and animals that had an allelic difference of one to six alleles. Two *S.* Telelkebir isolates from China had resistant genes [*fosA7* and *aac(6′)-Iaa*] but belonged to different types both by MLST and cgMLST. The results suggested that the antimicrobial resistance gene profile varied according to the cgMLST clusters (**Supplementary Table 1**). The fosfomycin resistance genes were observed in C1, C4, C5, C6, C7, C13, C14, C15, C16, C18, and C19 clusters. The *qnrB19* resistance gene was present among clusters C1 (28.57%, 2/7), C8 (100%, 3/3), and C10 (100%, 2/2) and it was mainly associated with plasmid Col440I.

DISCUSSION

There are earlier reports on the isolation of uncommon *Salmonella* Telelkebir serovar (Tankson et al., 2006; Octavia et al., 2019). However, little is known about its resistance

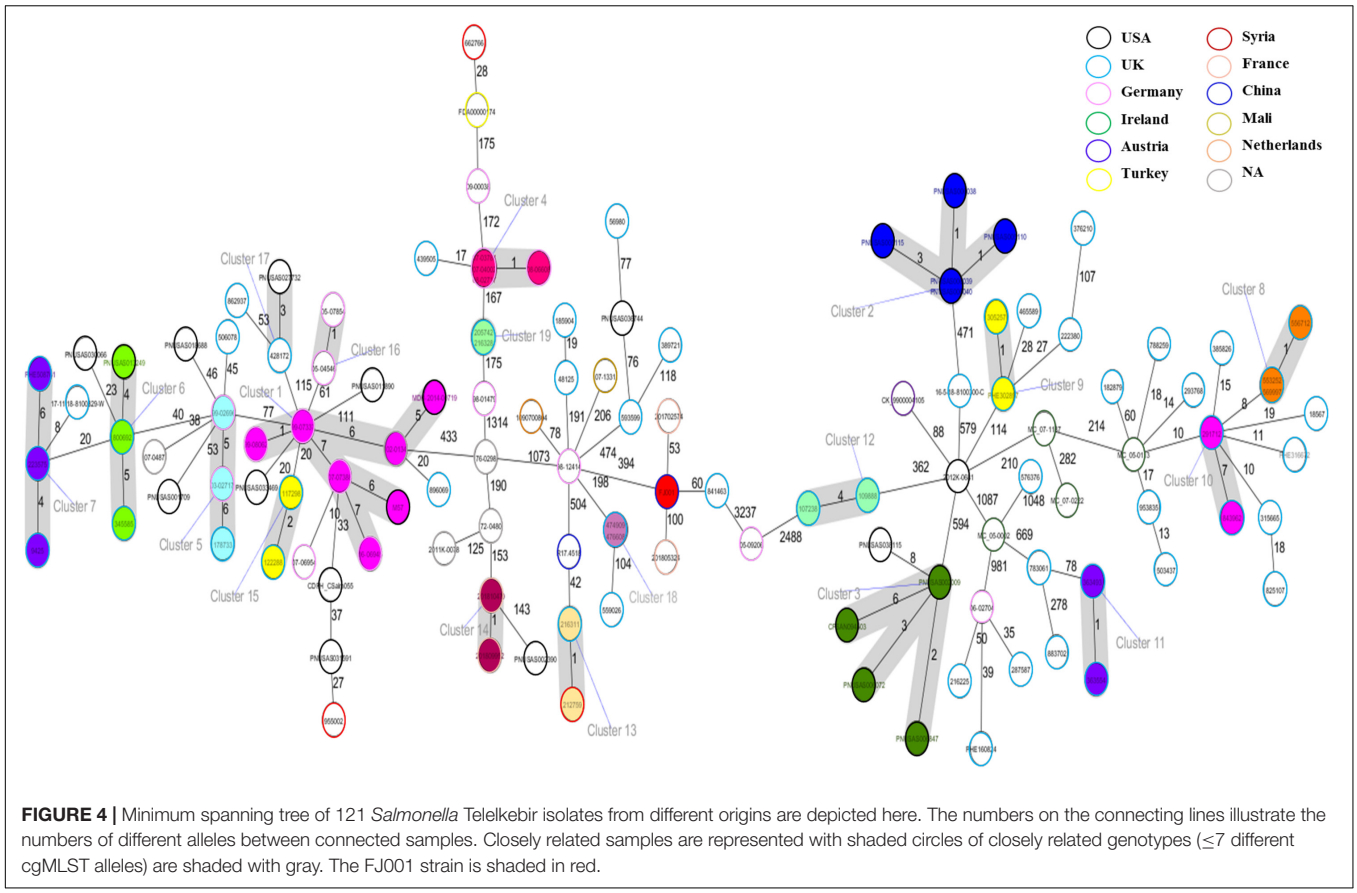


FIGURE 4 | Minimum spanning tree of 121 *Salmonella* Telelkebir isolates from different origins are depicted here. The numbers on the connecting lines illustrate the numbers of different alleles between connected samples. Closely related samples are represented with shaded circles of closely related genotypes (≤7 different cgMLST alleles) are shaded with gray. The FJ001 strain is shaded in red.

profile, epidemiology, and disease-causing potential. Previous reports suggested that *Salmonella* serotype Telelkebir is an infrequently reported variant that is mainly associated with exotic animal species, mostly reptiles (Berendes et al., 2007). Between 1995 and 2007, *S.* Telelkebir strains were obtained from animal-feed ingredients in Poland (Dera-Tomaszewska, 2012). Berendes et al. (2007) reported a case study where a 17-year-old girl had sepsis splenic abscesses caused by *Salmonella* serovar Telelkebir. The same variant was cultivated from the feces of the reptile pets that were held in the home of the patient. Our study demonstrates that the four global animal isolates had between one and six cgMLST allelic differences to the closest human isolates. In a previous study, three 2017 Bavarian *S.* Agona feed-origin outbreak strains were compared with a French outbreak isolate and 48 *S.* Agona isolates collected from 1993 to 2018, out of which 28 were epidemiologically outbreak related. The cgMLST analysis revealed that four most relevant clusters comprised 3 to 15 samples with a maximum within-cluster difference of zero to five alleles (Dangel et al., 2019). The study revealed that cgMLST can be used for reasonable, reproducible, and reliable high-resolution classification to monitor outbreak clusters and relationships among past or international cases, which could also be interpreted using representative public data.

Even though the FJ001 strain was pan-susceptible toward the tested antimicrobial agents, treatment with cefoperazone sulbactam alone or combined with levofloxacin failed when administrated to the patient. Interestingly, the WGS analysis revealed that the strain FJ001 had no antimicrobial resistance genes against quinolones and β-lactam class of antibiotics. One of the possible reason could be that for *Salmonella* spp. and *Shigella* spp., aminoglycosides, first- and second-generation cephalosporins and cephamycins may appear active *in vitro,* but are not effective clinically (CLSI, 2016). Also, according to "the 90–60 rule" coined by Rex and Pfaller (2002), a susceptible result in *in vitro* is associated with a favorable therapeutic response in 90–95% of patients. Various host or pathogen factors influence the efficiency of antibiotic therapy *in vivo*, such as the host immune system, site of infection (penetration of antimicrobial agents into the site), and bacterial virulence factors that may increase or hinder the immune response leading to poor clinical outcomes (Stratton, 2006; Ersoy et al., 2017). This could be the possible explanation of the ineffectiveness of these drugs in the patient.

The *fosA7* is a new antimicrobial resistance gene against fosfomycin that was recently identified in *S.* Heidelberg from broiler chickens in Canada (Rehman et al., 2017). The gene *fosA7* confers resistance to broad-spectrum antibiotic fosfomycin, which is extensively used to treat drug-resistant Gram-negative bacteria (Balbin et al., 2020). A total of 58.67% of the tested *S.* Telelkebir strains (71 strains) harbored the *fosA7* gene. The fluoroquinolones are the drugs of choice

for the treatment of iNTS due to their broad-spectrum antimicrobial activity. Nevertheless, the extensive use of these drugs has led to the appearance of resistant strains globally, mainly in Gram-negative bacterial species. The *qnrB19* gene is one the most frequent variants of *qnr* genes globally, and out of the 121 strains, 12 isolates are positive for *qnrB19* gene (**Supplementary Table 1**; Ciesielczuk et al., 2013). The *qnr* is often found in association with genes that impart resistance to other antibiotics classes, e.g., β-lactams and aminoglycosides (Robicsek et al., 2006). In the present study, the *qnr* genes were mostly associated with aminoglycosides which suggests that *Enterobacteriaceae* strains harboring *qnr* resistance may denote a serious threat to public health. Also, the presence of $bla_{\mathrm{TEM-1B}}$ gene was observed in three human *S.* Telelkebir isolates from the United Kingdom, Ireland, and Mali. The broad-spectrum β-lactamase enzymes can hydrolyze almost all β-lactams and are commonly linked with genes conferring resistance to several other classes of antibiotics (Bush and Bradford, 2016).

The screening of the virulence gene profile demonstrated that all *S.* Telelkebir isolates harbored various fimbrial genes (*bcf*, *fim*, *inv*, and *csg*) and secretion systems involved in cell invasion and bacterial viability in phagocytes. Previous results suggest that fimbriae are involved in differential intestinal colonization of animal species (Weening et al., 2005; Yue et al., 2012). The cytolethal distending toxin (CDT) or typhoid toxin is a bacterial genotoxin, which are encoded by several Gram-negative bacteria, including *S. enterica*. Our comparative genome analysis, for the first time, revealed that FJ001 strain and all the *Salmonella* Telelkebir strains harbored the gene *cdtB, pltA*, and *pltB* typhoid toxin gene, which highlights particular concern in public health. Typhoid toxin is recognized as a major virulence factor of *S.* Typhi, and it is reported to play a central role in the pathogenicity of *S.* Typhi. It has been observed that typhoid toxin is involved in the establishment of *S.* Typhi persistent infection most likely by altering the immune cell functions to its favor (Song et al., 2010; Chong et al., 2017). The genes encoding *Salmonella*-CDT (i.e., genes *pltA*, *pltB*, and *cdtB*) have been characterized in around 40 NTS serovars (den Bakker et al., 2011; Mezal et al., 2014a). Previous reports suggest that the amino acid alignments of CdtB, PltA, and PltB are extremely conserved among *S. enterica* serotypes (Rodriguez-Rivera et al., 2015). The BLASTp analysis of all the proteins CdtB, PltA, and PltB of *Salmonella* Telelkebir strains showed high sequence similarity (100–97%) to *S. enterica* strains. This suggests that the CdtB, PltA, and PltB are highly conserved among *S. enterica* serotypes. Also, cellular-level *Salmonella* CDT significantly alters the outcome of infection by inducing DNA damage, which is associated with a cell cycle arrest and activation of the host cell's DNA damage response (Chang et al., 2019). Earlier findings revealed that *S.* Javiana isolates harboring *cdtB*, *pltA*, and *pltB* caused cytoplasmic and nuclear enlargement together with cell cycle arrest in G2/M phase, increased invasion and cytotoxicity toward HeLa cells that are characteristic of CDT activity (Mezal et al., 2014a). Recently, Xu et al. (2020) reported the presence of the *cdtB, pltA*, *and pltB* genes in iNTS serovar *S.* Uzaramo, which showed a higher killing rate than the classic gastrointestinal infectious agent *S.* Typhimurium and suggested it to be a possible explanation of the invasiveness of these serovars. Ahn et al. (2021) demonstrated that even when host cells infected with *S.* Typhi are treated with antibiotics, typhoid toxin is continuously secreted by antibiotic-resistant *S.* Typhi. These NTS serovars might have acquired these unique virulence factors through horizontal gene transfer during homologous recombination between *S.* Typhi and non-typhoidal serovars, by a prophage that integrates into the bacterial DNA chromosome (Bakker et al., 2011), and it might have led to a higher capacity of invasion in NTS (Mezal et al., 2013; Mezal et al., 2014b).

The SNP-based phylogenomic analysis also revealed that the strains clustered together irrespective of the time of isolation, source, and geographical location. This could be the result of global dissemination, due to the convenient and regular global travels, there is an increase in the possibility for fast and extensive transmissions of bacterial pathogens, which permits these isolates to effortlessly cross geographical barriers. Regardless of the improved surveillance followed today in response to the outbreak of infectious diseases, transmission connections cannot be identified in the event of an infectious disease outbreak due to missed screening, antimicrobial therapy suppression, and difficulties recognizing contacts (Jia et al., 2019).

CONCLUSION

In summary, we reported a bloodstream infection caused by an uncommon NTS serovar Telelkebir in mainland China. These infrequently reported *Salmonella* serovars can possibly be misidentified in clinics, and the actual threat possessed by these serovars may be underestimated. Although the real threat of human diseases caused by this infrequently reported *Salmonella* serovar remains unknown, the combination of unique virulence factors and antimicrobial resistance genes carried by these minority NTS may lead to adverse clinical outcomes, demonstrating the necessity for an enhanced surveillance for the clinically important typhoid-toxin-containing serovars.

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.ncbi.nlm.nih.gov/, PRJNA666303.

AUTHOR CONTRIBUTIONS

YQ, RN, XX, and MY conceptualized the study, performed the data curation, and formulated the methodology. YQ, RN, and XX were in charge of the resources and performed the investigation. YQ, RN, XX, SW, and HP performed the formal analysis and validation. MY was in charge of the software and visualization. KZ and MY supervised the study,

were in charge of project administration, and acquisition of the funding. RN wrote the original draft of the study. RN and MY reviewed and edited the final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Program on Key Research Project of China (2017YFC1600103 and 2019YFE0103900) as well as the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 861917–SAFFI, the Zhejiang Provincial Natural Science Foundation of China (LR19C180001), the Zhejiang Provincial Key R&D Program of China (2020C02032 and 2021C02008), the Opening Fund of Key Laboratory of Microorganism Technology and Bioinformatics Research of Zhejiang Province (2017E10010), Construction of Fujian Provincial Scientific and Technological Innovation Platform (2019Y2001), and National Major Science and Technology Program of China (2017ZX10103008).

ACKNOWLEDGMENTS

We thank the staff from the Fujian and Shanghai CDCs for providing technical assistance during the clinical investigation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.704152/full#supplementary-material

Supplementary Table 1 | A complete analytic information for 121 the examined *Salmonella* Telelkebir isolates in this study.

Supplementary Table 2 | A summary table of the genomic assembly information for 121 the examined *Salmonella* Telelkebir isolates.

Supplementary Table 3 | cgMLST distance matrix table with a pairwise comparison of allelic differences among each two of 121 *S*. Telelkebir isolates.

Supplementary Table 4 | Minimum inhibitory concentration values for *S*. Telelkebir strain FJ001 under various different conditions.

REFERENCES

Ahn, C., Yang, Y.-A., Neupane, D. P., Nguyen, T., Richards, A. F., Sim, J. H., et al. (2021). Mechanisms of typhoid toxin neutralization by antibodies targeting glycan receptor binding and nuclease subunits. *iScience* 24:102454. doi: 10.1016/j.isci.2021.102454

Bakker, H. C., Switt, A. I. M., Cummings, C. A., Hoelzer, K., Degoricija, L., Rodriguez-Rivera, L. D., et al. (2011). A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. enterica serovar Montevideo pulsed-field gel electrophoresis type. *Appl. Environ. Microbiol.* 77, 8648–8655. doi: 10.1128/AEM.06538-11

Balbin, M. M., Hull, D., Guest, C., Nichols, L., Dunn, R., and Thakur, S. (2020). Antimicrobial resistance and virulence factors profile of *Salmonella* spp. and *Escherichia coli* isolated from different environments exposed to anthropogenic activity. *J. Glob. Antimicrob. Resist.* 22, 578–583. doi: 10.1016/j.jgar.2020.05.016

Berendes, T. D., Keijman, J. M. G., te Velde, L. F., and Oostenbroek, R. J. (2007). Splenic abscesses caused by a reptile-associated *Salmonella* infection. *Dig. Surg.* 24, 397–399. doi: 10.1159/000107718

Biswas, S., Li, Y., Elbediwi, M., and Yue, M. (2019). Emergence and dissemination of mcr-carrying clinically relevant *Salmonella* Typhimurium Monophasic Clone ST34. *Microorganisms* 7:298. doi: 10.3390/microorganisms7090298

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

Bush, K., and Bradford, P. A. (2016). β-Lactams and β-Lactamase inhibitors: an overview. *Cold Spring Harb. Perspect. Med.* 6:a025247. doi: 10.1101/cshperspect.a025247

Chand, Y., Alam, M. A., and Singh, S. (2020). Pan-genomic analysis of the species *Salmonella enterica*: identification of core essential and putative essential genes. *Gene Rep.* 20:100669. doi: 10.1016/j.genrep.2020.100669

Chang, S.-J., Jin, S. C., Jiao, X., and Galán, J. E. (2019). Unique features in the intracellular transport of typhoid toxin revealed by a genome-wide screen. *PLoS Pathog.* 15:e1007704. doi: 10.1371/journal.ppat.1007704

Chong, A., Lee, S., Yang, Y.-A., and Song, J. (2017). The role of typhoid Toxin in *Salmonella* Typhi Virulence. *Yale J. Biol. Med.* 90, 283–290.

Ciesielczuk, H., Hornsey, M., Choi, V., Woodford, N., and Wareham, D. W. (2013). Development and evaluation of a multiplex PCR for eight plasmid-mediated quinolone-resistance determinants. *J. Med. Microbiol.* 62, 1823–1827. doi: 10.1099/jmm.0.064428-0

CLSI (2016). *Performance Standards for Antimicrobial Susceptibility Testing Twenty-Sixth Informational Supplement M100-S26*. Wayne, PA: CLSI.

CLSI (2018). *Performance Standards for Antimicrobial Susceptibility Testing*, 28th Edn. Wayne, PA: CLSI, 608–608.

Dangel, A., Berger, A., Messelhäußer, U., Konrad, R., Hörmansdorfer, S., Ackermann, N., et al. (2019). Genetic diversity and delineation of *Salmonella* Agona outbreak strains by next generation sequencing, Bavaria, Germany, 1993 to 2018. *Euro Surveill.* 24:1800303. doi: 10.2807/1560-7917.ES.2019.24.18.1800303

den Bakker, H. C., Moreno Switt, A. I., Govoni, G., Cummings, C. A., Ranieri, M. L., Degoricija, L., et al. (2011). Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*. *BMC Genomics* 12:425. doi: 10.1186/1471-2164-12-425

Dera-Tomaszewska, B. (2012). *Salmonella* serovars isolated for the first time in Poland, 1995-2007. *Int. J. Occup. Med. Environ. Health* 25, 294–303. doi: 10.2478/S13382-012-0038-2

Dulbecco, R., and Freeman, G. (1959). Plaque production by the polyoma virus. *Virology* 8, 396–397. doi: 10.1016/0042-6822(59)90043-1

Elbediwi, M., Li, Y., Paudyal, N., Pan, H., Li, X., Xie, S. et al. (2019). Global burden of colistin resistant bacteria: mobilized colistin resistant genes study 1980-2018. *Microorganisms* 7:461. doi: 10.3390/microorganisms7100461

Ersoy, S. C., Heithoff, D. M., Barnes, L., Tripp, G. K., House, J. K., Marth, J. D., et al. (2017). Correcting a fundamental flaw in the paradigm for antimicrobial susceptibility testing. *EBioMedicine* 20, 173–181. doi: 10.1016/j.ebiom.2017.05.026

EUCAST (2018). *EUCAST Breakpoint Tables for Interpretation of MICs and Zone Diameters, Version*, 8th Edn. Växjö: The European Committee on Antimicrobial Susceptibility Testing.

FDA (2017). *The National Antimicrobial Resistance Monitoring System: NARMS Integrated Report, 2015*. Laurel, MD: U.S. Department of Health and Human Services.

Feasey, N. A., Dougan, G., Kingsley, R. A., Heyderman, R. S., and Gordon, M. A. (2012). Invasive non-typhoidal *salmonella* disease: an emerging and neglected tropical disease in Africa. *Lancet* 379, 2489–2499. doi: 10.1016/S0140-6736(11)61752-2

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

GBD 2017 Causes of Death Collaborators (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and

territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 392, 1736–1788. doi: 10.1016/S0140-6736(18)32203-7

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Jia, H., Chen, Y., Wang, J., Xie, X., and Ruan, Z. (2019). Emerging challenges of whole-genome-sequencing–powered epidemiological surveillance of globally distributed clonal groups of bacterial infections, giving *Acinetobacter baumannii* ST195 as an example. *Int. J. Med. Microbiol.* 309:151339. doi: 10.1016/j.ijmm.2019.151339

Jiang, Z., Anwar, T., Peng, X., Biswas, S., Elbediwi, M., Li, Y., et al. (2021). Prevalence and antimicrobial resistance of *Salmonella* recovered from pig-borne food products in Henan, China. *Food Control* 121:107535. doi: 10.1016/j.foodcont.2020.107535

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi: 10.1093/bioinformatics/btl529

Li, S., Zhou, Y., and Miao, Z. (2017). Prevalence and antibiotic resistance of non-typhoidal *Salmonella* isolated from raw chicken carcasses of commercial broilers and spent Hens in Tai'an, China. *Front. Microbiol.* 8:2106. doi: 10.3389/fmicb.2017.02106

Li, Y., Yang, Q., Cao, C., Cui, S., Wu, Y., Yang, H., et al. (2020). Prevalence and characteristics of *Salmonella* isolates recovered from retail raw chickens in Shaanxi Province, China. *Poult. Sci.* 99, 6031–6044. doi: 10.1016/j.psj.2020.07.038

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

Liu, Y., Jiang, J., -Dra, A., Li, X., Peng, X., Xia, L., et al. (2021). Prevalence and genomic investigation of *Salmonella* isolates recovered from animal food-chain in Xinjiang, China. *Food Res. Int.* 142:110198. doi: 10.1016/j.foodres.2021.110198

Mansour, M. N., Yaghi, J., El Khoury, A., Felten, A., Mistou, M.-Y., Atoui, A., et al. (2020). Prediction of *Salmonella* serovars isolated from clinical and food matrices in Lebanon and genomic-based investigation focusing on Enteritidis serovar. *Int. J. Food Microbiol.* 333:108831. doi: 10.1016/j.ijfoodmicro.2020.108831

Mellmann, A., Bletz, S., Böking, T., Kipp, F., Becker, K., Schultes, A., et al. (2016). Real-time genome sequencing of resistant bacteria provides precision infection control in an institutional setting. *J. Clin. Microbiol.* 54, 2874–2881. doi: 10.1128/JCM.00790-16

Mezal, E. H., Bae, D., and Khan, A. A. (2014a). Detection and functionality of the CdtB, PltA, and PltB from *Salmonella enterica* serovar Javiana. *Pathog. Dis.* 72, 95–103. doi: 10.1111/2049-632X.12191

Mezal, E. H., Sabol, A., Khan, M. A., Ali, N., Stefanova, R., and Khan, A. A. (2014b). Isolation and molecular characterization of *Salmonella enterica* serovar Enteritidis from poultry house and clinical samples during 2010. *Food Microbiol.* 38, 67–74. doi: 10.1016/j.fm.2013.08.003

Mezal, E. H., Stefanova, R., and Khan, A. A. (2013). Isolation and molecular characterization of *Salmonella enterica* serovar Javiana from food, environmental and clinical samples. *Int. J. Food Microbiol.* 164, 113–118. doi: 10.1016/j.ijfoodmicro.2013.03.021

Octavia, S., Zulaina, S., Seet, S. K., Tien, W. S., Thu, M., Ooi, P. L., et al. (2019). Whole-genome sequencing of the rare *Salmonella enterica* serovar Anfo isolated from food handlers. *J. Med. Microbiol.* 68, 429–431. doi: 10.1099/jmm.0.000934

Pan, H., Paudyal, N., Li, X., Fang, W., and Yue, M. (2018). Multiple food-animal-borne route in transmission of antibiotic-resistant *Salmonella* newport to humans. *Front. Microbiol.* 9:23. doi: 10.3389/fmicb.2018.00023

Parry, C. M., Hien, T. T., Dougan, G., White, N. J., and Farrar, J. J. (2002). Typhoid fever. *N. Engl. J. Med.* 347, 1770–1782. doi: 10.1056/NEJMra020201

Paudyal, N., Pan, H., Elbediwi, M., Zhou, X., Peng, X., Li, X., et al. (2019). Characterization of *Salmonella* Dublin isolated from bovine and human hosts. *BMC Microbiol.* 19:226. doi: 10.1186/s12866-019-1598-0

Paudyal, N., Pan, H., Liao, X., Zhang, X., Li, X., Fang, W., et al. (2018). A meta-analysis of major foodborne pathogens in Chinese food commodities between 2006 and 2016. *Foodborne Pathog. Dis.* 15, 187–197. doi: 10.1089/fpd.2017.2417

Pearce, M. E., Alikhan, N.-F., Dallman, T. J., Zhou, Z., Grant, K., and Maiden, M. C. J. (2018). Comparative analysis of core genome MLST and SNP typing within a European *Salmonella* serovar Enteritidis outbreak. *Int. J. Food Microbiol.* 274, 1–11. doi: 10.1016/j.ijfoodmicro.2018.02.023

Rehman, M. A., Yin, X., Persaud-Lachhman, M. G., and Diarra, M. S. (2017). First detection of a fosfomycin resistance Gene, fosA7, in *Salmonella enterica* Serovar Heidelberg isolated from broiler chickens. *Antimicrob. Agents Chemother.* 61, e00410-17. doi: 10.1128/aac.00410-17

Rex, J. H., and Pfaller, M. A. (2002). Has antifungal susceptibility testing come of age? *Clin. Infect. Dis.* 35, 982–989. doi: 10.1086/342384

Robicsek, A., Strahilevitz, J., Jacoby, G. A., Macielag, M., Abbanat, D., Hye Park, C., et al. (2006). Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nat. Med.* 12, 83–88. doi: 10.1038/nm1347

Rodriguez, A., Pangloli, P., Richards, H. A., Mount, J. R., and Draughon, F. A. (2006). Prevalence of *Salmonella* in diverse environmental farm samples. *J. Food Protect.* 69, 2576–2580. doi: 10.4315/0362-028X-69.11.2576

Rodriguez-Rivera, L. D., Bowen, B. M., den Bakker, H. C., Duhamel, G. E., and Wiedmann, M. (2015). Characterization of the cytolethal distending toxin (typhoid toxin) in non-typhoidal *Salmonella* serovars. *Gut Pathog.* 7:19. doi: 10.1186/s13099-015-0065-1

Rouzeau-Szynalski, K., Barretto, C., Fournier, C., Moine, D., Gimonet, J., and Baert, L. (2019). Whole genome sequencing used in an industrial context reveals a *Salmonella* laboratory cross-contamination. *Int. J. Food Microbiol.* 298, 39–43. doi: 10.1016/j.ijfoodmicro.2019.03.007

Ruan, Z., Yu, Y., and Feng, Y. (2020). The global dissemination of bacterial infections necessitates the study of reverse genomic epidemiology. *Brief. Bioinform.* 21, 741–750. doi: 10.1093/bib/bbz010

Simon, S., Trost, E., Bender, J., Fuchs, S., Malorny, B., Rabsch, W., et al. (2018). Evaluation of WGS based approaches for investigating a food-borne outbreak caused by *Salmonella enterica* serovar Derby in Germany. *Food Microbiol.* 71, 46–54. doi: 10.1016/j.fm.2017.08.017

Song, J., Willinger, T., Rongvaux, A., Eynon, E. E., Stevens, S., Manz, M. G., et al. (2010). A mouse model for the human pathogen *Salmonella* Typhi. *Cell Host Microbe* 8, 369–376. doi: 10.1016/j.chom.2010.09.003

Stratton, C. W. (2006). In vitro susceptibility testing versus in vivo effectiveness. *Med. Clin. North Am.* 90, 1077–1088. doi: 10.1016/j.mcna.2006.07.003

Suez, J., Porwollik, S., Dagan, A., Marzel, A., Schorr, Y. I., Desai, P. T., et al. (2013). Virulence gene profiling and pathogenicity characterization of non-typhoidal *Salmonella* accounted for invasive disease in humans. *PLoS One* 8:e58449. doi: 10.1371/journal.pone.0058449

Tang, P., Croxen, M. A., Hasan, M. R., Hsiao, W. W. L., and Hoang, L. M. (2017). Infection control in the new age of genomic epidemiology. *Am. J. Infect. Control* 45, 170–179. doi: 10.1016/j.ajic.2016.05.015

Tankson, J. D., Fedorka-Cray, P. J., Jackson, C. R., and Headrick, M. (2006). Genetic relatedness of a rarely isolated *Salmonella*: *Salmonella enterica* serotype Niakhar from NARMS animal isolates. *J. Antimicrob. Chemother.* 57, 190–198. doi: 10.1093/jac/dki439

Wang, X., Biswas, S., Paudyal, N., Pan, H., Li, X., Fang, W. et al. (2019). Antibiotic resistance in *Salmonella* typhimurium isolates recovered from the food Chain through National Antimicrobial Resistance Monitoring System between 1996 and 2016. *Front. Microbiol.* 10:985. doi: 10.3389/fmicb.2019.00985

Weening, E. H., Barker, J. D., Laarakker, M. C., Humphries, A. D., Tsolis, R. M., and Bäumler, A. J. (2005). The *Salmonella enterica* serotype Typhimurium lpf, bcf, stb, stc, std, and sth fimbrial operons are required for intestinal persistence in mice. *Infect. Immun.* 73, 3358–3366. doi: 10.1128/IAI.73.6.3358-3366.2005

World Health Organization (WHO) (2018). *Salmonella (Non-typhoidal)*. Available online at: https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal) (accessed March 10, 2021).

Xu, X., Chen, Y., Pan, H., Pang, Z., Li, F., Peng, X., et al. (2020). Genomic characterization of *Salmonella* uzaramo for human invasive infection. *Microb. Genom.* 6, 1–8. doi: 10.1099/mgen.0.000401

Yu, H., Elbediwi, M., Zhou, X., Shuai, H., Lou, X., Wang, H., et al. (2020). Epidemiological and genomic characterization of Campylobacter jejuni isolates

from a foodborne outbreak at Hangzhou, China. *Int. J. Mol. Sci.* 21:3001. doi: 10.3390/ijms21083001

Yue, M., Rankin, S. C., Blanchet, R. T., Nulton, J. D., Edwards, R. A., and Schifferli, D. M. (2012). Diversification of the *Salmonella* fimbriae: a model of macro- and microevolution. *PLoS One* 7:e38596. doi: 10.1371/journal.pone.0038596

Yu-feng, Q., Shun-tai, W., Xin-lan, H., Jian-hui, C., Jin-song, Y., Meng-yin, H., et al. (2020). *Salmonella* Telelkebir was first detected from clinical bloodstream infection in Fujian Province, China. *Chin. J. Zoonoses* 36, 643–645.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

Zhou, K., Wu, B., Pan, H., Paudyal, N., Jiang, J., Zhang, L., et al. (2020). ONE health approach to address zoonotic brucellosis: a spatiotemporal associations study between animals and humans. *Front. Vet. Sci.* 7:521. doi: 10.3389/fvets.2020.00521

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.

frontiers in Medicine

CASE REPORT
published: 10 September 2021
doi: 10.3389/fmed.2021.714916

Case Report: Diagnosis of *Klebsiella pneumoniae* Invasive Liver Abscess Syndrome With Purulent Meningitis in a Patient From Pathogen to Lesions

Sheng Zeng[1], *Wei-qian Yan*[2], *Xiao-mei Wu*[2*†] *and Hai-nan Zhang*[2*†]

[1] Department of Geriatrics, The Second Xiangya Hospital, Central South University, Changsha, China, [2] Department of Neurology, The Second Xiangya Hospital, Central South University, Changsha, China

OPEN ACCESS

Edited by:
Michael Kemp,
University of Southern Denmark, Denmark

Reviewed by:
Marcia Garnica,
Federal University of Rio de Janeiro, Brazil
Oana Sandulescu,
Carol Davila University of Medicine and Pharmacy, Romania

***Correspondence:**
Hai-nan Zhang
hainanzhang@csu.edu.cn
Xiao-mei Wu
wuxiaomei0923@csu.edu.cn

[†]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 Medicine

Received: 26 May 2021
Accepted: 11 August 2021
Published: 10 September 2021

Citation:
Zeng S, Yan W-q, Wu X-m and Zhang H-n (2021) Case Report: Diagnosis of Klebsiella pneumoniae Invasive Liver Abscess Syndrome With Purulent Meningitis in a Patient From Pathogen to Lesions. Front. Med. 8:714916. doi: 10.3389/fmed.2021.714916

As a determinant human pathogen, *Klebsiella pneumoniae* is known to cause rare *K. pneumoniae* liver abscess syndrome (KLAS) which was more common in Asia in early-stage and reported increasingly outside Asia now. Patients with KLAS who have septic metastatic ocular or central nervous system (CNS) lesions are associated with high morbidity and mortality. Relatively infrequent adult community-acquired *K. pneumoniae* meningitis have been documented and most were with poor prognosis. In this paper, we reported a case of KLAS presenting purulent meningitis as disease onset. While negative results were obtained in the bacterial culture of CSF, blood, or liver pus, metagenomic next-generation sequencing (mNGS) of CSF, and blood samples which were synchronously performed demonstrated *Klebsiella pneumoniae* as the pathogenic microorganism (13,470 and 5,318 unique reads, respectively). The ultimately cured patient benefited from rapid pathogen diagnosis, early percutaneous drainage of the abscess, and prompt appropriate antibiotic administration. Our case highlights the importance of clinicians using mNGS for early pathogen diagnosis of this disease.

Keywords: *K. pneumoniae* liver abscess syndrome, purulent meningitis, rapid pathogen diagnosis, metagenomic next-generation sequencing (mNGS), case report

INTRODUCTION

K. pneumoniae is a well-known human nosocomial pathogen although it appears as the normal flora of the human oral cavity and intestine in most situations (1). Most *Klebsiella pneumoniae* community-acquired infections result in pneumonia or urinary tract infections. However, it also causes rare *K. pneumoniae* liver abscess (KLA) in the absence of predisposing factors of hepatobiliary disease (2). KLA was reported most in Asia in early-stage and now it is increasing outside Asia (3, 4). Except for liver abscess, bacteremia dissemination may result in severe complications including endophthalmitis, meningitis, necrotizing fasciitis, and other illnesses. About 13% of patients with KLA have septic metastatic ocular or central nervous system (CNS) lesions which are associated with high morbidity and mortality (5). The rapid diagnosis followed by appropriate treatment may improve the patient's outcome.

Metagenomic next-generation sequencing (mNGS) is becoming a powerful pathogen detection tool with the advantages of high sensitivity, wide-coverage, and high efficiency, which can greatly improve pathogen identification (6). It has now been successfully used for pathogen detection in biological specimens including CSF and blood (7). Early diagnosis assisted by mNGS contributes

to timely treatment. In our report, we displayed a case of KLA manifested as purulent meningitis at the onset of the disease. Rapid pathogen diagnosis facilitated further lesion detection, proper treatment, and full recovery of the patient. Until now, few KLA cases with metastatic meningitis were reported, fewer patients achieved a good prognosis (8). Our case suggested the importance of early pathogen diagnosis.

