IJID One Health 4 (2024) 100031

Contents lists available at ScienceDirect

IJID One Health

journal homepage: www.sciencedirect.com/journal/ijid-one-health

Pathogen genomics and One Health: A scoping review of current practices in zoonotic disease research

Stefano Catalano[a,*], Francesca Battelli[b], Zoumana I. Traore[b], Jayna Raghwani[c], Christina L. Faust[a,1], Claire J. Standley[b,d,1]

[a] School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Glasgow, UK
[b] Center for Global Health Science and Security, Georgetown University, Washington, USA
[c] Department of Pathobiology and Population Sciences, Royal Veterinary College, University of London, London, UK
[d] Heidelberg Institute of Global Health, Heidelberg University, Heidelberg, Germany

ARTICLE INFO

Keywords:
Global health
Infectious disease surveillance
Next-generation sequencing
Phylogenetics
Zoonoses

ABSTRACT

Objectives: Whole-genome sequencing has revolutionized the field of infectious disease surveillance, enabling near real-time detection of pathogens and tracking how infections may spread. Our study aimed to characterize genomic applications to cross-domain zoonotic pathogen transmission at the human-animal and/or human-environment interfaces.

Methods: We performed a scoping review of studies that have applied genomic epidemiology to zoonotic disease transmission across One Health domains (human, animal, and environment). We identified 114 records published between 2005 and 2022 which reported multi-domain genomic data of zoonotic pathogens integrated into phylogenetic models.

Results: Most studies investigated bacterial pathogens, highlighting key knowledge gaps for other zoonotic agents, particularly arboviruses. Sampling and sequencing efforts varied greatly across domains: the median number and range of pathogen genomes analyzed were highest for humans (23; 1–29,586) and lowest for the environment domain (13; 1–956). Genomics was used to track zoonotic disease outbreaks and cross-domain transmission, to improve pathogen surveillance, and to disentangle evolutionary dynamics driving lineage diversification and virulence.

Conclusions: Our study highlights current practices and knowledge gaps to guide future study designs and genomic applications to multi-domain and cross-species transmission of zoonoses, with the potential to identify key infection sources and inform interventions for local and global health security.

Introduction

Zoonoses are infectious diseases that can be transmitted from animals to humans [1]. Public health estimates suggest that zoonoses are directly responsible for 2.5 billion infections and 2.7 million deaths each year, while approximately 60% of emerging infections are caused by zoonotic pathogens [2,3]. Managing and preventing zoonotic disease outbreaks requires interdisciplinary approaches and expertise from a variety of fields [1]. A One Health approach encourages close collaboration among different disciplines and sectors to recognize that human and animal health are interdependent and intricately connected to the health of the environment, or ecosystem, in which they coexist. The checklist for One Health epidemiological reporting of evidence (COHERE) was developed to guide the integration of knowledge across these three domains (i.e., human, animal, and environment) when designing and implementing interventions [4]. However, investigating multiple domains within a One Health context remains complex and requires overcoming many hurdles, including diagnostic capacity and supply chain limitations, policy and funding support, and meaningful equal participation from a wide variety of stakeholders [5].

At the beginning of the 21st century, the advent of the first high-throughput sequencing platforms ushered in the next-generation sequencing era, driving rapid and financially accessible sequencing of entire genomes [6]. The application of genomics to the transmission

* Corresponding author.
E-mail address: ste.cata85@gmail.com (S. Catalano).
[1] Christina L. Faust and Claire J. Standley contributed equally to this work as senior authors.

https://doi.org/10.1016/j.ijidoh.2024.100031
Received 1 March 2024; Received in revised form 10 June 2024; Accepted 14 June 2024
Available online 21 June 2024

dynamics of infectious diseases has enabled estimates of fine-scale epidemiological processes over relatively short timescales. The West African Ebola epidemic and the COVID-19 pandemic served as powerful reminders of the impact of zoonoses on human populations and underscored the importance of real-time surveillance to elucidate transmission pathways [7]. Therefore, the application of genomic tools to One Health research offers the exciting prospect of reconstructing transmission events via genomic epidemiology and phylodynamics. When combined with qualitative and quantitative epidemiological data, the analysis of pathogen genomes can help uncover outbreak origins, transmission routes, and/or potential super-spreading events [8,9]. Furthermore, mitigation of zoonotic diseases cannot bypass the identification of competent disease reservoirs, which benefits from progress in sampling, diagnostics, sequencing, and modeling techniques [10].

Leveraging genomic data and modeling approaches within fieldwork settings and across different domains remains logistically and analytically complex. As a result, the degree to which these approaches have been used to investigate the transmission of zoonotic diseases remains poorly understood. In this scoping review, we scanned the published literature and extracted data to the finest possible methodological, spatiotemporal, and phylogenetic level of detail to characterize sampling strategies, genomic approaches, and evolutionary models applied to zoonoses within One Health initiatives. Our objectives were to identify studies investigating cross-domain zoonotic pathogen transmission at the human-animal and/or human-environment interfaces using genomic epidemiology. Our study aimed to characterize and categorize genomic applications in One Health research and, in doing so, to highlight current practices and knowledge gaps to inform future studies.