CASE PRESENTATION

A 59-year-old female rural resident in Hunan province, China, was admitted to the Emergency Department (ED) of Second Xiangya Hospital, Central South University. She complained of headache with fever for 6 days and fatigue and somnolence for 1 day. The first symptoms were headache with fever. She experienced four episodes of diarrhea on the first day. No other gastrointestinal symptoms were presented since then. From the second day to the fourth day, she was admitted to a local hospital. A lumbar puncture was performed because of clinical suspicion of "intracranial infection." The intracranial pressure was 140 mmH_2O. Cerebrospinal fluid (CSF) was yellow and purulent, and revealed 1,080 white blood cells/μL with a monocyte's ratio of 56.9%. She was immediately treated with intravenous ceftriaxone 2,000 mg q12h. Although the history of diabetes was denied, her random blood glucose was as high as 14 mmol/L during hospitalization. On the fifth day, she was recommended to refer to our hospital due to the new appearance of fatigue and somnolence. Her past medical history showed no recent travel, tick bites, sick contact, alcohol, or drug use. Upon admission, her initial vital signs included a body temperature of 36.7°C, heart rate of 82 beats/min, blood pressure of 125/80 mmHg, respiratory rate of 23 breaths/min, and oxygen saturation of 99% on 2 L/min oxygen. Her consciousness was somnolence. No skin rash was observed, and no obvious abnormality was found on physical examination of the heart, lung, and abdomen. A neurologic examination revealed that the neck was stiff and Kernig's sign was positive. Laboratory test on ED showed a white blood cell count of 9.0×10^9/L with an elevated neutrophil ratio of 83.9% in blood routine test. Liver function test showed hypoalbuminemia (26.1 g/L). There was no significant change in the coagulation panel. The random blood glucose was 15.3 mmol/L. The blood ketone and lactate were 2.43 and 0.9 mmol/L, respectively. Lung computed tomography (CT) scan showed focal small patchy infiltrates in both lower lungs. Considering the severity of the patient's condition, she was sent to the intensive care unit of the department of neurology (NICU) for further treatment. Further laboratory tests and auxiliary examinations were arranged. The concentrations of procalcitonin, C-reactive protein, erythrocyte sedimentation rate (ESR) and interleukin 6 (IL-6) were 8.720 ng/ml, 134.0 mg/L, 45 mm/h, and 630.0 pg/ml, respectively. The result of glycosylated hemoglobin was 10.7%. Thyroid function test results showed that T3 was 0.56mol/L, FT3 was 1.58 mol/L, and T4 was 50.70 mol/L. Two sets of peripheral blood cultures were ordered. Lumbar puncture showed the intracranial pressure was 130 mmH_2O. Cerebrospinal fluid (CSF) revealed 1,000 white blood cells/μL with proportion of multinucleated cells of 95%, protein of 943 mg/L, and glucose of 3.97 mmol/L (synchronous blood glucose of 14.0 mmol/L). The CSF Gram staining, Ink staining, acid-fast staining, and bacterial culture were all negative. The above findings led to the diagnosis of purulent meningitis, pulmonary infection, diabetes mellitus (DM), and diabetic ketosis. To further identify the pathogen, the PACEseq mNGS test (Hugobiotech, Beijing, China) was also performed on CSF samples on the Nextseq 550 platform (Illumina, San Diego, CA). She has immediately received emergency management of ketosis and was treated with intravenous meropenem 2,000 mg q8h and vancomycin 500 mg q6h on admission (at this moment, the duration of ceftriaxone therapy was 3 days).

On the second day of hospitalization, a brain MRI scan was performed, and partial enhancement of the pia mater was recommended (**Figures 1A,B**). She experienced a chill with a body temperature of 38.6°C on this day. Peripheral blood culture was performed. Lumbar puncture and intrathecal injection of 20 mg vancomycin were arranged continuously on the second to the fourth day.

On the fourth day of hospitalization, the mNGS of CSF reported *Klebsiella pneumoniae*, with a total of 13,470 detected unique reads detected (**Figure 2A**). *Klebsiella pneumoniae* invasive liver abscess syndrome was suspected as the clinical features and unusual findings of the pathogen in this patient. An abdominal enhanced CT demonstrated a single abscess (67 mm × 62 mm) in the right lobe of the liver with separating enhancement (**Figures 1C,D**) and confirmed our suspicion. No other intra-abdominal pathologies, such as gallstones, were observed. To quickly ascertain the transmission route, mNGS of a blood sample was also arranged. After the diagnosis with *Klebsiella pneumoniae* invasive liver abscess syndrome, the antimicrobials of the patient were adjusted to meropenem 2,000 mg q8h intravenously. An emergency CT-guided percutaneous drainage of the liver abscess was performed, which drained 210 mL of yellow pus over the first 24 h. The liver aspirate was subjected to Gram staining and both aerobic and anaerobic culture. Repeated blood and liver pus cultures were required during episodes of fever.

On the seventh day of hospitalization (the third day after operation), only 20 mL yellow pus was drained over the third 24 h. Reexamination of abdominal CT suggested the liver abscess was reduced to 51 mm × 37 mm (**Figure 1E**). Lumbar puncture after operation showed the intracranial pressure was 140 mmH_2O. Cerebrospinal fluid (CSF) revealed 1 white blood cells/μL (mononuclear cells), protein 780 mg/L, and glucose 5.88 mmol/L (synchronous blood glucose of 8.1 mmol/L). Meanwhile, the blood mNGS results demonstrated 5,318 unique reads of *Klebsiella pneumoniae* (**Figure 2B**). Thus, purulent meningitis with *Klebsiella pneumoniae* was septic metastatic from invasive liver abscess through blood.

On the ninth day of hospitalization (the fourth day after operation), the patient recovered to good condition after treatment. However, the patients and their families strongly requested to return to the local hospital for further treatment. A telephone follow-up after 3 months showed the patient recovered and was discharged from the hospital after 29 days of treatment.

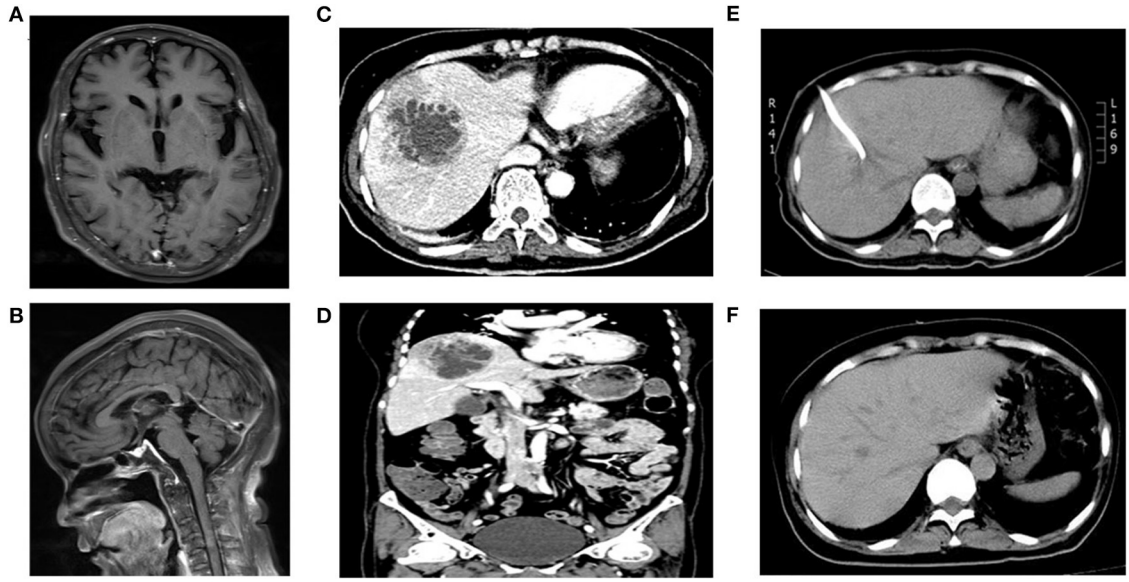


FIGURE 1 | MRI or CT scan images of the patient. **(A,B)** indicated partial enhancement of pia mater in brain enhanced MRI; **(C,D)** revealed an area of abnormal attenuation measuring 67 mm × 62 mm in the right lobe of the liver, indicative of a single large multi-loculated abscess in abdominal enhanced CT; **(E)** displayed the abscess reduced to 51 mm × 37 mm 3 days after emergency CT-guided percutaneous drainage of the liver abscess in abdominal plain CT; **(F)** showed no lesions in liver in the reexamination of abdominal CT after full recovery.

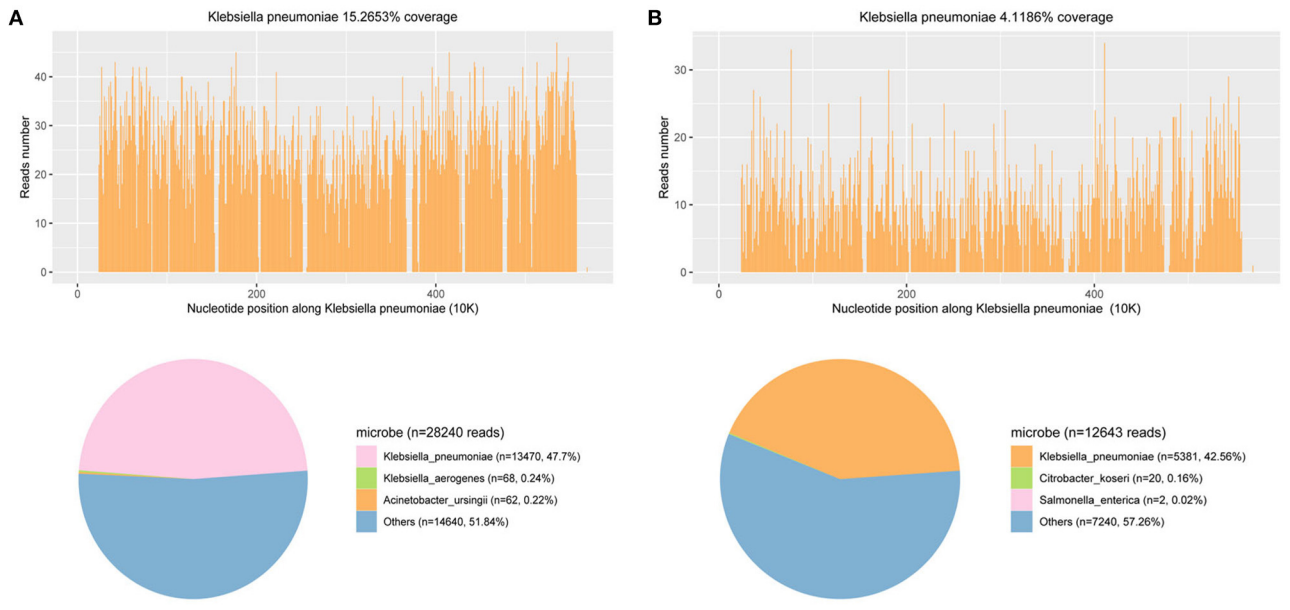


FIGURE 2 | mNGS results of pathogen identification: 47.7 and 42.56% of bacterial reads corresponded to *K. pneumoniae* with coverage of 15.2653 and 4.1186% in CSF **(A)** and blood **(B)**, respectively.

Thus, the total duration of treatment was nearly 40 days. The patient underwent abdominal CT again, which showed no lesions in her liver (**Figure 1F**) at the second follow-up after 9 months.

It should be noted that bacterial cultures of CSF, blood, and liver pus in the whole disease course were all negative.

DISCUSSION

Klebsiella pneumoniae liver abscess syndrome (KLAS) is clinically characterized by bacteremia, liver abscesses, and metastatic infection caused by hypervirulent strains harboring capsular

serotype K1 or K2. DM may be a risk factor for this syndrome (2). The most common clinical manifestations of patients with *K pneumoniae* liver abscesses are fever, chills, and abdominal pain. However, all these manifestations lack specificity. For patients, especially those who suffer from diabetes mellitus present with *K pneumoniae* bacteraemia, meningitis, endophthalmitis, or other extrahepatic infections, it is necessary to look for potential liver abscess. Thus, timely identification of pathogens is the key to better management. However, traditional methods for pathogen discovery such as bacterial culture are time-consuming and lack sensitivity (9). The positivity rate of traditional methods is influenced by quality and quantity of specimens, patient antibiotic administration, the severity of infection, and laboratory sufficiency. In our case, the effort was paid through the repeated bacterial cultures of CSF, blood, or liver pus. However, the results were poor and were most likely influenced by empirical antibiotic therapy. The patient's treatment could be delayed without the identification of the pathogen by mNGS.

mNGS emerges as an alternative and efficient molecular diagnostic method, which overcomes the limitations of traditional methods. It is now increasingly applied in the identification of pathogens in clinical practice, such as sepsis, meningitis, and acute respiratory infection (7, 10, 11). The state-of-the-art mNGS technology has advantages in identifying rare, novel, difficult-to-detect, and coinfected pathogens directly from clinical samples, providing timely diagnostic evidence to guide treatment plans (12). In the future, more validated workflows, lower cost, and simplified interpretation criteria would be needed to further routinely implement mNGS in clinical practice. Our patient was an elderly rural woman who had an unrecognized history of DM. She presented with purulent meningitis at the time of onset and was demonstrated as KLAS on the fourth day after admission. After consulting the literature, we found that adult community-acquired *K. pneumoniae* meningitis has been less recorded (8, 13). Unfortunately, few patients have survived this disease. In addition to the severity of the disease, the lack of timely diagnosis and standardized treatment may also be the important reasons.

Until now, there are no clear guidelines for the management of *Klebsiella* pneumoniae invasive liver abscess syndrome with purulent meningitis (2). The basic consensus is the combination of early percutaneous drainage or open (laparoscopic) surgical drainage of the abscess and prompts the appropriate antibiotic administration. The selection of antibiotics is based on *in-vitro* susceptibilities and clinical response. However, results of antibiotics susceptivity are time-consuming. Therefore, the empirical use of high-dose third-generation cephalosporins, including cefotaxime (up to 2,000 mg every 4 h) and ceftriaxone (2,000 mg, twice a day) are options for treatment of *K. pneumoniae* meningitis. Imipenem and meropenem can be given to patients when strains containing extended-spectrum beta-lactamase are suspected. Our patient benefited from the empirically and sufficiently intravenous use of meropenem and early percutaneous drainage of abscesses in our clinic.

In conclusion, we shared a case of KLA that presented with purulent meningitis at the onset of the disease. Rapid pathogen diagnosis through mNGS facilitated further lesion detection, proper treatment, and full recovery of the patient. Nowadays, adult community-acquired *K. pneumoniae* meningitis carries a very poor prognosis. Our case highlights the importance of early pathogen diagnosis by mNGS for the precise treatment of this disease. Meantime, it raises the potential of diagnosing severe infectious diseases from pathogens to lesions.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ETHICS STATEMENT

Ethical Committee of the Second Xiangya Hospital of the Central South University in China (equivalent to an Institutional Review Board) approved the study, and written informed consent was obtained from the patient for publication of this Case Report and any accompanying images.

AUTHOR CONTRIBUTIONS

SZ collected the materials, analyzed the data, and wrote the paper. Wq-Y assisted in clinical follow up. Xm-W and Hn-Z designed the concept and provided administrative support. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Program of National Natural Science Foundation of China (#81801123).

REFERENCES

1. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci USA.* (2015) 112:E3574–81. doi: 10.1073/pnas.1501049112
2. Siu LK, Yeh KM, Lin JC, Fung CP, Chang FY. *Klebsiella pneumoniae* liver abscess: a new invasive syndrome. *Lancet Infect Dis.* (2012) 12:881–7. doi: 10.1016/S1473-3099(12)70205-0
3. Chang FY CM, Fan RL, Shaio MF. A clinical study of Klebsiella liver abscess. *Taiwan Yi Xue Hui Za Zhi.* (1988) 87:282–87.
4. Vila A, Cassata A, Pagella H, Amadio C, Yeh KM, Chang FY, et al. Appearance of *Klebsiella pneumoniae* liver abscess syndrome in Argentina: case report and review of molecular mechanisms of pathogenesis. *Open Microbiol J.* (2011) 5:107–13. doi: 10.2174/1874285801105010107
5. Fang CT, Lai SY, Yi WC, Hsueh PR, Liu KL, Chang SC. *Klebsiella pneumoniae* genotype K1: an emerging pathogen that causes septic ocular or central nervous system complications from pyogenic liver abscess. *Clin Infect Dis.* (2007) 45:284–93. doi: 10.1086/519262

6. 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
7. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N Engl J Med.* (2019) 380:2327–40. doi: 10.1056/NEJMoa1803396
8. Sun R, Zhang H, Xu Y, Zhu H, Yu X, Xu J. Klebsiella pneumoniae-related invasive liver abscess syndrome complicated by purulent meningitis: a review of the literature and description of three cases. *BMC Infect Dis.* (2021) 21:15. doi: 10.1186/s12879-020-05702-3
9. Xing XW, Zhang JT, Ma YB, He MW, Yao GE, Wang W, et al. Metagenomic next-generation sequencing for diagnosis of infectious encephalitis and meningitis: a large, prospective case series of 213 patients. *Front Cell Infect Microbiol.* (2020) 10:88. doi: 10.3389/fcimb.2020.00088
10. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet.* (2019) 20:341–55. doi: 10.1038/s41576-019-0113-7
11. Chen Y, Feng W, Ye K, Guo L, Xia H, Guan Y, et al. Application of metagenomic next-generation sequencing in the diagnosis of pulmonary infectious pathogens from bronchoalveolar lavage samples. *Front Cell Infect Microbiol.* (2021) 11:541092. doi: 10.3389/fcimb.2021.541092
12. 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
13. Qian Y, Wong CC, Lai SC, Lin ZH, Zheng WL, Zhao H, et al. *Klebsiella pneumoniae* invasive liver abscess syndrome with purulent meningitis and septic shock: a case from mainland China. *World J Gastroenterol.* (2016) 22:2861–6. doi: 10.3748/wjg.v22.i9.2861

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fmicb.2021.715568

Whole-Genome Sequencing Elucidates the Epidemiology of Multidrug-Resistant *Acinetobacter baumannii* in an Intensive Care Unit

*Pu Mao, Xiaolong Deng, Leping Yan, Ya Wang, Yueting Jiang, Rong Zhang, Chun Yang, Yonghao Xu, Xiaoqing Liu and Yimin Li**

State Key Laboratory of Respiratory Diseases, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

The nosocomial pathogen *Acinetobacter baumannii* is a frequent cause of healthcare-acquired infections, particularly in critically ill patients, and is of serious concern due to its potential for acquired multidrug resistance. Whole-genome sequencing (WGS) is increasingly used to obtain a high-resolution view of relationships between isolates, which helps in controlling healthcare-acquired infections. Here, we conducted a retrospective study to identify epidemic situations and assess the percentage of transmission in intensive care units (ICUs). Multidrug-resistant *A. baumannii* (MDR-AB) were continuously isolated from the lower respiratory tract of different patients (at the first isolation in our ICU). We performed WGS, pulsed-field gel electrophoresis (PFGE), and multilocus-sequence typing (MLST) analyses to elucidate bacterial relatedness and to compare the performance of conventional methods with WGS for typing MDR-AB. From June 2017 to August 2018, *A. baumannii* complex strains were detected in 124 of 796 patients during their ICU stays, 103 of which were MDR-AB. Then we subjected 70 available MDR-AB strains to typing with WGS, PFGE, and MLST. Among the 70 *A. baumannii* isolates, 38 (54.29%) were isolated at admission, and 32(45.71%) were acquisition isolates. MLST identified 12 unique sequence types, a novel ST (ST2367) was founded. PFGE revealed 16 different pulsotypes. Finally, 38 genotypes and 23 transmissions were identified by WGS. Transmission was the main mode of MDR-AB acquisition in our ICU. Our results demonstrated that WGS was a discriminatory technique for epidemiological healthcare-infection studies. The technique should greatly benefit the identification of epidemic situations and controlling transmission events in the near future.

Keywords: ***Acinetobacter baumannii*, epidemiology, whole-genome sequencing, intensive care unit, multi-drug resistance**

OPEN ACCESS

Edited by:
Michael Kemp, University of Southern Denmark, Denmark

Reviewed by:
Dinesh Sriramulu, Independent Researcher, Chennai, India
Uga Dumpis, Pauls Stradiņš Clinical University Hospital, Latvia

***Correspondence:**
Yimin Li
dryiminli@vip.163.com

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *27 May 2021*
Accepted: *26 August 2021*
Published: *13 September 2021*

Citation:
Mao P, Deng X, Yan L, Wang Y, Jiang Y, Zhang R, Yang C, Xu Y, Liu X and Li Y (2021) Whole-Genome Sequencing Elucidates the Epidemiology of Multidrug-Resistant Acinetobacter baumannii in an Intensive Care Unit. Front. Microbiol. 12:715568. doi: 10.3389/fmicb.2021.715568

INTRODUCTION

Acinetobacter baumannii is a Gram-negative pathogen that causes serious nosocomial infections, especially in patients with advanced age, mechanical ventilation, respiratory failure, or a prolonged hospital stay in intensive care units (ICUs) (Vazquez-Lopez et al., 2020). In recent decades, *A. baumannii* has emerged as an important nosocomial pathogen that exhibits high levels of

resistance to antibiotics and has become endemic in some geographical regions. Multidrug-resistant *A. baumannii* (MDR-AB) increases hospital stays and costs and limits therapeutic options. Critically ill patients infected with MDR-AB strains have substantially higher mortality rates (Falagas et al., 2006).

Accurate strain classification is a key process necessary for managing antimicrobial-resistant strains in health care-related infections (Kalenic and Budimir, 2009; Nishi and Hidaka, 2016). Pulsed-field gel electrophoresis (PFGE) is still used for typing analysis of *A. baumannii*, although it has drawbacks in terms of the time requirement and difficulty in data interpretation (Thong et al., 2002; Singh et al., 2006). Multilocus-sequence typing (MLST), the partial sequence-based typing method, is usually considered as the gold standard for global epidemiological investigations. It focuses on the clustering of isolates worldwide and is not sensitive to short duration and small-scale outbreaks (Tomaschek et al., 2016). It is considered a complementary approach of PFGE since it connects isolates with global epidemiology (Tomaschek et al., 2016). In a decade, whole-genome sequencing (WGS) has become a promising method in microbiological laboratories for strain identification, molecular epidemiology and outbreak analysis (Makke et al., 2020). It has the highest discriminatory power and is likely to be widely used with cost reduction (Halachev et al., 2014; Salipante et al., 2015; Hwang et al., 2021).

In this study, we conducted a retrospective study to identify an epidemic situation and MDR-AB-transmission events. We used reference-based single-nucleotide polymorphism (SNP) approaches to map either processed reads or assembled contigs to a reference genome, followed by SNPs calling. Then, SNPs differences among the MDR-AB isolates were processed. We aimed to describe the prevalence and molecular epidemiology of MDR-AB in the ICU through WGS-based typing and compare the performance of conventional methods with WGS for typing MDR-AB.

MATERIALS AND METHODS

Setting

The First Affiliated Hospital of Guangzhou Medical University is a 1500-bed teaching hospital and a tertiary referral center for major respiratory disease in southern China. This study was conducted in the 27-bed adult ICU, which cares for patients with respiratory disease and those who have undergone thoracic surgical procedures. The ICU included 4 senior intensivists, 14 attendings, 6 residents, medical students, and nurses during the study period. Attendings and residents staffed the ICU 24 h per day and every day of the week. Three teams provided ICU medical coverage during the day: each including a senior intensivist, four attendings and two residents taking care of seven bed patients. The nurse-to-patient ratio was maintained at 1–1.5:1 during daytime and 1:3 at night.

Infection control measures mainly implemented in ICU have been previously described (Ye et al., 2015), mainly included: performing hygiene according to the WHO recommended and re-emphasis on the appropriate use of gloves, especially not allowed wear gloves to contract clean area items; implementing a cohort strategy, colonized/infected patients were admitted to a special room and treated by a dedicated team of healthcare workers; practice contact precautions for colonized/infected and newly admitted patients. Artificial airway management followed Critical Care Medicine Branch of Chinese Medical Association Guideline for diagnosis, prevention, and treatment of ventilator-associated pneumonia (Chinese Society of Critical Care Medicine, 2013).

Ethical approval was not required because the study was conducted as part of normal surveillance and management of healthcare-associated infections.

Bacterial Identification and Antimicrobial-Susceptibility Testing

Seventy-five MDR-AB complex isolates were collected from the lower respiratory tract of hospitalized ICU patients between June 2017 and August 2018 and were designated AB-1 to AB-75, based on admission date. The lower respiratory tract specimens were obtained by tracheal aspiration.

Strains were cryopreserved until needed for WGS. Initial identification was conducted using the VITEK2 Compact 30 System (bioMérieux). Antimicrobial susceptibility was determined *via* microdilution in accordance with guidelines of the Clinical and Laboratory Standards Institute (M100-S29). MDR was defined as resistance to at least one representative agent from three or more of the five antimicrobial-agent classes: cephalosporins (such as ceftazidime or cefepime), carbapenems (such as imipenem), β-lactamase inhibitors (such as cefoperazone/sulbactam), fluoroquinolones (such as ciprofloxacin), and aminoglycosides (such as amikacin) (Magiorakos et al., 2012).

PFGE Analysis

Isolates were inoculated in liquid Luria-Bertani medium and cultured overnight at 37°C. The bacterial supernatant was collected through centrifugation and resuspended in 1× Tris-ethylenediaminetetraacetic acid (EDTA) buffer (TE buffer) before an equal volume of 2% clean cut agarose was added. After immediate mixing, the liquid was placed in a mold and allowed to solidify. The block was then placed into 300 μL cell lysate (in 100 mmol/L Tris and 100 mmol/L EDTA), mixed with 5 μL proteinase K (20 mg/mL), and incubated overnight at 50°C. Next, the block was washed three times with 1 × TE, for 1 h/wash step, before being incubated overnight at 25°C with 50 U Apa I enzyme. Subsequently, PFGE was performed using a CHEF-DR III PFGE apparatus, under the following conditions: running buffer, 0.5 × TBE; 1% Pulsed-Field Certified Agarose; 14°C; 6 V/cm; angle of 120°; 0.5–20 s, overall time: 20 h. After the current ended, ethidium bromide staining was performed and images were obtained. Fingerprints were analyzed by BioNumerics software, using the unweighted pair group method and the arithmetic averages method, a 2% tolerance in strip position difference, and a 0.8% optimization value. Strains with ≥87% similarity were classified as the same subtype (representing the same clone), and those with <87% similarity

were classified as different genotypes (representing different clones) (Seifert et al., 2005).

WGS and Assembly

Genomic DNA was extracted from 3 mL overnight cultures of 75 *A. baumannii* complex isolates, using the SPARK DNA Sample Prep Kit-96 (Enzymatics, United States). A library was generated for each isolate using the NEB Next Ultra DNA Library Prep Kit (Illumina, San Diego, CA, United States). DNA library concentrations were quantified using a Qubit® 2.0 Fluorometer and a Qubit™ dsDNA HS Assay Kit, and 150-base pair (bp) paired-end reads were generated with a HiSeq®2500 instrument. Sequencing was performed at the Shanghai Biotechnology Corporation. We implemented a filtering pipeline that trimmed reads from the 3′ end with <20 sequencing quality, discarded adaptor sequences and reads <45 bp in length, removed reads containing Ns or where <50% of bases had >20 sequencing quality. Assembly statistics are displayed in **Supplementary Table S1**.

MLST Analysis

Conventional MLSTs were inferred *in silico* from WGS data. Seven MLST loci were selected using a sequence-extraction tool and then identified using a public *A. baumannii* MLST database.[1]

Single-Nucleotide Polymorphism Detection and Whole-Genome Phylogeny

Paired reads were mapped to *A. baumannii* strain XH386 (reference genome; GenBank accession number NZ_CP010779.1) using the Burrows–Wheeler Aligner (version 0.7.12). The original file was converted to a BAM file in SamTools. SNP and InDel results were processed simultaneously using the HaplotypeCaller tool, based on a real-time *de novo* algorithm with BAM files (reserved mutation sites with sequencing quality >30) and obtained VCF files. The SNPhylo program was used to remove low-quality SNPs; processed SNPs were then concatenated into a sequence and compared in the MUSCLE program. During this process, we discarded the following single-nucleotide variations (SNVs): (1) those in genes annotated as phages, transposases, or integrases, (2) those in genomic regions annotated as phages using Phage Finder, (3) those within 20 bp of the start or end of a contig, and (4) those with sequence quality <100. Tree construction was performed in PHYLIP. Trees were visualized in Evolview version 3.0.[2] This analysis was performed at the Shanghai Biotechnology Corporation.

Previously, *A. baumannii* was estimated to have a ∼5 SNPs/year mutation rate over the whole genome (Hawkey et al., 2018). Isolates were assigned as the same genotype if their SNV numbers were ≤10.

Nucleotide Sequence Accession Numbers

Generated sequence reads were submitted to the NCBI Sequence Read Archive under the accession number PRJNA639298.

[1]https://pubmlst.org/organisms/acinetobacter-baumannii

[2]https://evolgenius.info/

RESULTS

Bacterial Isolates and Patients

This study spanned the period from June 2017 to August 2018, when 868 admissions (involving 796 patients) to the ICU were recorded. After deleting repetitive samples from the same patients with same isolates, we identified 124 samples as *A. baumannii* complex or *A. baumannii* using the VITEK2 Compact 30 System, of which 103 showed MDR.

Among those 103 samples, the first 75 MDR-AB complex and *A. baumannii* isolates obtained were retrieved from storage; WGS successfully analyzed 71 isolates. Seventy isolates were determined to be *A. baumannii*, and one isolate was identified as *Acinetobacter calcoaceticus*.

The 70 MDR-AB strains analyzed using WGS were from 70 different patients. The clinical characteristics of these 70 patients are summarized in **Table 1**.

Presentation With MDR-AB Upon ICU Admission and MDR-AB Acquisition in the ICU

All patients were screened for lower-respiratory tract infections within 24 h of admission and assessed at least once/week. Acquisition with MDR-AB was defined by a change in the colonization status from culture-negative to positive or a change of strains based on epidemiological data. During the study period, 55 patients (53.4%, 55/103) carried MDR-AB at the time of admission, and 48 acquisitions were identified. Among the 70 patients whose isolates were analyzed with WGS, 54.28% (n = 38) carried MDR-AB at the time of admission. Thirty-two acquisitions were identified: 13 patients who were negative during the first screening showed MDR-AB in a later screening; 19 patients showed changes of MDR-AB strains.

Antibiotic-Susceptibility Testing

Drug-sensitivity test results of the 70 MDR-AB strains are shown in **Supplementary Table S2**. Almost all isolates were resistant to quinolones. We found that 100% of the MDR-AB strains (n = 70) were resistant to ciprofloxacin, 80% (n = 56) were resistant to levofloxacin, and 14 isolates showed intermediate resistance to levofloxacin. Resistance rates to third-generation cephalosporins (cefotaxime and ceftazidime) were 91.43 and 97.14%, respectively. In addition, 98.57% (n = 69) were resistant to imipenem or piperacillin–tazobactam, 97.14% (n = 68) were resistant to meropenem, and all tested isolates were sensitive to colistin (six strains were untested).

MLST Analysis

Ten different STs were identified using the Oxford scheme, whereas only two STs were identified using the Pasteur MLST scheme. Considering that the Oxford scheme shows a higher concordance with WGS phylogenies and better discriminates between strains with short evolutionary distances, we adopted the Oxford scheme for further analysis. We identified 12 unique STs among the 70 *A. baumannii* isolates, namely ST208 (n = 31), ST136 (n = 14), ST195 (n = 9), ST457 (n = 4), ST1633 (n = 3),

TABLE 1 | Clinical characteristics of the patients with MDR-AB sequenced in this study.

Characteristics[a]	
Age (years)	60.09 ± 17.01
Gender, male	45 (64.29)
Prior hospital stay (<1 year)	17 (24.29)
Location before ICU admission	
Internal[b]	35 (50.00)
Outside[c]	35 (50.00)
Days of stay in ICU	28.50 (11.75–48.75)
Days of ventilation	31.00 (18.75–66.00)
Days of ICU stay after MDR-AB isolated	22.50 (9.00–43.25)
APACHE II score upon ICU admission	18.71 ± 6.57
SOFA score upon ICU admission	8.66 ± 4.12
Primary ICU admission diagnosis	
Chronic pulmonary disease	25 (35.71)
Pneumonia	44 (62.85)
Diabetes mellitus	18 (25.71)
Hypertension	24 (34.29)
Malignancy	16 (22.86)
Surgery procedure (< 30 days)	17 (24.29)
Antibiotics used before ICU admission	66 (94.29)
Antibiotic class[d]	
Carbapenem	45 (64.29)
Penicillin	1 (1.43)
Piperacillin–tazobactam	42 (60.00)
Cephalosporin	17 (24.29)
Fluoroquinolones	23 (32.90)
Glycopeptide	26 (37.14)
Antifungal	43 (61.43)
Invasive procedures	
Ventilator	67 (95.71)
Blood transfusion	7 (10.00)
Hemodialysis	7 (10.00)
Nasogastric tube	2 (2.86)
Peripheral venous catheter	5 (7.14)
Central venous catheter	60 (85.71)
Urinary catheter	64 (91.43)
Outcome	
28-day mortality	3 (4.29)
Hospital mortality	17 (24.29)

[a] *All data are presented as the number, with the percentage in parenthesis, except for the age, APACHE II score, and SOFA score, which are presented as the mean ± SD. The number of days in the ICU, the number of days of ventilation, and the number of days in the ICU after MDR-AB was isolated are presented as median (IQR).*
[b] *Internal, patients were transferred from other departments in our hospital.*
[c] *Outside, patients transferred from the other hospitals.*
[d] *Use of antibiotics in patients: used within 30 days before the first MDR-AB isolate was discovered and the antibiotics were used for at least 72 h.*

ST1849 (n = 2), ST369 (n = 2), ST229 (n = 1), ST191 (n = 1), ST1486 (n = 1), ST1806 (n = 1), and a novel ST, ST2367(n = 1) (**Figure 1**). Except for the ST229 clone, belonging to complex 110 international clone VII (CC110 IC7), the remaining isolates were assigned to clonal complex 92 (also known as international clone II), widely epidemic in over 30 countries (Karah et al., 2012; Tomaschek et al., 2016). Our study identified three major STs, with the predominant population being ST208, accounting for 44.29% (31/70). ST136 was the second-most prevalent clone (20.00%, 14/70), followed by ST195 (12.86%, 9/70).

PFGE Analysis

During the period when these strains were collected, we also typed isolates using PFGE. Strains with ≥87% similarity were considered the same subtype, representing the same clone (Seifert et al., 2005). Accordingly, we identified 16 different pulsotypes (**Figure 1**). Notably, 27 strains were classified as clone 8, making it the most abundant epidemic clone, and it continued to spread through the ICU during the study period.

Whole-Genome SNP Phylogenetic Analysis

A phylogenetic tree based on SNPs was constructed to clarify relationships among the sequenced strains. The median of the minimum SNV number across different isolates during the research period was 426 (interquartile range: 275–948; range: 0–52,781). Five distinct primary clades were identified (clades A–E; **Figure 2**).

We found 38 genotypes, with 14 genotypes identified in more than one patient. Along the 14 genotypes identified in more than one patient, we found that five genotypes consist of 16 instances where patients carried MDR-AB at the time of admission, suggesting that a common source of infection existed outside of the ICU (**Figure 3**).

Transmission was defined as patient acquisition of a genotype cultured from a previous patient, irrespective of overlapping patient stay in the ICU. Under WGS, transmission occurred when SNVs ≤10 between isolated pathogens. We identified 23 transmissions for 11 genotypes, consisting of 16 transmissions where donors were patients who carried MDR-AB at the time of admission (for eight genotypes), and 7 transmissions where the donor was not identified (for three genotypes) (**Figure 3**). Genotypes 1 and 5 showed the highest number of transmissions (four cases in each genotype).These results demonstrate that the major source of transmissions in our ICU was the admission isolates, also suggested that MDR-AB spread in our ICU mainly through transmission.

According to PFGE results, 27 MDR-AB strains formed clone 8, the largest epidemic clone, which was associated with 13 acquisitions. We reexamined this prevalent clone using WGS to determine whether "outbreak" isolates and transmissions identified with this technique actually reflect the relationship between different pathogens. The 13 acquisitions of MDR-AB isolates belonging to clone 8 were divided into five WGS clades, involving seven genotypes (**Figures 2**, **3**). Mean number of SNVs among them was 279, ranging from 0 to 602 (92–377; **Figure 4**), and six isolates had <10 SNVs.

DISCUSSION

In this study, we used WGS to retrospectively investigate the epidemiology of clinical MDR-AB isolates in an adult ICU

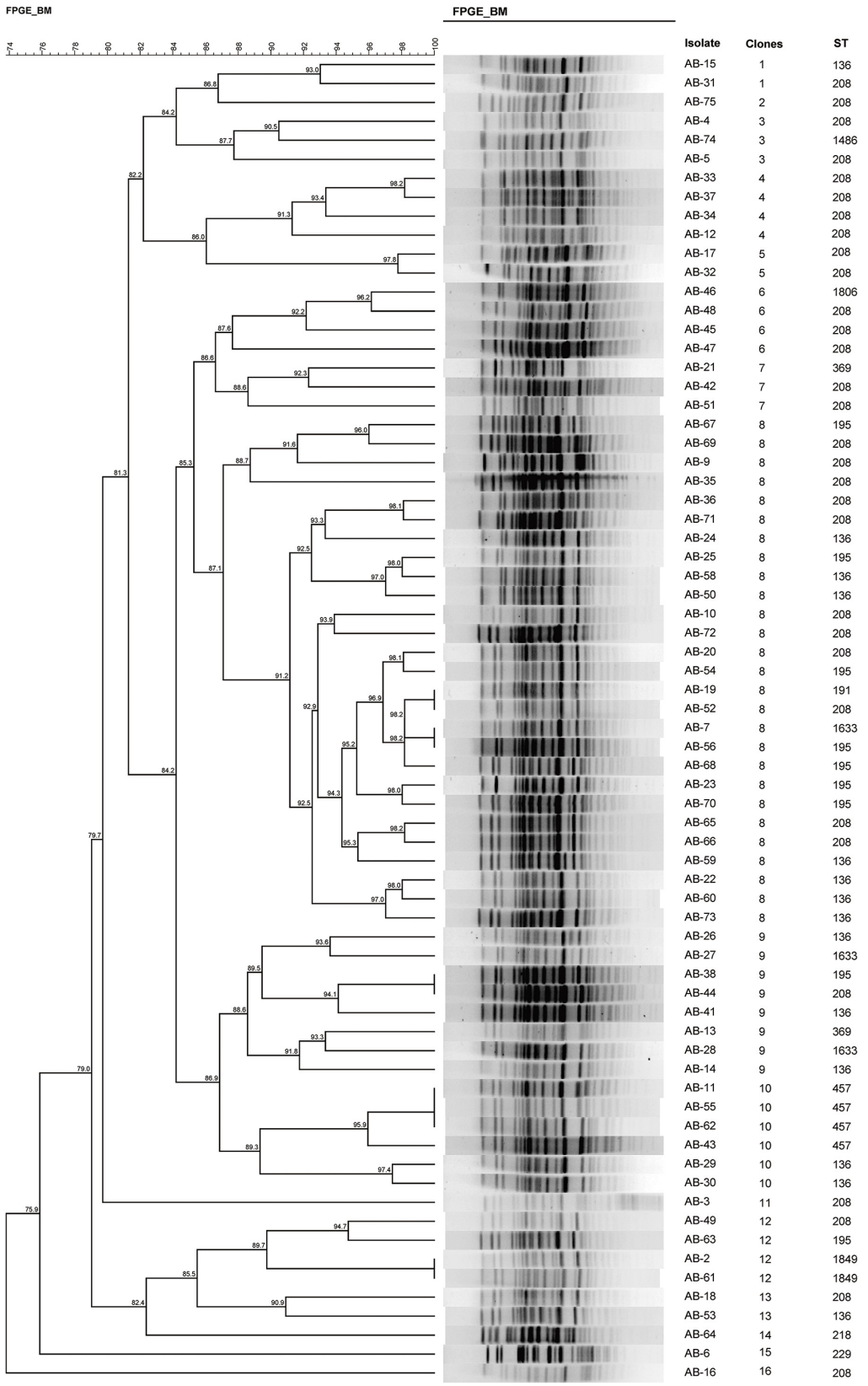


FIGURE 1 | Pulsed-field gel electrophoresis profiles of MDR-AB and dendrogram, and MLST.

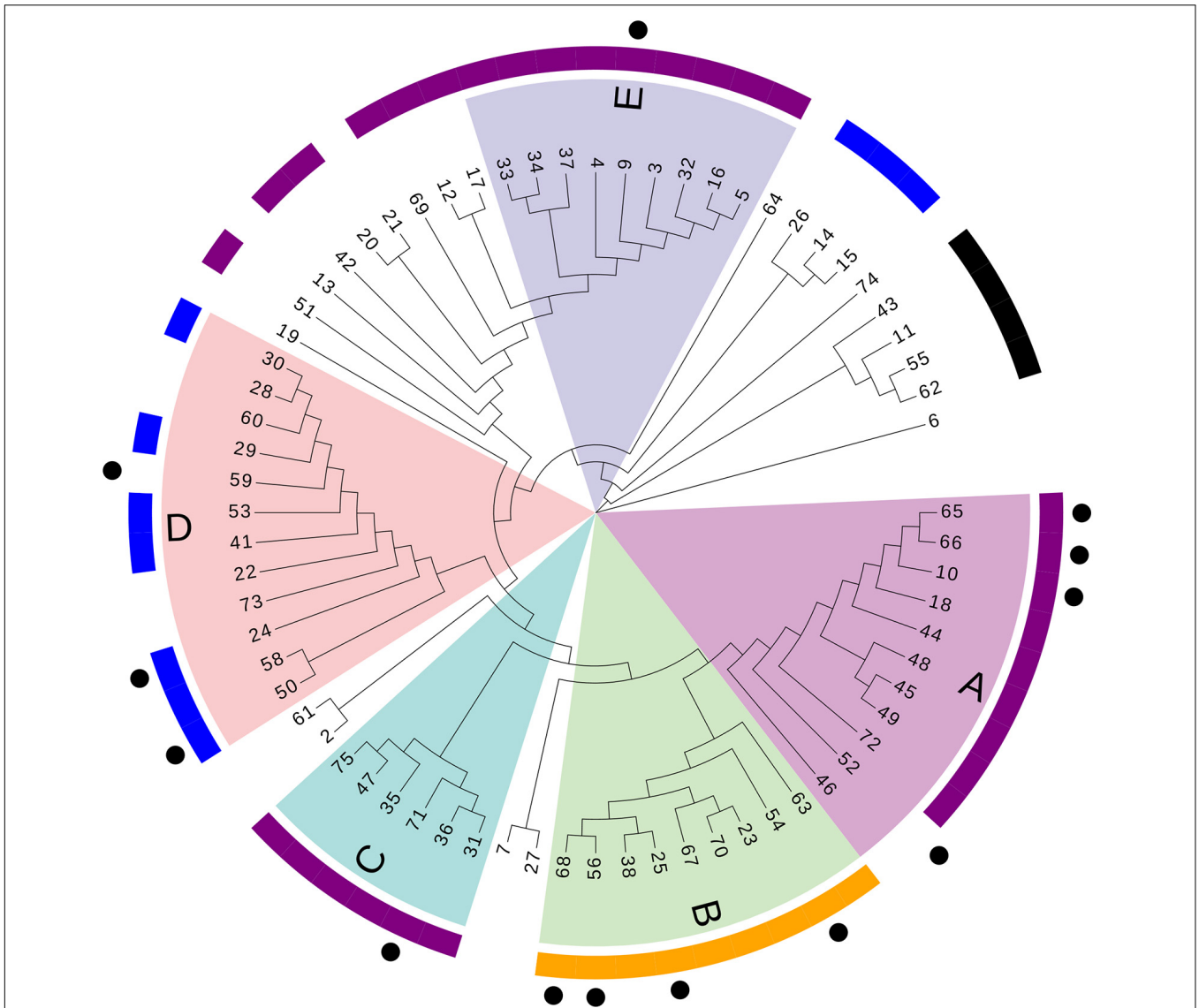


FIGURE 2 | Phylogenetic tree based on SNPs. Numbers refer to the ID number of isolates. Different letters (A–E) and colors refer to branches of the five major clades. Outer purple ring, ST208; outer yellow ring, ST195; outer blue ring, ST136; outer black ring, ST457. Dots represent acquisitions of MDR-AB isolates of clone 8 with the same PFGE type.

and to determine how frequently MDR-AB is transmitted between ICU patients.

Data from a previous epidemiological WGS study of *Staphylococcus aureus* suggested that, under standard infection-control measures, transmissions contributed to a small part of *S. aureus* acquisitions (Price et al., 2014). Price et al. (2014) reported that just 14% of patients were potentially involved in transmission at an adult ICU and high-dependency unit. Our findings demonstrated that, among the 868 ICU admissions, 23 of 103 patients with lower-respiratory tract infections acquired MDR-AB through transmission. Several factors may have contributed to the higher frequency of transmission in our ICU. First, the primary diagnosis upon ICU admission was mainly pneumonia (44/70, 62.86%). In 27 of these patients MDR-AB were cultured. Second, the samples collected also different from those of previous studies, as diagnosing lower-respiratory tract infections was the main objective of our study. Third, although the number of beds in our ICU and the study duration were similar, length of stay in our ICU was significantly longer than others reported.

Since our hospital is a center of respiratory disease in southern China, one of the most common reasons for admission to ICU is critical patients cause by respiratory diseases. Half of the patients were transferred from other hospitals where they had been treated, and 38/70 patients (56.52%) carried MDR-AB at admission. Preventing the spread of MDR-AB was a big challenge for our ICU. In our

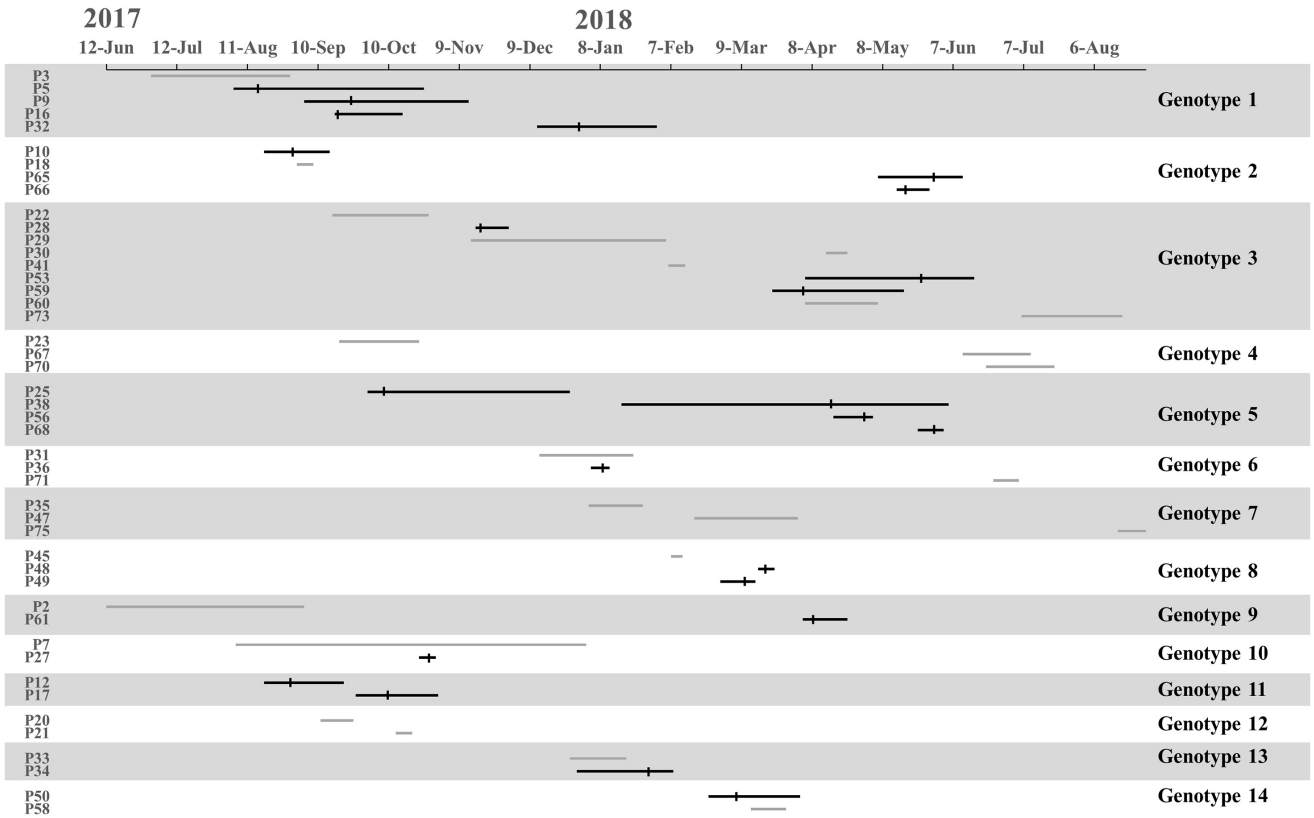


FIGURE 3 | Timelines and overlap of major MDR-AB genotypes. The line indicates stay periods in the ICU. The gray line represents patients in whom MDR-AB isolates were detected before ICU admission. Vertical bars indicate the time at which MDR-AB-positive samples were detected.

AB-10	AB-10												
AB-24	345	AB-24											
AB-9	377	526	AB-9										
AB-25	92	361	451	AB-25									
AB-36	269	384	286	257	AB-36								
AB-50	421	104	602	437	460	AB-50							
AB-52	25	352	378	111	272	428	AB-52						
AB-54	92	361	451	18	257	437	111	AB-54					
AB-59	335	18	516	351	374	94	342	351	AB-59				
AB-56	92	361	451	0	257	437	111	18	351	AB-56			
AB-66	4	345	377	92	269	421	25	92	335	92	AB-66		
AB-65	5	346	378	93	270	422	26	93	336	93	1	AB-65	
AB-68	92	361	451	0	257	437	111	18	351	0	92	93	AB-68

FIGURE 4 | Single-nucleotide polymorphisms matrices for the 13 MDR-AB acquisitions belong to clone-8, as typed by PFGE.

study, only two patients did not require ventilation, and most patients used antibiotics before admission into the ICU. Mechanical ventilation frequently requires tracheal intubation, which increases the risk of inhaling pathogenic bacteria or gastrointestinal tract bacteria (Coffin et al., 2008; De Waele et al., 2010). Prior use of antibiotics was demonstrated as an independent risk factor for colonization/infections with multi-resistant Gram-negative bacteria (Vasudevan et al., 2013). Along the ICU patients, bronchoscopy is a common diagnostic and therapeutic procedure which could cause nosocomial outbreak (McGrath et al., 2017).Those factors all pose risks for acquisition of MDR-AB and may contribute to transmission.

Among the 38 admission isolates, 16 MDR-AB isolates were highly related (SNV <10) and belonged to four genotypes, comprising seven isolates of patients transferred from another hospital and nine isolates from a different department in our hospital. These findings suggest that MDR-AB had disseminated in our hospital and spread widely

throughout the local Guangzhou area. The data also imply that MDR-AB adapted for persistence and transmission in a hospital environment.

Acinetobacter baumannii can survive in the environment for a long time and is potentially transmissible (Marchaim et al., 2007). Contaminated environments and equipment can act as reservoirs for MDR-AB (Chapartegui-Gonzalez et al., 2018). In our study, strains belonging to the same genotype isolated from different patients (such as genotype 1 and 5) were sustained for a long period, and the patients did not share the same time in the ICU. This indicated that some transmission events might have occurred indirectly *via* the contaminated environment or healthcare workers. These findings are consistent with previous research conducted in our ICU (Ye et al., 2015). Except for genotype 1 and 5, the remain genotypes involved in more than one patient did not persist for a longer time. In our study period, a bundle of infection-prevention measures was implemented, including hand hygiene, contact precautions, and cohorting with dedicated healthcare staff. All of the components of the infection prevention bundle played a role in controlling MDR-AB dissemination. Cohorting with dedicated healthcare staff was firstly implemented in 2015 in our ICU to control MDR-AB outbreak (Ye et al., 2015) and is still implemented up to now. This measure was recommended by the Centers for Disease Control (WHO, 2017) and Prevention and the European Centre for Disease Prevention and Control (Magiorakos et al., 2017) and was shown to be efficient (Abboud et al., 2016; Perez-Blanco et al., 2018). We suggested that cohort strategy maybe contribute to reduce the MDR-AB dissemination in our ICU.

Classifying pathogens to elucidate epidemiology of pathogenic bacteria and hospital outbreaks relies on typing techniques with higher discriminatory power. Using conventional epidemiological data, our study identified 32 potential acquisitions that involved 29 clones from more than two isolates identified through PFGE. When WGS data were included, we reduced the number of transmissions to 23. In comparison with PFGE ability to type MDR-AB, WGS separated isolates belonging to clone 8 into five different clades. Then, we examined the typing ability of MLST. Although each clade typed by WGS corresponded to one of the three predominant STs, ST208 was interspersed into three different clades. PFGE and MLST both show poor agreement with WGS. Our data indicated that WGS offers the advantage of resolving differences between closely related isolates (Zarrilli et al., 2013). Therefore, WGS was more suitable for typing and identifying transmission between patients, especially when the same pandemic ST isolates are identified regionally.

The PubMLST database assigned all strains to 12 STs that could be grouped into CC92 clonal complexes (also known as international clone II), except for ST229. The majority of isolates belonged to ST-208, which had spread throughout other provinces in China (Jiang et al., 2016; Ning et al., 2017). Between 2012 and 2015, ST457 was reported as a prevalent pathogen with enhanced virulence in five hospitals in southern China (Zhou et al., 2018). Our study identified just four strains belonging to ST457, three of which were isolated at the time of admission, and they were not closely related (>25 SNPs). These findings suggest our ICU was not associated with the new emerging clones.

This study had two major limitations. First, we did not test any environmental samples or healthcare workers' hands, sources that may have contributed to the spread of MDR-AB. Second, considering the sensitivity of bacterial culture, some changes in colonization status from culture-negative to culture-positive may have been false-positives. Hence, rate of acquisition may have been overestimated.

CONCLUSION

In conclusion, transmission mainly contributed to MDR-AB acquisition in our ICU; thus, prevention and control of MDR-AB hospital infections must be strengthened. Our study provides a high-resolution genome-wide perspective on MDR-AB epidemiology in a healthcare setting, while contributing to the development of appropriate intervention and prevention strategies.

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

PM and XD participated in data analysis and drafted the manuscript. LY and YW carried out the molecular genetic studies. YJ participated in the clinical sample isolation and antibiotic testing. RZ and CY managed the data collection. YX, XL, and YL participated in the study design. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (Grant Number 81770077) and the Natural Science Foundation of Guangdong Province, China (Grant Number 2017A030313712).

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.cn) for English language editing and Shanghai Biotechnology Corporation for whole-genome sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.715568/full#supplementary-material

Supplementary Table S1 | Whole-genome sequencing and assembly statistics of 70 isolates.

Supplementary Table S2 | General patient information and the results of drug-sensitivity testing.

REFERENCES

Abboud, C. S., de Souza, E. E., Zandonadi, E. C., Borges, L. S., Miglioli, L., Monaco, F. C., et al. (2016). Carbapenem-resistant *Enterobacteriaceae* on a cardiac surgery intensive care unit: successful measures for infection control. *J. Hosp. Infect.* 94, 60–64. doi: 10.1016/j.jhin.2016.06.010

Chapartegui-Gonzalez, I., Lazaro-Diez, M., Bravo, Z., Navas, J., Icardo, J. M., and Ramos-Vivas, J. (2018). *Acinetobacter baumannii* maintains its virulence after long-time starvation. *PLoS One* 13:e0201961. doi: 10.1371/journal.pone.0201961

Chinese Society of Critical Care Medicine (2013). Guideline for the prevention, diagnosis and treatment of ventilator-associated pneumonia (2013). *Zhonghua Nei Ke Za Zhi* 52, 524–543. doi: 10.3760/cma.j.issn.0578-1426.2013.06.024

Coffin, S., Klompas, M., Classen, D., Arias, K., Podgorny, K., Anderson, D., et al. (2008). Strategies to prevent ventilator-associated pneumonia in acute care hospitals. *Infect. Control Hosp. Epidemiol.* 29(Suppl. 1), S31–S40. doi: 10.1086/591062

De Waele, J., Ravyts, M., Depuydt, P., Blot, S., Decruyenaere, J., and Vogelaers, D. (2010). De-escalation after empirical meropenem treatment in the intensive care unit: fiction or reality? *J. Crit. Care* 25, 641–646. doi: 10.1016/j.jcrc.2009.11.007

Falagas, M. E., Bliziotis, I. A., and Siempos, I. I. (2006). Attributable mortality of *Acinetobacter baumannii* infections in critically ill patients: a systematic review of matched cohort and case-control studies. *Crit. Care* 10:R48. doi: 10.1186/cc4869

Halachev, M. R., Chan, J. Z. M., Constantinidou, C. I., Cumley, N., Bradley, C., Smith-Banks, M., et al. (2014). Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant *Acinetobacter baumannii*in Birmingham, England. *Genome Med.* 6:70. doi: 10.1186/s13073-014-0070-x

Hawkey, J., Ascher, D., Judd, L., Wick, R., Kostoulias, X., Cleland, H., et al. (2018). Evolution of carbapenem resistance in *Acinetobacter baumannii* during a prolonged infection. *Microb. Genom.* 4:e000165. doi: 10.1099/mgen.0.000165

Hwang, S. M., Cho, H. W., Kim, T. Y., Park, J. S., Jung, J., Song, K. H., et al. (2021). Whole-genome sequencing for investigating a health care-associated outbreak of carbapenem-resistant *Acinetobacter baumannii*. *Diagnostics* 11:201. doi: 10.3390/diagnostics11020201

Jiang, M., Liu, L., Ma, Y., Zhang, Z., Li, N., Zhang, F., et al. (2016). *Acinetobacter baumannii* molecular epidemiology of multi-drug resistant isolated in Shandong, China. *Front. Microbiol.* 7:1687. doi: 10.3389/fmicb.2016.01687

Kalenic, S., and Budimir, A. (2009). The role of microbiology laboratory in healthcare-associated infection prevention. *Int. J. Infect. Control* V5:i2. doi: 10.3396/ijic.V5i2.014.09

Karah, N., Sundsfjord, A., Towner, K., Samuelsen, Ø, antimicrobial, c.i, and chemotherapy, a (2012). Insights into the global molecular epidemiology of carbapenem non-susceptible clones of *Acinetobacter baumannii*. *Drug Resist. Updat.* 15, 237–247. doi: 10.1016/j.drup.2012.06.001

Magiorakos, A. P., Burns, K., Rodríguez Baño, J., Borg, M., Daikos, G., Dumpis, U., et al. (2017). Infection prevention and control measures and tools for the prevention of entry of carbapenem-resistant *Enterobacteriaceae* into healthcare settings: guidance from the European Centre for Disease Prevention and Control. *Antimicrob. Resist. Infect. Control* 6:113. doi: 10.1186/s13756-017-0259-z

Magiorakos, A. P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18, 268–281. doi: 10.1111/j.1469-0691.2011.03570.x

Makke, G., Bitar, I., Salloum, T., Panossian, B., Alousi, S., Arabaghian, H., et al. (2020). Whole-genome-sequence-based characterization of extensively drug-resistant *Acinetobacter baumannii* hospital outbreak. *mSphere* 5:e00934-19. doi: 10.1128/mSphere.00934-19

Marchaim, D., Navon-Venezia, S., Schwartz, D., Tarabeia, J., Fefer, I., Schwaber, M. J., et al. (2007). Surveillance cultures and duration of carriage of multidrug-resistant *Acinetobacter baumannii*. *J. Clin. Microbiol.* 45, 1551–1555. doi: 10.1128/JCM.02424-06

McGrath, B. A., Ruane, S., McKenna, J., and Thomas, S. (2017). Contamination of single-use bronchoscopes in critically ill patients. *Anaesthesia* 72, 36–41. doi: 10.1111/anae.13622

Ning, N., Liu, X., Bao, C., Chen, S., Cui, E., Zhang, J., et al. (2017). Molecular epidemiology of bla -producing carbapenem-resistant *Acinetobacter baumannii* in a single institution over a 65-month period in north China. *Infect. Drug Resist.* 17:14. doi: 10.1186/s12879-016-2110-1

Nishi, I., and Hidaka, Y. (2016). [The role of the microbiology laboratory in healthcare-associated infection control]. *Rinsho Byori* 64, 338–344.

Perez-Blanco, V., Redondo-Bravo, L., Ruiz-Carrascoso, G., Pano-Pardo, J. R., Gomez-Gil, R., Robustillo-Rodela, A., et al. (2018). Epidemiology and control measures of an OXA-48-producing *Enterobacteriaceae* hospital-wide oligoclonal outbreak. *Epidemiol. Infect.* 146, 656–662. doi: 10.1017/S0950268818000249

Price, J. R., Golubchik, T., Cole, K., Wilson, D. J., Crook, D. W., Thwaites, G. E., et al. (2014). Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of Staphylococcus aureus in an intensive care unit. *Clin. Infect. Dis.* 58, 609–618. doi: 10.1093/cid/cit807

Salipante, S. J., SenGupta, D. J., Cummings, L. A., Land, T. A., Hoogestraat, D. R., and Cookson, B. T. (2015). Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *J. Clin. Microbiol.* 53, 1072–1079. doi: 10.1128/JCM.03385-14

Seifert, H., Dolzani, L., Bressan, R., van der Reijden, T., van Strijen, B., Stefanik, D., et al. (2005). Standardization and interlaboratory reproducibility assessment of pulsed-field gel electrophoresis-generated fingerprints of *Acinetobacter baumannii*. *J. Clin. Microbiol.* 43, 4328–4335. doi: 10.1128/JCM.43.9.4328-4335.2005

Singh, A., Goering, R. V., Simjee, S., Foley, S. L., and Zervos, M. J. (2006). Application of molecular techniques to the study of hospital infection. *Clin. Microbiol. Rev.* 19, 512–530. doi: 10.1128/CMR.00025-05

Thong, K. L., Goh, Y. L., Radu, S., Noorzaleha, S., Yasin, R., Koh, Y. T., et al. (2002). Genetic diversity of clinical and environmental strains of *Salmonella enterica* serotype Weltevreden isolated in Malaysia. *J. Clin. Microbiol.* 40, 2498–2503. doi: 10.1128/jcm.40.7.2498-2503.2002

Tomaschek, F., Higgins, P. G., Stefanik, D., Wisplinghoff, H., and Seifert, H. (2016). Head-to-Head comparison of two multi-locus sequence typing (MLST) schemes for characterization of *Acinetobacter baumannii* outbreak and sporadic isolates. *PLoS One* 11:e0153014. doi: 10.1371/journal.pone.0153014

Vasudevan, A., Mukhopadhyay, A., Goh, E. Y., Li, J., and Tambyah, P. A. (2013). Risk factors for infection/colonization caused by resistant Gram negative bacilli in critically ill patients (an observational study of 1633 critically ill patients). *Prev. Med.* 57, S70–S73. doi: 10.1016/j.ypmed.2012.12.003

Vazquez-Lopez, R., Solano-Galvez, S. G., Juarez Vignon-Whaley, J. J., Abello Vaamonde, J. A., Padro Alonzo, L. A., Rivera Resendiz, A., et al. (2020). *Acinetobacter baumannii* resistance: a real challenge for clinicians. *Antibiotics* 9:205. doi: 10.3390/antibiotics9040205

WHO (Eds) (2017). "WHO Guidelines Approved by the Guidelines Review Committee," in *Guidelines for the Prevention and Control of Carbapenem-Resistant Enterobacteriaceae, Acinetobacter baumannii and Pseudomonas aeruginosa in Health Care Facilities*, (Geneva: World Health Organization Copyright © World Health Organization 2017).

Ye, D., Shan, J., Huang, Y., Li, J., Li, C., Liu, X., et al. (2015). A gloves-associated outbreak of imipenem-resistant *Acinetobacter baumannii* in an intensive care unit in Guangdong, China. *BMC Infect. Dis.* 15:179. doi: 10.1186/s12879-015-0917-9

Zarrilli, R., Pournaras, S., Giannouli, M., and Tsakris, A. (2013). Global evolution of multidrug-resistant *Acinetobacter baumannii* clonal lineages. *Int. J. Antimicrob. Agents* 41, 11–19. doi: 10.1016/j.ijantimicag.2012.09.008

Zhou, K., Tang, X., Wang, L., Guo, Z., Xiao, S., Wang, Q., et al. (2018). An emerging clone (ST457) of *Acinetobacter baumannii* clonal complex 92 with enhanced virulence and increasing endemicity in South China. Clin. Infect. Dis. 67, S179–S188. doi: 10.1093/cid/ciy691

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.

frontiers in Medicine

BRIEF RESEARCH REPORT
published: 27 October 2021
doi: 10.3389/fmed.2021.733676

A New Tool for Analyses of Whole Genome Sequences Reveals Dissemination of Specific Strains of Vancomycin-Resistant *Enterococcus faecium* in a Hospital

*Lis Marbjerg[1], Caroline Louise Stougaard[1], Sophie-Amalie Grønhøj Sørensen[1], Amalie Vørs Thomsen[1], Lis Wang[1], Lise Andersen[2], Thomas Emil Andersen[1,2], Birgitte Kallipolitis[3] and Michael Kemp[4,5]**

[1] Department of Clinical Research, University of Southern Denmark, Odense, Denmark, [2] Department of Clinical Microbiology, Odense University Hospital, Odense, Denmark, [3] Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark, [4] Department of Regional Health Research, University of Southern Denmark, Odense, Denmark, [5] Regional Department of Clinical Microbiology, Zealand University Hospital, Slagelse, Denmark

OPEN ACCESS

Edited by:
Yann Felix Boucher, National University of Singapore, Singapore

Reviewed by:
Paulo J. M. Bispo, Harvard Medical School, United States
Andrei Nicoli Gebieluca Dabul, University of São Paulo, Brazil

***Correspondence:**
Michael Kemp
mkemp@regionsjaelland.dk

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

Received: 30 June 2021
Accepted: 05 October 2021
Published: 27 October 2021

Citation:
Marbjerg L, Stougaard CL, Sørensen S-AG, Thomsen AV, Wang L, Andersen L, Andersen TE, Kallipolitis B and Kemp M (2021) A New Tool for Analyses of Whole Genome Sequences Reveals Dissemination of Specific Strains of Vancomycin-Resistant Enterococcus faecium in a Hospital. Front. Med. 8:733676. doi: 10.3389/fmed.2021.733676

A new easy-to-use online bioinformatic tool analyzing whole genome sequences of healthcare associated bacteria was used by a local infection control unit to retrospectively map genetic relationship of isolates of *E. faecium* carrying resistance genes to vancomycin in a hospital. Three clusters of isolates were detected over a period of 5 years, suggesting transmission between patients. Individual relatedness between isolates within each cluster was established by SNP analyses provided by the system. Genetic antimicrobial resistance mechanisms to antibiotics other than vancomycin were identified. The results suggest that the system is suited for hospital surveillance of *E. faecium* carrying resistance genes to vancomycin in settings with access to next Generation Sequencing without bioinformatic expertise for interpretation of the genome sequences.

Keywords: vancomycin resistant enterococci, whole genome sequencing, bioinformatics, genomic epidemiology, infection control, surveillance

INTRODUCTION

Surveillance of specific microorganisms in hospitals and evaluation of genetic relationship of isolates is fundamental in detecting and interrupting transmission. Whole Genome Sequencing (WGS) offers a unique tool for analyzing relationship of vancomycin resistant enterococci (1–3), and may serve as a method for typing the bacteria for local surveillance. While bioinformatic analyses of WGS data have generally been hampered by complicated procedures, tools requiring no bioinformatic skills are now available to support infection control. Here, a local infection control unit retrospectively analyzed all available genome sequences of *Enterococcus faecium* carrying resistance genes to vancomycin isolated at a hospital from 2014 to 2018 by a new online tool, 1928 Diagnostics. Based on WGS data 1928 Diagnostics analyze core genome genetic relationship and genetic antibiotic resistance mechanisms in individual bacterial isolates (4).

MATERIALS AND METHODS

Clinical Isolates

The study was carried out at Odense University Hospital, which is a >1,000 bed university hospital placed at two geographical locations. Isolates of *E. faecium* carrying *vanA* and/or *vanB* gene from clinical (non-screening) samples were analyzed. The first isolate from each patient in the years 2014 to 2018 was included. Date and place of sampling was recorded. No patient information was included. Core genome Multi Locus Sequencing Typing (cgMLST) of some of the included isolates have previously been reported (5, 6).

Whole Genome Sequencing Data Analyses

WGS was carried out by 2 × 150 bp paired end sequencing using a MiSeq instrument (Illumina, San Diego, CA, USA). Compressed unassembled sequence files were uploaded to the 1,928 Diagnostics pipeline (https://www.1928diagnostics.com/). The system initially performed a quality check based on sequencing depth/coverage. Results for individual isolates processed by 1,928 Diagnostics were presented with MLST sequence type (ST) and antibiotic resistance gene profiles. A cgMLST based Unweighted Pair Group Method with Arithmetic mean (UPGMA) dendrogram showing number of allelic differences was generated by the system. One thousand nine hundred twenty-eight Diagnostics uses the resqu database for identification of horizontally transferred antibiotic resistance genes (https://www.1928diagnostics.com/resdb/).

Isolates with 20 or less allelic differences were considered genetically related (7). Clusters of more than three genetically related isolates were studied in detail by the single nucleotide polymorphism (SNP) analysis function in 1928 Diagnostics. Closest Genbank complete chromosome sequence, identified by KmerResistance (https://cge.cbs.dtu.dk/services/kmerresistance/) using sequence files from a randomly selected isolate from each cluster, served as reference. The exclusion distance was set to 10.

RESULTS

MLST

A total of 64 genomes from *E. faecium* were available for analyses.

Fifty-eight isolates carried the *vanA* gene, five carried *vanB* gene and one isolate from December 2018 carried both *vanA* and *vanB* gene (**Figure 1**). One isolate was not typeable by MLST. This was confirmed by MLST-2.0 (https://cge.cbs.dtu.dk/services/MLST/), which failed to detect hits in the pstS locus. The most common MLST type was ST80, including 30 isolates of which 26 carried *vanA* gene, three carried *vanB* gene, and the one isolate carrying both *vanA* and *vanB* genes. The UPGMA dendrogram showed large allelic variation in the core genome of the ST80 isolates (**Figure 1**). Twelve isolates were ST203 and eight were ST1421.

Core Genome MLST, Epidemiology, and Antibiotic Resistance Gene Profiles

Three clusters were evident form the cgMLST based UPGMA dendrogram (**Figure 1**). Cluster 1 consisted of eight isolates of *vanA* gene positive *E. faecium* ST1421. All isolates in this cluster were from the second half of 2018 (**Figure 2**). The SNP analyses (**Figure 3**) showed few SNP differences between isolates. In addition to the *vanA* gene, the isolates all carried resistance genes to aminoglycosides [*aac(6')*, *ant(9')-la*, *aph(3')-IIIa*, macrolide-lincosamide-streptogramin B (MLS), *erm(A)*, *erm(B)*, *msr(C)*, and trimethoprim *dfr(G)*]. All resistance genes had more than 98.5% amino acid match and 100% length match compared to the reference sequence. Six isolates were cultured from patients at five hospital departments and two isolates were from patients consulting General Practitioners.

Cluster 2 included 10 ST203 isolates carrying *vanA* gene. The bacteria were cultured from first half of 2017 onwards (**Figure 2**). The isolates were cultured from patients at seven hospital departments. The isolates in Cluster 2 SNP analyses showed larger variation compared to the isolates in cluster 1 (**Figure 3**). All isolates in this cluster carried resistance genes to aminoglycosides [*aac(6')*, *ant(6')-la*, MLS *erm(B)*, *msr(C)*, trimethoprim (*dfrG*), and tetracycline *tet(M)*, *tet(U)*]. Except for one isolate with a length fraction of 94.3% of the reference sequences, all resistance genes had more than 98.5% amino acid identity and 100% length to the references. The *catA* gene, an effector of chloramphenicol resistance, was detected in seven of the isolates. The amino acid sequence and length were both 100% identical to the reference sequence.

Cluster 3 included 17 *vanA* gene positive isolates. Sixteen of the isolates were typed as ST80, while one was not typeable. MLST 2.0 (https://cge.cbs.dtu.dk/services/MLST/) also failed to type this isolate. The isolates in cluster 3 were cultured from 2016 onwards and showed numbers of SNP differences between those detected in Cluster 1 and Cluster 2. The isolates carried resistance genes to aminoglycosides [*aac(6')*, *aph(3')-IIIa*, MLS *erm(B)*, *msr(C)*, and tetracyclines *tet(M)*]. With three exceptions both identity and fragment length were more than 98% of the reference sequences for all of these resistance genes. The *E. faecium* in this cluster were collected from six departments and one GP from 2016 onwards.

DISCUSSION

WGS offers a unique method for typing and monitoring healthcare associated bacteria such as enterococci carrying resistance genes to vancomycin. For the technique to be useful in clinical practice, the bioinformatic handling of sequence data should be easy and fast. Access to external sequencing facilities may provide high quality genome sequences without data analyses and interpretation. We used 1928 Diagnostics, a new online tool for bioinformatic analyses of transmissible bacteria, to review the genetic characteristics of *E. faecium* carrying genes encoding resistance to vancomycin at the hospital during a period of 5 years.