Materials and methods

Search strategy

We followed published guidance on conducting and reporting evidence synthesis [11], including the Preferred Reporting Items for Systematic Reviews (PROSPERO) and Meta-Analyses extension for Scoping Reviews (Supplementary Table 1 adapted from [12]). First, we searched PROSPERO database to determine whether our research questions had not been already addressed by a registered review. After the identification and refinement of the terminology applied to our search string (Supplementary Table 2), we queried the following search engines based on their large, multidisciplinary spectrum and their classification as principal resources [13]: PubMed and Web of Science (Web of Science Core Collections selected within the platform). Intentionally, we did not use the Medical Subject Headings (MeSH) database in PubMed to include any non-indexed articles and to restrict our query to the exact search string. The final search string was: (ecolog* OR evolution* OR epidemiolog*) AND (“transmission” OR “surveillance”) AND (zoono* OR “disease” OR infect*) AND (“molecular” OR genetic* OR genom* OR metagenom*) AND (phylogen* OR phylodynamic* OR phylogeograph*) AND (“reads” OR librar* OR align* OR polymorph* OR “next generation”) NOT (Sanger OR microsatellite*).

Record screening

The search was completed within one day on September 27, 2022. Records were exported to EndNote X9.3.3 (Clarivate, Philadelphia, USA) and combined into a single library. Screening of articles was performed using a three-stage process. In the first stage, the library was de-duplicated using EndNote, followed by a visual check of the record list sorted by digital object identifier. In the second stage, two reviewers independently screened titles and abstracts in Rayyan [14]; 100 randomly selected records were initially screened to ensure an agreement rate of at least 80% between reviewers before proceeding with title/abstract review of all records. In the third stage, two reviewers independently screened the full text of each retained article; 10 randomly selected records were first screened to ensure an agreement rate of at least 80% between reviewers before proceeding with full-text review of included records. At each stage, the reviewers followed a decision tree, which was defined by the inclusion and exclusion criteria listed below (Supplementary Figure 1). The resolution of any conflicting classifications was addressed by a discussion between reviewers; if needed, the full paper was retrieved and re-screened to resolve the disagreement.

Records that complied with the following inclusion criteria were retained: (i) the infectious agent(s) is classified as zoonotic or deemed a potentially emerging zoonotic disease by the publication and/or co-authors of this scoping review (Supplementary Table 3 adapted from [15]); (ii) the record includes sampling activities (or handling of samples for nucleic acid extraction, library preparation, and sequencing) of human hosts in addition to the animal and/or environment One Health domains (in other words, the record includes genomic data produced directly by the study from the human domain in addition to the animal and/or environment One Health domains) (Supplementary Table 4 adapted from [16]); (iii) genomic data are integrated into evolutionary models of transmission dynamics; and (iv) articles’ publication date goes from January 01, 2005 to present (this criterion was based on the commercial release of the first high-throughput sequencing platforms [6]).

The following studies were excluded from our scoping review: (i) scientific work focusing on SARS-CoV-2; (ii) articles that do not incorporate original genomic data (in other words, we excluded studies that only collated data deposited in publicly accessible databases); (iii) methodologies exclusively based on Sanger sequencing and amplified/restriction fragment length polymorphism; (iv) literature reviews, perspective articles, and commentaries; and (v) gray literature and literature whose full text is not available in English.

Data extraction and analysis

After full-text screening, each retained record was subjected to data extraction to understand its overarching aim, sampling effort, laboratory methodologies, and analytical approach (Table 1). We were also interested in the reproducibility and accessibility of results, and therefore collated data on whether genetic data and open-source code were submitted to public repositories, and if software used in each study was licensed or open-source. Based on the extracted data, geographic localities where sampling was carried out were aggregated based on income status (i.e., low/lower-middle income countries and upper-middle/high-income countries) as reported by the Organisation for Economic Co-operation and Development in 2022 (https://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/daclist.htm accessed on May 10, 2023). Furthermore, the biological agents included in our review were categorized based on hazard group definitions by Health and Safety Executive (https://www.hse.gov.uk/pubns/misc208.htm accessed on January 22, 2024). These categories reflect infectiousness, morbidity, and available vaccines or treatments, which were translated into laboratory containment levels required to work with the listed pathogens.

To understand factors influencing sample size, we constructed a generalized linear model with total sample size of each record (log-transformed with Poisson family links) as the response variable and pathogen type, biocontainment level, sequencing platform, sampling geographic origin, income status stratification, overarching aim of the study, number of surveyed domains, and year of publication as predictors. Data were modeled using quasi-Poisson and negative binomial families and all models were tested for overdispersion. Data were analyzed and visualized in R version 4.3.2 [17].

Table 1
Summary of the data extracted from each record included after full-text screening.

Data categories	Description of the extracted data
Publication	
First author's surname	Not applicable
Scientific journal	ISO4 abbreviation of the scientific journal and publication's year
Epidemiology	
Country or geographic region	Geographic origin of the study and where samples were sequenced
Sampling period	Sampling years spanning from most dated to most recent sample
Infectious agent(s)	Pathogen(s) included in the study
Study aim(s)	Overarching aim(s) of the study as specified by its Introduction's final paragraph and Conclusions sections
One Health domain(s)	Counts of samples from the human, animal, and environment domains that were collected (or analyzed) and sequenced by the study
Animal domain	Subdivided into livestock (i.e., farmed domestic pigs, ruminants, horses, and fish), pet (i.e., domestic dogs, cats, pet birds, pet rodents, and exotic pets), poultry (i.e., farmed chickens, ducks, geese, guinea fowl, and turkeys), and wildlife (i.e., any non-domesticated vertebrates)
Environment domain	Subdivided into abiotic (i.e., water, soil, wastewater, housing/transport/market/slaughterhouse environment, and unspecified environmental samples), Arachnida (i.e., ticks and mites), biotic (i.e., vegetation, animal feed, fecal matter collected from the environment, and unspecified items for human consumptions of non-animal origin), and Insecta (i.e., mosquitoes, flies, fleas, and lice)
Laboratory	
Isolation by cell culture	Pathogen isolation prior to genome sequencing
Detection/Characterization	Pathogen identification by serological, molecular, or other methods prior to genome sequencing
Nucleic acid	Targeted nucleic acid for automated or manual extraction using commercial kits or in-house methods
Genome sequencing	Library preparation kits and genome sequencing platform(s)
Data analysis	
Assembly of reads	Software used for *de novo* assembly or mapping to a reference genome
Evolutionary model	Software used to align consensus sequences, identify recombination events, and construct phylogenetic trees
Phylogenetic output	Subdivided into genetic distance, phylogenetic relatedness, and temporally resolved models
Data repository	
Public archive	Format and public repository for the generated genomic data and bioinformatics code