The genomes of the ST80 isolates were highly heterogenous, including isolates carrying *vanA* gene and isolates with *vanB* gene. One isolate of ST80 *E. faecium* carried both *vanA* and *vanB* genes as did 5% of vancomycin-resistant *E. faecium* in Denmark

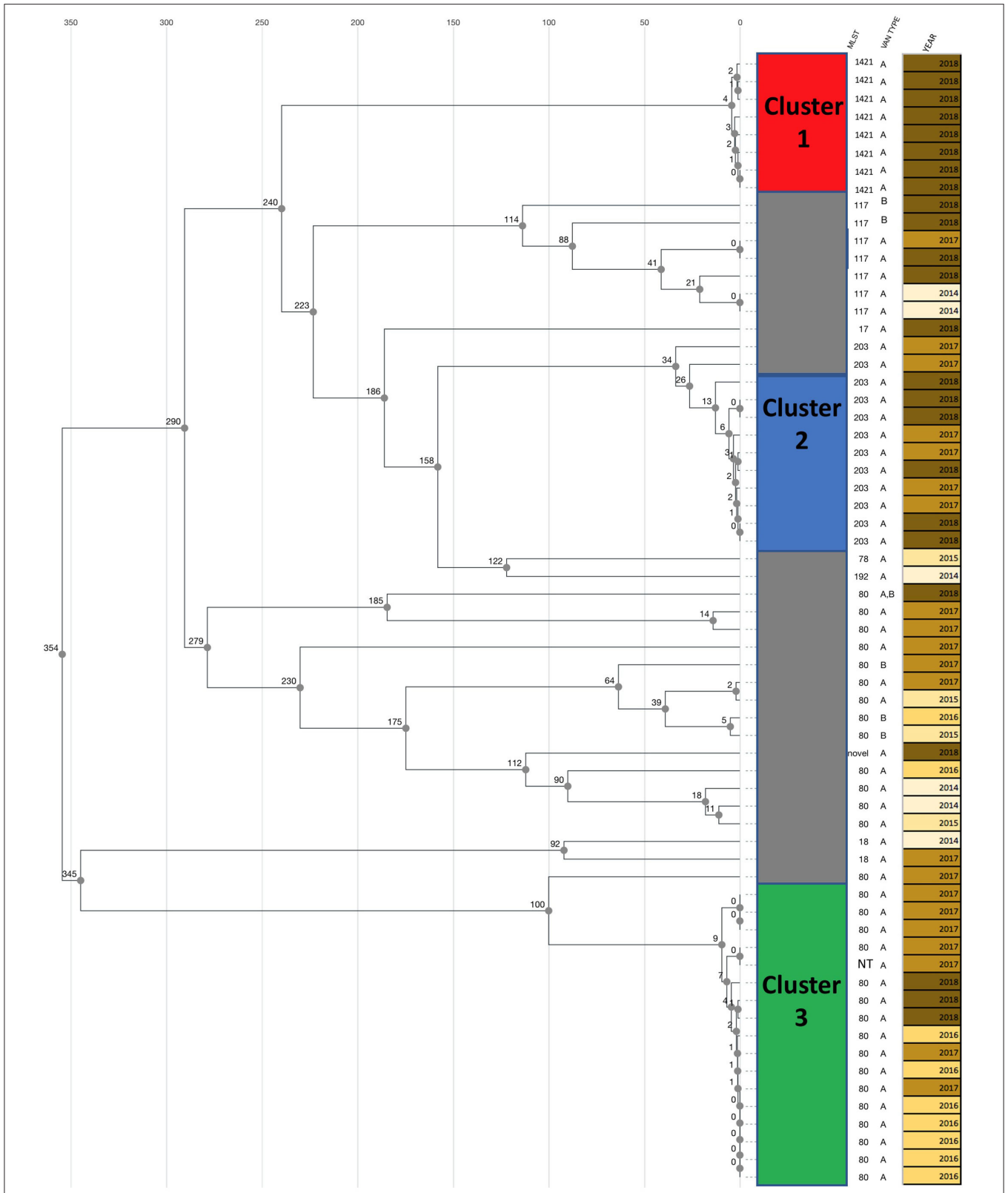


FIGURE 1 | cgMLST based UPGMA clustering dendrogram of clinical (non-screening) *E. faecium* carrying resistance genes to vancomycin isolated 2014–2018 isolated at a large university hospital. Three clusters of genetically related isolates were identified. The number of allelic differences between individual isolates are indicated at each node. The MLST sequence type and the van gene (*vanA* or *vanB* gene) is indicated for each isolate next to the dendrogram. One isolate was not typable (NT). One isolate carried both *vanA* and *vanB* gene. Year of isolation is indicated for individual isolates.

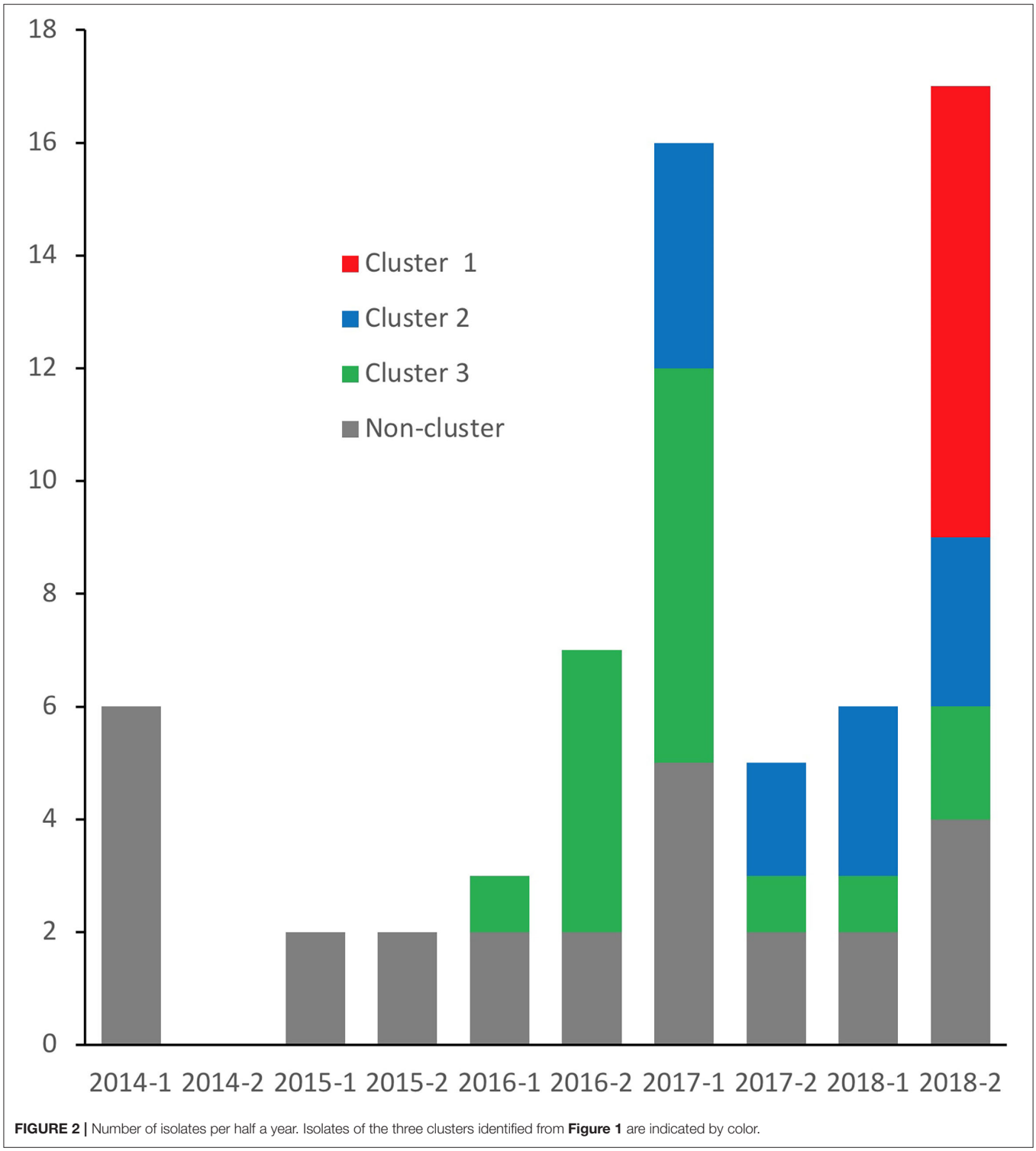


FIGURE 2 | Number of isolates per half a year. Isolates of the three clusters identified from **Figure 1** are indicated by color.

in 2018 (5). Heterogenous ST80 vancomycin-esistant *E. faecium* has been reported from Denmark (8). Recent studies of ST80 vancomycin-resistant *E. faecium* in Sweden showed comparable epidemiological discrimination by 1,928 Diagnostics cgMLST and SNP analyses (9).

Three major clusters of related isolates were evident. The clusters replaced each other over time as the most frequently isolated strain and isolates of the strains within each cluster continued to be detected afterwards. The Cluster 1 strain of *E. faecium* ST1421 carried the *vanA* gene without necessarily

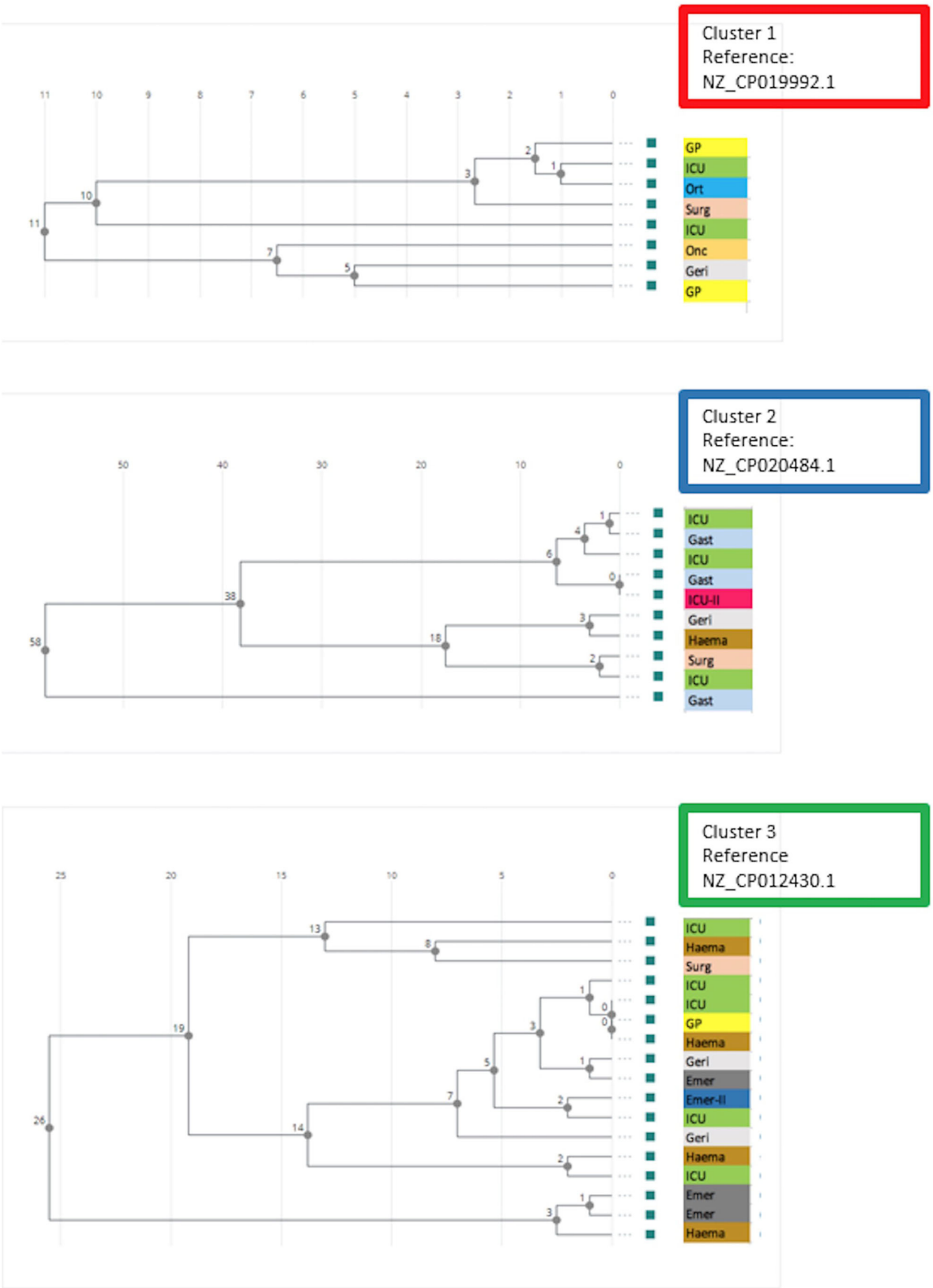


FIGURE 3 | Phylogenetic trees based on SNPs in the core genome of *E. faecium* carrying *vanA* gene in the three clusters from **Figure 1**. Reference chromosome genome used for the analysis is indicated for each cluster. Location for sampling (General Practitioner or hospital department) is indicated for each isolate. ICU, Intensive Care Unit (hospital location 1); Ort, orthopedic surgery; Surg, abdominal surgery; Onc, oncology; Geri, geriatrics; GP, General Practitioner; Gast, gastroenterology; ICU-II, Intensive Care Unit at second hospital location; Haema, hematology; Emer, emergency; Emer-II, emergency at second location.

expressing resistance to vancomycin *in vitro* (10). This strain was frequently detected in Denmark in 2018 and 2019 (5). Cluster 2 and Cluster 3 also reflected the nation-wide occurrence of dominating strains of *E. faecium* carrying *vanA* gene during the 5 years (5). Genes encoding resistance to antibiotics other than vancomycin varied between clusters and were largely identical in isolates within each cluster. Knowledge on mechanisms of resistance to antimicrobials other than vancomycin may potentially contribute to treatment guidance and to tracking routes of transmission of antimicrobial resistance.

The detailed relatedness between isolates within each cluster was established by SNP analyses. No association between place of sampling and relatedness of isolates was evident. However, location of sampling does often not reflect place of transmission. It is noteworthy, that bacteria of all three clusters were isolated from patients referred to the same Intensive Care Unit and that 10 of the isolates were from patients admitted to this unit. This may be due to patient factors, use of broad-spectrum antimicrobials, extensive use of indwelling foreign body materials, and frequent sampling practice.

In conclusion, this retrospective study showed that a rapid and easy-to-use bioinformatic pipeline clearly separated isolates of *E. faecium* with resistance genes to vancomycin into clusters and sporadic strains and allowed detailed analyses of clusters of related isolates. Prospective studies should establish the usefulness of 1928 Diagnostics and similar systems for real time surveillance of vancomycin resistant *E. faecium* and other hospital-associated bacteria.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found at: https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA767758.

AUTHOR CONTRIBUTIONS

MK, LM, and LA conceived and designed the study. MK, LM, CS, S-AS, AT, and LW analyzed the data. MK, LM, CS, S-AS, AT, LA, TA, and BK critically revised the manuscript. All authors read and approved the manuscript.

FUNDING

The study was supported by a research grant from Odense University Hospital (Grant No.: 72-A3862).

ACKNOWLEDGMENTS

Statens Serum Institut, Copenhagen, Denmark is thanked for sharing genome sequence data.

REFERENCES

1. Neumann B, Bender JK, Maier BF, Wittig A, Fuchs S, Brockmann D, et al. Comprehensive integrated NGS-based surveillance and contact-network modeling unravels transmission dynamics of vancomycin-resistant enterococci in a high-risk population within a tertiary care hospital. *PLoS ONE.* (2020) 15:e0235160. doi: 10.1371/journal.pone.0235160
2. Xanthopoulou K, Peter S, Tobys D, Behnke M, Dinkelacker AG, Eisenbeis S, et al. Vancomycin-resistant *Enterococcus faecium* colonizing patients on hospital admission in Germany: prevalence and molecular epidemiology. *J Antimicrob Chemother.* (2020) 75:2743–51. doi: 10.1093/jac/dkaa271
3. Pinholt M, Gumpert H, Bayliss S, Nielsen JB, Vorobieva V, Pedersen M, et al. Genomic analysis of 495 vancomycin-resistant *Enterococcus faecium* reveals broad dissemination of a vanA plasmid in more than 19 clones from Copenhagen, Denmark. *J Antimicrob Chemother.* (2017) 2017:7240–7. doi: 10.1093/jac/dkw360
4. Slott Jensen ML, Chen M, Klitgaard JK, Andersen TE, Kemp M. Whole genome sequence analyses by a new easy-to-use software solution support suspicion of a neonatal ward outbreak of MRSA ST22 and transmission between hospitals. *Infect Control Hospital Epidemiol.* (2021) 2021:123. doi: 10.1017/ice.2021.123
5. Hammerum AM, Justesen US, Pinholt M, Roer L, Kaya H, Worning P, et al. Surveillance of vancomycin-resistant enterococci reveals shift in dominating clones and national spread of a vancomycin-variable *vanA Enterococcus faecium* ST1421-CT1134 clone, Denmark, 2015 to March 2019. *Euro Surveill.* (2019) 24:1900503. doi: 10.2807/1560-7917.ES.2019.24.34.1900503
6. Hansen SK, Andersen L, Detlefsen M, Holm A, Roer L, Antoniades P, et al. Use of core genomes MLST for infection control during a Danish outbreak of vancomycin-resistant *Enterococcus faecium*. *J Glob Antimicrob Resist.* (2021) 24:418–23. doi: 10.1016/j.jgar.2021.02.007
7. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, et al. Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. *J Clin Microbiol.* (2015) 53:3788–97. doi: 10.1128/JCM.01946-15
8. Pinholt M, Bayliss SC, Gumpert H, Worning P, Jensen VVS, Pedersen M, et al. WGS of 1058 *Enterococcus faecium* from Copenhagen, Denmark, reveals rapid clonal expansion of vancomycin-resistant clone ST80 combined with widespread dissemination of a vanA-containing plasmid and acquisition of a heterogeneous accessory genome. *J Antimicrob Chemother.* (2019) 74:1776–85. doi: 10.1093/jac/dkz118
9. Dyrkell F, Giske CG, Fang H. Epidemiological typing of ST80 vancomycin-resistant *Enterococcus faecium*: core-genome multi-locus sequence typing vs. single nucleotide polymorphism-based typing. *J Glob Antimicrob Resist.* (2021) 25:119–23. doi: 10.1016/j.jgar.2021.03.005
10. Hansen TA, Pedersen MS, Nielsen LG, Ma CMG, Søes LM, Worning P, et al. Emergence of a vancomycin-variable *Enterococcus faecium* ST1421 strain containing a deletion in vanX. *J Antimicrob Chemother.* (2018) 73:2936–40. doi: 10.1093/jac/dky308

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 November 2021
doi: 10.3389/fmicb.2021.739621

Cultivation and Genomic Characterization of the Bile Bacterial Species From Cholecystitis Patients

Qiulong Yan[1,2†], Siyi Zhang[1†], Shenghui Li[3†], Guangyang Wang[1,4], Aiqin Zhang[3], Taiyang Jin[1], Yue Zhang[3], Qingbo Lv[3], Manchun Xiao[1], Yuanyuan Sun[1], Xiang Li[1], Song Cui[1], Rui Li[1], Xiaochi Ma[5], Chao Wang[5], Xiangge Tian[5], Xiaohui Duan[2], Yi Xin[4], Xianhai Mao[2]* and Yufang Ma[1]**

[1] Department of Microbiology, College of Basic Medical Sciences, Dalian Medical University, Dalian, China, [2] Department of Hepatobiliary Surgery, Hunan Provincial People's Hospital, The First Affiliated Hospital of Hunan Normal University, Changsha, China, [3] Puensum Genetech Institute, Wuhan, China, [4] Department of Biotechnology, College of Basic Medical Sciences, Dalian Medical University, Dalian, China, [5] College of Pharmacy, Dalian Medical University, Dalian, China

The microbes in human bile are closely related to gallbladder health and other potential disorders. Although the bile microbial community has been investigated by recent studies using amplicon or metagenomic sequencing technologies, the genomic information of the microbial species resident in bile is rarely reported. Herein, we isolated 138 bacterial colonies from the fresh bile specimens of four cholecystitis patients using a culturome approach and genomically characterized 35 non-redundant strains using whole-genome shotgun sequencing. The bile bacterial isolates spanned 3 classes, 6 orders, 10 families, and 14 genera, of which the members of *Enterococcus*, *Escherichia–Shigella*, *Lysinibacillus*, and *Enterobacter* frequently appeared. Genomic analysis identified three species, including *Providencia* sp. D135, *Psychrobacter* sp. D093, and *Vibrio* sp. D074, which are not represented in existing reference genome databases. Based on the genome data, the functional capacity between bile and gut isolates was compared. The bile strains encoded 5,488 KEGG orthologs, of which 4.9% were specific to the gut strains, including the enzymes involved in biofilm formation, two-component systems, and quorum-sensing pathways. A total of 472 antibiotic resistance genes (ARGs) were identified from the bile genomes including multidrug resistance proteins (42.6%), fluoroquinolone resistance proteins (12.3%), aminoglycoside resistance proteins (9.1%), and β-lactamase (7.2%). Moreover, *in vitro* experiments showed that some bile bacteria have the capabilities for bile salt deconjugation or biotransformation (of primary bile acids into secondary bile acids). Although the physiological or pathological significance of these bacteria needs further exploration, our works expanded knowledge about the genome, diversity, and function of human bile bacteria.

Keywords: bile bacteria, genomic characterization, antibiotic resistance genes, virulence factors, culturome, cholecystitis patients

OPEN ACCESS

Edited by:
Martin Maiers, National Marrow Donor Program, United States

Reviewed by:
Babak Haghshenas, Kermanshah University of Medical Sciences, Iran
Maryam Tidjani Alou, Aix Marseille Université, France

***Correspondence:**
Yi Xin
jimxin@hotmail.com
Xianhai Mao
maoxianhai@hunnu.edu.cn
Yufang Ma
yufangma@dmu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *11 July 2021*
Accepted: *07 October 2021*
Published: *01 November 2021*

Citation:
Yan Q, Zhang S, Li S, Wang G, Zhang A, Jin T, Zhang Y, Lv Q, Xiao M, Sun Y, Li X, Cui S, Li R, Ma X, Wang C, Tian X, Duan X, Xin Y, Mao X and Ma Y (2021) Cultivation and Genomic Characterization of the Bile Bacterial Species From Cholecystitis Patients. Front. Microbiol. 12:739621. doi: 10.3389/fmicb.2021.739621

INTRODUCTION

Cholecystitis leads to gallbladder inflammation or even perforation, tissue death, gangrene, fibrosis, and shrinking of the gallbladder, which is a hospitalized disease with increasing medical and financial burden (Wadhwa et al., 2017; Wang et al., 2021). The cardinal symptoms were fever, pain, and nausea with increasing levels of clinical indicators, such as serum transaminases, alkaline phosphatase, and total bilirubin (Chisholm et al., 2019). Laparoscopic cholecystectomy was commonly applied for treating this disease (Loozen et al., 2018).

Although the pathogenicity in cholecystitis still needs further exploration, bacteria, as one of the important risk factors, were closely correlated with its poor operative outcomes (Liu et al., 2015; Yun and Seo, 2018). Moreover, positive bile culture rates from laparoscopic cholecystectomy patients ranged from 25.1 to 60.3%, suggesting the presence of bacteria in the cholecystitis patients' bile (Nitzan et al., 2017; Yun and Seo, 2018). Due to the limitation of culture conditions, most bile bacterial isolates from clinical samples are aerobic bacteria especially Gram stain-negative Enterobacteriaceae, such as *Escherichia coli*, *Klebsiella* spp., *Enterobacter* spp., and less frequently anaerobic bacteria, including *Bacteroides* and *Clostridium* spp., as well as microaerophilic *Helicobacter pylori* (Nitzan et al., 2017; Cen et al., 2018). For culture-independent methods, denaturing gradient gel electrophoresis (DGGE) was applied for the investigation of microbiota composition (Liu et al., 2015); however, low throughput and complex operation were its limitations. With the advent of high-throughput sequencing platforms and bioinformatic analysis pipelines, metagenome sequencing analysis including 16S rRNA amplicon sequencing and whole genomic shotgun sequencing have revealed the abundant diversity composition of the bile microbiome (Wu et al., 2013; Liu et al., 2015). Meanwhile, owing to nearly 80% of the human intestinal bacteria were unculturable (Tajeddin et al., 2016), with the advancement in bacterial culturomics technology (Tidjani Alou et al., 2020), it is reasonable to infer that abundant bile bacteria are waiting to be isolated.

Proper identification method of the isolated bacteria was a highly essential step in the culture-dependent workflow. The traditional polyphasic taxonomic strategy aims to generate phenotypic and phylogenetic information of a bacterium, such as its shape, color, size, staining properties, host range, pathogenicity, indole test, gelatin liquefaction test, motility test, and assimilation of carbon sources (Das et al., 2014). In addition to biochemical identification, flow cytometry and MALDI-TOF mass spectrometry can also be used for rapid identification for cultured isolates (Das et al., 2014; Jang and Kim, 2018). Despite each approach having its own advantages and limitations, such as higher cost or rough results of flow cytometer and equipment popularity of MALDI-TOF mass spectrum hindered their application. Considerable advancement in culturable bacterial identification and taxonomy methods using genotypic and phylogenetic information has taken place such as chemotaxonomy, numerical taxonomy, and DNA–DNA hybridization, housekeeping gene amplification, and sequencing (16S rRNA gene, *gyrB*, *rpoABCD*, etc.) and whole-genome sequencing (Das et al., 2014). As the primary gene target for bacterial identification and comprehensive consideration of cost performance, the 16S rRNA gene was often used for preliminary screening. Then whole-genome sequencing was used for in-depth exploration of functional genes. In recent years, cultivation-dependent microbiota studies of the human digestive tract (Browne et al., 2016; Lagier et al., 2016), respiratory tract (Fonkou et al., 2018), genitourinary tract (Thomas-White et al., 2018), oral cavity (Fonkou et al., 2018; Martellacci et al., 2020), and even in other model animals (Lagkouvardos et al., 2016; Liu et al., 2020) mostly adopted this strategy and had isolated hundreds of previously unknown bacteria inhabiting the human body. In addition, live bacteria could facilitate the follow-up mechanism researches regarding their impact on physiological and pathological functions, such as the bile salt (salt ion form of bile acid) metabolism test. Bile acids are cholesterol-derived natural surfactants and are regarded as important digestive hormones that are produced in the liver and secreted into the duodenum via the gallbladder to regulate numerous physiological processes in the host (Harris et al., 2018). They are known to affect lipid digestion, antibacterial defense, glucose metabolism, host metabolism, cancer progression, and innate and adaptive immunity (Shapiro et al., 2018; Hang et al., 2019; Song et al., 2020). The biotransformation of bile acids (BAs) by intestinal bacteria may have an important role in cholesterol gallstone formation and colon carcinogenesis via TGR and FXR (farnesoid-X receptor) receptors (Salvioli et al., 1982; Funabashi et al., 2020). So far, bacteria are mainly involved in two categories of bile acid metabolism. Bile salt hydrolase (BSH) with deconjugation activity hydrolyzes the amide bond that links the bile acid side chain to glycine or taurine and takes part in the transformation process from conjugated bile acid to unconjugated bile acid (Ridlon et al., 2006). Besides, the dehydroxylase activity participates in the transformation process from primary bile acid to secondary bile acid (White et al., 1980). These metabolic activities have mostly been found in intestinal bacteria (Kuipers et al., 2020). However, the ability of bile acid biotransformation by bile bacteria has not been reported to date.

Altogether, the objectives of the study are to (1) cultivate and isolate a large number of bile bacteria of cholecystitis patients and whole-genome sequence parts of isolates, (2) conduct comparative analyses between our isolates and gut isolates in public databases, which revealed different pathways in our isolates, and (3) evaluate the bile salt metabolic capacity of biliary bacteria *in vitro*. This work provided new insights into the bile bacterial communities based on genomes and the biotransformation potentiality of human bile bacteria.

MATERIALS AND METHODS

Sample Collection

Informed consents were obtained from all subjects, and this study was approved by the Medical Ethics Committee of Hunan Provincial People's Hospital. The study was conducted according to the principles of the Declaration of Helsinki of 1964 and later versions (2013) and was approved by the Medical

Ethics Committee of the Hunan Provincial People's Hospital (Ethical approval number: 2020-54). All the experiments were performed in accordance with relevant guidelines and regulations. The bile samples collected from four postoperative (laparoscopic cholecystectomy) acute cholecystitis patients in Hunan Provincial People's Hospital were kept in an anaerobic gas bag (Mitsubishi, Japan), respectively, and rapidly transferred to the laboratory for cultivation. To address the potential contamination that might have occurred before and after the samples reached our laboratory, sterile physiological saline control samples without bile were applied in the study.

Cultivation and Identification of Bile Species

All bile samples were centrifuged at 3,000 × *g* for 5 min and carefully removed most of the supernatant of the bile sample but left around 200 μl of supernatant on the bottom to resuspend the pellet to keep bacteria with the highest density. These bacteria were then spread on different solid media for culture. [BBL—Brain Heart Infusion Agar (Becton Dickinson, United States), *Bifidobacterium* Medium (Becton Dickinson, United States), Lactose Bile (Becton Dickinson, United States), Blood Agar Base Infusion Agar (Becton Dickinson, United States), Columbia Blood Agar Base (Becton Dickinson, United States), BBE Agar (Becton Dickinson, United States), and Gifu Anaerobic Medium (Beijing Land Bridge Technology Co., Ltd, China)] under aerobic and anaerobic conditions at 37°C. Cultivation and identification methods of bile species mainly referred to two studies (Browne et al., 2016; Lagier et al., 2016). Briefly, colonies were picked up and transferred on a fresh medium for the purification of bacterial colony. Isolated colonies were transferred into an 8-ml liquid medium for enrichment culture and were identified using PCR amplification of the 16S rRNA gene (primers: 7F 5′-AGAGTTTGATYMTGGCTCAG-3′; 1510R 5′-ACGGYTACCTTGTTACGACTT-3′) (Browne et al., 2016). The PCR products were applied on Sanger sequencing at ABI 3730XL platform (Applied Biosystems, United States) (Browne et al., 2016). Each 16S rRNA sequence was blasted against the rRNA/ITS databases of the National Center of Biotechnology Information (NCBI) for bacterial strain identification. The identified aerobic strains were preserved in 30% glycerol and anaerobic strains in 30% glycerol and 0.1% cysteine at a −80°C freezer.

Whole-Genome Shotgun Sequencing, Assembly, and Annotation

The DNA of bacterial isolates was extracted using the Qiagen DNA extraction kit (Qiagen, Germany) according to the protocols of the manufacturer. The DNA concentration and purity were determined by NanoDrop2000. DNA quality was examined with a 1% agarose gel electrophoresis. Bacterial DNA was fragmented to an average size of ∼300 bp using Covaris M220 (Gene Company Limited, China). Paired-end libraries were prepared by using a TruSeq DNA sample prep kit (Illumina, United States). Adapters containing the full complement of sequencing primer hybridization sites were ligated to blunt-end fragments. Paired-end whole-genome shotgun sequencing was performed on the Illumina HiSeq platform. High-quality reads were extracted based on the FASTQ (Chen et al., 2018), with default parameters. High-quality reads were used for *de novo* assembly via SPAdes (Bankevich et al., 2012), using different k-mer sizes (k = 21, 33, 55, 77). The shortest scaffolds were filtered with a minimum length threshold of 200 bp. Gene identification was performed from the assembled genome using Prodigal (Hyatt et al., 2010), and the other genomic contents (e.g., rRNA and tRNA sequences) were annotated using the Prokka (Seemann, 2014) pipeline. Protein-coding genes were further annotated to the KEGG (Kyoto Encyclopedia of Genes and Genomes, downloaded on February 2020) (Kanehisa et al., 2017) databases using BLASTP (identity threshold of 35%, covering > 70% of the gene length).

IDENTIFICATION OF ANTIBIOTIC RESISTANCE GENES AND VIRULENCE FACTORS

Identification and characterization of antibiotic resistance genes (ARGs) and virulence factors of pathogens are crucial in understanding bacterial pathogenesis and their interactions with the host, and in the development of novel drugs, vaccines, and molecular diagnostic tools. Furthermore, detecting virulence or resistance markers improved outbreak monitoring and therapeutic management. Currently, the use of the next-generation sequencing platforms has allowed great progress in this field (Bakour et al., 2016). ARGs were identified based on ABRicate.[1] ABRicate searched on the databases, including NCBI Bacterial Antimicrobial Resistance Reference Gene Database, CARD (Jia et al., 2017), ARG-ANNOT (Gupta et al., 2014), and ResFinder (Zankari et al., 2012) for predicting ARGs. The virulence factors of the genome were identified based on the VFDB database (Chen et al., 2016).

Comparison Genome Analyses

Average nucleotide identity (ANI) between two genomes was calculated using the fastANI algorithm (Jain et al., 2018). An ANI threshold of 95% was used for species delineation for prokaryotic genomes (Jain et al., 2018). The phylogenetic tree of bacterial species was generated using PhyloPhlAn2 (Segata et al., 2013) and visualized using iTOL (Letunic and Bork, 2019).

In vitro Bile Acid Transformations by Bile Bacteria

Bile acids were dissolved in DMSO for preparation of a highly concentrated stock solution. Then, the autoclaved GAM liquid medium was applied to dilute the stock solution (volume ratio of DMSO in a working solution <1‰). The bile acid experimental group [containing 200 ng/ml of final concentration: TCA (taurocholic acid), CA (cholic acid), or GCA (glycocholic acid)] and the control group (without bile acid) were inoculated

[1] https://github.com/tseemann/abricate

with isolated bile bacteria in 200 μl of liquid media reaction system at 37°C. A control without bile acid (but with the DMSO vehicle) and a sterile control with an equal volume of bile acid (200 ng/ml) were prepared. Every reaction was prepared in parallel. This experiment was done in triplicates. Bile acids of internal standards were prepared at 200 ng/ml. After 12 h, the cultures were centrifuged (10,000 × *g*, 5 min), and the supernatants were collected. Then, the double volume of acetonitrile was added, and the aqueous phase was collected after centrifuging at 20,000 × *g* for 10 min. The detection was achieved using a Phenomenex Kinetex Polar C18 ODS (2.1 × 100 mm, 2.6 μm) analytical column. The mobile phase consisted of 0.1% ammonium hydroxide water (A) and acetonitrile (B) at a flow rate of 0.3 ml/min, and the following gradient condition was used: 0–1 min 65% A; 1–4 min 65–60% A; 4–5.5 min 60–35% A; 5.5–6.5 min 35–20% A, 6.5–7.0 20–10% A, 7.0–8.0 10% A, 8.0–8.5 10–65% A). An Applied Biosystems AB Sciex Qtrap5500 Mass Spectrometer (MS/MS) equipped with an electron spray ionization (ESI) source was used to analyze target metabolites at negative ionization mode (Scherer et al., 2009). The optimized ion spray voltage and drying gas temperature were set at -4,500 V and 550°C, respectively. The curtain gas (CUR) flow was 35 L/min; gas1 and gas2 (nitrogen) were set at 45 and 55 psi, respectively, and the retention time was 70 ms. Quantification assays were performed using multiple reaction monitoring.

Statistical Analyses

The value of metabolite output is equal to the output of the experimental group minus the average output of the control group. All statistical analyses were performed based on the R platform. Histograms were performed using GraphPad Prism 7 software (GraphPad Software) and were indicated using the mean with the standard deviation.