Results

Record screening

The literature search yielded 2094 and 1637 results for PubMed and Web of Science, respectively, for a total of 3731 results. Automatic de-duplication produced a list of 3030 records and was followed by manual screening, which removed a further 36 records. The list of records subject to title/abstract screening contained 2992 results. Screening of titles and abstracts led to 272 conflicting decisions out of 2992 records (90.9% agreement rate). At this stage, we observed a high number of records focusing on SARS-CoV-2 (n = 164) that we excluded. In agreement with inclusion and exclusion criteria and after conflict resolution, a total of 2723 records were excluded while 269 articles were included for full-text screening. Screening based on full texts led to the exclusion of 155 records (57.6%) while 114 (42.4%) were included for data extraction (Supplementary Figure 2).

Overall, most studies targeted bacterial pathogens (83.3% [n = 95]) while viral and parasitic organisms were less represented (Table 2). The 114 studies included for data extraction covered 36 different families, genera, or species of infectious disease agents (i.e., 23 bacteria, seven ribonucleic acid (RNA) viruses, three protozoa, one fungus, one nematode, and one DNA virus). The distribution of the studies among pathogen taxa was highly skewed with 17 infectious disease agents represented by one single study. Of the 95 studies on bacterial pathogens, almost half (48.4% [n = 46]) focused on one of three species: *Salmonella enterica, Escherichia coli,* or *Staphylococcus aureus*. RNA viruses were represented by 13 studies (11.4%) focusing on seven genera (i.e., *Alphainfluenzavirus*, *Flavivirus*, *Kobuvirus*, *Orthohantavirus*, *Orthonairovirus*, *Phlebovirus*, and *Rotavirus*), although four of these genera were each represented by one single study.

Geographic, temporal, and motivation trends of genomics applied to One Health

Our dataset included 92 records originating from 33 different countries, whereas 22 publications collected and/or analyzed multiple samples from at least three different countries. Most records (74.6% [n = 85]) based sampling activities exclusively in upper-middle/high-income countries, while only 10.5% (n = 12) focused solely on low/lower-middle income countries (sampling and sequencing from both income status groups was included in 13.2% [n = 15] of the studies). The People's Republic of China was the most represented country in studies that focused sampling efforts on a single country (n = 15), followed by the USA (n = 9) and Australia (n = 9).

We observed an increasing trend in the average number of published studies annually between 2011 and 2019, followed by a decline in the last 3 years (2020–2022 but our search stopped in September 2022) (Figure 1). For all included studies, the average time lag between sampling end date and publication year was 3.8 years (median 3; range 0–21). Most records (79.8% [n = 91]) had a time lag of 5 years or less, whereas fewer records (9.6% [n = 11] and 4.4% [n = 5]) had 6–10 years and 11–21 years as time lags, respectively (Supplementary Figure 3).

The application of different sequencing systems over time shows a consistent delay of at least 2 years between the commercialization of the technology and the first scientific publication(s) in One Health studies (Figure 2). The only exception is PacBio RS II, an instrument that started appearing in scientific publications soon after its commercial release. Illumina sequencing platforms represented the most widely used systems, particularly MiSeq and HiSeq (deployed as the primary platform in 33.3% [n = 38] and 29.8% [n = 34] of the studies, respectively). Long-read sequencing with PacBio RS II and/or Oxford Nanopore Technologies MinION was used by 14.0% (n = 16) of the records as either the main platform or as support to short-read sequencing of bacteria (n = 14), *Zika virus* (n = 1), and *Plasmodium* spp. (n = 1).

Multi-domain analysis of infectious disease agents

In our dataset, the human, animal, and environment domains were simultaneously surveyed by 37.7% (n = 43) of the records. Most of these three-domain studies investigated bacterial pathogens (n = 39), particularly *E. coli*, *S. enterica*, and *S. aureus*, while a limited number of three-domain studies focused on RNA viruses (n = 3) or protozoa (n = 1). Almost two-thirds of the studies analyzed and sequenced samples either at the human-animal (38.6% [n = 44]) or human-environment interface (23.7% [n = 27]) (Table 2). Studies focusing on

Table 2
Zoonotic pathogens investigated in the 114 studies included for data extraction subdivided by the surveyed One Health domains (human [H], animal [A], and environment [E]). The total number of records for each taxon is reported in parentheses.

	One Health domains		
Pathogen type	H-A-E	H-A	H-E
Gram-negative bacteria (60)	**28**	**17**	**15**
Salmonella enterica (22)	13	8	1
Escherichia coli (13)	8	3	2
Burkholderia pseudomallei (5)	1	0	4
Campylobacter spp. (3)	0	3	0
Francisella tularensis holartica (3)	1	1	1
Yersinia pestis (3)	2	0	1
Coxiella burnetii (2)	1	1	0
Vibrio parahaemolyticus (2)	0	0	2
Acinetobacter baumannii (1)	0	0	1
Chlamydia psittaci (1)	0	1	0
Enterobacteriaceae family (1)	0	0	1
Klebsiella pneumoniae (1)	0	0	1
Leptospira spp. (1)	1	0	0
Orientia tsutsugamushi (1)	1	0	0
Rickettsia japonica (1)	0	0	1
Gram-positive bacteria (35)	**11**	**17**	**7**
Staphylococcus aureus (11)	2	7	2
Listeria monocytogenes (5)	2	1	2
Mycobacterium spp. (5)	2	3	0
Bacillus anthracis (4)	3	1	0
Enterococcus spp. (3)	1	1	1
Streptococcus spp. (3)	0	3	0
Bacillus cereus (2)	1	0	1
Clostridium spp. (2)	0	1	1
RNA viruses (13)	**3**	**6**	**4**
Flavivirus genus (3)	1	0	2
Influenza A virus (3)	1	1	1
Orthohantavirus genus (3)	0	3	0
CCHF *orthonairovirus* (1)	1	0	0
Kobuvirus genus (1)	0	1	0
Phlebovirus genus (1)	0	0	1
Rotavirus A (1)	0	1	0
Other organisms (7)	**1**	**5**	**1**
Babesia microti (1)	0	0	1
Cowpox virus (1)	0	1	0
Giardia duodenalis (1)	1	0	0
Plasmodium spp. (1)	0	1	0
Sporothrix spp. (1)	0	1	0
Strongyloides stercoralis (1)	0	1	0

zoonotic diseases at the animal-environment interface were excluded during abstract/title and full-text screening.

Sample sizes were highly variable across domains. The median number of samples analyzed and sequenced was highest in humans (23; range 1–29,586) and closely followed by animals (21; range 1–2004). Environmental samples had the smallest median sample size and range (13; range 1–956). Of all publications that included environmental samples (n = 70), 52.9% (n = 37) analyzed and sequenced abiotic samples. In contrast, a focus on vectors was only included in 22.9% (n = 16) of these studies (Figure 3a). Arachnids had the highest median number of samples (14; range 1–175), followed by abiotic sources (12; range 1–559), biotic sources (5.5; range 1–956), and insects (2.5; range 1–18). Among studies that sequenced samples from the animal domain (n = 87), livestock had the highest representation (63.2% [n = 55]), followed by wildlife (36.8% [n = 32]), poultry (34.5% [n = 30]), and pets (24.1% [n = 21]) (Figure 3a). Of these 87 studies, 66.7% (n = 58) included sampling of multiple non-human vertebrate species. The median number of samples was highest for both livestock (13; range 1–760) and poultry (13; range 1–1244), followed by wildlife (7; range 1–116) and pets (3; range 1–18).

We used a generalized linear model to investigate how the total sample size of each record was influenced by different variables (i.e., pathogen type, biocontainment level, sequencing platform, sampling geographic origin, income status stratification, overarching aim of the study, number of domains surveyed, and year of publication). The models were all over-dispersed and, therefore, we could not construct a model that sufficiently described sample size distributions even when using negative binomial families. Only publications focusing on bacterial pathogens had significantly larger sample sizes in all domains (n = 95 studies; median 68.5; range 2–31,292) when compared to viruses (n = 14 studies; median 8; range 6–219), or parasites (i.e., fungi, helminths, and protozoa) (n = 5 studies; median 18; range 2–89) (Figure 3b).

Analytical trends of genomics applied to One Health

We extracted data on genome assembly software/pipelines from each paper and identified 23 unique toolkits used by the 114 records (Supplementary Table 5). SPAdes was the most widely used tool for genome assembly, implemented by 28.9% (n = 33) of the records. SPAdes is free software that supports a wide array of data types and pipelines, which may contribute to making it the most selected tool for both *de novo* assembly and mapping to a reference genome. CLC Genomics Workbench (QIAGEN, Aarhus), a licensed software suite for integrated genomic data analytics, was also frequently used, almost exclusively for *de novo* assembly of short-read data (14.0% [n = 16] of the studies). A variety of software was used to build phylogenetic trees (Supplementary Table 6). Randomized Axelerated Maximum

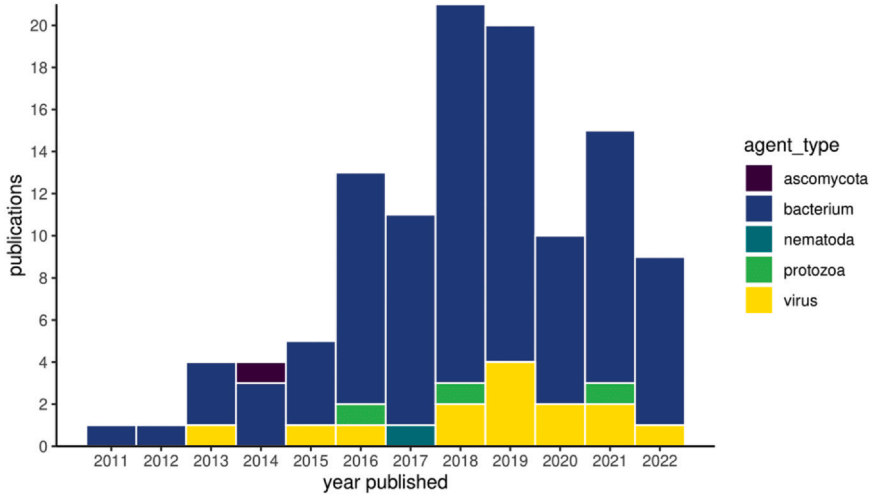

Figure 1. Number of records included for data extraction (n = 114) grouped by year of publication and zoonotic pathogen type.