RESULTS

Cultivation and Genome Sequencing of Bacteria in Bile Specimens

To explore the bile bacterial community, the fresh bile specimens collected from cholecystitis patients were cultivated using seven different media under both aerobic and anaerobic conditions. In total, we obtained 138 bacterial colonies and performed full-length 16S rRNA gene sequencing to enable taxonomic assignment. These isolates included 70 (50.7%) members of Firmicutes, 65 (47.1%) members of Proteobacteria, and 3 (2.2%) members of Actinobacteria; meanwhile, they expanded 3 classes, 6 orders, 10 families, and 14 genera (**Supplementary Table 1**). At the genus level, *Enterococcus* (31.9%), *Escherichia–Shigella* (22.9%), *Lysinibacillus* (9.4%), and *Enterobacter* (8%) were the most frequent members in the biliary community, while the remaining 10 genera represented 28.2% (39/138) of the cultivated isolates (**Figure 1A**). *Proteus mirabilis* (24 strains) and *Enterococcus faecalis* (17 strains) were the most abundant species, which may be due to the property of facultative anaerobes (Portela et al., 2014; Jamil et al., 2021). *Enterobacteriaceae* had the largest isolated strains (28 strains) under aerobic conditions (**Supplementary Table 1**). In addition, 63 strains were obtained under anaerobic conditions (63/138), and most of them were also facultative anaerobes, such as the *Enterococcus* genus.

We clustered the isolates into 35 operational taxonomic units (OTUs) based on 99% nucleotide similarity of their 16S rRNA gene sequences and performed whole-genome sequencing for the representative isolate for each cluster. After *de novo* assembly of the sequencing reads, we obtained 35 high-quality draft genomes that exceeded 95% genomic completeness for each strain (**Figure 1B**). A phylogenetic tree of the draft genomes was shown in **Figure 1C**. The genome sizes of these biliary bacterial isolates were on average 4.0 Mbp, ranging from 1.9 to 7.1 Mbp (**Supplementary Table 2**), and the G + C contents of these genomes ranged from 37.4 to 63.3% (on average, 47.4%). The majority (32/35) of the biliary genomes could be assigned into known species, showing > 95% ANI to at least one sequenced genome in the NCBI database, whereas the other 3 biliary isolates were "novel species" as they previously had no available whole-genome information (**Supplementary Table 2**). Moreover, despite the 16S rRNA gene sequences of these biliary strains seemed diverse, 26 genomes were further grouped into 8 species-level genome bins based on 95% ANI.

Characteristics of the Novel Species

We analyzed the genomes of three novel species to investigate the distinctive features of those biliary bacterial species. Biliary strain *Providencia* sp. D135 consisted of 27 contigs with a total length of 4.45 Mbp (N50 length: 315 kbp). This strain was moderate homology with the genome of *Providencia* sp. WCHPr000369, a clinical strain that was isolated from the human rectum (proposed name: *Providencia huaxiensis*) (Hu et al., 2019), with 92.3% ANI (**Figure 2A**). Biliary strain *Psychrobacter* sp. D093 consisted of 99 contigs with a total length of 2.69 Mbp and showed 82.9% ANI with the genome of *Psychrobacter cryohalolentis* K5 (**Figure 2B**). Biliary strain *Vibrio* sp. D074 consisted of 87 contigs with a total length of 3.93 Mbp and showed 79.6% ANI with the genome of *Vibrio anguillarum* VIB43 (**Figure 2C**).

Comparison of Bile and Gut Isolates

To investigate the specificity of bile bacterial species, we compared our bile isolates with 1,520 bacterial strains cultivated from the gastrointestinal tract of healthy adults (Zou et al., 2019). Based on whole-genome pairwise ANI, 13 strains were specifically detected in bile samples. These bile-specific genomes were mainly distributed in Firmicutes, including *Lactobacillus murinus*, *Enterococcus casseliflavus*, *Paenibacillus lautus*, *Staphylococcus pettenkoferi*, *Lysinibacillus* sp. D060, and *Solibacillus silvestris*, and several Proteobacteria clades, including *Providencia* sp. D135, *Morganella morganii*, *Vibrio* sp. D074, and *Psychrobacter* sp. D093 (**Supplementary Figure 1**).

The functional roles of the members of bile and gut strains were compared based on their profiles of the KEGG orthologs (KOs). The bile strains encoded 5,488 KOs, of which 95.1%

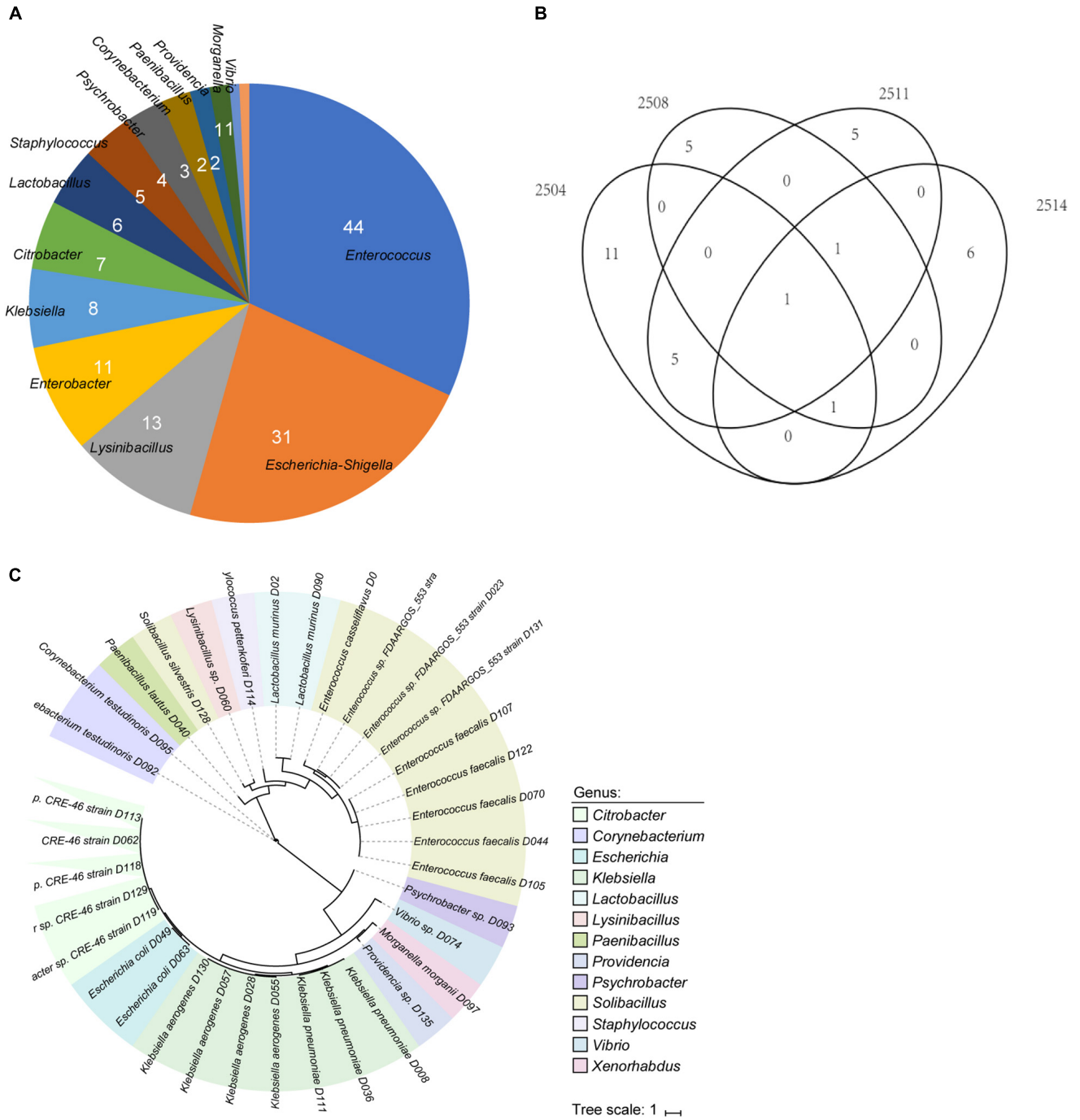


FIGURE 1 | Overview of the cultivated bacteria from bile specimens. **(A)** Pie plot shows the composition of 138 bacterial colonies. **(B)** Venn diagram of the distribution of isolates in the four bile samples. **(C)** Phylogenetic tree of 35 whole-genome sequenced isolates. Color represents the genus assignment of the isolates.

were also encoded by the gut strains (**Figure 3A**). The bile-specific KOs were involved in pathways including biofilm formation, two-component systems, quorum sensing, and ABC transporters (**Figure 3B**). We also undertook an enrichment analysis and identified 3,740 statistically enriched KOs and 695 reduced KOs in bile bacteria (Benjamini–Hochberg-adjusted $p < 0.05$, Fisher's exact test). The bile-enriched KOs had frequently performed some key functions, including ABC transporters, two-component system, quorum sensing, purine metabolism, biofilm formation, oxidative phosphorylation, pyruvate metabolism, and phosphotransferase system, while the bile-reduced KOs were involved in functions such as

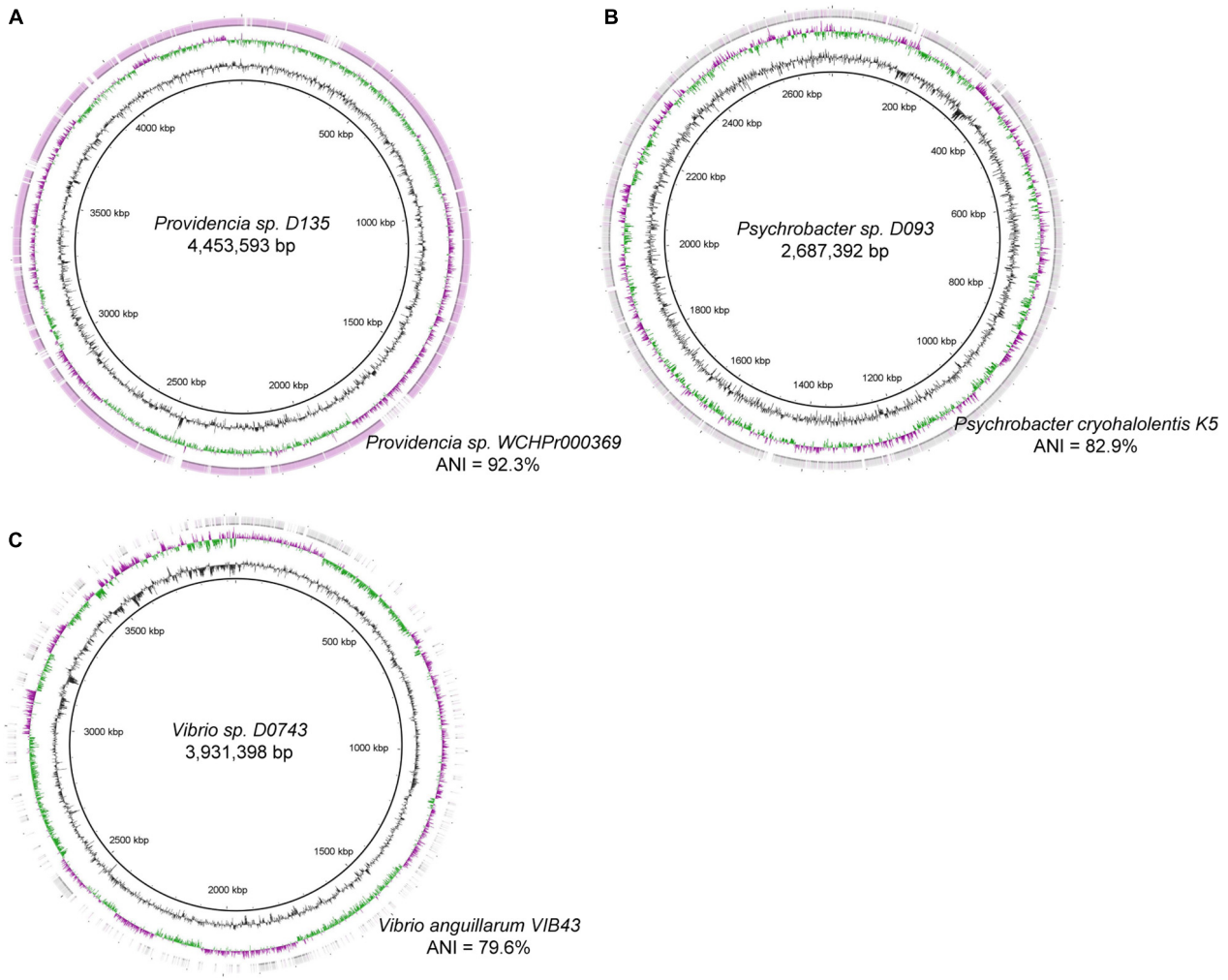


FIGURE 2 | Circular representation of the genomes of four novel species, including **(A)** *Providencia* sp. *D135*, **(B)** *Psychrobacter* sp. *D093*, and **(C)** *Vibrio* sp. *D074*. For each genome, the inner three circles represent the genome scale, G + C skew, and G + C content; the outer circle(s) shows the genome region that have close orthologs with the closest strains in the National Center of Biotechnology Information (NCBI) database. Average nucleotide identity (ANI) of the bile genomes and its closest strains are shown.

the ribosome, two-component system, amino sugar and nucleotide sugar metabolism, and starch and sucrose metabolism (**Figures 3C,D**).

Antibiotic Resistance Genes, Virulence Factors, and Bile Acid Metabolism Genes

Next, we examined the existence of ARGs and virulence factors in the genomes of bile species, aiding to extend their potential clinical pathologic features. A total of 472 genes from 27 genomes were identified as ARGs based on annotations in the available antibiotic resistance databases (**Supplementary Table 3**). Most of the ARGs were related to multidrug resistance (MDR) (42.6%), fluoroquinolone resistance (12.3%), aminoglycoside resistance (9.1%), and β-lactamase (7.2%). The remaining genes were involved in resistance against various types of antibiotics such as tetracycline (4.4%), glycopeptide (4%), peptide (4%), and trimethoprim (4%). At the species level, *Escherichia coli*, *Citrobacter* sp. CRE-46, *Klebsiella aerogenes*, and *Klebsiella pneumoniae* encoded a higher number of ARGs than other bile species, especially that their ARGs were enriched in types including MDR, fluoroquinolone resistance, and β-lactamase (**Figure 4A**). *Enterococcus faecalis* frequently encoded the genes involving trimethoprim, macrolide, and tetracycline resistances, while *Enterococcus casseliflavus* and *Enterococcus* sp. FDAARGOS_553 encoded the genes of glycopeptide resistance. In addition, we identified 354 genes involving virulence factors from 20 bile strains (**Supplementary Table 4**). The virulence factors were mainly distributed in species including *Escherichia coli*, *Enterococcus faecalis*, *Klebsiella pneumoniae*, *Citrobacter* sp. CRE-46, and *Klebsiella aerogenes* (**Figure 4B**). In addition, we identified a total of 12 genes encoding the BSH and 319 genes involved in the transformation of primary bile acids to secondary bile acids (i.e., the multiple steps

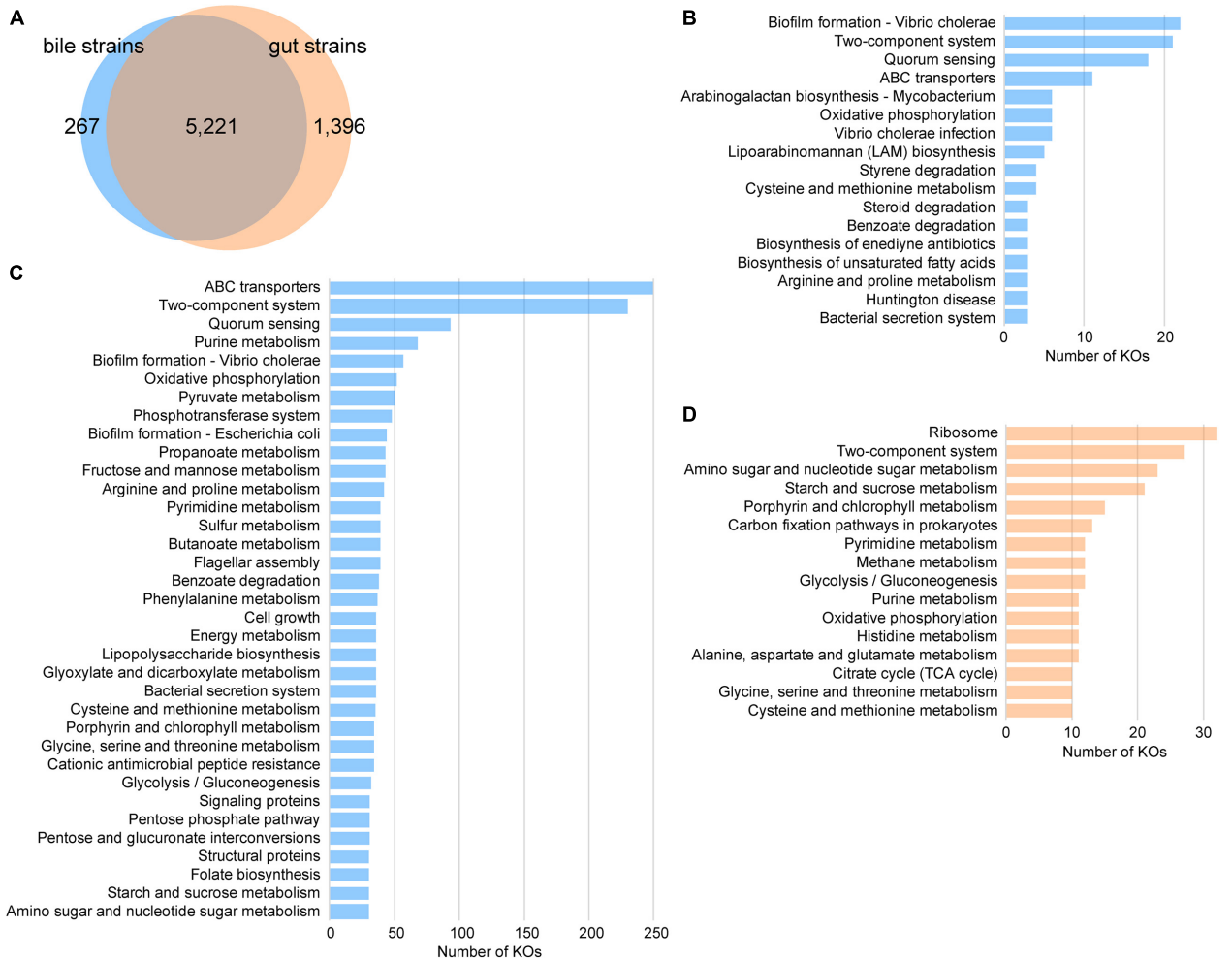


FIGURE 3 | Comparison of the functions of bile and gut bacterial strains. **(A)** Venn diagram shows the comparison of KEGG orthologs (KOs) between bile and gut strains. **(B–D)** Distribution of KEGG level C pathways of the bile-specific KOs **(B)**, the bile-enriched KOs **(C)**, and the gut-enriched KOs **(D)**.

7α/β-dehydroxylation) (**Supplementary Table 5**). These genes spanned all 35 genome-sequenced strains, suggesting that the bile bacteria have widely participated in bile acid metabolism. At the species level, *Lysinibacillus* sp. D060, *Solibacillus silvestris*, *Paenibacillus lautus*, and the members of *Klebsiella* (i.e., *K. aerogenes* and *K. pneumoniae*) had the largest number of bile acid metabolism genes.

Evaluation of Bile Salt Metabolism Potentiality of Bile Bacteria

We evaluated the potential BSH and dehydroxylase activities in 24 strains, which were selected by phylogenetic tree based on 16S rRNA gene amplicon sequence by measuring the changed concentration of free bile acids or secondary bile acids in the supernatant of cultures. *Paenibacillus* sp. H203 (Ps) and *Lysinibacillus xylanilyticus* Gute33 (Lx) were detected with strong bile salt deconjugation capabilities. Cholic acid formed with GCA substrate [586.1 ± 135 ng/ml (Ps), 445.4 ± 14.1 ng/ml (Lx)] and 1,500.6 ± 42 ng/ml (Ps), 1,525.4 ± 49 ng/ml (Lx) cholic acid formed with TCA substrate (**Figure 5A**). *Lysinibacillus* sp. MJJ-11 and *Corynebacterium testudinoris* DSM 44614 strain showed the stronger capabilities of forming secondary bile acid, which converted CA to DCA (**Figure 5B**).

DISCUSSION

The bile bacteria, including the *Vibrio* genus especially *Vibrio cholerae* (Asnis et al., 1996), *Escherichia coli*, *Klebsiella* spp., *Enterococcus* (Yun and Seo, 2018), *Staphylococcus*, *Corynebacterium* (Backert et al., 2018), and *Lactobacillus salivarius*, have been studied extensively (Woo et al., 2002). Tajeddin et al. (2016) found that enterobacteria were the dominant biliary bacteria based on cultivate-dependent methods. As an alternative animal model of humans, most of the bile and gall bladder bacteria of healthy pigs were *Streptococcus alactolyticus*, *Staphylococcus epidermidis*, and *Corynebacterium testudinoris*, which suggested that healthy host bile may also

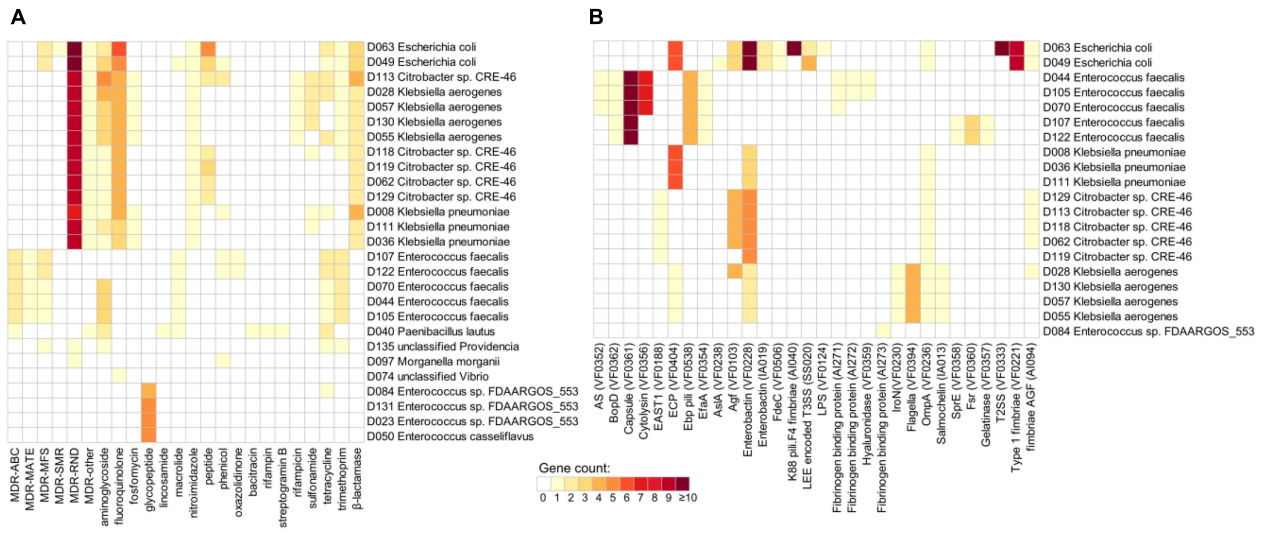


FIGURE 4 | Distribution of antibiotic resistance genes (ARGs) and virulence factors in the bile bacterial strains. Heat maps show the number of ARGs **(A)** and virulence factors **(B)** for each strain.

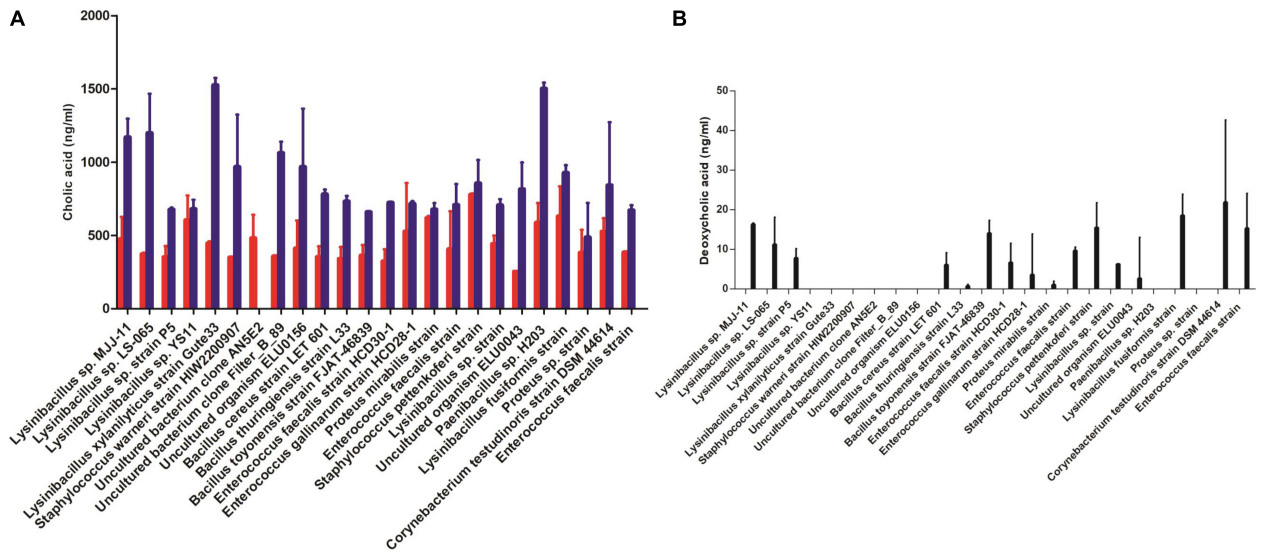


FIGURE 5 | Detection of bile salt metabolism by bile bacterial strains. The concentration of metabolites **(A)** converted glycocholic acid (GCA) to cholic acid (CA) (red bar), or taurocholic acid (TCA) to CA (blue bar) **(B)** converted CA to deoxycholic acid (DCA).

contain bacteria (Jimenez et al., 2014). In our previous research and the work of others, the bile pathogenic bacteria were considered a predisposing factor in gallbladder disorders (Liu et al., 2015; Shen et al., 2015; Tajeddin et al., 2016). Furthermore, metagenomic results indicated that many potential bile bacteria still cannot be isolated and cultivated (Liu et al., 2015). With more cultivated and detected conditions termed "culturomics" emerging (Lagier et al., 2016), many previously unavailable "dark matter bacteria" from bile would be isolated.

In this study, we characterized the isolated bacterial genome from the fresh bile specimens from four cholecystitis patients and evaluated its bile acid biotransformation ability *in vitro*. Our results showed that the number of cultivable bile bacteria exceeds expectations, spanning 138 bacterial isolates that were distributed in 14 genera. In particular, three bacterial species (i.e., *Providencia* sp. D135, *Psychrobacter* sp. D093, and *Vibrio* sp. D074) were not represented in existing reference genome databases. Enterobacteriaceae had the most isolated strains (28 strains) under aerobic conditions, in agreement with previous studies (Tajeddin et al., 2016). Although there was little research on bile anaerobes, we obtained 63 strains under anaerobic conditions, however, most of them were facultative anaerobes.

The possible reasons were the small sample size (n = 4) and the nutrient composition of the medium, which was not specifically optimized or not rich enough for bile bacteria. For instance, Lagier et al. (2016) reported that liquid media precultivation operation can effectively promote the diversity of isolated bacteria. In addition, the centrifugal operation may lose some lighter bacteria; therefore, after centrifugation, the supernatant can be aseptically filtered (Spin-X Centrifuge Tube Filter, Corning Inc.) and concentrated, and the concentrated part will be used for cultivation to obtain more bacteria. The isolated bacterial genomes could extend the databases, which are the basis for metagenomics analysis (Lagier et al., 2016).

Whole-genome sequencing was performed on 35 bacterial species and compared with gut isolates. Our results showed that many pathways, such as biofilm formation, ABC transporters, and quorum sensing, were enriched in bile bacteria, which may contribute to communicate with each other and resist the antibacterial high concentration of bile salts (Davidson et al., 2008; Moons et al., 2009; Rutherford and Bassler, 2012). Bile salts promoted the biofilm formation of bile bacteria, and the biofilms of bacteria-laden gallstones influenced the illness severity (Stewart et al., 2007). Especially, the biofilm formation ability seems to be a necessary condition for bile resistance and polymicrobial infection in this organ (Tajeddin et al., 2016; Zheng et al., 2017). Therefore, this high abundance of functional pathways of bile bacteria, such as biofilm formation, was related to its living environment and physiological characteristics.

Enterobacteriaceae had the largest isolated strains (28 strains) under aerobic conditions in this study which was consistent with the previous studies (Yun and Seo, 2018) and had been identified to be closely related to cholecystitis in our previous research (Liu et al., 2015). ARGs and virulence factors analyses of the bacterial isolates further revealed that important Gram-negative opportunistic pathogen of the Enterobacteriaceae, such as *Escherichia coli*, *Citrobacter* sp. CRE-46, *Klebsiella aerogenes*, and *Klebsiella pneumoniae*, encoded the higher number of ARGs and the virulence factors than other bile species. This result was inconsistent with those of previous studies that less than 5% resistance was observed in bile isolated bacteria against carbapenem, beta-lactam antibiotics, glycopeptide antibiotics, and linezolid via the antibiotic susceptibility tests (Yun and Seo, 2018). This difference may be caused by different test methods and their sensitivity. Therefore, the combination of the bioinformatics and the verification of conventional drug-susceptibility testing method on the pivotal isolates in the next step may have positive significance in the quicker selection of the most suitable antibiotics for clinical applications.

In humans, the major bile salts included taurocholic acid (TCA) and glycocholic acid (GCA) derived from cholic acid (CA), while taurochenodeoxycholic acid (TCDCA) and glycochenodeoxycholic acid (GCDCA) derived from chenodeoxycholic acid (CDCA). The salts of their 7-alpha-dehydroxylated derivatives, deoxycholic acid (DCA), and lithocholic acid (LCA), are also found, with derivatives of cholic, chenodeoxycholic, and deoxycholic acids accounting for over 90% of human biliary bile acids (Hofmann, 1999). Bile acids are closely related to human health and can be metabolized by gut bacteria, which generated free bile acid by deconjugation and generated secondary bile acids by 7α-and 7β-dehydroxylation, such as deoxycholic acid and lithocholic acid (Ridlon et al., 2016; Wang and Jia, 2016; Marion et al., 2019; Song et al., 2019; Funabashi et al., 2020). Secondary bile acids, especially DCA, have been listed as carcinogens and promoters of CRC (Yang and Yu, 2018). High-fat diet-induced high levels of DCA in the intestine are also considered an important risk factor for CRC. Herein, based on the phylogenetic tree of isolates, we selected 24 bile bacterial strains and detected their bile salt deconjugation and transformation abilities. *Lysinibacillus* genus had strong bile acid biotransformation ability (**Figure 5**), which was consistent with previous reports that *Paenibacillus* and *Lysinibacillus* widely carried BSH (Song et al., 2019). This ability of bile bacteria may allow them to overcome the bile salt damaging effect. Although this study provided direct evidence *in vitro* for the bile bacteria involved in the biotransformation of bile acid, however, we are not sure that the bacteria-mediated transformations occur in the gallbladder or not so far. The reason is that the presence of *in vivo* bile acid transformations performed by certain isolated microorganisms lacks rigor since the same strain may display completely different functional performance between its *in vivo* and *in vitro* activity (Marion et al., 2019). For instance, *Clostridium hiranonis* did not deconjugate tauro-conjugated bile acids in germ-free mice, although its ability to deconjugate taurocholic acid (TCA) was shown *in vitro* (Narushima et al., 1999). Importantly, our research results provided new insights for screening bile salt metabolizing bacteria, which many studies focused on gut bacteria (Lucas et al., 2021).

Our data helps in understanding the genetic basis of the physiology, biochemical pathways, and evolution of the isolates and provides a preliminary exploration in bile acid transformations by bile bacteria; however, the sample size and culture conditions were relatively small. Therefore, larger-scale bacterial isolation and identification, more in-depth screening isolates with bile acid transformation ability *in vivo* by metabolite detection method, and verifying the impact of metabolite on gallbladder and even whole-body health are to be investigated in the next step.

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.ncbi.nlm.nih.gov/bioproject/PRJNA726166, PRJNA726166.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of the Hunan Provincial People's Hospital (Ethical approval number: 2020-54). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YM, QY, SL, YX, and XHM contributed to conception and design of the study. SL, SZ, YZ, QL, and AZ organized the formal analysis. GW, RL, MX, TJ, YS, XL, SC, XT, and XD performed methodology. YM, XCM, CW, and YX were responsible for project administration. QY, SL, and SZ wrote the first draft of the manuscript. All authors were involved in preparing the manuscript and contributed to manuscript revision, read, and approved the submitted version.

FUNDING

YM was supported by the National Natural Science Foundation of China (81930112 and 81573469). QY was supported by the National Natural Science Foundation of China (81902037) and the Natural Science Foundation of Liaoning Province, China (20180530086). XHM was supported by the Natural Science Foundation of Hunan Province, China (2019JJ80007) and the Key Scientific Research Project of Hunan Provincial Education Department, China (20A312).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.739621/full#supplementary-material

Supplementary Figure 1 | Phylogenetic tree of 35 bile isolates and 1,520 bacterial strains cultivated from the feces. Outer color bars represent the phylum assignment of the isolates.

REFERENCES

Asnis, D. S., Golub, R., and Bresciani, A. (1996). *Vibrio cholerae* 01 isolated in the gallbladder of a patient presenting with cholecystitis. *Am. J. Gastroenterol.* 91, 2241–2242.

Backert, S., Tegtmeyer, N., Oyarzabal, O. A., Osman, D., Rohde, M., Grutzmann, R., et al. (2018). Unusual manifestation of live *Staphylococcus saprophyticus*, *Corynebacterium urinapleomorphum*, and *Helicobacter pylori* in the gallbladder with cholecystitis. *Int. J. Mol. Sci.* 19:1826. doi: 10.3390/ijms19071826

Bakour, S., Sankar, S. A., Rathored, J., Biagini, P., Raoult, D., and Fournier, P. E. (2016). Identification of virulence factors and antibiotic resistance markers using bacterial genomics. *Future Microbiol.* 11, 455–466. doi: 10.2217/fmb.15.149

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., et al. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546. doi: 10.1038/nature17645

Cen, L., Pan, J., Zhou, B., Yu, C., Li, Y., Chen, W., et al. (2018). *Helicobacter Pylori* infection of the gallbladder and the risk of chronic cholecystitis and cholelithiasis: a systematic review and meta-analysis. *Helicobacter* 23:e12457. doi: 10.1111/hel.12457

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239

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

Chisholm, P. R., Patel, A. H., Law, R. J., Schulman, A. R., Bedi, A. O., Kwon, R. S., et al. (2019). Preoperative predictors of choledocholithiasis in patients presenting with acute calculous cholecystitis. *Gastrointest. Endosc.* 89, 977–983 e972. doi: 10.1016/j.gie.2018.11.017

Das, S., Dash, H. R., Mangwani, N., Chakraborty, J., and Kumari, S. (2014). Understanding molecular identification and polyphasic taxonomic approaches for genetic relatedness and phylogenetic relationships of microorganisms. *J. Microbiol. Methods* 103, 80–100. doi: 10.1016/j.mimet.2014.05.013

Davidson, A. L., Dassa, E., Orelle, C., and Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* 72, 317–364. doi: 10.1128/MMBR.00031-07

Fonkou, M. D., Dufour, J. C., Dubourg, G., and Raoult, D. (2018). Repertoire of bacterial species cultured from the human oral cavity and respiratory tract. *Future Microbiol.* 13, 1611–1624. doi: 10.2217/fmb-2018-0181

Funabashi, M., Grove, T. L., Wang, M., Varma, Y., McFadden, M. E., Brown, L. C., et al. (2020). A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 582, 566–570. doi: 10.1038/s41586-020-2396-4

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., et al. (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* 58, 212–220. doi: 10.1128/AAC.01310-13

Hang, S., Paik, D., Yao, L., Kim, E., Trinath, J., Lu, J., et al. (2019). Bile acid metabolites control TH17 and Treg cell differentiation. *Nature* 576, 143–148. doi: 10.1038/s41586-019-1785-z

Harris, S. C., Devendran, S., Mendez-Garcia, C., Mythen, S. M., Wright, C. L., Fields, C. J., et al. (2018). Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243(T). *Gut Microbes* 9, 523–539. doi: 10.1080/19490976.2018.1458180

Hofmann, A. F. (1999). The continuing importance of bile acids in liver and intestinal disease. *Arch. Intern. Med.* 159, 2647–2658. doi: 10.1001/archinte.159.22.2647

Hu, Y., Feng, Y., Zhang, X., and Zong, Z. (2019). *Providencia huaxiensis* sp. nov., recovered from a human rectal swab. *Int. J. Syst. Evol. Microbiol.* 69, 2638–2643. doi: 10.1099/ijsem.0.003502

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Jamil, R. T., Foris, L. A., and Snowden, J. (2021). *Proteus Mirabilis Infections*. Treasure Island, FL: StatPearls.