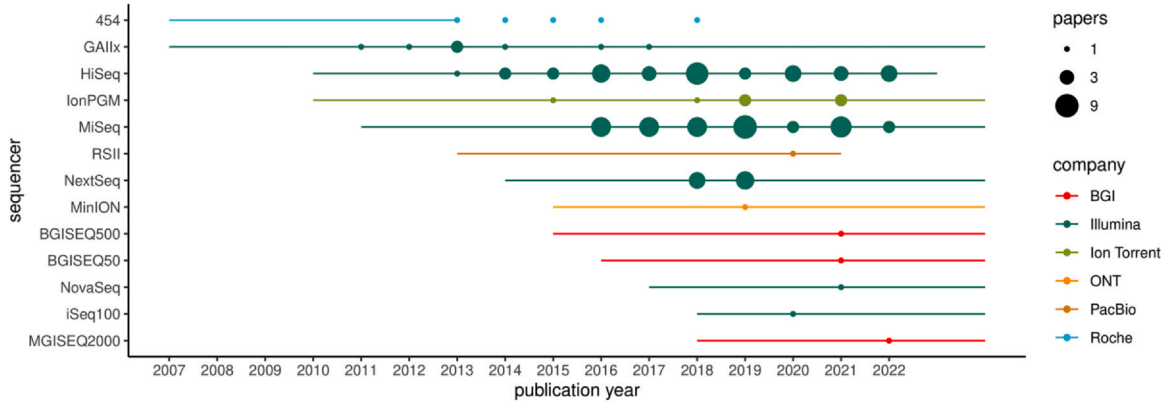

Figure 2. Sequencing platforms and frequency of use in One Health studies over time. For each technology, the line shows the year of commercialization and discontinuation (collated until 2023) when available. The point size reflects the year of publication for each of the 114 records included in this study and the specific technologies used for genomic sequencing. 454, 454 Sequencing; GAIIx, Genome Analyzer IIx; ONT, Oxford Nanopore Technologies.

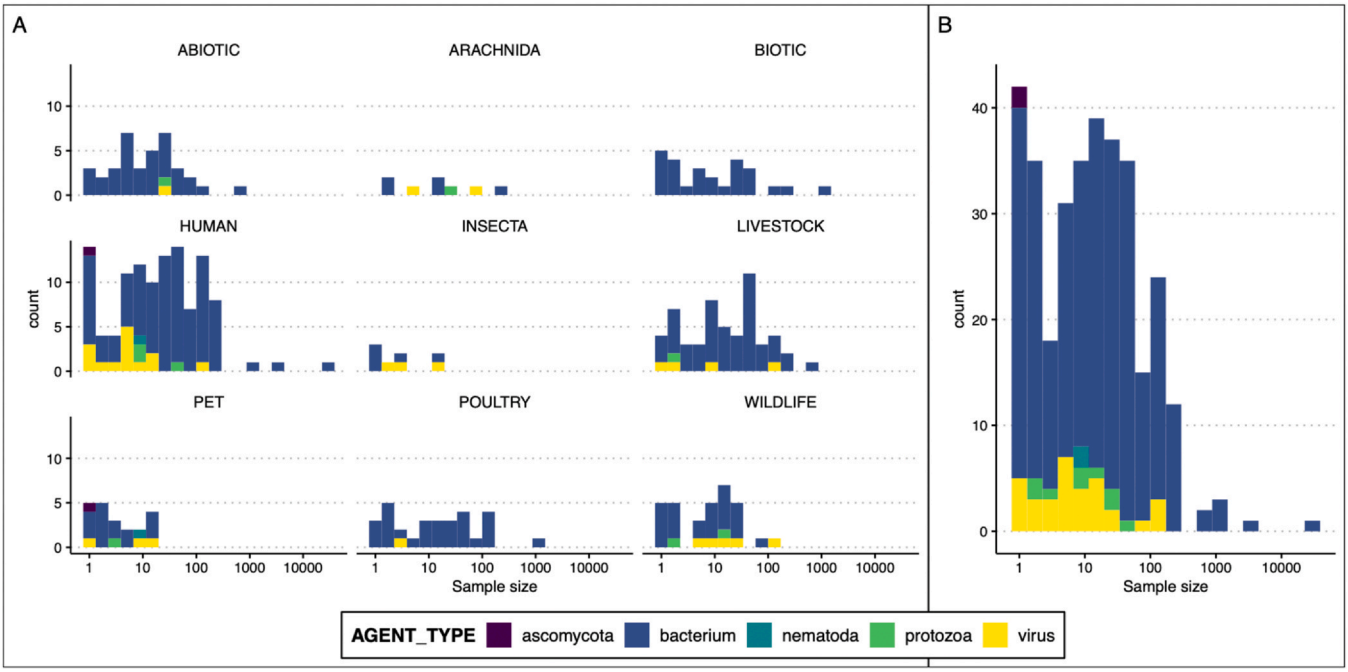

Figure 3. Counts of samples collected (or analyzed) and sequenced by each study shaded by the type of infectious disease agents. The histograms represent the sample size reported by each publication (x-axis) and the number of studies (y-axis) for each One Health domain (i.e., human, animal (subdivided into livestock, pet, poultry, and wildlife), and environment (subdivided into abiotic, Arachnida, biotic, and Insecta)) from which the analyzed samples originated (a). Total sample size for each publication (x-axis) and number of studies (y-axis) (b).

Likelihood, a freely available program for maximum likelihood estimation, was the most common (28.1% [n = 32] of the studies), followed by Bayesian Evolutionary Analysis Sampling Trees (BEAST and BEAST 2) (16.7% [n = 19]) and molecular evolutionary genetics analysis software (15.8% [n = 18]).

Furthermore, we extracted details of statistical approaches and tools employed for building phylogenetic trees using sequencing data from multi-domain pathogens. Overall, maximum likelihood estimation of phylogenetic relatedness was the most widely used inference method (72.8% [n = 83 studies]). Bayesian inference was only applied in 19.3% (n = 22) of the records, although it was the dominant method for estimating species divergence timescales (used by 18 of the 19 studies that implemented either internal or external molecular clock calibrations) (Supplementary Figure 4). Almost one-third of the studies (26.3% [n = 30]) carried out phylogenetic model validation using a combination of approaches: 12 of them applied both maximum likelihood and Bayesian inference, 10 applied both neighbor-joining and maximum likelihood, and only five studies validated phylogenetic models using three different inference methods.

Discussion

The depth of information provided by genomic data already plays a critical role in pathogen characterization, disease surveillance, and preventive strategies, while genome databases continue to grow vastly and are becoming extensive big data resources in infectious disease research [18]. Our scoping review provides epidemiological and analytical insight into genomic studies investigating transmission and diversity of zoonotic pathogens across multiple One Health domains. We retrieved 114 studies whose sampling and research efforts were highly variable between infectious agents, domains, geographies, sequencing technologies, bioinformatic toolkits, and phylogenetic modeling. Below, we discuss such heterogeneities and highlight opportunities for addressing current knowledge gaps.

Our findings strongly indicated that bacterial pathogens are a major focus of genomic studies of zoonotic diseases across domains (95 out of

114 publications). On the contrary, viruses and other pathogens with complex transmission pathways and multi-host life cycles (i.e., helminths and protozoa) were poorly represented. To date, genomic epidemiology, and phylodynamics in particular, appeared to be more restricted to single-stranded RNA viruses due to their short generation times, rapidly mutating genomes, and large population sizes [8,19]. However, the underrepresentation of viruses in One Health zoonotic studies may reflect the difficulty in isolating and amplifying their genomes in multi-domain contexts. Historical trends demonstrate that spillover events of zoonotic viruses are increasingly frequent and leading to more severe epidemics [20]. Therefore, improving our understanding of the evolution and ecology of viral communities is crucial to establishing spatial, temporal, and environmental traits that can support cross-species transmission forecast and public health risk mitigation.

More than one-third of the records that were included for data extraction explored zoonotic transmission across all One Health domains by integrating human, animal, and environmental components. This is an exciting finding since zoonotic disease control is increasingly recognized as a complex, multi-factorial issue requiring concerted responses across different sectors, including public health, environmental management, veterinary medicine, and agriculture [4,21]. Nevertheless, investigating all One Health domains may not always be a priority, or even necessary, for certain pathogens or research questions. For example, infectious disease organisms such as *Campylobacter* bacteria and *Orthohantavirus* RNA virus are characterized by relatively short environmental persistence [22,23], which may lead investigators to disregard the inclusion of environmental specimens. Moreover, the epidemiological role of non-human vertebrate hosts remains a work in progress for many infectious disease organisms [10]. Consequently, One Health research may not feel the urge to investigate potential environmental sources or animal reservoirs of zoonotic diseases until serendipitous findings shed light on their competence in pathogen transmission and spillover mechanisms [24].

Our objective was to identify, in a transparent and reproducible manner, relevant records that modeled genomic data to track zoonotic pathogen transmission across One Health domains. However, we defined One Health merely using a human-centric perspective on zoonotic disease epidemiology without including other socio-ecological aspects shaping community and ecosystem health [25]. We deliberately chose to limit our screening to studies that included genomic sequencing of samples from human hosts. Therefore, we omitted the substantial body of work integrating genomics to decipher pathogen transmission among multiple non-human vertebrate hosts and at the animal-environment interface. Nevertheless, our work revealed key gaps in sampling efforts within multi-domain initiatives, which may have implications for pathogen surveillance. Overall, less effort has been put into applying genomic tools to environmental samples. Therefore, there is limited understanding of the value of environmental monitoring for the detection of infectious disease agents. For vector-borne pathogens, only a small proportion of publications combined human data with vector surveys (i.e., arachnids and insects), while the median number of samples analyzed from insects was the lowest among the sub-categories across all domains. This knowledge gap is striking given that vector-borne zoonotic diseases such as dengue, Rift Valley fever, and West Nile fever are mosquito-borne public health priorities in many regions across the globe [26]. A clearer understanding of the transmission dynamics between animal reservoirs, arthropod vectors, and human hosts is essential for control strategies and interventions. Excitingly, metagenomic sequencing of individual or pooled arthropod vectors offers a potential single assay to comprehensively identify vector species, vector-borne pathogens, and animal hosts that define their transmission cycle [27].

We noted a rising trend in the number of published studies over time, which sharply declined between 2020 and 2022. This decrease may likely be attributed to the COVID-19 pandemic that has captured most of the scientific attention in recent years. The large number of records on SARS-CoV-2 that were excluded during our title/abstract screening (n = 164) supports this observation. By excluding these articles, we aimed to maintain a balanced and focused review, ensuring a manageable screening process.