Jang, K. S., and Kim, Y. H. (2018). Rapid and robust MALDI-TOF MS techniques for microbial identification: a brief overview of their diverse applications. *J. Microbiol.* 56, 209–216. doi: 10.1007/s12275-018-7457-0

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004

Jimenez, E., Sanchez, B., Farina, A., Margolles, A., and Rodriguez, J. M. (2014). Characterization of the bile and gall bladder microbiota of healthy pigs. *Microbiologyopen* 3, 937–949. doi: 10.1002/mbo3.218

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Kuipers, F., de Boer, J. F., and Staels, B. (2020). Microbiome modulation of the host adaptive immunity through bile acid modification. *Cell Metab.* 31, 445–447. doi: 10.1016/j.cmet.2020.02.006

Lagier, J. C., Khelaifia, S., Alou, M. T., Ndongo, S., Dione, N., Hugon, P., et al. (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat. Microbiol.* 1:16203. doi: 10.1038/nmicrobiol.2016.203

Lagkouvardos, I., Pukall, R., Abt, B., Foesel, B. U., Meier-Kolthoff, J. P., Kumar, N., et al. (2016). The mouse intestinal bacterial collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol.* 1:16131. doi: 10.1038/nmicrobiol.2016.131

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Liu, C., Zhou, N., Du, M. X., Sun, Y. T., Wang, K., Wang, Y. J., et al. (2020). The mouse gut microbial Biobank expands the coverage of cultured bacteria. *Nat. Commun.* 11:79. doi: 10.1038/s41467-019-13836-5

Liu, J., Yan, Q., Luo, F., Shang, D., Wu, D., Zhang, H., et al. (2015). Acute cholecystitis associated with infection of *Enterobacteriaceae* from gut microbiota. *Clin. Microbiol. Infect.* 21, 851 e851–e859. doi: 10.1016/j.cmi.2015.05.017

Loozen, C. S., van Santvoort, H. C., van Duijvendijk, P., Besselink, M. G., Gouma, D. J., Nieuwenhuijzen, G. A., et al. (2018). Laparoscopic cholecystectomy versus percutaneous catheter drainage for acute cholecystitis in high risk patients (CHOCOLATE): multicentre randomised clinical trial. *BMJ* 363:k3965. doi: 10.1136/bmj.k3965

Lucas, L. N., Barrett, K., Kerby, R. L., Zhang, Q., Cattaneo, L. E., Stevenson, D., et al. (2021). Dominant bacterial phyla from the human gut show widespread ability to transform and conjugate bile acids. *mSystems* 6:e0080521. doi: 10.1128/mSystems.00805-21

Marion, S., Studer, N., Desharnais, L., Menin, L., Escrig, S., Meibom, A., et al. (2019). In vitro and in vivo characterization of clostridium scindens bile acid transformations. *Gut Microbes* 10, 481–503. doi: 10.1080/19490976.2018.1549420

Martellacci, L., Quaranta, G., Fancello, G., D'Addona, A., Sanguinetti, M., Patini, R., et al. (2020). Characterizing peri-implant and sub-gingival microbiota through culturomics. first isolation of some species in the oral cavity. a pilot study. *Pathogens* 9:365. doi: 10.3390/pathogens9050365

Moons, P., Michiels, C. W., and Aertsen, A. (2009). Bacterial interactions in biofilms. *Crit. Rev. Microbiol.* 35, 157–168. doi: 10.1080/10408410902809431

Narushima, S., Itoh, K., Takamine, F., and Uchida, K. (1999). Absence of cecal secondary bile acids in gnotobiotic mice associated with two human intestinal bacteria with the ability to dehydroxylate bile acids in vitro. *Microbiol. Immunol.* 43, 893–897. doi: 10.1111/j.1348-0421.1999.tb01224.x

Nitzan, O., Brodsky, Y., Edelstein, H., Hershko, D., Saliba, W., Keness, Y., et al. (2017). Microbiologic data in acute cholecystitis: ten years' experience from bile cultures obtained during percutaneous cholecystostomy. *Surg. Infect. (Larchmt.)* 18, 345–349. doi: 10.1089/sur.2016.232

Portela, C. A., Smart, K. F., Tumanov, S., Cook, G. M., and Villas-Boas, S. G. (2014). Global metabolic response of *Enterococcus faecalis* to oxygen. *J. Bacteriol.* 196, 2012–2022. doi: 10.1128/JB.01354-13

Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D. J., and Hylemon, P. B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* 7, 22–39. doi: 10.1080/19490976.2015.1127483

Ridlon, J. M., Kang, D. J., and Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* 47, 241–259. doi: 10.1194/jlr.R500013-JLR200

Rutherford, S. T., and Bassler, B. L. (2012). Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb. Perspect. Med.* 2:a012427. doi: 10.1101/cshperspect.a012427

Salvioli, G., Salati, R., Bondi, M., Fratalocchi, A., Sala, B. M., and Gibertini, A. (1982). Bile acid transformation by the intestinal flora and cholesterol saturation in bile. Effects of *Streptococcus faecium* administration. *Digestion* 23, 80–88. doi: 10.1159/000198691

Scherer, M., Gnewuch, C., Schmitz, G., and Liebisch, G. (2009). Rapid quantification of bile acids and their conjugates in serum by liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 877, 3920–3925. doi: 10.1016/j.jchromb.2009.09.038

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Segata, N., Bornigen, D., Morgan, X. C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4:2304. doi: 10.1038/ncomms3304

Shapiro, H., Kolodziejczyk, A. A., Halstuch, D., and Elinav, E. (2018). Bile acids in glucose metabolism in health and disease. *J. Exp. Med.* 215, 383–396. doi: 10.1084/jem.20171965

Shen, H., Ye, F., Xie, L., Yang, J., Li, Z., Xu, P., et al. (2015). Metagenomic sequencing of bile from gallstone patients to identify different microbial community patterns and novel biliary bacteria. *Sci. Rep.* 5:17450. doi: 10.1038/srep17450

Song, X., Sun, X., Oh, S. F., Wu, M., Zhang, Y., Zheng, W., et al. (2020). Microbial bile acid metabolites modulate gut RORgamma(+) regulatory T cell homeostasis. *Nature* 577, 410–415. doi: 10.1038/s41586-019-1865-0

Song, Z., Cai, Y., Lao, X., Wang, X., Lin, X., Cui, Y., et al. (2019). Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome* 7:9. doi: 10.1186/s40168-019-0628-3

Stewart, L., Griffiss, J. M., Jarvis, G. A., and Way, L. W. (2007). Gallstones containing bacteria are biofilms: bacterial slime production and ability to form pigment solids determines infection severity and bacteremia. *J. Gastrointest. Surg.* 11, 977–983; discussion983–974. doi: 10.1007/s11605-007-0168-1

Tajeddin, E., Sherafat, S. J., Majidi, M. R., Alebouyeh, M., Alizadeh, A. H., and Zali, M. R. (2016). Association of diverse bacterial communities in human bile samples with biliary tract disorders: a survey using culture and polymerase chain reaction-denaturing gradient gel electrophoresis methods. *Eur. J. Clin. Microbiol. Infect. Dis.* 35, 1331–1339. doi: 10.1007/s10096-016-2669-x

Thomas-White, K., Forster, S. C., Kumar, N., Van Kuiken, M., Putonti, C., Stares, M. D., et al. (2018). Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat. Commun.* 9:1557. doi: 10.1038/s41467-018-03968-5

Tidjani Alou, M., Naud, S., Khelaifia, S., Bonnet, M., Lagier, J. C., and Raoult, D. (2020). State of the art in the culture of the human microbiota: new interests and strategies. *Clin. Microbiol. Rev.* 34:e00129–19. doi: 10.1128/CMR.00129-19

Wadhwa, V., Jobanputra, Y., Garg, S. K., Patwardhan, S., Mehta, D., and Sanaka, M. R. (2017). Nationwide trends of hospital admissions for acute cholecystitis in the United States. *Gastroenterol. Rep. (Oxf.)* 5, 36–42. doi: 10.1093/gastro/gov015

Wang, B. F., Yi, S., Keshavamurthy, J., Williams, H., and Pucar, D. (2021). Gallbladder perforation into the peritoneal cavity. *Clin. Nucl. Med.* 46, e84–e85. doi: 10.1097/RLU.0000000000003317

Wang, J., and Jia, H. (2016). Metagenome-wide association studies: fine-mining the microbiome. *Nat. Rev. Microbiol.* 14, 508–522. doi: 10.1038/nrmicro.2016.83

White, B. A., Lipsky, R. L., Fricke, R. J., and Hylemon, P. B. (1980). Bile acid induction specificity of 7 alpha-dehydroxylase activity in an intestinal *Eubacterium* species. *Steroids* 35, 103–109. doi: 10.1016/0039-128x(80)90115-4

Woo, P. C., Fung, A. M., Lau, S. K., and Yuen, K. Y. (2002). Identification by 16S rRNA gene sequencing of *Lactobacillus salivarius* bacteremic cholecystitis. *J. Clin. Microbiol.* 40, 265–267. doi: 10.1128/jcm.40.1.265-267.2002

Wu, T., Zhang, Z., Liu, B., Hou, D., Liang, Y., Zhang, J., et al. (2013). Gut microbiota dysbiosis and bacterial community assembly associated with cholesterol gallstones in large-scale study. *BMC Genomics* 14:669. doi: 10.1186/1471-2164-14-669

Yang, J., and Yu, J. (2018). The association of diet, gut microbiota and colorectal cancer: what we eat may imply what we get. *Protein Cell* 9, 474–487. doi: 10.1007/s13238-018-0543-6

Yun, S. P., and Seo, H. I. (2018). Clinical aspects of bile culture in patients undergoing laparoscopic cholecystectomy. *Medicine (Baltimore)* 97:e11234. doi: 10.1097/MD.0000000000011234

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

Zheng, B., Jiang, X., Cheng, H., Guo, L., Zhang, J., Xu, H., et al. (2017). Genome characterization of two bile-isolated *Vibrio fluvialis* strains: an insight into pathogenicity and bile salt adaption. *Sci. Rep.* 7:11827. doi: 10.1038/s41598-017-12304-8

Zou, Y., Xue, W., Luo, G., Deng, Z., Qin, P., Guo, R., et al. (2019). 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat. Biotechnol.* 37, 179–185. doi: 10.1038/s41587-018-0008-8

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fmicb.2021.740068

Genetic Characterization of *Brucella* spp.: Whole Genome Sequencing-Based Approach for the Determination of Multiple Locus Variable Number Tandem Repeat Profiles

OPEN ACCESS

Edited by:
Martin Maiers, National Marrow Donor Program, United States

Reviewed by:
Maryam Dadar, Razi Vaccine and Serum Research Institute, Iran
Marcela Suárez-Esquivel Universidad Nacional de Costa Rica, Costa Rica

***Correspondence:**
João Paulo Gomes
j.paulo.gomes@insa.min-saude.pt

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *12 July 2021*
Accepted: *18 October 2021*
Published: *12 November 2021*

Citation:
Pelerito A, Nunes A, Grilo T, Isidro J, Silva C, Ferreira AC, Valdezate S, Núncio MS, Georgi E and Gomes JP (2021) Genetic Characterization of Brucella spp.: Whole Genome Sequencing-Based Approach for the Determination of Multiple Locus Variable Number Tandem Repeat Profiles. Front. Microbiol. 12:740068. doi: 10.3389/fmicb.2021.740068

***Ana Pelerito*[1†], *Alexandra Nunes*[2,3,4†], *Teresa Grilo*[1], *Joana Isidro*[2], *Catarina Silva*[5,6], *Ana Cristina Ferreira*[4,7], *Sylvia Valdezate*[8], *Maria Sofia Núncio*[1], *Enrico Georgi*[9] *and João Paulo Gomes*[2,4*]**

[1]Emergency Response and Biopreparedness Unit, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal, [2]Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal, [3]CBIOS – Universidade Lusófona's Research Center for Biosciences & Health Technologies, Lisbon, Portugal, [4]Faculty of Veterinary Medicine, Lusófona University, Lisbon, Portugal, [5]Technology and Innovation Unit, Department of Human Genetics, National Institute of Health, Lisbon, Portugal, [6]Centre for Toxicogenomics and Human Health (ToxOmics), Faculdade de Ciências Médicas, Nova Medical School, Universidade Nova de Lisboa, Lisbon, Portugal, [7]National Institute for Agrarian and Veterinary Research, I.P. (INIAV, IP), Oeiras, Portugal, [8]ISCIII Reference and Research Laboratory for Taxonomy, National Centre of Microbiology, Instituto de Salud Carlos III, Madrid, Spain, [9]Bundeswehr Institute of Microbiology, Munich, Germany

Brucellosis is an important zoonosis that is emerging in some regions of the world, gaining increased relevance with the inclusion of the causing agent *Brucella* spp. in the class B bioterrorism group. Until now, multi-locus VNTR Analysis (MLVA) based on 16 loci has been considered as the gold standard for *Brucella* typing. However, this methodology is laborious, and, with the rampant release of *Brucella* genomes, the transition from the traditional MLVA to whole genome sequencing (WGS)-based typing is on course. Nevertheless, in order to avoid a disruptive transition with the loss of massive genetic data obtained throughout the last decade and considering that the transition timings will vary considerably among different countries, it is important to determine WGS-based MLVA alleles of the nowadays sequenced genomes. On this regard, we aimed to evaluate the performance of a Python script that had been previously developed for the rapid *in silico* extraction of the MLVA alleles, by comparing it to the PCR-based MLVA procedure over 83 strains from different *Brucella* species. The WGS-based MLVA approach detected 95.3% of all possible 1,328 hits (83 strains × 16 loci) and showed an agreement rate with the PCR-based MLVA procedure of 96.4% for MLVA-16. According to our dataset, we suggest the use of a minimal depth of coverage of ~50x and a maximum number of ~200 contigs as guiding "boundaries" for the future application of the script. In conclusion, the evaluated script seems to be a very useful and robust tool for the *in silico* determination of MLVA profiles of *Brucella* strains, allowing retrospective and prospective molecular

epidemiological studies, which are important for maintaining an active epidemiological surveillance of brucellosis.

Keywords: *Brucella* spp., MLVA, whole-genome sequencing, zoonosis, genotyping, Python script

INTRODUCTION

Brucellosis is one of the most common bacterial zoonosis causing great damage to the farming industry and public health (Franc et al., 2018). The brucellosis burden specifically on low-income countries has led the World Health Organization (WHO) to classify it as one of the world's leading neglected zoonotic diseases (WHO, 2020). However, given the absence of specific signs and symptoms, the disease is commonly underdiagnosed (Valdezate et al., 2010).

Brucellosis is transmitted to humans by ingestion of unpasteurized dairy products or by direct contact with infected animals, placentas, or aborted fetuses (Young, 2005). It can constitute a severely debilitating illness, with diverse symptoms ranging from fever, sweating, fatigue, weight loss, headache, and joint pain. Neurological complications, such as personality changes, meningitides, encephalitis, and peripheral neuropathy, can also occur (Dean et al., 2012).

The interest in human brucellosis has been boosted due to its recent re-emergence and enhanced surveillance worldwide and from the inclusion of the causing agent *Brucella* spp. in the group of class B bioterrorism agent (Franco et al., 2007). A low infectious dose of 100–1,000 organisms is sufficient to cause an infection. The mechanisms of transmission, through aerosols or food chains, make them easily transmissible to both humans and animals (Tan et al., 2015). Thus, the distinction between natural outbreaks and/or intentional release of microorganisms may be of fundamental importance in the context of the bioterrorism.

Studies enrolling DNA–DNA hybridization procedures and comparative genomics revealed that *Brucella* species are characterized by >80% interspecies homology and >98% sequence similarity (Whatmore et al., 2006; Kattar et al., 2008). Indeed, the sequencing of *16S rRNA* gene showed 100% identity between all *Brucella* spp. (Gee et al., 2004). The genus presently encloses 12 genetically highly related species. Human brucellosis can be caused by various *Brucella* species; however, *Brucella melitensis* is the most virulent and by far the most frequently observed causative agent of human infection (Young, 2005; Georgi et al., 2017). On this regard, the identification of the circulating *Brucella* species, biovar, and genotype is very important, mainly for tracking back infectious sources and monitoring transmission routes (Pisarenko et al., 2018). The species identification by PCR assays is sufficient for the purposes of diagnosis of human/animal disease or the detection of food contamination but not for the tracing of outbreaks or bioterrorism events (De Santis et al., 2011).

To achieve the goal of sub-species discrimination, Variable Number Tandem Repeats (VNTR) have been investigated in multi-locus VNTR analysis (MLVA) by various research groups since 2003 (Bricker et al., 2003; Le Fleche et al., 2006; Whatmore et al., 2006). This MLVA *Brucella* typing scheme has proved to have the ability to differentiate *Brucella* species, biovar, and even some isolates. This is facilitated do to the creation of an online database of MLVA-16 profiles (MLVA Bank, n.d.)[1] that is available to all laboratories, allowing the comparison of *Brucella* strains worldwide (Le Fleche et al., 2006; Mambres et al., 2017; Sun et al., 2017). The recent implementation of whole genome single nucleotide polymorphism (SNP)-based typing, associated with its decreasing costs, has led to substantial improvements of both molecular subtyping and phylogenetic analyses in microbiology. The development of core- and whole-genome multilocus sequence typing (MLST) schemes has been focused on the restrict number of bacterial pathogens, including *Brucella* spp. but their application may be tricky (Tan et al., 2015; Janowicz et al., 2018; Sankarasubramanian et al., 2019). In fact, the creation of universal intra- or inter-species schemes needs to overcome some genetic hurdles such as the existence of paralogous genes, annotation issues, the accessory genome, and nomenclature-associated difficulties. Nevertheless, public databases for molecular typing and microbial genome diversity (PubMLST) are already available,[2] allowing the use of whole genome sequences for typing purposes of multiple bacterial species. The same scenario is seen in the viral field as bioinformatics platforms were already developed, allowing the genotype determination from viral complete genomes, as for influenza virus and SARS-CoV-2 (Borges et al., 2018).[3] Meanwhile, until whole genome data is fully established and accepted by the scientific community for classification/typing purposes in *Brucella*, the *in silico* determination of MLVA schemes can be of extreme utility. In fact, not only it overcomes the laborious PCR-based MLVA assessment but it also allows the dynamic cross-comparison with the typing-associated genetic data determined during the last decade. On this regard, a Python script has been developed focusing on the *in silico* determination of *Brucella* MLVA schemes taking advantage of the increasing number of sequenced genomes (Georgi et al., 2017). As no experimental validation of such script was performed, we now aimed to evaluate the agreement among the MLVA profiles determined through PCR- and WGS-based approaches for strains from several *Brucella* species, in order to check the validity of such technological transition underlying the genetic characterization of *Brucella*.

MATERIALS AND METHODS

Samples

Eighty-three *Brucella* strains isolated in Portugal, Spain, Germany, Hungary, and Belgium were used in this study. This set comprises

[1]https://microbesgenotyping.i2bc.paris-saclay.fr
[2]https://pubmlst.org/
[3]https://insaflu.insa.pt/

essentially not only *B. melitensis* but also some representatives of *B. suis*, *B. abortus*, and *B. ovis* (**Supplementary Table S1**). Unfortunately, we had no access to the isolates of low prevalent species such as *B. ceti*, *B. canis*, and *B. pinnipedialis*. *Brucella melitensis* 16M strain (NC_003317 and NC_003318) was used as a reference strain.

All samples were handled in a BSL-3 biocontainment laboratory at the Portuguese National Institute of Health. *Brucella* isolates were cultured on blood agar for 3–5 days at 37°C under 5% CO_2, and total DNA was extracted from fresh cultures on the NucliSens easyMAG platform (Biomerieux), according to the manufacturer's instructions.

All strains were identified as *Brucella* species by real-time PCR, using a previously published assay (Pelerito et al., 2017). The molecular methods used up to species differentiation were performed in a tandem fashion. Firstly, an "in house" real-time PCR using hydrolysis probes was used to detect and identify *Brucella* genus. Secondly, for species differentiation, primers and Taqman probes were designed within the BMEII0466 gene for *B. melitensis* and BruAb2_0168 gene for *B. abortus* (Gopaul et al., 2008; Pelerito et al., 2017).

MLVA Assays

Single locus amplification of the eight minisatelite loci (panel 1) and eight microsatelite loci (panels 2A and 2B), that constitute the MLVA-16 assay, was performed as describe by Le Fleche (Le Fleche et al., 2006), with modifications by Garofolo (2015). MLVA PCRs were performed in four multiplex reactions in a final volume of 10 µl. The reactions contained: 1× Type-it Multiplex PCR Master Mix (Qiagen), 0.5× solution buffer, primers at appropriate concentrations, and 5 µl of DNA. The thermocycling conditions were as follows: 96°C for 5 min followed by either 30 (for multiplex 1, 3, and 4) or 24 cycles (for multiplex 2) of: 95°C for 30 s, 60°C for 90 s, and 72°C for 30 s; followed by 60°C for 30 min. Multiplex 2 was run for 24 cycles in order to contain VNTR amplification artifacts (Garofolo et al., 2013). A 1.1 µl of each MLVA PCR product was mixed with 15 µl of formamide-diluted GeneScan 500 LIZ dye or GeneScan 1,200 LIZ dye size standards (Applied Biosystems), depending on the expected size of the fragments, and denatured at 96°C for 3 or 5 min, respectively. The mixtures were electrophoresed on an 8-capillary 3500 Genetic Analyzer equipped with 50 cm-long capillaries and POP7 polymer (Applied Biosystems). Estimation of molecular sizes of PCR products was obtained using GeneMapper software 6 (Applied Biosystems) with default analysis parameters. The reference *B. melitensis* 16 M strain, for which the expected size is known for each VNTR locus, was used as control for allele's assignment.

Whole Genome Sequencing

For each strain, WGS was performed as previously described (Pinto et al., 2018). Briefly, quantification and quality assessment of the purified DNA was performed using the DNA HS Assay Kit (Thermo Fisher Scientific) in the Qubit Fluorometer and agarose gel electrophoresis (0.8%), respectively. High-quality DNA samples were then used to prepare dual-indexed Nextera XT Illumina libraries that were subsequently subjected to cluster generation and paired-end sequencing (2 × 250 bp and 2 × 300bp) on a MiSeq Illumina platform (Illumina Inc.), according to the manufacturer's instructions.

Reads quality control and bacterial *de novo* assembly were performed using the INNUca v4.0.1 pipeline,[4] which consists of several integrated modules for reads QA/QC, *de novo* assembly, and post-assembly optimization steps. Briefly, after reads' quality analysis (FastQC v0.11.5)[5] and cleaning (Trimmomatic v0.36; Bolger et al., 2014), genomes were assembled with SPAdes 3.11 (Bankevich et al., 2012) and subsequently improved using Pilon v1.18 (Walker et al., 2014), with genome coverage being monitored and reported after each processes. In order to evaluate the impact of the "post-assembly polishing" on the assembled genomes and subsequently on the *in silico* MLVA analyses, the SPAdes assemblies were also performed skipping the Pilon step. A final check was also performed. Considering that the *in silico* extraction of loci may be influenced by the quality of the assembled genomes, another largely used *de novo* assembler – Velvet (Zerbino and Birney, 2008) was applied through VelvetOptimiser v.2.2.5,[6] for comparative purposes, with and without Pilon. The VelvetOptimiser script was run using trimmed reads for odd k-mer values ranging from 31 to 127 (highest k-mer used in SPAdes), with all program default settings unchanged apart from the minimum output contig size, which was the same as used by SPAdes.

WGS-Based MLVA

Bacterial draft genomes were subjected to a Python script for *in silico* extraction of *Brucella* MLVA scheme (with 16 loci) as previously described (Georgi et al., 2017).[7] This script is based on the count of repetitive DNA stretches contained within conserved DNA boxes that are upstream and downstream to the repetitions. As determining numbers of repeated stretches from WGS data may be error-prone, we carefully checked each locus in respect to the expected total length, internal repeat homogeneity or probability to get collapsed VNTRs during the assembly. All resulting MLVA 16 genotypes were compared to a public database with 2,215 entries of *B. melitensis* strains that can be assessed online (Grissa et al., 2008).[8]

Accuracy Evaluation of the WGS-Based MLVA Approach

To assess the performance of the WGS-based MLVA approach, we determined the percentage of agreement between PCR- and WGS-based MLVA methods by calculating the number of identical results (i.e., identical called alleles), divided by the total number of hits that were detected simultaneously by both approaches. Although the PCR-based MLVA approach is considered as the gold standard (with obvious accuracy increment after the optimization of the multiplex PCR with fluorescent

[4]https://github.com/B-UMMI/INNUca
[5]http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[6]https://github.com/tseemann/VelvetOptimiser
[7]https://github.com/dinergate42/in-silico-typing
[8]http://microbesgenotyping.i2bc.paris-saclay.fr/

dyes for capillary electrophoresis), we believe the use of the "total number of hits that were detected simultaneously by both approaches" as the denominator, as the most reasonable and cautious procedure. In fact, although some hits were exclusively detected by the PCR-based MLVA approach, the opposite scenario was also observed. The maximum number of possible hits is 1,328 (i.e., 83 strains × 16 loci).

Finally, for all strains, the performance of the bioinformatics script in extracting all MLVA loci was also evaluated by taking into account the quality of the draft genome generated by two assemblers (SPAdes and VelvetOptimiser) with and without "post-assembly polishing." Briefly, for each condition, both the mean coverage depth and the number of contigs of each draft genome were correlated with the number of extracted alleles. Pearson's coefficients (r) were measured to see potential linear associations. Nevertheless, as these final evaluations were done as complements of the major strategy, for the sake of clarity, whenever the text refers "WGS-based MLVA approach" it refers to the approach that used SPAdes with Pilon.

RESULTS

The performance of the WGS-based MLVA approach was assessed through the determination of the percentage of agreement with the results obtained by using the gold standard PCR-based MLVA. For the sake of clarity, we defined as "shared hits" the ones that were simultaneously detected by both approaches regardless their correct allele assignment. In this regard, the number of "matching alleles" was estimated using the number of "shared hits" as denominator.

Overall, the WGS-based MLVA approach detected 1,265 of the 1,328 possible hits (95.3%). One, two, and ≥3 loci yielded no results for 18 (21.6%), 7 (8.4%), and 7 (8.4%) of the strains, respectively (**Figure 1**). Regarding the PCR-based MLVA approach, it detected 1,269 (95.6%) hits, whereas one, two, and ≥3 loci yielded no results for 18 (21.6%), 8 (9.6%), and 6 (7.2%) strains, respectively.

The allelic profiles obtained by both approaches are presented in **Supplementary Table S2**. Of the total 1,328 possible hits, 1,222 (92.0%) were simultaneously detected by both approaches ("shared hits"), whereas 6.8% were differently detected solely by one approach. These discrepancies were mostly observed for loci Bru07 and Bru09 (both from panel 2B) and Bru11 (from panel 1; **Figure 2**). Both MLVA approaches simultaneously failed the detection of 16 out of the 1,328 hits (1.2%), where 12 fall in panel 2B, specially focusing loci Bru04 and Bru07.

The analysis *per* locus showed an agreement rate of MLVA profiles (ratio of "matching alleles" *per* number of "shared hits") determined through PCR- and WGS-based approaches ranging from 89.9 to 100.0% (**Figure 2**). The discrepancies involved loci from all three panels, being more pronounced in six loci (Bru06, Bru12, Bru55, Bru19, Bru21, and Bru04). In general, the mean of agreement of MLVA profiles for all 16 loci was 96.4%, revealing a high allele concordance between the two approaches.

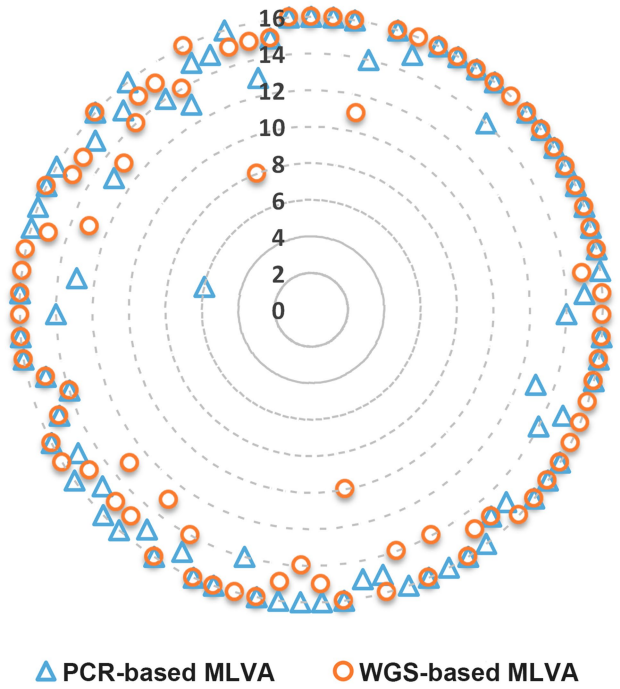


FIGURE 1 | Number of MLVA loci detected per strain through PCR- and WGS-based MLVA approaches. The data for each approach are disposed along the imaginary radius of the circle graph (one radius *per* strain). The WGS-based results are relative to the MLVA loci extraction using draft genomes assembled with SPAdes 3.11 (Walker et al., 2014) and subsequently improved with Pilon v1.18 (Bankevich et al., 2012) from the INNUca v4.0.1 pipeline (see methods for details).

We also inspected the quality of the draft sequences used as input because the Python script for *in silico* extraction of *Brucella* MLVA schemes is applied after the genome assembly and thus may be dependent on the quality of the "reads." The influence of the mean depth of coverage and number of contigs on the efficacy of the bioinformatics script is illustrated in **Figure 3**. As expected, a negative linear correlation was observed among the efficacy of the *in silico* extraction and the number of assembled contigs, with less partitioned genomes allowing the detection of a higher number of alleles. On the other hand, higher genome mean coverage depth seem to favour the *in silico* extraction of MLVA loci.

As a final assessment, the performance of the bioinformatics script was also evaluated by using as input draft genome sequences assembled with different assemblers (SPAdes versus VelvetOptimiser; **Supplementary Figure S1**). Curiously, although no significant differences were observed regarding the number of loci extracted both with and without "post-assembly polishing" (data not shown), we observed that VelvetOptimiser was particularly affected by a low depth of coverage and genome fragmentation. In fact, for a mean depth of coverage <50 as well as for high fragmented genomes, the number of detected MLVA loci decreased more sharply when we used VelvetOptimiser than when SPAdes assemblies were used.

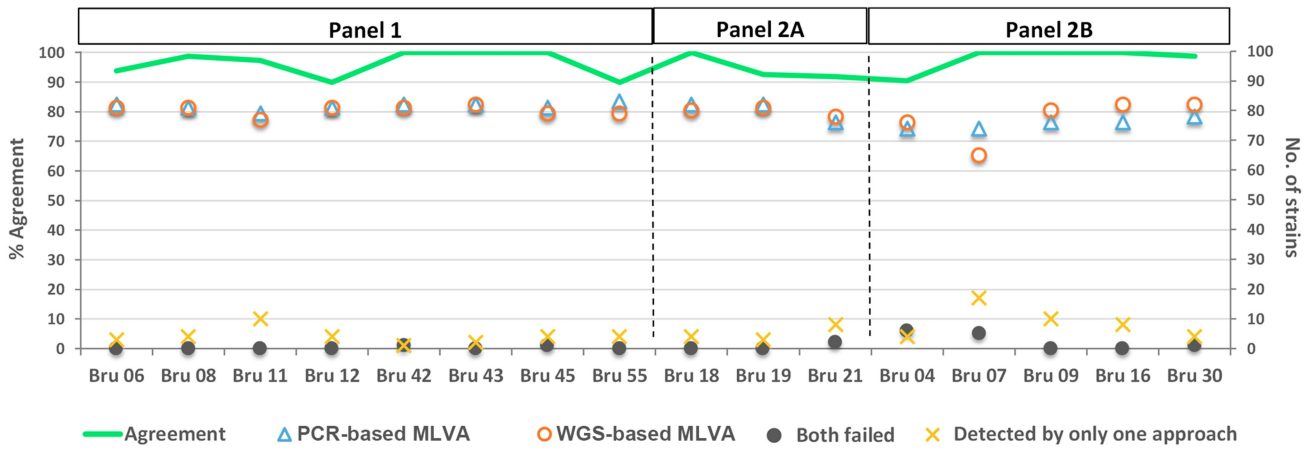


FIGURE 2 | Performance of the PCR- and WGS-based MLVA approaches *per* locus. The graph represents the number of strains (right YY scale), for which it was possible to determine an allele *per* locus using each approach. The green line shows the percentage of agreement (left YY scale) *per* locus between both approaches (i.e., when identical alleles were called between the two methodologies). The loci are grouped according to the MLVA-16 panel they belong to (i.e., Panel 1, Panel 2A, and Panel 2B). The WGS-based results are relative to the MLVA loci extraction using draft genomes assembled with the INNUca v4.0.1 pipeline using Pilon v1.18 (Bankevich et al., 2012).

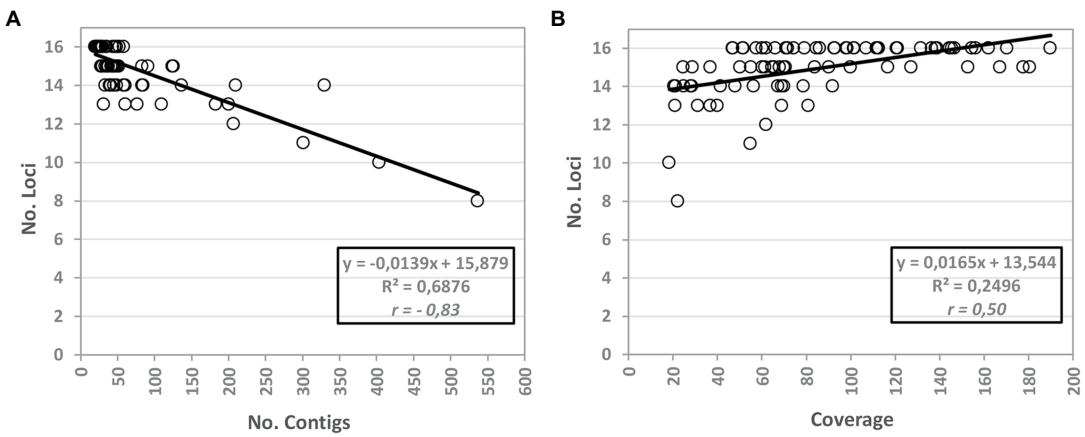


FIGURE 3 | Influence of the mean coverage depth and number of contigs on the efficacy of the WGS-based MLVA extraction. The graphs show the correlation of the efficacy (measured by the number of loci for which an allele was called) of the bioinformatics script with the number of assembled contigs **(A)** and with the depth of coverage **(B)** after quality improvement. For **(A,B)**, the tendency lines are shown with the respective equations and the Pearson coefficient (*r*). The WGS-based results are relative to the MLVA loci extraction using draft genomes assembled with the INNUca v4.0.1 pipeline using Pilon v1.18 (Bankevich et al., 2012).