We observed a clear socio-economic disparity regarding leading countries where studies in pathogen genomics and One Health are undertaken. Only 27 out of 114 studies were based in low/lower-middle income countries, further highlighting the challenges that low/lower-middle income countries face in embracing the so-called genome sequencing revolution. Numerous obstacles are intrinsic to resource-scarce settings such as access to education and retention of skilled personnel, availability of sequencing platforms, reagents, and maintenance service, financial sustainability, analytical bottlenecks, and access to research initiatives and data [28]. A further obstacle in resource-scarce settings may be implementing genomic surveillance beyond the few centralized hubs that currently exist. Across the African continent, approximately 71% of sequencing systems reside in just five countries, most of them at laboratories with no affiliation with national public health institutes [29]. Nevertheless, low/lower-middle income countries are precisely where genomic pathogen surveillance applied to One Health initiatives is more appropriately deployed given the current public health challenges that these regions face [26].

Finally, we observed a gap in the practice of evolutionary model validation by combining different inference methods. Model validation is a crucial step in establishing that phylogenetic trees depict ancestral divergencies that are reasonably accurate, adequately supported, and reproducible [30]. Our study confirmed that Illumina systems are the most used short-read sequencers, particularly MiSeq which was employed by one-third of the studies as the main next-generation sequencing technology and remains the most common sequencing platform for infectious disease research and public health [6,31]. The popularity of Illumina platforms can be explained by several factors such as high-throughput sequencing, cost-effectiveness, and accessibility, allowing scientists to improve the detection and characterization of pathogens, even when present at low titers, and undertake larger surveys. Most publications that used PacBio RS II and/or Oxford Nanopore Technologies MinION long-read sequencing platforms focused on bacterial pathogens, which often have highly complex genome structures including long, repetitive, and mobile elements requiring validation via long-read data. Studies that integrate multiple sequencing approaches with complementary strengths are a positive example of avoiding potential systemic biases in the produced data [32,33]. Genome assemblies generated through integrated approaches generally exhibit higher base-level accuracy and coverage due to high-throughput sequencing afforded by short-read platforms, while long-read sequencing enhances the scaffolding of high-quality contigs [34].

Conclusion

Our scoping review identified several key areas for future progress in the application of genomic technologies to infectious disease research across multiple One Health domains. First, the need to integrate multi-domain surveillance of arboviruses and their vectors is clear. Furthermore, high-throughput sequencing in low- and middle-income countries remains not sufficiently leveraged to study zoonotic diseases in all domains. Our work highlights the clear need for additional studies exploring the human-animal-environment triad using genomic epidemiology for infectious disease detection and characterization. To address the threat of emerging pathogens and inform public health policy development, scientific communities, and health authorities must join forces to develop a roadmap for ensuring research capacity strengthening and sustainability applied to whole-genome sequencing technologies in One Health [9,28]. Within interdisciplinary research, and more specifically One Health, the migration to whole-genome studies has recently begun. A large expansion is still expected and, therefore, it

is essential that future initiatives invest time and effort to establish collaborations across disciplines and sectors whose research has long remained siloed.

Funding

This study was supported by the Defense Threat Reduction Agency, United States Department of the Defense (grant number HDTRA12110028) (the content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred); a Natural Environment Research Council Independent Research Fellowship (grant number NE/V014730/1); the Biotechnology and Biological Sciences Research Council (grant number BB/Y006879/1); and the UK Research and Innovation Global Challenges Research Fund One Health Poultry Hub (grant number BB/S011269/1).

Author contributions

SC, CLF, and CJS conceptualized the study, curated the data, and drafted the manuscript. FB, ZIT, and JR curated the data, performed data analysis, and revised the manuscript.

Data availability

The code used for data analysis and visualization is available at https://github.com/cfaustus/ngs_scoping_review.git. The data used in this study are deposited in an open-access repository of the University of Glasgow (http://dx.doi.org/10.5525/gla.researchdata.1534).

Declaration of Competing Interest

The authors have no competing interests to declare.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ijidoh.2024.100031.

References

[1] World Health Organization. Taking a multisectoral, one health approach: a tripartite guide to addressing zoonotic diseases in countries. Geneva: WHO Press; 2019.
[2] Grace D, Mutua FK, Ochungo P, Kruska RL, Jones K, Brierley L, et al. Mapping of poverty and likely zoonoses hotspots. Zoonoses Project 4. Report to the UK Department for International Development. Nairobi: ILRI; 2012.
[3] Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature 2008;451:990–3. https://doi.org/10.1038/nature06536.
[4] Davis MF, Rankin SC, Schurer JM, Cole S, Conti L, Rabinowitz P, et al. Checklist for One Health epidemiological reporting of evidence (COHERE). One Health 2017;4:14–21. https://doi.org/10.1016/j.onehlt.2017.07.001.
[5] Dos S, Ribeiro CDS, van de Burgwal LHM, Regeer BJ. Overcoming challenges for designing and implementing the One Health approach: a systematic review of the literature. One Health 2019;7:100085. https://doi.org/10.1016/j.onehlt.2019.100085.
[6] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17:333–51. https://doi.org/10.1038/nrg.2016.49.
[7] Keita AK, Koundouno FR, Faye M, Düx A, Hinzmann J, Diallo H, et al. Resurgence of Ebola virus in 2021 in Guinea suggests a new paradigm for outbreaks. Nature 2021;597:539–43. https://doi.org/10.1038/s41586-021-03901-9.
[8] Ingle DJ, Howden BP, Duchene S. Development of phylodynamic methods for bacterial pathogens. Trends Microbiol 2021;29:788–97. https://doi.org/10.1016/j.tim.2021.02.008.
[9] Urban L, Perlas A, Francino O, Martí-Carreras J, Muga BA, Mwangi JW, et al. Real-time genomics for One Health. Mol Syst Biol 2023;19:e11686. https://doi.org/10.15252/msb.202311686.
[10] Becker DJ, Seifert SN, Carlson CJ. Beyond infection: integrating competence into reservoir host prediction. Trends Ecol Evol 2020;35:1062–5. https://doi.org/10.1016/j.tree.2020.08.014.
[11] Foo YZ, O'Dea RE, Koricheva J, Nakagawa S, Lagisz M. A practical guide to question formation, systematic searching and study screening for literature reviews in ecology and evolution. Methods Ecol Evol 2021;12:1705–20. https://doi.org/10.1111/2041-210X.13654.
[12] 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. https://doi.org/10.7326/M18-0850.
[13] Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods 2020;11:181–217. https://doi.org/10.1002/jrsm.1378. (, PubMed).
[14] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4.
[15] Rees EM, Minter A, Edmunds WJ, Lau CL, Kucharski AJ, Lowe R. Transmission modelling of environmentally persistent zoonotic diseases: a systematic review. Lancet Planet Health 2021;5:e466–78. https://doi.org/10.1016/S2542-5196(21)00137-6.
[16] Cavalerie L, Wardeh M, Lebrasseur O, Nanyingi M, McIntyre KM, Kaba M, et al. One hundred years of zoonoses research in the Horn of Africa: a scoping review. PLoS Negl Trop Dis 2021;15:e0009607. https://doi.org/10.1371/journal.pntd.0009607.
[17] R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2023.
[18] Stockdale JE, Liu P, Colijn C. The potential of genomics for infectious disease forecasting. Nat Microbiol 2022;7:1736–43. https://doi.org/10.1038/s41564-022-01233-6.
[19] Hill V, Ruis C, Bajaj S, Pybus OG, Kraemer MUG. Progress and challenges in virus genomic epidemiology. Trends Parasitol 2021;37:1038–49. https://doi.org/10.1016/j.pt.2021.08.007.
[20] Meadows AJ, Stephenson N, Madhav NK, Oppenheim B. Historical trends demonstrate a pattern of increasingly frequent and severe spillover events of high-consequence zoonotic viruses. BMJ Glob Health 2023;8:e012026. https://doi.org/10.1136/bmjgh-2023-012026.
[21] Standley CJ, Fogarty AS, Miller LN, Sorrell EM. One Health systems assessments for sustainable capacity strengthening to control priority zoonotic diseases within and between countries. Risk Manag Healthc Policy 2023;16:2497–504. https://doi.org/10.2147/RMHP.S428398.
[22] Bronowski C, James CE, Winstanley C. Role of environmental survival in transmission of *Campylobacter jejuni*. FEMS Microbiol Lett 2014;356:8–19. https://doi.org/10.1111/1574-6968.12488.
[23] Guterres A, de Lemos ERS. Hantaviruses and a neglected environmental determinant. One Health 2018;5:27–33. https://doi.org/10.1016/j.onehlt.2017.12.002.
[24] Tomori O, Oluwayelu DO. Domestic animals as potential reservoirs of zoonotic viral diseases. Annu Rev Anim Biosci 2023;11:33–55. https://doi.org/10.1146/annurev-animal-062922-060125.
[25] Gallagher CA, Keehner JR, Hervé-Claude LP, Stephen C. Health promotion and harm reduction attributes in One Health literature: a scoping review. One Health 2021;13:100284. https://doi.org/10.1016/j.onehlt.2021.100284.
[26] Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. Nat Rev Microbiol 2022;20:193–205. https://doi.org/10.1038/s41579-021-00639-z.
[27] Batson J, Dudas G, Haas-Stapleton E, Kistler AL, Li LM, Logan P, et al. Single mosquito metatranscriptomics identifies vectors, emerging pathogens and reservoirs in one assay. eLife 2021;10:e68353. https://doi.org/10.7554/eLife.68353.
[28] Yek C, Pacheco AR, Vanaerschot M, Bohl JA, Fahsbender E, Aranda-Díaz A, et al. Metagenomic pathogen sequencing in resource-scarce settings: lessons learned and the road ahead. Front Epidemiol 2022;2:926695. https://doi.org/10.3389/fepid.2022.926695.
[29] Inzaule SC, Tessema SK, Kebede Y, Ogwell Ouma AEO, Nkengasong JN. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. Lancet Infect Dis 2021;21(9):e281. https://doi.org/10.1016/S1473-3099(20)30939-7.
[30] Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. Science 2015;347:aaa4339. https://doi.org/10.1126/science.aaa4339.
[31] Segerman B. The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and RefSeq genome databases. Front Cell Infect Microbiol 2020;10:527102. https://doi.org/10.3389/fcimb.2020.527102.
[32] Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. NAR Genom Bioinform 2021;3:lqab019. https://doi.org/10.1093/nargab/lqab019.
[33] Wick RR, Judd LM, Holt KE. Assembling the perfect bacterial genome using Oxford nanopore and Illumina sequencing. PLOS Comput Biol 2023;19:e1010905. https://doi.org/10.1371/journal.pcbi.1010905.
[34] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 2020;21:30. https://doi.org/10.1186/s13059-020-1935-5.