DISCUSSION

The control of brucellosis requires an accurate surveillance and the use of high discriminatory methods to characterize outbreak strains and determine the infection source and transmission routes. For many years, multiple typing methods were used for *Brucella* characterization at both species and biovar levels. These relied on host specificity, growth features, biochemical reactions, serotyping and bacteriophage typing, but they lacked discriminatory power (Sun et al., 2017). There are some studies focusing the application of WGS-based approaches to *Brucella* spp., where strains from multiple species and locations are commonly included (Pisarenko et al., 2018; Ashford et al., 2020; Esquivel et al., 2020; Pelerito et al., 2020). Nevertheless, currently, PCR-based MLVA is the most widely used approach for outbreak investigations and is still considered the gold standard for *Brucella* typing. Although the traditional MLVA assay relies on singleplex PCR followed by gel electrophoresis, some laboratories have already adopted a less error-prone approach based on multiplex PCRs and multicolor capillary electrophoresis (Garofolo, 2015; Vergnaud et al., 2018). The sixteen markers enrolled in the MLVA-16 scheme are a combination of moderately variable (minisatellites, panel 1) and highly discriminatory (microsatellites, panels 2A and 2B) *loci* (Al Dahouk et al., 2007). A MLVA typing

assay depends on the selection of markers which individually would not provide a relevant clustering. Taken separately, the Tandem Repeat markers are either not informative enough, are too variable or show a high level of homoplasy. As such, the combination of well selected independent loci may be highly discriminatory as previously shown for other species (Le Fleche et al., 2006).

On behalf of the unavoidable transition from the classical typing to the WGS-based approaches, Python scripts were recently developed for the rapid *in silico* extraction of the *Brucella* MLVA alleles (Georgi et al., 2017; Vergnaud et al., 2018). This will allow the assignment of the MLVA types in the genomic era, avoiding the undesirable loss of genetic information that has been provided during more than 10 years by using the gold standard PCR-based MLVA typing. This is also important because the timings for the technology transition will vary considerably among different countries. As no experimental evaluation had been performed so far for one of those scripts (Georgi et al., 2017), our main goal was to access its performance against the gold standard method.

Overall, the WGS-based MLVA approach detected 95.3% of all possible 1,328 hits and the agreement with the experimental method in detecting the correct alleles was found to be 96.4%. Despite this high level of agreement, some loci could not be detected either by both or by solely a single approach. Although most of the loci did not simultaneously fail for several strains, it is worth highlighting Bru04 and Bru07 (both from panel 2B), for which the PCR-based MLVA approach failed for nine out of the 83 strains under study. We may hypothesize the occurrence of experimental difficulties associated with a less optimized PCR for these specific loci. However, we observed that Bru04 and Bru07 were also the loci for which the WGS-based MLVA approach most failed (for seven and 18 strains, respectively). We speculate that this relies on the fact that these loci are among the ones with the shortest repetitive sequence stretches. This could hypothetically be an obstacle for a more precise experimental distinction between the number of repeats. Bioinformatically, short repetitive stretches may eventually impact MLVA loci determination mainly when high genome fragmentation occurs.

According to the results obtained for the present dataset, we observed that the performance of the WGS-based MLVA approach does not seem to be dependent on the post-assembly polishing but is clearly dependent on the depth of coverage and the degree of assembly fragmentation (where SPAdes performed better). Still, a minimal depth of coverage of ~50x and a maximum number of ~200 contigs (a range where both assemblers behaved similarly) seem to constitute guiding "boundaries" for the future application of the script. Of note, some failures underlying the WGS-based MLVA approach could hypothetically be solved by the long-reads technology as it usually overcomes the assembly problems associated with the repeated sequences, although this technology is more prone to yield sequencing errors. On the other hand, the use of very short reads (e.g., 100 nt) may yield highly fragmented genomes and thus a more erroneous determination of the repeat number. In this study, we used the Illumina technology with reads of 250–300 nt, which, together with a minimum depth of coverage of ~50x, yielded highly solid results.

In conclusion, although this study focused the most prevalent species (with emphasis in *B. melitensis*) the evaluated script seems to be a very useful and robust tool for *in silico* extraction of MLVA types of *Brucella* strains, dealing with a large number of samples in a short time period, and allowing retrospective and prospective molecular epidemiological studies. This allows a continuous and non-disruptive transition to a new typing era by putting the newly sequenced strains in the frame of the genetic characterization obtained for thousands of isolates collected worldwide throughout the last decade. This will certainly be important for public health reference laboratories to maintain an active epidemiological surveillance of brucellosis.

DATA AVAILABILITY STATEMENT

All raw sequence reads used in the present study were deposited in the European Nucleotide Archive under the study accession number PRJEB30030.

AUTHOR CONTRIBUTIONS

AP, AN, and JG conceived and designed the experiments, performed the analysis, interpreted the results, and wrote the manuscript. AP, JI, CS, AF, SV, and MN performed the experiments. EG performed the script optimization and implementation. All authors contributed to the article and approved the submitted version.

FUNDING

This work is a result of the GenomePT project (POCI-01-0145-FEDER-022184), supported by the COMPETE 2020 – Operational Program for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Program (Lisboa2020), Algarve Portugal Regional Operational Program (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by the Fundação para a Ciência e a Tecnologia (FCT). The studies have arisen from the Project QUANDHIP (Chafea Grant Agreement no. 2010 21 02), which has been funded by the European Commission in the framework of the Health Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.740068/full#supplementary-material

REFERENCES

Al Dahouk, S., Flèche, P. L., Nöckler, K., Jacques, I., Grayon, M., Scholz, H. C., et al. (2007). Evaluation of *Brucella* MLVA typing for human brucellosis. *J. Microbiol. Methods* 69, 137–145. doi: 10.1016/j.mimet.2006.12.015

Ashford, R. T., Muchowski, J., Koylass, M., Scholz, H. C., and Whatmore, A. M. (2020). Application of whole genome sequencing and pan-family multi-locus sequence analysis to characterize relationships within the family Brucellaceae. *Front. Microbiol.* 11:1329. doi: 10.3389/fmicb.2020.01329

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

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

Borges, V., Pinheiro, M., Pechirra, P., Guiomar, R., and Gomes, J. P. (2018). INSaFLU: an automated open web-based bioinformatics suite "from-reads" for influenza whole-genome-sequencing-based surveillance. *Genome Med.* 10:46. doi: 10.1186/s13073-018-0555-0

Bricker, B. J., Ewalt, D. R., and Halling, S. M. (2003). *Brucella* 'HOOF-Prints': strain typing by multi-locus analysis of variable number tandem repeats (VNTRs). *BMC Microbiol.* 3:15. doi: 10.1186/1471-2180-3-15

De Santis, R., Ciammaruconi, A., Faggioni, G., Fillo, S., Gentile, B., di Giannatale, E., et al. (2011). High throughput MLVA-16 typing for *Brucella* based on the microfluidics technology. *BMC Microbiol.* 11:60. doi: 10.1186/1471-2180-11-60

Dean, A. S., Crump, L., Greter, H., Schelling, E., and Zinsstag, J. (2012). Global burden of human brucellosis: a systematic review of disease frequency. *PLOS Negl. Trop. Dis.* 6:e1865. doi: 10.1371/journal.pntd.0001865

Franc, K. A., Krecek, R. C., Hasler, B. N., and Arenas-Gamboa, A. M. (2018). Brucellosis remains a neglected disease in the developing world: a call for interdisciplinary action. *BMC Public Health* 18:125. doi: 10.1186/s12889-017-5016-y

Franco, M. P., Mulder, M., Gilman, R. H., and Smits, H. L. (2007). Human Brucellosis. *Lancet Infect. Dis.* 7, 775–786. doi: 10.1016/S1473-3099(07)70286-4

Garofolo, G. (2015). Multiple-locus variable-number tandem repeat (VNTR) analysis (MLVA) using multiplex PCR and multicolor capillary electrophoresis: application to the genotyping of *Brucella* species. *Methods Mol. Biol.* 1247, 335–347. doi: 10.1007/978-1-4939-2004-4_24

Garofolo, G., Ancora, M., and Giannatale, E. (2013). MLVA – 16 loci panel on *Brucella* spp. using multiplex PCR and multicolor capillary electrophresis. *J. Microbiol. Methods* 92, 103–107. doi: 10.1016/j.mimet.2012.11.007

Gee, J. E., de, B. K., Levett, P. N., Whitney, A. M., Novak, R. T., and Popovic, T. (2004). Use pf 16s rRNA gene sequencing for rapid confirmatory identification of *Brucella* isolates. *J. Clin. Microbiol.* 42, 3649–3654. doi: 10.1128/JCM.42.8.3649-3654.2004

Georgi, E., Walter, M. C., Pfalzgraf, M. T., Northoff, B. H., Holdt, L. M., Scholz, H. C., et al. (2017). Whole genome sequencing of *Brucella melitensis* isolated from 57 patients in Germany reveals high diversity in strains from Middle East. *PLoS One* 12:e0175425. doi: 10.1371/journal.pone.0175425

Gopaul, K. K., Koylass, M. S., Smith, C. J., and Whatmore, A. M. (2008). Rapid identification of *Brucella* isolates to the species level by real time PCR based single nucleotide polymorphism (SNP) analysis. *BMC Microbiol.* 8:14. doi: 10.1186/1471-2180-8-86

Grissa, I., Bouchon, P., Pourcel, C., and Vergnaud, G. (2008). On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. *Biochimie* 90, 660–668. doi: 10.1016/j.biochi.2007.07.014

Hanot Mambres, D., Boarbi, S., Michel, P., Bouker, N., Escobar-Calle, L., Desqueper, D., et al. (2017). Imported human brucellosis in Belgium: BIO and molecular typing of bacterial isolates, 1996-2015. *PLoS One* 12:e0174756. doi: 10.1371/journal.pone.0174756

Janowicz, A., de Massis, F., Ancora, M., Cammà, C., Patavino, C., Battisti, A., et al. (2018). Core genome multilocus sequence typing and single nucleotide polymorphism analysis in the epidemiology of *Brucella melitensis* infections. *J. Clin. Microbiol.* 56, e00517–e00518. doi: 10.1128/JCM.00517-18

Kattar, M. M., Jaafar, R. F., Araj, G. F., Le Flèche, P., Matar, G. M., Abi Rached, R., et al. (2008). Evaluation of a multilocus variable number tandem repeat analysis scheme for a typing human *Brucella* isolates in a region of brucellosis endemicity. *J. Clin. Microbiol.* 46, 3935–3940. doi: 10.1128/JCM.00464-08

Le Flèche, P., Jacques, I., Grayon, M., Al Dahouk, S., Bouchon, P., Denoeud, F., et al. (2006). Evaluation and selection of tandem repeat loci for a *Brucella* MLVA typing assay. *BMC Microbiol.* 6:9. doi: 10.1186/1471-2180-6-9

MLVA Bank (n.d.). *MLVA Bank for Microbes Genotyping*. Available at: http://microbesgenotyping.i2bc.paris-saclay.fr/

Pelerito, A., Matos, R., Santos, M. A., Soeiro, S., Santos, J., Manita, C., et al. (2017). Human brucellosis in Portugal – Retrosptective analysis of suspected clinical cases of infection from 2009 to 2016. *PLoS One* 12:e0179667. doi: 10.1371/journal.pone.0179667

Pelerito, A., Nunes, A., Núncio, M. S., and Gomes, J. P. (2020). Genome-scale approach to study the genetic relatedness among Brucella melitensis strains. *PLoS One* 15:e0229863. doi: 10.1371/journal.pone.0229863

Pinto, M., González-Díaz, A., Machado, M. P., Duarte, S., Vieira, L., Carriço, J. A., et al. (2018). Insights into the population structure and pan-genome of *Haemophilus influenzae*. *Infect. Genet. Evol.* 67, 126–135. doi: 10.1016/j.meegid.2018.10.025

Pisarenko, S. V., Kovalev, D. A., Volynkina, A. S., Ponomarenko, D. G., Rusanova, D. V., Zharinova, N. V., et al. (2018). Global evolution and phylogeography of *Brucella melitensis* strains. *BMC Genomics* 19:353. doi: 10.1186/s12864-018-4762-2

Sankarasubramanian, J., Vishnu, U. S., Gunasekaran, P., and Rajendhran, J. (2019). Development and evaluation of a core genome multilocus sequence typing (cgMLST) scheme for *Brucella* spp. *Infect. Genet. Evol.* 67, 38–43. doi: 10.1016/j.meegid.2018.10.021

Suárez-Esquivel, M., Hernández-Mora, G., Ruiz-Villalobos, N., Barquero-Calvo, E., Chacón-Díaz, C., Ladner, J. T., et al. (2020). Persistence of Brucella abortus lineages revealed by genomic characterization and phylodynamic analysis. *PLoS Negl. Trop. Dis.* 14:e0008235. doi: 10.1371/journal.pntd.0008235

Sun, M., Jing, Z., di, D., Yan, H., Zhang, Z., Xu, Q., et al. (2017). Multiple locus variable – number tandem – repeat and single – nucleotide polymorphism – based *Brucella* typing reveals multiple lineages in *Brucella melitensis* currently endemic in China. *Front. Vet. Sci.* 4:215. doi: 10.3389/fvets.2017.00215

Tan, K. K., Tan, Y. C., Chang, L. Y., Lee, K. W., Nore, S. S., Yee, W. Y., et al. (2015). Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genomics* 16:93. doi: 10.1186/s12864-015-1294-x

Valdezate, S., Navarro, A., Villalón, P., Carrasco, G., and Saéz-Nieto, J. A. (2010). Epidemiological and phylogenetic analysis of spanish human *Brucella melitensis* strains by multiple locus variable number tandem repeat typing, hypervariable octameric oligonucleotide fingerprinting, and rpoB typing. *J. Clin. Microbiol.* 48, 2734–2740. doi: 10.1128/JCM.00533-10

Vergnaud, G., Hauck, Y., Christiany, D., Daoud, B., Pourcel, C., Jacques, I., et al. (2018). Genotypic expansion within the population structure of classical *Brucella* species revealed by MLVA16 typing of 1404 *Brucella* isolates from different animal and geographic origins. *Front. Microbiol.* 9:1545. doi: 10.3389/fmicb.2018.01545

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Whatmore, A. M., Shankster, S. J., Perrett, L. L., Murphy, T. J., Brew, S. D., Thirlwall, R. E., et al. (2006). Identification and characterization of VARIABLE-number tandem – REPEAT MARKERS for typing of *Brucella* spp. *J. Clin. Microbiol.* 44, 1982–1993. doi: 10.1128/JCM.02039-05

WHO (2020). *Neglected Tropical Diseases*. Available at: http://www.who.int/neglected_diseases/zoonoses/other_NZDs/en/ (Accessed October 06, 2021).

Young, E. J. (2005). "*Brucella* species," in *Principles and Practice of Infectious Diseases. 6th Edn*. eds. G. L. Mandell, J. E. Bennet and R. Dolin (Philadelphia: Churchill Livingstone), 2669–2772.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

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.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 April 2022
doi: 10.3389/fmicb.2022.687387

Transcriptome Analysis of *Otodectes cynotis* in Different Developmental Stages

*Ran He[1], Qian Zhang[1], Xiaobin Gu[1], Yue Xie[1], Jing Xu[1], Xuerong Peng[2] and Guangyou Yang[1]**

[1] Department of Parasitology, College of Veterinary Medicine, Sichuan Agricultural University, Chengdu, China, [2] Department of Chemistry, College of Life and Basic Science, Sichuan Agricultural University, Chengdu, China

The mite *Otodectes cynotis* is distributed worldwide and parasitism the ear canals of cats and dogs, causing otitis externa. Molecular biology of *O. cynotis* is poorly understood, with only a few genes being deposited in public databases. In the present study, we aimed to perform transcriptome analysis of *O. cynotis* using SMRT and Illumina sequencing of RNA from different development stages. SMRT-Seq of *O. cynotis* demonstrated 5,431 final transcripts, including 406 long non-coding RNAs and 2,698 differentially expressed genes (DEGs), including 1,357 up-regulated genes and 1,341 down-regulated genes between adult mites and nymph/larva. A total of 397 putative allergen genes were detected, 231 of which were DEGs. Among them, 77 were homologous of known mite allergens. The expression level of allergen genes hints at the pathogenicity of mites in different life stages, and the protein interaction network analysis could identify possible key genes in the pathogenic mechanism. Intriguingly, Gene Ontology analysis showed that most of the (DEGs) were associated with the terms hydrolase activity and proteolysis. Kyoto Encyclopedia of genes and genomes (KEGG) analysis identified drug metabolism-cytochrome P450 signal pathway as one of the top pathways. SMRT-Seq of the full-length transcriptome of *O. cynotis* was performed first, and a valuable resource was acquired through the combination analysis with the Illumina sequencing data. The results of our analyses provide new information for further research into *Otodectes cynotis*.

Keywords: *Otodectes cynotis*, SMRT, Illumina, different expressed genes, allergen

OPEN ACCESS

Edited by:
Michael Kemp, University of Southern Denmark, Denmark

Reviewed by:
Stewart Burgess, Moredun Research Institute, United Kingdom
Guanghui Zhao, Northwest A&F University, China

***Correspondence:**
Guangyou Yang
guangyou1963@aliyun.com

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *29 March 2021*
Accepted: *25 February 2022*
Published: *04 April 2022*

Citation:
He R, Zhang Q, Gu X, Xie Y, Xu J, Peng X and Yang G (2022) Transcriptome Analysis of Otodectes cynotis in Different Developmental Stages. Front. Microbiol. 13:687387. doi: 10.3389/fmicb.2022.687387

INTRODUCTION

The ear mite *O. cynotis* is the most common etiological agent of otitis externa in cats and dogs (Bosco et al., 2019; Silva et al., 2020). It also parasitizes in the ear canal of ferrets, red foxes, and terrestrial carnivores (Taenzler et al., 2017; Briceño et al., 2020; Fanelli et al., 2020; Huang et al., 2020). Animals can be infected by direct or indirect contact with infected animals (Six et al., 2016). A brown waxy substance can be found in the ear canal of mildly infested cats, which then form a crust on the surface of the ear canal. As the irritation intensifies, itching becomes more and more obvious. Cats keep shaking their heads, scratching their ears, and rubbing their ears on objects due to the itching, resulting to hematomas and ulcers (Montoya, 2018). Cats may cramp or move in circles in severe cases. In addition, purulent otitis externa can be caused by the occurrence

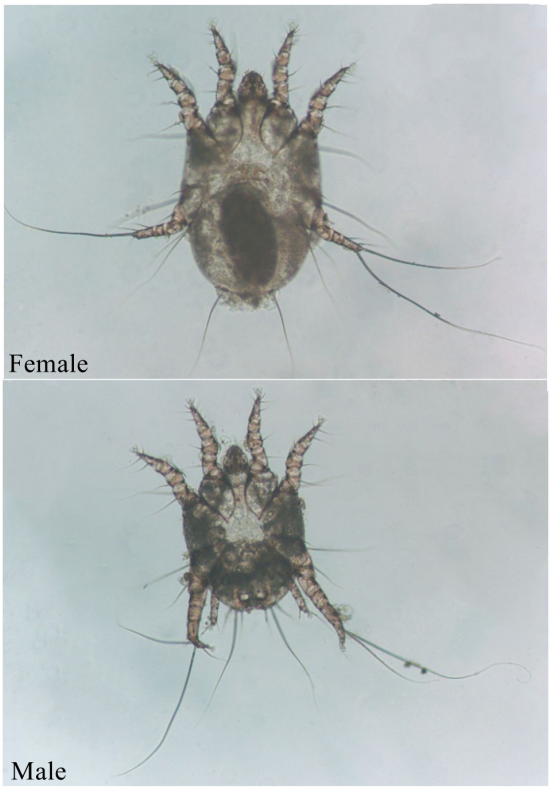


FIGURE 1 | Adult *O. cynotis* mites.

of secondary bacterial infections (Roy et al., 2011). The life cycle of *O. cynotis* is incomplete metamorphosis (egg, larva, nymph, and adult) (Tonn, 1961). The mite is long-legged and can be easily recognized as whitish spots in the ear canal. Underside of its body, the chitinous bars behind the either front legs meet to form a V shape, and the female mites have suckers only on the front two pairs of legs (**Figure 1**). *O. cynotis* mites live on the surface of the ear canal of cats and dogs (Curtis, 2004). They pierce the host skin with its mouthparts, and feed on the lymph, tissue fluid, and blood of the host (Powell et al., 1980), they cause stimulation of the parasitic site, leading to dermatitis or allergic reactions (Montoya, 2018), causing excessive keratinization and proliferation of epithelial cells (Powell et al., 1980). Meanwhile, inflammatory cells, especially mast cells and macrophages increase, and subcutaneous blood vessels dilate (Kelleher, 2001).

Current research on *O. cynotis* focuses mainly on case reports, drug treatments, and *in vitro* acaricidal (Carithers et al., 2016; Yang and Huang, 2016). Only a few studies of its molecular biology (Salib and Baraka, 2011; Taenzler et al., 2017). As a result, there are only a limited number genes from *O. cynotis* mites deposited in public databases (Hu et al., 2019). In the present study, we aim to use single-molecule real-time sequencing (SMRT-Seq) and Illumina based RNA-sequencing (RNA-Seq) to obtain sequence information of *O. cynotis* mites, analyze the expression of allergen genes at different life stages, and determine the protein-protein interaction (PPI) network of *O. cynotis* mites.

MATERIALS AND METHODS

Ethics Approval and Consent to Participate

This study was approved by the Animal Care and Use Committee of Sichuan Agricultural University (SYXK2019-187). All animal procedures used in this study were carried out in strict accordance with the guide for the Care and Use of Laboratory Animals (National Research Council, Bethesda, MD, United States) and the recommendations of the ARRIVE guidelines.[1] All methods were carried out in accordance with relevant guidelines and regulations.

Sample Collection

O. cynotis mites were isolated from the crust in the ear canal of naturally infested cats. Cats were obtained from the cattery in Chengdu City. The veterinarian treated cats accordingly after the crusts were taken. Crusts were incubated in petri dishes at 35°C, and mites were collected every half an hour. Live mites (larvae, nymphs, and adults) were identified depending on their morphological criteria (Ahn et al., 2013). Samples that included larva, nymph, and adult mites (Ocy_A_N_L) were utilized for SMRT-Seq Meanwhile, nymph/larva mite samples (Ocy_N_L) and adult mite samples (Ocy_A) were prepared for Illumina sequencing. Each group comprised three samples.

Total RNA was isolated using RNeasy Plus Mini Kit (Qiagen, Valencia, CA, United States), the mites were homogenized by full grinding. The purity and concentration of the RNA were evaluated using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Rockland, DE, United States) using the OD260/280 ratio. The integrity of RNA was determined using agarose gel electrophoresis on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States).

Single-Molecule Real-Time Sequencing

The Iso-Seq library of the Ocy_A_N_L group was prepared according to the Isoform Sequencing protocol using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PacBio) (PN 100-092-800-03; Pacific Biosciences, Menlo Park, CA, United States). Full length cDNAs were obtained, subjected to DNA damage repair and end repair, and then ligated to sequencing adapters and digested with exonuclease. Libraries that met the quality criteria were sequenced on the PacBio Sequel platform according to the effective concentration and data output requirements of the library.

Illumina Sequencing

For the Ocy_N_L and Ocy_A groups, a total amount of 1.5 μg RNA per sample was used as input material for the RNA

[1] http://www.nc3rs.org.uk/arrive-guidelines

sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H^-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends *via* exonuclease/polymerase activities. After adenylation of 3′ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 250∼300 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, United States). Then 3 μl USER Enzyme (NEB, United States) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq platform NovaSeq 6000 (Illumina, San Diego, CA, United States) and 125 bp/150 bp paired-end reads were generated. The fastp (0.19.7, -g -q 5 -u 50 -n 15 -l 150) was used to conduct quality control on Illumina raw data. Raw reads of fastq format were firstly processed through in-house perl scripts. In this step, clean reads were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads from raw reads. At the same time, Q20, Q30, and GC content the clean data were calculated. All the downstream analyses were based on the clean data with high quality.

Single-Molecule Real-Time Reads Processing

Sequence data were processed using the SMRTlink 7.0 software (PacBio). Circular consensus sequences (CCSs) were generated from the subread BAM files. The CCS.BAM files were used as the output, which were then classified into full-length and non-full-length reads using Lima (PacBio), and polyA sequences were removed using refine (PacBio). The full-length fasta files produced were then subjected to isoform-level clustering [n*log(n)]. Additional nucleotide errors in consensus reads were corrected using the Illumina RNAseq data using the software LoRDEC (V0.7;-k 23;-s3) (Salmela and Rivals, 2014). Any redundancy in the corrected consensus reads was removed using CD-HIT (v4.6.8; -c0.95; -T6; -G0; -aL0.00; -aS0.99; -AS30) (Fu et al., 2012) to obtain final transcripts for subsequent analysis.

Functional Annotation of Genes

Non-redundant transcript sequence obtained after CD-HIT deduplication were regarded as genes and were grouped and mapped to seven protein and nucleic acid databases to obtain gene annotation information. These databases included the NCBI non-redundant protein sequence database (NR) (Li et al., 2002), NCBI nucleotide sequences (Nt), Protein family (Pfam) (Finn et al., 2016), Clusters of Orthologous Groups of proteins and euKaryotic Ortholog Groups (KOG/COG) (Tatusov et al., 2003), a manually annotated and reviewed protein sequence database (Swiss-prot) (Bairoch and Apweiler, 2000), Gene Ontology (GO) (Ashburner et al., 2000) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2004). For BLAST searching, we set the e-value to "1e-5" in the NT database analysis and used the Diamond BLASTX software (v0.8.36) and set the e-value to "1e-5" for the NR, KOG, Swiss-Prot, and KEGG analyses.

Long Non-coding RNA (lncRNA) Prediction

PLEK (version 1.2)[2] (Li et al., 2014) and CNCI (version 2)[3] (Sun et al., 2013) were used to predict the gene coding potential of transcripts obtained after CD-HIT deduplication. Then, genes predicted from PLEK and CNCI were blast searched against databases of known proteins using CPC (2.0)[4] (Kang et al., 2017). The transcripts predicted by PLEK, CNCI, and CPC were subjected to hmmscan homologous searching (PfamScan, 1.6)[5] in the Pfam database. Transcripts that predicted by PLEK, CNCI, and CPC software, and homologous searched with Pfam database were putatively identified as lncRNAs.

Identification of Differentially Expressed Genes and Functional Categorization

The clean reads from Illumina RNA-Seq were mapped back onto the final transcript sequences from SMRT-Seq by using Bowtie (version V2.3.4, -q;–phred33;–sensitive;–dpad 0;–gbar 99999999;–mp 1,1;–np 1;–score-min L,0,-0.1;-I 1;-X 1000;–no-mixed;–no-discordant;-p 8;-k 30) (**Supplementary Table 1**). Gene expression levels were calculated by RSEM (V1.3.0, –phred33;-quals;–forward-prob 0.5;–time) with FPKM (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) method. Differential expression analysis of the Ocy_N_L group and the Ocy_A group was carried out using the DESeq2 R package (1.16.1) (Love et al., 2014). DESeq2 provides statistical routines to determine differential expression in digital gene expression data using a model based on the negative binominal distribution.

The resulting *P*-values were adjusted using the Benjamini and Hochberg's approach to control the false discovery rate. Genes with an adjusted *P*-value < 0.05 found by DESeq2 were assigned as differentially expressed.

[2]https://sourceforge.net/projects/plek/

[3]https://github.com/www-bioinfo-org/CNCI

[4]http://cpc2.gao-lab.org/download.php

[5]https://www.ebi.ac.uk/seqdb/confluence/display/THD/PfamScan

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis of Differentially Expressed Genes

Gene Ontology (GO) enrichment analysis of Differentially Expressed Genes (DEGs) was implemented by the GOseq (version 1.10.1),[6] in which gene length bias was corrected. GO terms with corrected *P*-value less than 0.05 were considered significantly enriched by the DEGs. KEGG is a database resource for understanding high-level function and utilities of the biological systems, such as cells, organisms, and ecosystems, from molecular level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.[7] We got KOBAS software (version 3.0)[8] to test the statistical enrichment of DEGs in KEGG pathways.

Protein-Protein Interaction Network Construction and Hub Gene Identification

The PPI network for the *O. cynotis* gene products was predicted using Search Tool for the Retrieval of Interacting Genes (STRING, version 11.0)[9] online database. The PPI network was analyzed using Cytoscape software (version 3.5.1),[10] which also calculated the degree of each protein node. The genes with the top 10-degree scores were identified as hub genes.

Analysis of Allergen Genes

The 5,431 final transcripts of *O. cynotis* demonstrated from SMRT-Seq were BLAST (e-vaule;10e-5) searched against allergen sequences from the allergome database website[11] to obtain homologous sequence. These homologous sequences can be regarded as putative allergen genes (Chruszcz et al., 2018). The sequence homologs of known mite allergens and the DEGs were filtered from the putative allergen genes.

Quantitative Real-Time Reverse Transcription PCR Identification

Sixteen genes were selected to validate their expression levels using quantitative real-time reverse transcription PCR (qRT-PCR). Among these genes, five were hub genes from PPI network analysis, and eleven genes are putative allergen genes with high FPKM values in the Ocy_N_L and Ocy_A groups. The qRT-PCR analysis was carried according to previous research (He et al., 2019).

[6]http://www.bioconductor.org/packages/release/bioc/html/goseq.html

[7]http://www.genome.jp/kegg/

[8]http://bioinfo.org/kobas/download/

[9]http://string-db.org/

[10]https://cytoscape.org

[11]http://www.allergome.org/

RESULTS

Transcriptome Analysis Using PacBio Sequel

The full-length transcriptome of *O. cynotis* mite was generated by PacBio Sequel. In total, 12,842,815 subreads with an average of 1,417 bp from 18.2 Gb of data were obtained. Then 191,632 CCS were generated after removing adapters and artifacts, including 113,064 full-length non-chimeric (FLNC) reads. The length of the FLNCs ranged from 50 to 14,775 nt, with an average of 1,679 nt. The number of polished consensus sequences was 14,210, and the mean length of polished consensus sequences was 1,653 nt (range = 57–8,193 nt). After removing redundant sequences from the corrected consensus reads by CD-HIT, 5,431 final transcripts with an average of 1,774 nt were obtained for subsequent analysis.

Functional Annotation of Genes

A Venn diagram of the gene functional annotation from five databases is given in **Supplementary Figure 1**. The diagram shows the transcript numbers annotated in the different databases, and the overlapping relationship between these five databases. A total of 1,260 transcripts were annotated in the Nt database; 3,683 in the Pfam database; 3,692 in the KOG/COG database; 3,908 in the Swiss-prot database, and 3,684 and 4,128 in GO and KEGG databases (**Supplementary Figure 2**), respectively. In the GO annotation, 5,510; 7,723; and 4,369 genes were annotated in the cellular component, biological process, and molecular function categories, respectively (**Figure 2**). The homologous species of *O. cynotis* were predicted by sequence alignment based on hits from the NR database. Of all the gene hits to NR from BLASTx, the transcripts had the highest number of hits to *Sarcoptes scabiei* (3,138 genes), followed by *Rhagoletis zephyria* (711 genes), and *Tetranychus urticae* (172 genes) (**Supplementary Figure 3**).

LncRNA Prediction

We identified 406 lncRNAs of 5,431 genes (S1), as showed in the Venn diagram of the number of LncRNAs predicted according to CNCI, Pfam, Plek, and CPC analyses (**Supplementary Figure 4**). Two of these lncRNAs were nearly 3,000 nt. The functions of these lncRNAs will be characterized in a future study.

Gene Expression Level Analysis

To assess the changes in gene expression levels in the different life stages of *O. cynotis*, Illumina sequencing clean reads were compared to the set of final transcripts obtained from PacBio sequencing. Genes expression levels were quantified using Fragments Per Kilo-base of transcript per Million mapped reads (FPKM). The number of genes with different FPKM values from six samples is shown in **Figure 3**. Over 40% genes in the adult mites had a highest FPKM of more than 60, whereas, the FPKM values of more than 40% of the nymph/larva mites genes were located between 15 and 60. To evaluate the reproducibility of the RNA-Seq data from the three biological replicates for each sample, the Pearson correlation analysis was conducted using

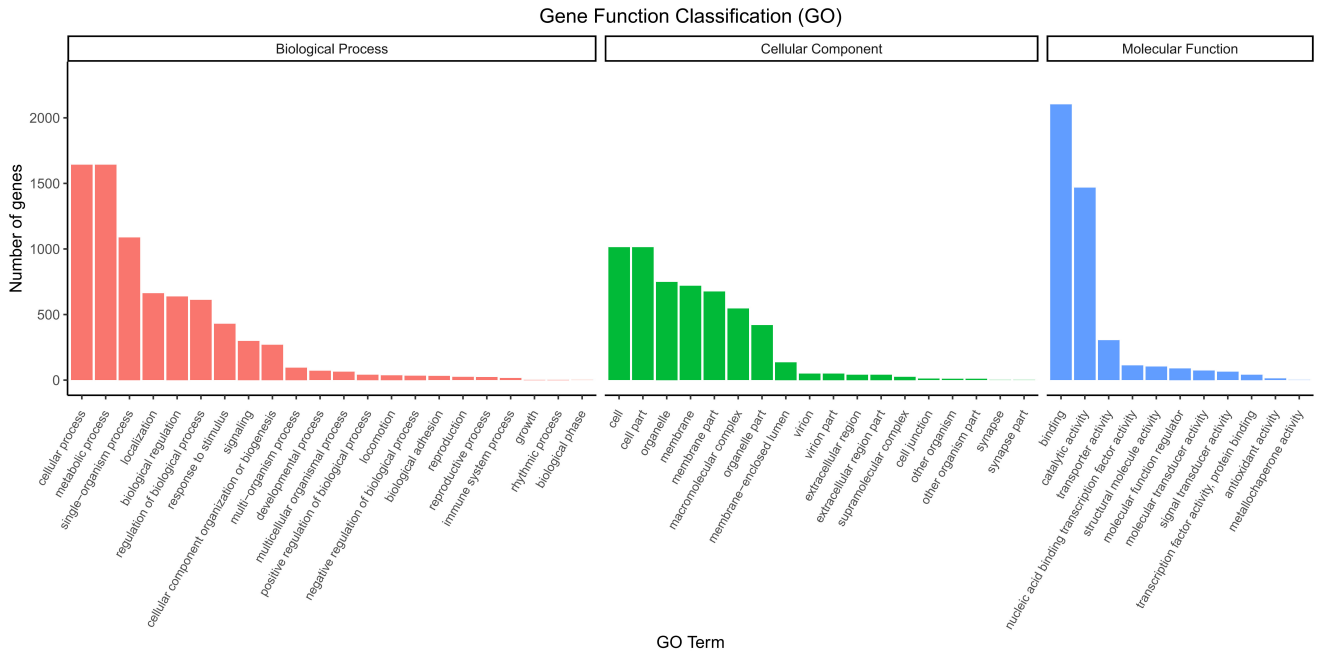


FIGURE 2 | Gene Ontology (GO) functional annotation of *O. cynotis* genes.

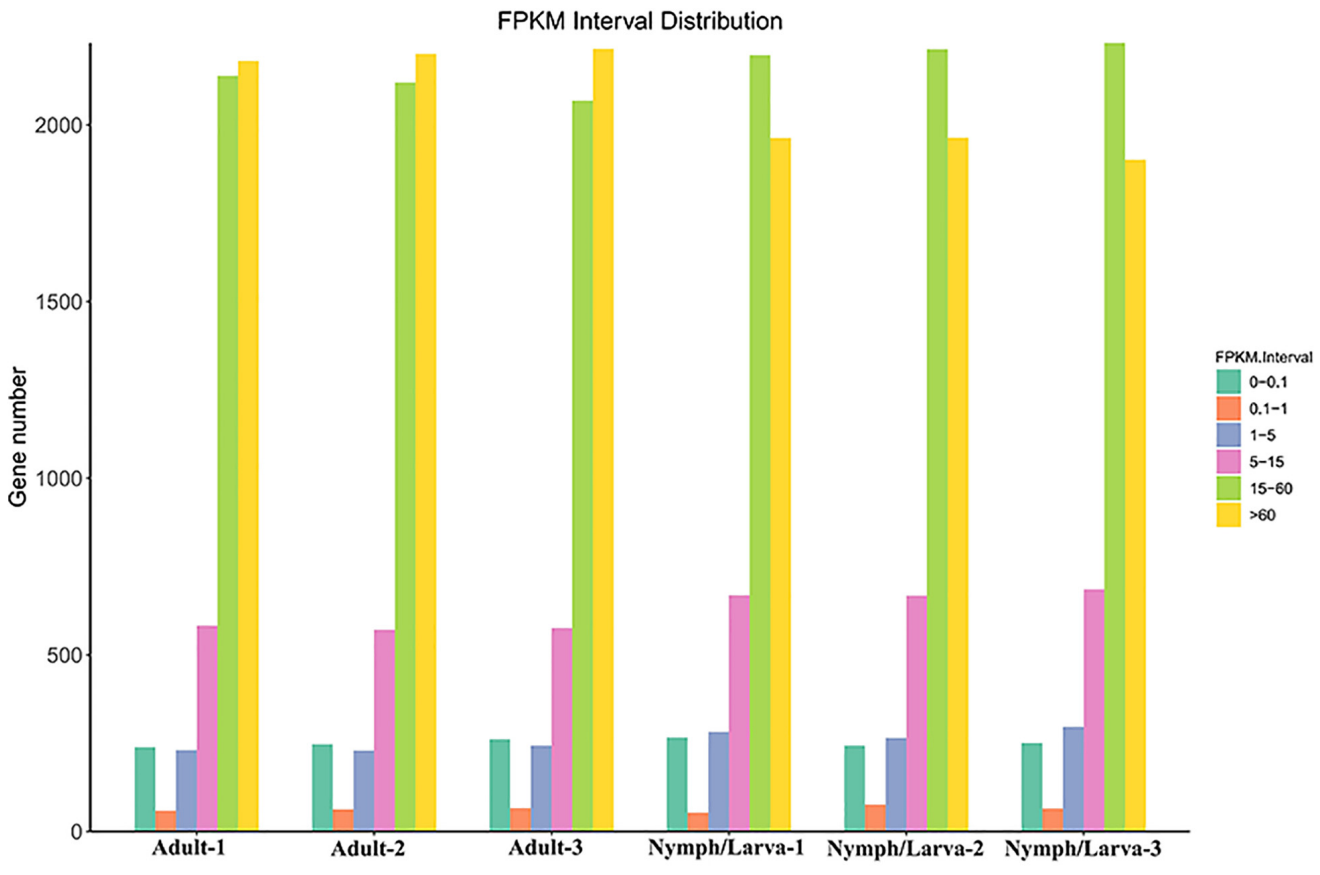


FIGURE 3 | The gene numbers for different Fragments Per Kilo-base of transcript per Million mapped reads (FPKM) values of adult mites *vs*. nymph/larva mites.

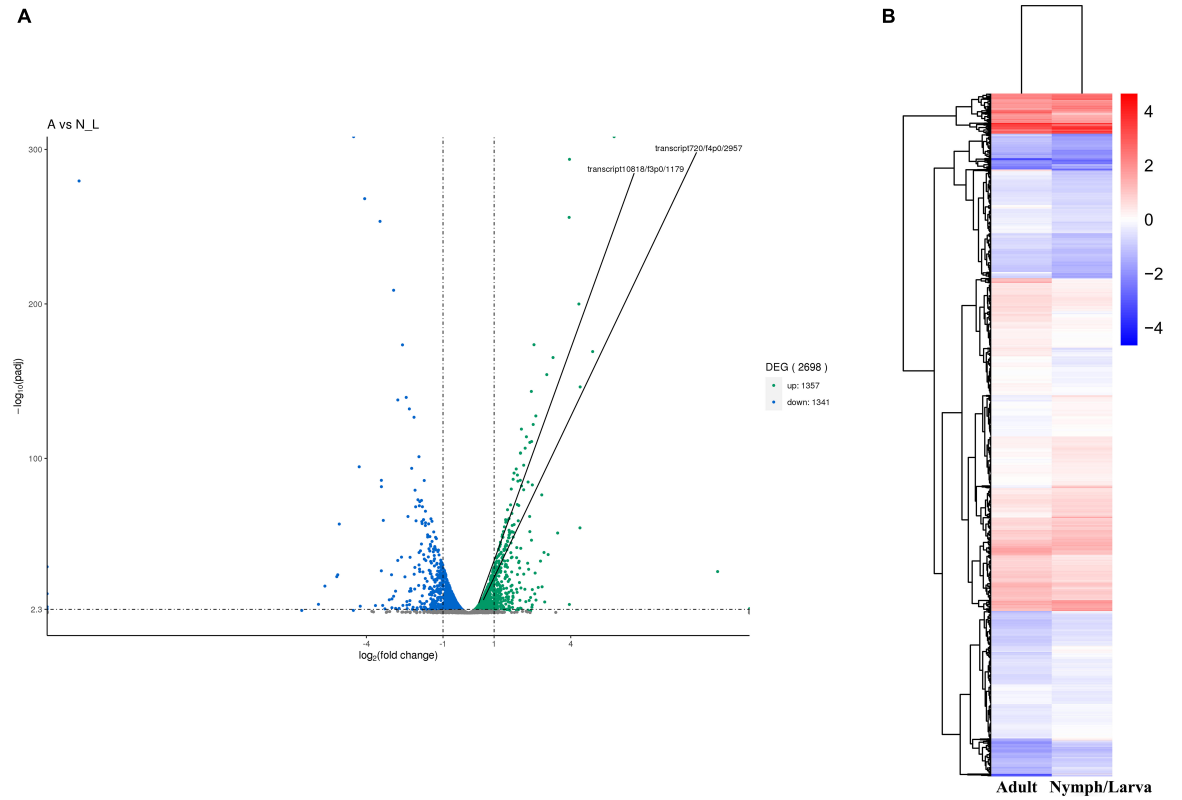


FIGURE 4 | (A) Map of the number of differentially expressed genes. **(B)** Cluster map of DEGs.

the R^2 values of all six samples. The average correlation of gene expression level among biological replicates were 0.9733 and 0.9697 in adult and nymph/larva mites, respectively, which demonstrated good reproducibility of the gene expression results.

Identification of Differentially Expressed Genes and Functional Categorization

We identified 2,698 DEGs, among which 1,357 genes were up-regulated and 1,341 genes were down-regulated (**Figure 4A**). A heat map of all the DEGs was established (**Figure 4B**). A Venn diagram showed that 5,191 and 5,201 genes were expressed in adult mites and nymphs/larvae mites, respectively. Moreover, 23 and 33 genes were expressed only in adult mites and nymphs/larvae, respectively (**Figure 5**).

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis of Differentially Expressed Genes

Gene Ontology analysis of DEGs indicated that 1,471 DEGs were enriched in the biological process category, 384 DEGs were enriched in the cellular process category, and 785 DEGs were enriched in the molecular process category. Intriguingly, in the molecular process category, more than 500 DEGs were associated with the term "catalytic activity" and over 200 DEGs were associated with the term "hydrolase" activity. In the biological process category, DEGs were mainly associated with the terms "single-organism metabolic process" and "proteolysis." The main KEGG pathways associated with the DEGs were "Lysosome," "Phagosome," "Drug metabolism -cytochrome P450" and "Rheumatoid Arthritis" signaling pathways. DEGs-enriched KEGG pathways scatter diagram is shown in **Figure 6**.

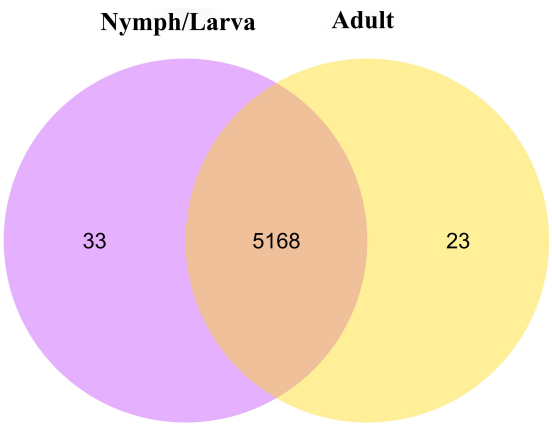


FIGURE 5 | A Venn diagram of differentially expressed genes (the criteria for differential gene screening are |log2(FoldChange)| > 0&qvalue < 0.05).

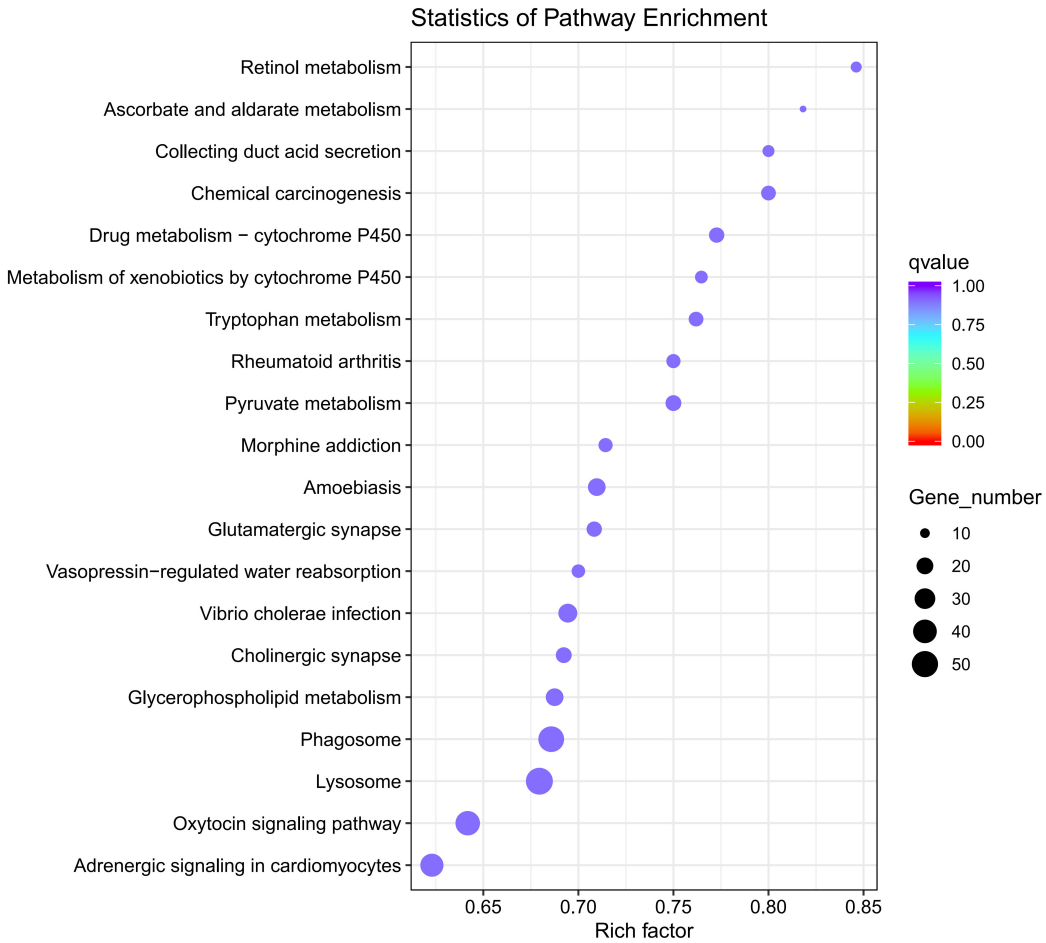


FIGURE 6 | DEGs-enriched Kyoto Encyclopedia of genes and genomes (KEGG) pathways scatter diagram (*q*-value is the *p*-value corrected by multiple hypothesis test, the value range of *q*-value is [0,1]).

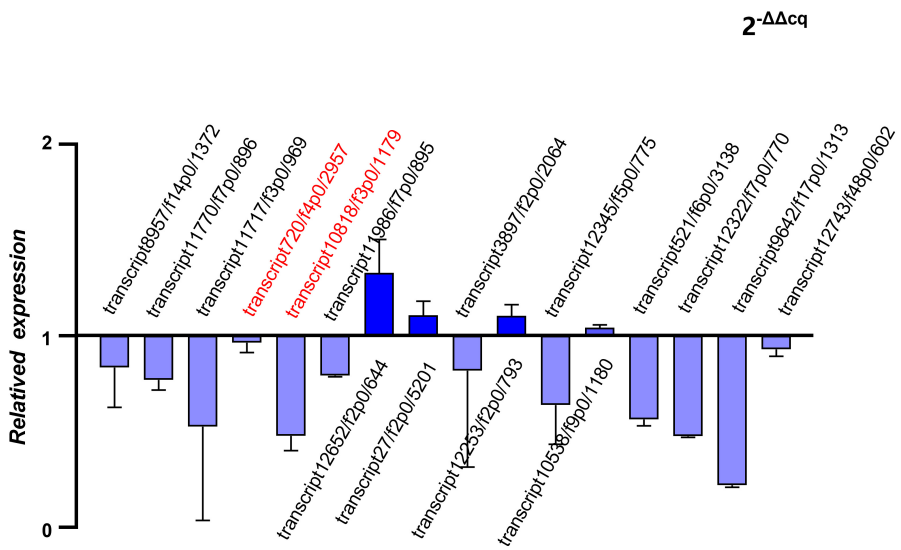


FIGURE 7 | Quantitative real-time reverse transcription PCR (qRT-PCR) of hub genes and putative allergen genes (relative expression of Ocy_N_L groups to Ocy_A groups).

Protein-Protein Interaction Network Construction and Hub Gene Identification

The PPIs among final transcripts from PacBio sequencing and DEGs were predicted using STRING tools. The analysis identified 49 nodes and 147 edges among the final transcripts; and 26 nodes and 47 edges among the DEGs. The results showed that FBpp0079187 and FBpp0085265 were both hub genes among final transcripts and DEGs. The hub genes are listed in **Supplementary Table 2**.

Analysis of Allergen Genes

According to the results from the putative allergen analysis, 397 putative allergen genes were obtained from the final transcripts, among which 231 were DEGs. Seventy-seven sequences were homologs of known mite allergens from *Dermatophagoides farinae*, *Dermatophagoides pteronyssinus*, *Tyrophagus putrescentiae*, *Euroglyphus maynei*, and *Lepidoglyphus destructor*.

Quantitative Real-Time Reverse Transcription PCR Validation of Differentially Expressed Genes

Five hub genes and 11 putative allergen genes were selected to validate their expression levels using qRT-PCR. Furthermore, a comparison between the expression value of qRT-PCR and the FPKM was conducted. Two of these hub genes were all outstanding genes with a high degree of connectivity in the PPI network among the final transcripts and DEGs. Selected putative allergen genes were from those with the top 30 FPKMs in both of adult mites and nymphs/larvae. The relative expression levels of these sixteen genes are presented in **Figure 7**. The results of the analysis indicated that the expression levels of selected genes determined using RNA-Seq were reliable.

DISCUSSION

In this study, a combination of SMRT and Illumina sequencing was applied to identify DEGs between the different development stages of *O. cynotis*. Illumina sequencing is limited to the assembly of full-length transcripts, whereas SMRT-Seq overcomes these limitations by mostly generating the full-length transcript as one read (Rhoads and Au, 2015). Bioinformatic tools, such as LoRDEC and CD-HIT could rectify the errors of SMRT reads (Li et al., 2017; Xiao et al., 2017). After SMRT sequencing and processing, the maximum length of final transcripts was 8,193 nt and the N50 of these transcripts were 2,059 nt, with an average of 1,774 nt. Functional annotation analysis showed that the transcripts had the highest number of hits to *S. scabiei* in NR BLASTx, probably because *S. scabiei* and *O. cynotis* are mites that both belong to Arthropoda, Arachida, Arcarina. However, *S. scabiei* belongs to the Sarcoptidae and *O. cynotis* belongs to the Psoroptidae.

LncRNAs are key molecules that regulate gene expression and have become a hot topic in biology (Luo et al., 2017; Jia et al., 2018). To accurately predict lncRNAs, PLEK and CNCI were used to predict the gene coding potential of the final transcript sequences, identifying 4,831 and 4,624 transcript sequences, respectively. Then, 707 and 1,600 transcript sequences were obtained from CPC and Pfam database analyses. In the Pfam database, Pfam-A database records the high-quality domains of most known proteins and the Pfam-B database covers the domain families more comprehensively, which made the gene coding potential prediction more accurate. Ultimately, 406 lncRNA sequences were predicted. **Supplementary Figure 5** shows that the average length of the mRNAs was 8,193 nt, whereas that of the lncRNAs was 3,377 nt. The functions of these putative lncRNAs require further study.

DEGs analysis identified 23 genes that were unique to adult mites and 33 genes that were unique to nymphs/larvae. Among the genes unique to adults, functional annotation predicted that three of them encoded autophagy protein 6, colincin immunity protein, and acetylcholinesterase. Among the genes unique to nymphs/larvae, functional annotation predicted that three of them encoded protein kinase domain, cystathionine beta-synthase-like, and major facilitator like protein 5. However, some of these unique genes had no annotation in the seven databases. These genes might have crucial functions in the development, reproduction, and pathogenic mechanism of the different life stages of *O. cynotis* mites.

Differential gene expression analysis identified 2,698 DEGs (1,357 up-regulated and 1,341 down-regulated) between adult mites and nymphs/larvae. Gene Ontology analysis of the DEGs indicated that genes involved in catalytic activity, hydrolase activity and proteolysis might play a significant role in the life stages of mites. Proteases are the main allergens of the house dust mite, and the hydrolysis of these proteases is reported to promote and aggravate the allergic reaction and inflammatory responses of the host (Stewart and Robinson, 2003). In *Psoroptes ovis* (*P. ovis*) infections, proteases can degrade fibrinogen to provide a flow of serous exudate from the host (Kenyon and Knox, 2002). For *S. scabiei* infections, aspartic protease can digest serum molecules and skin from the host (Mahmood et al., 2013). Equally, hydrolases might also help to mite survival and invasion. The enrichment of the DEGs in proteolysis and hydrolase activity could explain how the mites in different life stages have different abilities to invade host skin, i.e., the hydrolases and proteases might participate in the invasion of *O. cynotis* mites.

In the KEGG pathway analysis of DEGs, the up-regulated genes were enriched in 243 pathways and the down-regulated genes were enriched in 231 pathways. The KEGG pathways are illustrated in **Figure 6**, including both up-regulated and down-regulated genes. These pathways were "Lysosome," "Phagosome," and "Drug metabolism Cytochrome P450" signaling pathways; however, the top pathway enriched for the down-regulated genes was the oxytocin signaling pathway (30 DEGs). The hub genes selected in the PPI network might have pivotal roles in the development of *O. cynotis* mites. FBpp0079187 and FBpp0085265 were hub genes defined in both the final transcripts and DEGs. FBpp0079187 was annotated as receptor of activated protein C kinase 1-like (RACK1-like) in the NR database. RACK1 is a candidate interacting protein for establishing infection of

Bombyx mori cypovirus (Zhang et al., 2017). RACK1 participates in physiological process, such as development, system function, and circadian rhythm. Moreover, it plays a significant role in shuttling proteins around the cell, anchoring proteins at particular locations, and stabilizing protein activity (McCahill et al., 2002; Adams et al., 2011). FBpp0085265 was annotated as elongation factor 2-like (eEF2-like) in the NR database. eEF2 has an essential role in protein synthesis, in which it catalyzes the translocation of tRNA and mRNA (Carlberg et al., 1990; Kaul et al., 2011). These two hub genes were the top two DEGs among the hub genes, which suggested that the main difference between adult mites and nymph/larva mites was in the life processes of development system function and circadian rhythm.

O. cynotis mites are the most common etiological agent of otitis externa in cats and dogs, which feed on lymph, tissue fluid, and blood of the host. When mites invade a host, dermatitis or allergic reactions occur. Allergic reactions can be protective, but can also be harmful to the host, leading to tissue damage and immunological hypersensitivity. The ability of an antigen to induce allergic sensitization is called allergenicity (Thepen et al., 1992; Hales et al., 2006; Chapman et al., 2007; Tsolakis et al., 2018). Allergens of mites could promote inflammation and allergic reactions, including rhinitis, asthma, and dermatitis (Aalberse, 2000; Arlian, 2002; Stewart and Robinson, 2003). In the present study, 231 transcript sequences were predicted to be homologous genes of allergens, and 77 of them were homologous genes of known mite allergens. These genes might be the main participants in the pathogenesis of *O. cynotis* mites. The role played by allergens in house dust mite infections has made them a hot topic in mite research. House dust mite allergens were identified to increase the allergic immune response of the host by cleaving IgE receptors and boosting the innate immune response (Gough et al., 1902; Trompette et al., 2008; Mizuma et al., 2016). Twenty-one house dust mite allergens have been widely studied, including their structure, chemical and biological properties (Thomas et al., 2010; Fischer and Walton, 2014). Several other allergens were predicted in the genome or transcriptome of *S. scabiei* and *P. ovis*, such as triosephosphate isomerase, chitinase-like protein, *Pso* 1, *Pso* 2, *Pso* 10, and *Pso* 11 (Temeyer et al., 2002; Nisbet et al., 2006, 2007; He et al., 2017, 2018; Korhonen et al., 2020). Moreover, a series of sequences homologous to known mite allergens were predicted in the transcriptome of *Chorioptes texanus* (He et al., 2019). These findings highlight that mite allergens might also play a crucial role in the pathogenesis of chorioptic mange. In the present study, the allergens selected for qRT-PCR verification were homologous to mite allergens, such as Der p 2, Ferritin, Fatty acid-binding protein, Tropomyosin, High molecular weight allergen M-177, Der f 30 allergen, Group 14 allergen protein, HDM allergen, Der f 13 allergen, Eur m 1, and Der f 39. In total, we predicted 77 genes in *O. cynotis* mites that were homologous to known allergen sequences; however, the functions of these sequences require a further investigation. Research concerning allergens from the house dust mite, *S. scabiei*, *P. ovis* and other mites will provide the basis for further research into *O. cynotis* allergens.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI repository, accession number PRJNA713931.

ETHICS STATEMENT

This study was approved by the Animal Care and Use Committee of Sichuan Agricultural University (SYXK2019-187).

AUTHOR CONTRIBUTIONS

GY and RH conceived and designed the experiments. RH and QZ performed the experiments. RH and GY analyzed the data. RH wrote the manuscript. GY, RH, QZ, XG, YX, JX, and XP prepared the samples. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Chengdu Research of Giant Panda Breeding Research Fund (Project No. CPF2014-17) and the Sichuan Science and Technology Program (2020YJ0114). The funder had no role in study design, sample collection, data analysis, preparation or publication of the manuscript.

ACKNOWLEDGMENTS

We thank Song Liu, Yuejun Luo, Lang Xiong, Jie Xiao, Ke Guan, Ruiqi Hua, Yanting Xu, Ziyi Xu, and Lei Shi for their technical help and suggestions on this work. We would be grateful to thank the native English-speaking scientists of Elixigen Company (Huntington Beach, California) for editing our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.687387/full#supplementary-material

Supplementary Figure 1 | A Venn of gene functional annotation.

Supplementary Figure 2 | Gene function annotation using Kyoto Encyclopedia of genes and genomes (KEGG) pathway classification.

Supplementary Figure 3 | Homologous species distribution of *O. cynotis* annotated in the non-redundant (NR) database.

Supplementary Figure 4 | Venn diagram of the number of long non-coding RNAs (LncRNAs) predicted by CNCI, Pfam, Plek, and CPC.

Supplementary Figure 5 | Long non-coding RNA (LncRNA) and mRNA length distribution map.

REFERENCES

Aalberse, R. C. (2000). Structural biology of allergens. *J. Allergy Clin. Immunol.* 106, 228–238.

Adams, D. R., Ron, D., and Kiely, P. A. (2011). RACK1, A multifaceted scaffolding protein: structure and function. *Cell Commun. Signal.* 9:22. doi: 10.1186/1478-811X-9-22

Ahn, A. J., Oh, D. S., Ahn, K. S., and Shin, S. S. (2013). First feline case of otodectosis in the Republic of Korea and successful treatment with imidacloprid/moxidectin topical solution. *Korean J. Parasitol.* 51:125. doi: 10.3347/kjp.2013.51.1.125

Arlian, L. G. (2002). Arthropod allergens and human health. *Annu. Rev. Entomol.* 47, 395–433. doi: 10.1146/annurev.ento.47.091201.145224

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45–48. doi: 10.1093/nar/28.1.45

Bosco, A., Leone, F., Vascone, R., Pennacchio, S., Ciuca, L., Cringoli, G., et al. (2019). Efficacy of fluralaner spot-on solution for the treatment of *Ctenocephalides felis* and *Otodectes cynotis* mixed infestation in naturally infested cats. *BMC Vet. Res.* 15:28. doi: 10.1186/s12917-019-1775-2

Carithers, D., Crawford, J., De, V. C., Lotriet, A., and Fourie, J. (2016). Assessment of afoxolaner efficacy against *Otodectes cynotis* infestations of dogs. *Parasit. Vectors* 9:635. doi: 10.1186/s13071-016-1924-4

Carlberg, U., Nilsson, A., and Nygard, O. (1990). Functional properties of phosphorylated elongation factor 2. *Eur. J. Biochem.* 191, 639–645. doi: 10.1111/j.1432-1033.1990.tb19169.x

Chapman, M. D., Pomes, A., Breiteneder, H., and Ferreira, F. (2007). Nomenclature and structural biology of allergens. *J. Allergy Clin. Immunol.* 119:414. doi: 10.1016/j.jaci.2006.11.001

Chruszcz, M., Kapingidza, A. B., Dolamore, C., and Kowal, K. (2018). A robust method for the estimation and visualization of IgE cross-reactivity likelihood between allergens belonging to the same protein family. *PLoS One* 13:e0208276. doi: 10.1371/journal.pone.0208276

Briceño, C., Daniel, G., Jaime, E. J., Bornscheuer, M. L., Funk, S. M., and Knapp, L. A. (2020). Ear mites, *Otodectes cynotis*, on wild foxes (pseudalopex spp.) In Chile. *J. Wildl. Dis.* 56, 105–112. doi: 10.7589/2018-10-247

Curtis, C. F. (2004). Current trends in the treatment of *Sarcoptes*, *Cheyletiella* and *Otodectes* mite infestations in dogs and cats. *Vet. Dermatol.* 15, 108–114.

Fanelli, A., Domenech, G., Alongso, F., Martinez, C. F., Tizzani, P., Martinez, C. C., et al. (2020). *Otodectes cynotis* in urban and peri-urban semi-arid areas: a widespread parasite in the cat population. *J. Parasit. Dis.* 44, 481–485. doi: 10.1007/s12639-020-01215-7

Finn, R. D., Coggili, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Fischer, K., and Walton, S. (2014). Parasitic mites of medical and veterinary importance–is there a common research agenda?. *Int. J. Parasitol.* 44, 955–967. doi: 10.1016/j.ijpara.2014.08.003

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gough, L., Schulz, O., Sewell, H. F., and Shakib, F. (1902). The cysteine protease activity of the major dust mite allergen Der p 1 selectively enhances the immunoglobulin E antibody response. *J. Exp. Med.* 190, 1897–1902. doi: 10.1084/jem.190.12.1897

Hales, B. J., Martin, A. C., Pearce, L. J., Laing, I. A., Hayden, C. M., Goldblatt, J., et al. (2006). IgE and IgG anti-house dust mite specificities in allergic disease. *J. Clin. Immunol.* 118, 361–367. doi: 10.1016/j.jaci.2006.04.001

He, R., Gu, X. B., Xie, Y., Peng, X. R., Angel, C., and Yang, G. Y. (2019). Transcriptome-based analysis of putative allergens of *Chorioptes texanus*. *Parasit. Vectors* 12:587. doi: 10.1186/s13071-019-3843-7

He, R., Shen, N., Zhang, H., Ren, Y., He, M., Xu, J., et al. (2017). Molecular characteristics and serodiagnostic potential of chitinase-like protein from *Sarcoptes scabiei*. *Oncotarget* 8:83995. doi: 10.18632/oncotarget.21056

He, R., Zhang, H., Shen, N., Guo, C., Ren, Y., Xie, Y., et al. (2018). Molecular characterization and allergenicity potential of triosephosphate isomerase from *Sarcoptes scabiei*. *Vet. Parasitol.* 257, 40–47. doi: 10.1016/j.vetpar.2018.05.016

Hu, L., Zhao, Y. E., Niu, D. L., Yang, R., and Zeng, J. H. (2019). The construction of full-length cDNA library for *Otodectes cynotis*. *Acta Parasitol.* 64, 251–256. doi: 10.2478/s11686-019-00034-y

Huang, B. M., Bassini, S. R., Rolim, L. S., Oconnor, B., Ochoa, R., Barros, B., et al. (2020). *Otodectes cynotis* (Sarcoptiformes: Psoroptidae): new Records on Wild Carnivores in Brazil with a Case Report. *J. Med. Entomol.* 57, 1090–1095. doi: 10.1093/jme/tjaa042

Jia, D., Wang, Y., Liu, Y., Hu, J., Guo, Y., Gao, L., et al. (2018). SMRT sequencing of full-length transcriptome of flea beetle *Agasicles hygrophila* (Selman and Vogt). *Sci. Rep.* 8:2197. doi: 10.1038/s41598-018-20181-y

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi: 10.1093/nar/gkh063

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Kaul, G., Pattan, G., and Rafeequi, T. (2011). Eukaryotic elongation factor-2 (eEF2): its regulation and peptide chain elongation. *Cell Biochem. Funct.* 29, 227–234. doi: 10.1002/cbf.1740

Kelleher, S. A. (2001). Skin diseases of ferrets. *Vet. Clin. North Am. Exot. Anim. Pract.* 4, 565–572. doi: 10.1016/s1094-9194(17)30043-9

Kenyon, F., and Knox, D. (2002). The proteinases of *Psoroptes ovis*, the sheep scab mite—their diversity and substrate specificity. *Vet. Parasitol.* 105, 317–325. doi: 10.1016/s0304-4017(02)00015-8

Korhonen, P. K., Gasser, R. B., Ma, G., Wang, T., Stroehlein, A. J., Young, N. D., et al. (2020). High-quality nuclear genome for *Sarcoptes scabiei*-A critical resource for a neglected parasite. *PLoS Negl. Trop. Dis.* 14:e0008720. doi: 10.1371/journal.pntd.0008720

Li, A., Zhang, J., and Zhou, Z. (2014). PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* 15:311. doi: 10.1186/1471-2105-15-311

Li, W., Jaroszewski, L., and Godzik, A. (2002). Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* 18, 77–82. doi: 10.1093/bioinformatics/18.1.77

Li, Y., Dai, C., Hu, C., Liu, Z., and Kang, C. (2017). Global identification of alternative splicing *via* comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. *Plant J.* 90, 164–176. doi: 10.1111/tpj.13462

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Luo, Y., Ding, N., Shi, X., Wu, Y., Wang, R., Pei, L., et al. (2017). Generation and comparative analysis of full-length transcriptomes in sweetpotato and its putative wild ancestor I. trifida. *BioRxiv* [Preprint]. doi: 10.1101/112425

Mahmood, W., Viberg, L. T., Fischer, K., Walton, S. F., and Holt, D. C. (2013). An aspartic protease of the scabies mite *Sarcoptes scabiei* is involved in the digestion of host skin and blood macromolecules. *PLoS Negl. Trop. Dis.* 7:e2525. doi: 10.1371/journal.pntd.0002525

McCahill, A., Warwicker, J., Bolger, G. B., Houslay, M. D., and Yarwood, S. J. (2002). The RACK1 scaffold protein: a dynamic cog in cell response mechanisms. *Mol. Pharmacol.* 62, 1261–1273. doi: 10.1124/mol.62.6.1261

Mizuma, H., Tanaka, A., Uchida, Y., Fujiwara, A., Manabe, R., Furukawa, H., et al. (2016). Influence of Omalizumab on Allergen-Specific IgE in Patients with Adult Asthma. *Int. Arch. Allergy Immunol.* 168:165. doi: 10.1159/000442668

Montoya, J. (2018). Otitis and miliary dermatitis in a cat infected with *Otodectes cynotis*. *Argos Inf. Vet.* 202, 18–21.

Nisbet, A., Mackellar, A., Wright, H., Brennan, G., Chua, K., Cheong, N., et al. (2006). Molecular characterization, expression and localization of tropomyosin and paramyosin immunodominant allergens from sheep scab mites (*Psoroptes ovis*). *Parasitology* 133, 515–523. doi: 10.1017/S0031182006000631

Nisbet, A., Mackllar, A., Mclean, K., Brennan, G., and Huntley, J. (2007). Eukaryotic expression of recombinant *Pso o* 1, an allergen from *Psoroptes ovis*, and its localization in the mite. *Parasitology* 134, 83–89. doi: 10.1017/S0031182006001235

Powell, M., Weisbroth, S., Roth, L., and Wilhelmsen, C. (1980). Reaginic hypersensitivity in *Otodectes cynotis* infestation of cats and mode of mite feeding. *Am. J. Vet. Res.* 41, 877–882.

Rhoads, A., and Au, K. F. (2015). PacBio sequencing and its applications. *Genom. Proteom. Bioinform.* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002

Roy, J., Bedard, C., and Moreau, M. (2011). Treatment of feline otitis externa due to *Otodectes cynotis* and complicated by secondary bacterial and fungal infections with Oridermyl auricular ointment. *Can. Vet. J.* 52:277.

Salib, F. A., and Baraka, T. A. (2011). Epidemiology, genetic divergence and acaricides of *Otodectes cynotis* in cats and dogs. *Vet. World* 4:109.

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538

Silva, J. T., Ferreira, L. C., Fernandes, M. M., Sousa, L. N., Feitosa, T. F., Braga, F. R., et al. (2020). Prevalence and clinical aspects of *Otodectes cynotis* infestation in dogs and cats in the Semi-arid region of Paraíba, Brazil. *Acta Sci. Vet.* 48:1.

Six, R. H., Becskei, C., Mazaleski, M. M., Fourie, J. J., Mahabir, S. P., Myers, M. R., et al. (2016). Efficacy of sarolaner, a novel oral isoxazoline, against two common mite infestations in dogs: Demodex spp. and *Otodectes cynotis*. *Vet. Parasitol.* 222, 62–66. doi: 10.1016/j.vetpar.2016.02.027

Stewart, G., and Robinson, C. (2003). The immunobiology of allergenic peptidases. *Clin. Exp. Allergy* 33, 3–6. doi: 10.1046/j.1365-2222.2003.01601.x

Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., and Zhang, C. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166. doi: 10.1093/nar/gkt646

Taenzler, J., De, V. C., Roepke, R. K., Frenais, R., and Heckeroth, A. R. (2017). Efficacy of fluralaner against *Otodectes cynotis* infestations in dogs and cats. *Parasit. Vectors* 10:30. doi: 10.1186/s13071-016-1954-y

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., and Koonie, E. V. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41

Temeyer, K. B., Soileau, L. C., and Pruett, J. H. (2002). Cloning and sequence analysis of a cDNA encoding *Pso o* II, a mite group II allergen of the sheep scab mite (Acari: Psoroptidae). *J. Med. Entomol.* 39, 384–391. doi: 10.1603/0022-2585-39.2.384

Thepen, T., Mcmenamin, C., Girn, B., Kraal, G., and Holt, P. (1992). Regulation of IgE production in pre-sensitized animals: *in vivo* elimination of alveolar macrophages preferentially increases IgE responses to inhaled allergen. *Clin. Exp. Allergy* 22, 1107–1114. doi: 10.1111/j.1365-2222.1992.tb00137.x

Thomas, W. R., Hales, B. J., and Smith, W. A. (2010). House dust mite allergens in asthma and allergy. *Trends Mol. Med.* 16, 321–328.

Tonn, R. J. (1961). Studies on the ear mite *Otodectes cynotis*, including life cycle. *Ann. Entomol. Soc. Am.* 54, 416–421.

Trompette, A., Divanovic, S., Visntin, A., Blanchard, C., Hegde, R. S., and Madan, R. (2008). Allergenicity resulting from functional mimicry of a Toll-like receptor complex protein. *Nature* 457, 585–588.

Tsolakis, N., Malinovschi, A., Nordvall, L., Mattsson, L., Lidholm, J., Pedroletti, C., et al. (2018). Sensitisation to minor cat allergen components is associated with type-2 biomarkers in young asthmatics. *Clin. Exp. Allergy* 48, 1186–1194. doi: 10.1111/cea.13135

Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., et al. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14:1072. doi: 10.1038/nmeth.4432

Yang, C., and Huang, H. P. (2016). Evidence-based veterinary dermatology: a review of published studies of treatments for *Otodectes cynotis* (ear mite) infestation in cats. *Vet. Dermatol.* 27, 221–e56. doi: 10.1111/vde.12340

Zhang, Y., Cao, G., Zhu, L., Chen, F., Zar, M. S., Wang, S., et al. (2017). Integrin beta and receptor for activated protein kinase C are involved in the cell entry of *Bombyx mori cypovirus*. *Appl. Microbiol. Biotechnol.* 101, 3703–3716. doi: 10.1007/s00253-017-8158-z

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership