

Review

Cite this article: Morton LC, Forshey BM, Klinkhammer KE, Romaner A, Traore Z, Hartman LJ, Standley CJ and Roess AA (2026). Local genomic epidemiology investigations of SARS-CoV-2 during the early pandemic response: A global systematic review. *Epidemiology and Infection*, **154**, e56, 1–12 <https://doi.org/10.1017/S0950268826101381>

Received: 19 July 2025

Revised: 20 December 2025

Accepted: 13 March 2026

Keywords:

COVID-19; epidemiology; genomic surveillance; outbreak investigation; SARS-CoV-2; sequencing

Corresponding author:

Lindsay C. Morton;

Email: lcmorton@gwu.edu

Local genomic epidemiology investigations of SARS-CoV-2 during the early pandemic response: A global systematic review

Lindsay C. Morton¹ , Brett M. Forshey², Katharina E. Klinkhammer³, Andrew Romaner², Zoumanna Traore⁴, Laurie J. Hartman^{5,6}, Claire J. Standley^{5,7} and Amira A. Roess⁸

¹The George Washington University Milken Institute of Public Health, USA; ²US Public Health Service Commissioned Corps, USA; ³The George Washington University Milken Institute School of Public Health, USA; ⁴Georgetown University, USA; ⁵Cherokee Nation Businesses, USA; ⁶Defense Health Agency, Global Emerging Infections Surveillance, USA; ⁷Coalition for Epidemic Preparedness Innovations, USA and ⁸George Mason University, USA

Abstract

Genomic epidemiology was essential for characterizing SARS-CoV-2 transmission during the early COVID-19 pandemic. This systematic review examined how whole-genome sequencing was used in local outbreak investigations published between March 2020 and March 2021. Searches of PubMed, Scopus, and Web of Science identified 32 studies from 18 countries that integrated genomic and epidemiological data for local outbreak investigations. Most studies were conducted in healthcare settings or in high-income countries. A limited number of studies were conducted in low- and middle-income countries, except for China and Vietnam. Illumina or Oxford Nanopore platforms and tiled-amplicon protocols were the most common sequencing methods. Phylogenetic trees were the most common genomic epidemiology analytical approach. Genomic data enabled confirmation of suspected transmission links, detection of multiple introductions, and identification of asymptomatic or presymptomatic transmission. Important enablers of early implementation included open-access genomics databases, standardized protocols (e.g. ARTIC), open-source tools (e.g. Nextstrain), and cross-sector partnerships and funding. Study quality and adherence to common observational study reporting guidelines varied widely. Familiarity with the STROME-ID guidelines for molecular epidemiology studies would have improved overall quality. These findings highlight the utility of genomic epidemiology in outbreak response and support its continued integration into public health surveillance systems.

Introduction

In recent decades, genome sequencing technologies have matured and spread rapidly across nearly all areas of the biological sciences [1]. Pathogen genomic information can enhance traditional epidemiology methods and disease surveillance programmes. At the international, national, or regional level, genomic surveillance can be used to study broad evolutionary trends in infectious diseases over time and space and to monitor changes in the genome that potentially impact medical countermeasures. However, local outbreak investigations can also benefit from additional genomic information to elucidate potential sources of infection, identify disease clusters, study individual transmission events, and guide the deployment of public health interventions [2, 3].

Previous investments by governments, academia, and industry in genome sequencing for pathogen surveillance – such as foodborne pathogen surveillance in Europe and the Americas, pandemic influenza monitoring globally, and large-scale outbreak responses like those to the Zika virus in the Americas and Ebola virus in West Africa – were made in both high-income countries (HICs) and low- to middle-income countries (LMICs) [4–7]. However, by 2019, the analysis and interpretation of genomic sequence data within the context of local public health were not widely practiced globally. Additionally, the field of genomic (or molecular) epidemiology had not yet been fully implemented by traditional epidemiologists working in infectious disease outbreaks, particularly in applied or lower-resource settings. As a result, when the coronavirus disease 2019 (COVID-19) pandemic emerged, there was significant untapped potential for the widespread application of genomics in local outbreak investigations.

Studying genomic epidemiology practices during the early phase of the COVID-19 pandemic is necessary for understanding how novel technologies were integrated into the public health response. This review focuses on the application of genomic epidemiology during a period when non-pharmaceutical interventions and contact tracing were the primary tools for controlling

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.



SARS-CoV-2 transmission. The findings will inform future pandemic preparedness for novel pathogens and illustrate opportunities to strengthen local public health capacity and support implementation in lower resourced settings.

Purpose of the review

This review describes the use of SARS-CoV-2 whole-genome sequencing and genomic epidemiology as part of local outbreak investigations and response during the first year of the COVID-19 pandemic. We examined the study setting, methods used, study results, quality and potential facilitators of the use of genomics in these settings. Finally, we used thematic analysis to illustrate the added value of genomic epidemiology for understanding SARS-CoV-2 transmission in the early pandemic response.

Methods

The study followed common systematic review standards, including the Cochrane Handbook for Systematic Reviews of Interventions and updated Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines (Supplementary Material 2) [8, 9]. The protocol was registered in the international prospective register of systematic reviews before abstract screening (<https://www.crd.york.ac.uk/prospero/>, CRD42022255562).

Systematic search strategy

A search of PubMed, Scopus, and Web of Science was conducted on 23–24 April 2021, using terms for COVID-19, genomics, and epidemiology (Supplementary Material 2). The search was limited to papers published within one year of the World Health Organization (WHO) pandemic declaration (11 March 2020) [10]. Searches of PubMed used keywords and MeSH terms, while searches of Scopus and Web of Science used only keywords. The search results were uploaded to Covidence for screening (Covidence,

Veritas Health Innovation, Melbourne, Australia; www.covidence.org).

Abstract and full-text screening

An interdisciplinary team of reviewers with expertise in genomic epidemiology, global public health, virology, laboratory science, and global health security conducted abstract screening in Covidence. Two independent reviewers evaluated each title and abstract using a priori inclusion criteria (Supplementary Table S2). After achieving an inter-rater reliability of 0.80 on a pilot set of ten papers, the team screened the entire set of abstracts. Full-text screening followed the same protocol, and the primary justification for excluding any full-text article was documented. Any disagreements were resolved by consensus or adjudicated by a third reviewer.

Data abstraction and synthesis

Articles remaining after full-text screening underwent data extraction and thematic synthesis. Numeric and qualitative data elements were manually extracted using a standardized template uploaded into Covidence (Supplementary Material 4). Extracted data were cleaned, coded, and analysed using Microsoft Excel, RStudio 2024.09.0, and Datawrapper.de. World regions and country income level designations were determined using World Bank 2020 definitions. Further thematic analysis of extracted genomic epidemiology conclusions was conducted in NotebookLM (Gemini 2.0).

Qualitative framework for study settings

A general qualitative framework was developed to define local-level genomic epidemiology studies, focusing on 'local outbreak investigations' or discrete, setting-specific clusters (e.g. healthcare facilities, workplaces, congregate environments, ships, events) where genomic and epidemiologic data were analysed together to understand transmission (Figure 1). Studies involving multiple institutions across broader geographic areas (e.g. multiple hospitals with

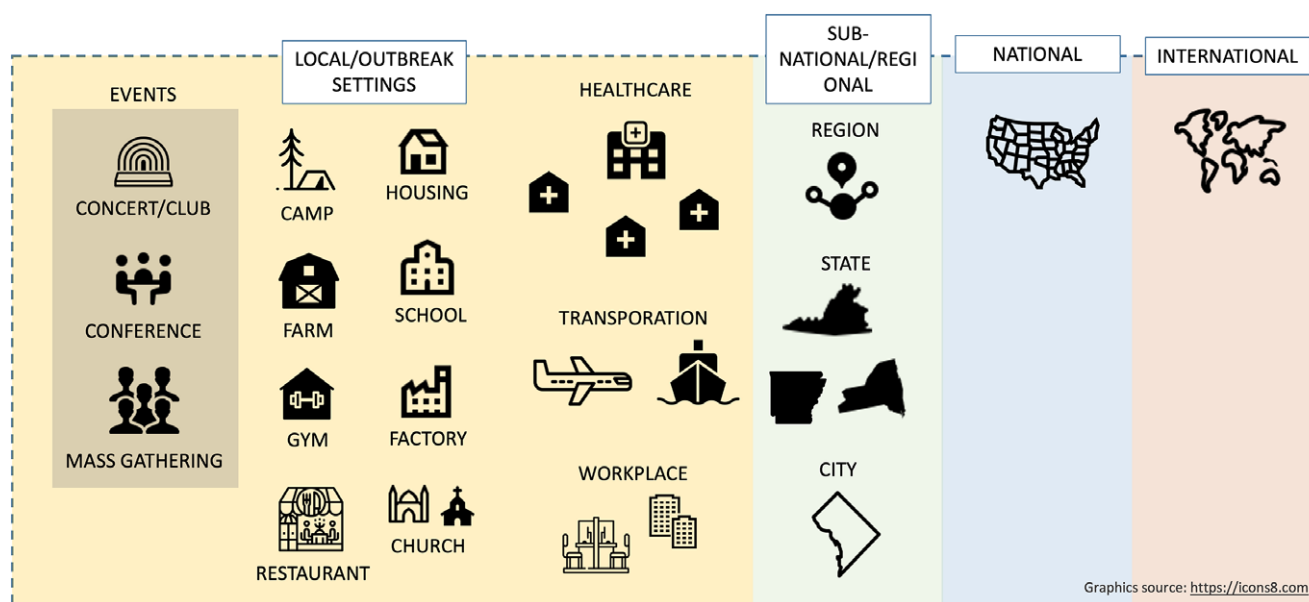


Figure 1. Framework for grouping settings and level of analysis for genomic epidemiology studies conducted early in the COVID-19 pandemic. *Note:* Local outbreak response settings can include event-based, institutional, congregate, workplace, and transportation-related exposures. Several institutions grouped together, or larger geographic units, can inform city, state, and regional-level analyses (source: <https://icons8.com>).

separate patient populations) were excluded, as they often informed state- or regional-level analyses rather than local contexts.

Quality assessment

Risk of bias was evaluated using a custom tool incorporating complementary criteria from the National Institutes of Health (NIH) Study Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies and the Joanna Briggs Institute (JBI) Critical Appraisal Tools Checklist for Analytical Cross-Sectional Studies [11, 12] (Supplementary Material 5). Studies were also assessed for adherence to the Strengthening of Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID) reporting guidelines [13] (Supplementary Material 5). Sub-analyses compared the mean quality and reporting scores between studies from HICs and LMICs, as well as by the extent of genomic epidemiology methods used.

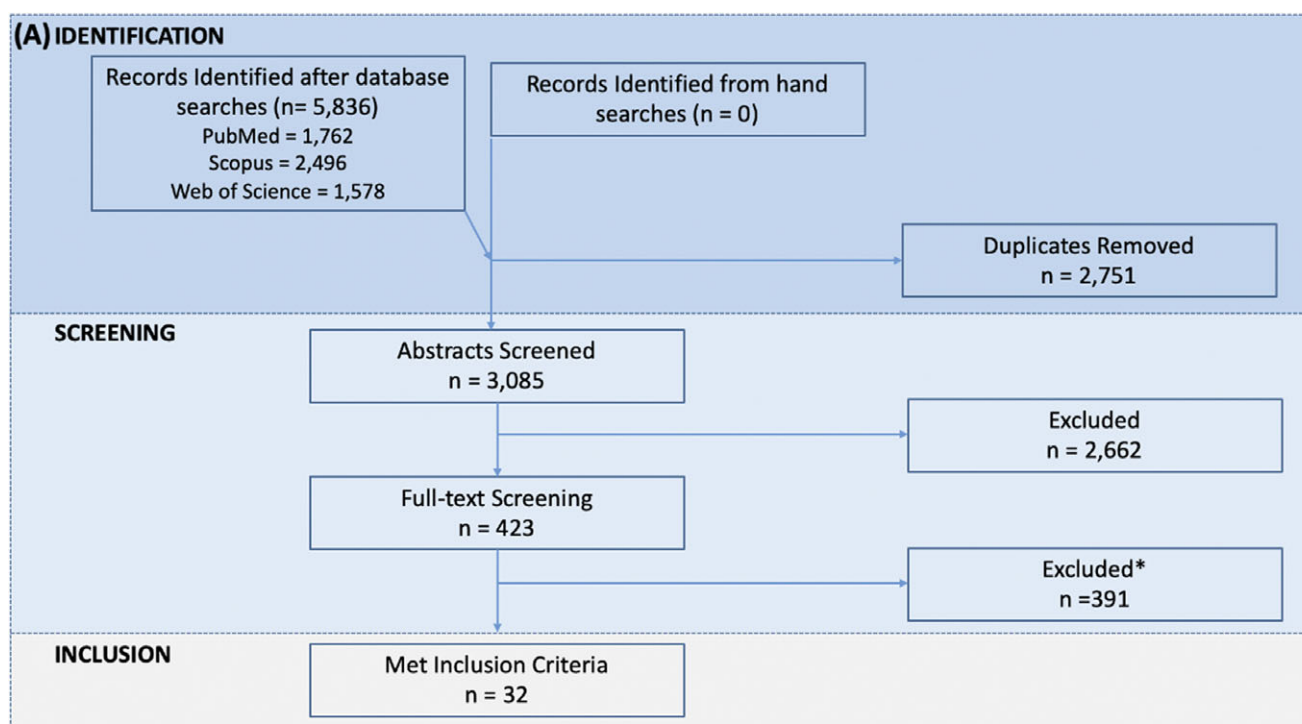
Results

Systematic search and screening results

The systematic search identified 5836 abstracts across three databases: 1762 from PubMed, 2496 from Scopus, and 1578 from Web of Science. After removing 2751 duplicate articles, 3085 abstracts were screened with an inter-rater reliability of 88% during the pilot phase and a Cohen's Kappa of 0.5–0.8 during full screening [14]. Following exclusion of 2662 abstracts and 391 full-text articles, 32 studies met the full inclusion criteria for final review (Figure 2 and Supplementary Material 6).

Reasons for exclusion

The most common reason for full-text exclusion was a lack of focus on local, institutional, or outbreak settings (225, 58%) (Figure 2). Excluded studies ranged from intra-host viral dynamics to global



(B) *Exclusion criteria (n = 391)

	n
Not local, institutional, or outbreak setting (includes multi-institutional and regional-level studies)	225
No primary genomic data collected	44
No genomic epidemiology analysis	25
Not a research study (e.g., commentary, review, genome announcement)	12
Case study	25
Case series among unconnected cases (i.e., no putative epi links, including returned travelers)	20
Limited epidemiological data (includes wastewater surveillance, fomites)	11
Methods focused	11
Epizootic (i.e., primarily a wild or domesticated animal study)	5
Full-text not available in English	4
Not in publication date range	3
Only partial SARS-CoV-2 genome sequencing	3
Pre-print	2
Human genome sequencing (no viral sequencing)	1

Figure 2. PRISMA flow diagram. Note: (a) Depicts studies at each step from identification, screening, to final inclusion. (b) Primary reasons for excluding full-text articles.

surveillance trends. Multi-institutional and regional-level analyses were also excluded under this criterion. Additional exclusion reasons included use of only secondary genomic data (44, 11%) and absence of genomic epidemiology methods (25, 6%), which meant the study did not integrate genomic and epidemiological data in the analysis or interpretation. Other excluded articles were case studies (25, 6%), case series among sporadic cases (20, 5%), and articles that contained little to no epidemiological data (11, 3%).

Location and setting

The 32 included studies were conducted in 18 countries (Figure 3). Three of these studies (9%) were conducted in multiple countries (e.g. primary exposure/outbreaks occurred in one country and the investigation occurred in another). Most studies occurred in HICs (29, 91%). Only six studies (19%) were conducted in LMICs (Table 1). In some cases (e.g. Egypt, India, and Niger), SARS-CoV-2 exposure occurred locally, but sample processing and study analysis were conducted elsewhere (Figure 3).

Geographically, most studies were conducted in Europe and Central Asia (13, 41%), East Asia and the Pacific (9, 28%), and North America (8, 28%) (Table 1). The United States had the most studies ($n = 8$) and the greatest diversity in settings such as healthcare facilities [15–17], a fishing vessel [18], a mass gathering [19], a prison [20], a summer camp [21], and a military training location [22] (Figure 3). Studies were also conducted in Japan (4), Germany (3), the Netherlands (3), China (2), and the United Kingdom (2).

In terms of setting, nearly half of the studies (14 of 32, 44%) investigated healthcare-associated outbreaks, including hospitals, long-term care facilities, and a psychiatric ward [15–17, 23–32]. Congregate living environments were the second most common (5, 16%), including military barracks [22, 33, 34], a prison [20], and a sleep-away camp [21]. Six studies (19%) focused on transmission during travel or in transit-related environments, including cruise ships [35–37], commercial flights [38, 39], and fishing vessels [18]. Event-based (4, 13%) and workplace-related outbreaks (4, 13%) were less common but represented diverse contexts like nightclubs [40], mass gatherings [19, 30, 40], a meat-processing plant [41], and a laboratory [42] (Table 1).

Major characteristics of LMIC studies

Many studies from LMICs were excluded during full-text screening. Common reasons included reliance solely on secondary genomic data, a focus on broader national or international trends; the studies were an evolutionary analysis of SARS-CoV-2 and other coronaviruses or were a ‘snapshot’ of the first imported cases in a country. The limited number of included papers from LMICs prevented extensive sub-analyses for country-income-level effects.

Six studies were either conducted entirely in LMICs (3, 9%) or focused on major SARS-CoV-2 exposure events in LMICs (3, 9%). For the three LMIC-led studies, a study in Vietnam investigated a potential superspreading event at a bar and leveraged resources from Oxford University and the Wellcome Trust [43]. Two studies were conducted in China, entirely utilizing local institutions and resources. One analysed nosocomial transmission of SARS-CoV-2 in a Wuhan hospital [31], while the other examined community transmission in Beijing [44]. For studies of groups of individuals with extended travel in LMICs, this included Belgian military personnel training in Niger, Japanese tourists on Nile River cruises in Egypt, and an international flight from India to New Zealand [33, 36, 38]. Notably, the authors, institutions, and funding sources for these three LMIC-travel-related studies were primarily based in HICs (Supplementary Material S3-1).

Study type and characteristics

Outbreak investigations were half (16, 50%) of the included studies, often employing contact tracing, snowball sampling, or elements of multiple observational epidemiology study designs (Table 1). A quarter of papers (8, 25%) were cross-sectional studies, including point prevalence and period prevalence studies. Cohort (6, 19%), case–control (1, 3%), and case series (1, 3%) comprised the remaining quarter of publications.

Nearly half of the studies (14, 44%) started sample collection in March 2020, coinciding with the declaration of the COVID-19 pandemic (Table 1). The earliest collection period began in January, and the latest was September 2020. While a quarter of studies ended sample collections in May 2020, end dates were more evenly distributed across the year compared to start dates. Most studies

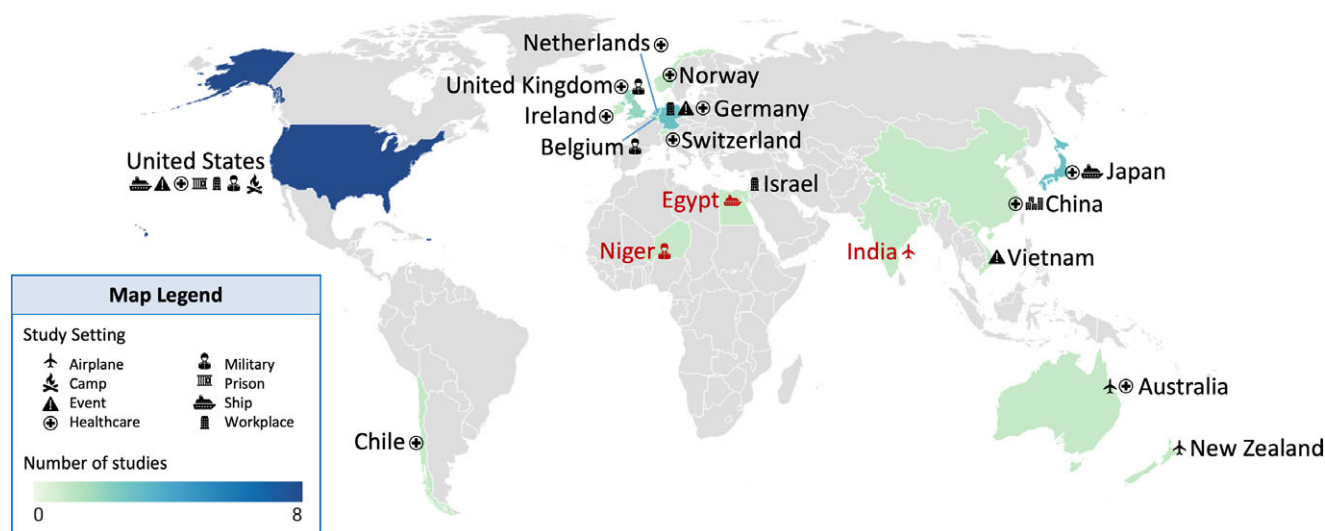


Figure 3. Global map of study locations. *Note:* Icons indicate the study settings represented within each country, and the colour ramp indicates the number of published studies. A red icon and country name indicate that major exposure or travel history data were from these countries, and sample collection and analysis occurred elsewhere. Figure developed using Datawrapper (source: <https://www.datawrapper.de/>).

Table 1. Study characteristics (*n* = 32)

Variable	Studies	
	<i>n</i> = 32	% ^a
Setting		
Healthcare	14	44
Hospital	11	79
Skilled nursing/long-term care	3	21
Event/mass gathering	4	13
Bar/nightclub	2	50
Church service	1	25
Other event	1	25
Workplace	4	13
Office	2	50
Meat packing plant	1	25
Laboratory	1	25
Transit-related	6	19
Ship	4	67
Airplane	2	33
Congregate	5	16
Military	3	60
Camp	1	20
Prison	1	20
Other/multiple^b	3	9
Study type		
Outbreak investigation	16	50
Cross-sectional ^c	8	25
Cohort ^d	6	19
Case-control	1	3
Case series	1	3
Geographical region		
East Asia and Pacific	9	28
Europe and Central Asia	13	41
Latin America and Caribbean	1	3
Middle East and North Africa	2	6
North America	8	25
South Asia	1	3
Sub-Saharan Africa	1	3
Country income level		
High	29	91
Upper middle	2	6
Lower middle	3	9
Low	1	3
Study sample collection start (month in 2020)		
January	1	3
February	4	13
March	14	44

(Continued)

Table 1. (Continued)

Variable	Studies	
	<i>n</i> = 32	% ^a
April	2	6
May	5	16
June	1	3
August	3	9
September	2	6
Study sample collection end (month in 2020)		
February	3	9
March	6	19
April	4	13
May	8	25
June	1	3
July	5	16
September	3	9
October	2	6
Length of sample collection		
≤ 2 weeks	5	15
2–8 weeks ^e	20	62
> 8 weeks	7	22

^aDenominator is 32 unless indicated; some categories will sum to over 32 if not mutually exclusive.^bOther, includes housing during an event, outbreak investigation, or quarantine, such as a quarantine hotel, apartment building, or household.^cCross-sectional includes point and period prevalence studies.^dCohort includes retrospective and prospective cohort studies.^eFifteen to sixty days, inclusive.

collected samples over 2 weeks to 2 months (22, 68%), with durations ranging from 2 to 113 days.

Sequencing methods

Eighteen studies used Illumina short-read sequencing platforms (56%); the most common were MiSeq and NextSeq. Nine studies used long-read sequencing (28%), primarily the Oxford Nanopore Technologies (ONT) MinION. Less commonly used platforms included the Ion Torrent (1, 3%) and CapitalBio BioelectronSeq (1, 3%), while five studies (16%) did not specify a sequencer. Two studies (6%) used more than one sequencing technology (Table 2).

Most studies used a whole-genome-tiled-amplicon method to sequence SARS-CoV-2 (26, 81%), while two studies (6%) used either shotgun metagenomics or RNA-enrichment metagenomics [18, 44]. Notably, four studies (13%) either did not report a sequencing protocol or the specific method could not be determined from the reference (Table 2). Six studies (19%) used in-house protocols or cited those developed by a coauthor (Table 2) [33, 45–48]. Two studies (6%) did not specify a sequencing protocol. Most studies (22, 69%) referred to common or previously published protocols, while two studies (6%) reported adapting a published protocol. Twenty-one studies (66%) used the ARTIC Network tiled-amplicon-whole-genome-sequencing protocol for SARS-CoV-2 published in January 2020 [49], while three studies (9%) used methods similar or derived from ARTIC (PrimalSeq [50, 51] (Table 2).

Table 2. Study laboratory methods ($n = 32$)

Variable	Studies	
	n	% ^a
Sequencer		
Illumina MiSeq	11	34
Illumina NextSeq	5	16
Illumina (not specified)	2	6
Nanopore MinION	5	16
Nanopore GridION	1	3
Nanopore (not specified)	3	9
BioelectronSeq	1	3
IonTorrent	1	3
Not reported	5	16
Sequencing method		
Tiled amplicon	26	81
Metagenomics ^b	2	6
Not reported	4	13
Lab developed protocol		
Yes ^c	6	19
Partially adapted from existing protocol	2	6
No	22	69
Protocol not reported	2	6
Sequencing protocol reference		
ARTIC Network/Quick 2020 ^{d,e}	21	66
PrimaSeq/Grubaugh 2019, Quick 2017 ^e	3	9
Oude et al.	3	9
Eden et al.	2	6
Paden et al.	2	6
AmpliSeq	1	3
TruSeq	1	3
MINERVA	1	3
Not reported/not applicable	3	9
Study sequence data shared		
Yes	20	63
GISAID ^f	17	85
NCBI ^g	4	20
ENA ^h	1	5
CNCB ⁱ	1	5
Institutional website	1	5
No/not reported	12	38

^aDenominator is 32 unless indicated otherwise; some categories will sum to over 32 if not mutually exclusive.

^bIncludes shotgun metagenomics and RNA-enrichment metagenomics.

^cYes, indicates the study was the first publication of the sequencing protocol, or the lead author of the sequencing protocol was a co-author on the study.

^dARTIC Network's SARS-CoV-2 tiled-amplicon sequencing primer scheme, Versions 1, 2, and 3; Quick is the primary author on the first ARTIC SARS-CoV-2 protocols published in early 2020.

^eAlthough closely related, PrimaSeq (Grubaugh 2019, Quick 2017) refers to the broader precursor tiled-amplicon-sequencing framework for viral genomes, while ARTIC (Quick

2020) refers specifically to the ARTIC Network's SARS-CoV-2 sequencing protocols built on the same tiled-amplicon idea.

^fGISAID; Global Initiative on Sharing All Influenza Data.

^gNCBI; National Center for Biotechnology Information, U.S. National Institutes of Health, includes both GenBank and SRA (sequence read archive) databases.

^hENA; European Nucleotide Archive.

ⁱCNCB; China National Center for Bioinformation, National Genomics Data Center.

Twenty studies (63%) reported openly sharing genomic data, while 12 (37%) did not. Of the 20 studies that shared data, most deposited sequences in the Global Initiative on Sharing All Influenza Data (GISAID) (17, 85%) or the National Center for Biotechnology Information (NCBI) (4, 20%). Other databases included European Nucleotide Archive (ENA) (1, 5%), Chinese National Center for Bioinformation (CNCB) (1, 5%), and an institutional website (1, 5%) (Table 2).

Genomic epidemiology methods

The use of genomic epidemiology methods was grouped into three categories: primary/extensive, moderate, and minor. Only ten studies (31%) included genomic epidemiology as a primary objective of the study. These studies extensively utilized genomics and genomic epidemiology in the methods and analyses and often had genomic epidemiology in the title or abstract. Fifteen studies (47%) used genomic epidemiology to a moderate extent, meaning that while these methods were used, they were not the primary purpose of the study. Finally, seven studies (22%) used genomics as a minor component of a more traditional epidemiological study (e.g. seroconversion study) (Table 3).

Nearly all studies used a phylogenetic tree to integrate genomic and epidemiological data (28, 88%) (Table 3). The most common models were maximum-likelihood (17, 53%) and time-scaled maximum-likelihood trees (7, 22%). The most sophisticated method was Bayesian time-scaled tree-building, used in only one paper [43]. Notably, a fifth of studies either did not report the phylogenetic method used (2, 6%) or did not produce a phylogenetic tree (4, 14%). Other genomic epidemiology methods included cluster diagrams (6, 19%), haplotype or single-nucleotide variant (SNV) networks (6, 19%), genotype classification tables (e.g. SNVs, lineages, clusters) (11, 34%), and location maps annotated with genomic information (7, 22%) (Table 3).

Eight studies (25%) reported using the open-source software Nextstrain for genomic epidemiology analyses (<https://docs.nextstrain.org/projects/ncov/en/latest/>) [52]. Twenty-four studies (75%) used location information, while 16 (50%) used sample collection date. Location data were commonly used to colour nodes on phylogenetic trees or cluster diagrams, while collection dates informed time-scaled phylogenetic trees (Table 3).

Synthesis of genomic epidemiology study findings

Thematic analysis showed that, across studies, genomic epidemiology supported multiple, recurring applications in early SARS-CoV-2 outbreak response (Table 4). First, genomic and phylogenetic analyses were used to confirm epidemiologically suspected transmission clusters by showing close genetic relatedness among viral genomes from individuals within a cluster, often differing by only a small number of single-nucleotide polymorphisms (SNPs) [15, 19, 21, 43]. Second, genomic data revealed transmission patterns not identifiable through traditional epidemiologic

Table 3. Study genomic epidemiology methods ($n = 32$)

Variable	Studies	
	<i>n</i>	% ^a
Primary purpose of the study		
Primarily a genomic epidemiology study	10	31
Partially genomics/genomic epi study	15	47
Minor genomic epi component	7	22
Genomic epidemiology methods		
Phylogenetic tree	27	84
Table with genomic information ^b	11	34
Location map with genomic information	7	22
Haplotype or SNV network diagram ^c	6	19
Cluster diagram	6	19
Other ^d	2	6
Phylogenetic model		
Maximum-likelihood	17	53
Time-scaled maximum-likelihood	7	22
Bayesian	1	3
Neighbour-joining	1	3
Other/not reported	2	6
Not performed	4	13
Use of Nextstrain software^e		
Yes	8	25
No/not reported	24	75
Epidemiological fields used		
Location	24	75
Date of sample collection or symptom onset	16	50

Note: SNV, single-nucleotide variant.

^aDenominator is 32 unless indicated otherwise; some categories will sum to over 32 if not mutually exclusive.

^bResults included a table(s) with genomic comparison methods such as genetic distances, SNPs, lineage information, SARS-CoV-2 variant information, or genomic cluster assignment.

^cIncludes haplotype networks, minimum-spanning trees using SNV differences, or median-joining networks using haplotype differences.

^dIncludes narrative comparisons or other qualitative comparisons.

^eNextstrain indicates study used Nextstrain open-source software resources, including Auspice and Augur, for bioinformatic or phylogenetic analyses.

methods alone, including cryptic linkages and missed transmission events [24, 30, 32, 39].

Third, several studies used viral genomics to distinguish multiple independent introductions of SARS-CoV-2 from subsequent local transmission, clarifying whether observed case increases reflected ongoing spread within a setting or repeated importations [24, 25, 34, 37]. Fourth, when combined with contact tracing and sampling of asymptomatic individuals, genomics was also used to detect transmission chains involving asymptomatic, presymptomatic, and subclinical carriers, inferred through near-identical viral genomes (≤ 1 –2 SNP differences) among individuals whose epidemiologic timelines suggested exposure prior to symptom onset [43, 44].

Finally, genomic data were used to evaluate the effectiveness of infection prevention and control interventions by assessing whether

genetically linked transmission clusters persisted or ceased following implementation of non-pharmaceutical measures such as universal masking, physical distancing, or cohorting [16, 23, 24, 26]. In several investigations, the absence of further spread within defined genomic clusters after intervention implementation was interpreted as molecular evidence consistent with intervention effectiveness.

Quality assessment and adherence to reporting guidelines

Across all studies, the average quality score was 55%, and the average completeness of STROME-ID reporting elements was 57% (Table 5). Nearly half of the studies ($n = 15$) were rated as moderate quality, while only five studies achieved a high-quality score ($\geq 70\%$). The remaining 12 studies were considered lower quality due to potential bias in study design, analysis, or insufficient methodology details. Only four studies demonstrated high STROME-ID reporting completeness. Six showed moderate (50–70%), and 22 had low completeness scores ($< 50\%$).

Quality scores were lower for studies from LMICs compared to HICs, though the difference was not statistically significant by two-sample t-test. STROME-ID completeness did not differ between HIC and LMIC studies. Interestingly, studies with limited genomic epidemiology analysis had higher quality scores than those with partial or extensive genomic epidemiology analysis (t-test, $p = 0.02$). However, these same studies scored significantly lower on STROME-ID completeness compared to studies with more extensive genomic epidemiology analysis (t-test, $p < 0.001$) (Table 6).

Discussion

Features of genomic epidemiology studies published during the early pandemic

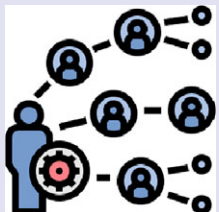
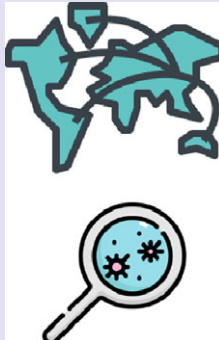
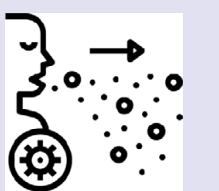
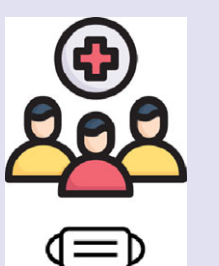
The relatively small number of included studies likely reflects the early-pandemic challenges of generating and then integrating genomic and epidemiologic data, as well as the reality that most genomic analyses at the time were designed for broader surveillance purposes rather than outbreak investigation. Reviewing genomic epidemiology studies published early in the COVID-19 pandemic offers insights into the state of infrastructure and local outbreak response capacity at that time. This also limits our focus to early outbreak response practices before widespread availability of vaccines, the emergence of variants of concern, and additional global investments in sequencing [53, 54]. Many studies (44%) involved healthcare-associated outbreaks, reflecting the urgent need early in the pandemic to understand transmission in high-risk settings.

Most local-level genomic epidemiology studies published in the first year of the pandemic were conducted in HICs (91%). In contrast, six studies were considered LMIC-related, and three of those only had exposure events, not the final study analysis, occurring in the LMIC. This imbalance may reflect the limited sequencing infrastructure and bioinformatics expertise in LMICs at the start of the pandemic [55]. Additionally, the diversion of resources to urgent response activities like clinical testing and case ascertainment may have disproportionately impacted LMICs.

Common laboratory methods

The Illumina MiSeq and ONT MinION were the most frequently used sequencing instruments. This observation likely reflects prior

Table 4. Thematic analysis identified benefits of viral genomics in understanding the transmission of SARS-CoV-2

1. Confirming or refuting transmission clusters or putative epidemiological links	
	Sequencing identified a cluster of closely related cases originating from a group of newly transferred prisoners, with all subsequent cases in the prison being of the same SARS-CoV-2 lineage [20]. Phylogenetic analysis confirmed transmission events among roommates and within platoons in a marine recruit quarantine setting by identifying monophyletic transmission clusters [22]. Genomics was used to rule out five epidemiologically linked cases from an acute care hospital outbreak, based on 10–30 SNP differences from the main cluster [17]. Cases of sporadic infections within a hospital were determined not to be part of a main outbreak cluster due to the absence of two defining mutations (SNPs) [29].
2. Revealing unexpected transmission routes or sources	
	In a hospital, genomic epidemiology helped uncover previously unrecognized HCW-to-patient and HCW-to-HCW transmission [24]. Genomic data uncovered a complex picture of introductions of SARS-CoV-2 into a nursing home, challenging initial assumptions that a church service was the primary source of transmission. Instead, the analysis revealed broader community spread and several entry points into the facility, including, but not limited to, the church event [30]. In a flight-related study, genomics indicated that overseas travellers acquired SARS-CoV-2 in Australia, not in the United States, like the investigators had assumed [39]. Genomics was used to determine that a cluster of nosocomial infections unexpectedly originated from a patient being treated for a non-COVID-19 illness in a regular ward [32].
3. Distinguishing between separate introductions and local transmission	
	Sequencing of viral genomes from infected HCWs in a single clinic suggested at least five different introductions, indicating they likely became infected in the community and not in the clinic [25]. In a military barracks setting, genome sequence data indicated multiple different introductions of the virus, with only two leading to sustained transmission [34]. Phylogenetic analysis can place local cases within a global or regional context, illustrating the diverse genetic origins of the virus within an institution or a vessel [24, 37].
4. Detecting and understanding asymptomatic and presymptomatic transmission	
	High sequence similarity and close phylogenetic clustering among viral genomes, along with known contact history, supported asymptomatic transmission of secondary infections tied to a superspreading event [43]. Phylogenetic evidence and the close relatedness of viruses sequenced from symptomatic and asymptomatic individuals demonstrated possible SARS-CoV-2 transmission involving asymptomatic carriers [43]. Potential transmission from presymptomatic individuals was demonstrated when samples tested positive for SARS-CoV-2 and were nearly identical using whole-genome sequencing [15].
5. Evaluating the impacts of infection control and non-pharmaceutical interventions	
	Papers that analysed the relatedness of viruses before and after the implementation of non-pharmaceutical interventions provided insights into the effectiveness of infection prevention and control practices [16, 23, 24, 26]. In one teaching hospital, after implementing universal masking and physical distancing, no further genetic evidence of ongoing transmission from patient-to-HCW or HCW-to-HCW from the outbreak cluster was found [26].

Note; HCW, healthcare worker; SNP, single-nucleotide polymorphism.

*Graphics source: <https://www.freepik.com/icons>.

global investments in Illumina platforms for pathogen surveillance programmes (e.g. PulseNet [7]) and the active user community and relatively low price of the MinION [56].

During the early pandemic response, only a small proportion of genomic epidemiology studies (6%) employed shotgun metagenomic

sequencing to recover the SARS-CoV-2 genome. In contrast, most (81%) used tiled-amplicon whole-genome sequencing, typically following the open-source ARTIC protocol. This approach is more cost-effective and accessible, particularly for laboratories in resource-limited settings or those newly implementing viral sequencing.

Table 5. Bias assessment and adherence to reporting guidelines

Variable	Studies	
	<i>n</i>	%
Quality score (<i>n</i> = 32)		
High >70%	5	16
Moderate 50–70%	15	47
Low <50%	12	38
Mean score ^a	55%	23–92
STROME-ID reporting completeness (<i>n</i> = 32)		
High >70%	4	13
Moderate 50–70%	6	19
Low <50%	22	69
Mean score ^a	57%	25–83

Note: STROME-ID, Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases.

^aAverage quality and STROME-ID completeness scores for all 32 studies. Minimum and maximum scores.

Table 6. Sub-analysis of mean quality and STROME-ID completeness scores by country income level and extent of genomic epidemiology analysis

	HIC (<i>n</i> = 26)	LMIC (<i>n</i> = 6)	<i>p</i> -value
	Mean ± st. dev.	Mean ± st. dev.	
Quality score	56.5 ± 7.3	49.3 ± 14.5	0.31
STROME-ID	57.7 ± 4.1	55.7 ± 18.4	0.79
	GE Study ^a (<i>n</i> = 25)	Minor GE ^b (<i>n</i> = 7)	<i>p</i> -value
	Mean ± st. dev.	Mean ± st. dev.	
Quality score	51.3 ± 6.8	68.9 ± 13.1	0.02
STROME-ID	60.0 ± 4.8	48.0 ± 3.1	<0.001

Note: HIC, high-income country; LMIC, low- and middle-income country; STROME-ID, Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases.

H₀: mean difference = 0; H_A: mean difference ≠ 0.

^aGE Study means primarily a genomic epidemiology study, or genomics was a major component.

^bMinor GE means the study only had a minor genomics or genomic epidemiology component.

However, sequencing protocols were often inadequately described or referenced, limiting the ability to evaluate methodological rigor and the use of quality control measures (see [Supplementary Material 5](#)).

Evaluation of genomic epidemiology methods

SARS-CoV-2 genomic data are most useful for public health actions when analysed and visualized in the context of epidemiological data [57]. During the early pandemic, commonly used methods for integrating these data included phylogenetic trees, phylodynamic models, and haplotype and network diagrams. Among the studies reviewed, maximum-likelihood phylogenetic trees were the most frequently employed. This is likely due to the availability of open-source software, like Nextstrain, for exploring SARS-CoV-2 genomic epidemiology data, which utilizes maximum-likelihood methods for rapid tree construction (Nextstrain, <https://nextstrain.org/sars-cov-2>) [52].

However, many studies lacked clarity in their sampling frames, making it difficult to assess the representativeness of sequenced SARS-CoV-2 infections relative to the underlying population ([Supplementary Material S3-2](#)). While a few studies followed a well-defined cohort of individuals over time, most relied on contact tracing to identify potential COVID-19 cases. Whole-genome sequencing was typically performed on a fraction of the total PCR-positive samples from a study. Authors rarely reported whether these samples were representative of all SARS-CoV-2 infections in a study or the underlying sample population (See [Supplementary Material 5](#)).

Few studies examined potential sources of bias, including confounding or ascertainment bias. Factors influencing which samples were sequenced, the quality of those sequences, and which were ultimately included in analyses were often not reported. For some studies, the number of samples in the phylogenetic tree or other genomic epidemiology analyses did not match the number of successfully sequenced samples. Additionally, the choice of phylogenetic modelling approaches was not consistently described, though a few studies cited tools such as jModelTest.

Several common reasons for excluding samples or sequence data from final genomic epidemiology analyses were noted across studies. These included specimen quality, restrictions on sample sharing, and high CT values indicating low viral load. Often, no specific rationale was provided, or it remained unclear why some samples failed sequencing or were omitted from genomic epidemiology analysis. Each of these issues reduced the representativeness of SARS-CoV-2 cases and the reliability of inferences made from genomic epidemiology analyses.

The extent and quality of accompanying epidemiological metadata also varied. Some studies, typically traditional observational designs, employed comprehensive exposure and demographic questionnaires. Unfortunately, many studies collected minimal metadata or experienced substantial non-response. Even when rich epidemiological metadata were available, only sample collection dates and geographic locations were consistently used in genomic epidemiology analyses.

Study quality and reporting adherence

Overall, studies using traditional observational epidemiology methods, rather than outbreak investigation or genomic analyses, were of higher quality. Most did not provide information on how missing data were handled, whether and how confounding was addressed, and how sources of bias may have impacted the validity of overall findings ([Supplementary Material 5A](#)). These quality issues may reflect how common study appraisal tools assess risk of bias using criteria that are more easily addressed in conventional epidemiology study designs.

We also evaluated adherence to established reporting standards for molecular epidemiology studies. Low adherence to STROME-ID guidelines may be due to accelerated publication timelines during the pandemic, restrictive word limits, or lack of familiarity with these guidelines among authors and journal editors. Despite some criteria being outdated or not well-suited for outbreak-focused genomic studies, greater adherence would have improved overall methods and interpretability across publications ([Supplementary Material 5B](#)). More detailed reporting on sample selection, genomic representativeness, laboratory quality control procedures, and justification for analytic approaches would have strengthened the robustness and reproducibility of the genomic epidemiology findings included in this systematic review.

Limitations

This review assumes that the included peer-reviewed studies, authors, and institutions are broadly representative of local or outbreak-related genomic epidemiology investigations conducted during the first year of the COVID-19 pandemic. However, several limitations may have impacted generalizability or comprehensiveness:

Publication bias

Potential publication bias may be high for reasons common to all systematic reviews and reasons unique to the COVID-19 pandemic. Null or seemingly less impactful findings from local outbreak investigations may not have been prioritized for publication. Additionally, the significant demands of the pandemic response may have limited the capacity of public health practitioners to prepare manuscripts. Finally, some countries or institutions may have imposed data and sample sharing restrictions that delayed or prevented peer-reviewed publication of outbreak results altogether.

Language restrictions

Although no language restrictions were applied during the abstract screening stage, only full-text articles available in English were included in the final review. Four full-text studies (two Chinese and two Spanish) were excluded, though they failed other inclusion criteria (e.g. case studies or non-local investigations). The use of English search terms could have missed some locally published studies in LMICs, particularly China. However, ad hoc searches of a common Chinese scientific database (CNKI, <https://oversea.cnki.net/index/>) did not yield additional eligible publications in English or Chinese. To capture additional local-level investigations, future reviews could include peer-reviewed and grey literature and be conducted in Chinese, Spanish, French, or other common scientific publication languages.

Narrow inclusion criteria

The specificity of our inclusion criteria excluded many valuable broader genomic studies published early in the pandemic. This narrow focus, however, allowed for a meaningful comparison across genomic epidemiology studies conducted in similar operational contexts and enabled the synthesis of the discrete ways genomics informed local outbreak responses.

Future directions

To further inform future outbreak preparedness, we recommend qualitative studies with public health professionals to determine how SARS-CoV-2 genomic data were used in decision-making and what barriers or facilitators influenced the uptake of genomic epidemiology.

Conclusion

Our analysis revealed that genomic epidemiology enhanced outbreak investigations during the early pandemic and likely transformed local public health response [58]. Whole-genome sequencing enabled high-resolution analysis to confirm or reject suspected transmission links, uncover hidden transmission chains, and distinguish between single and multiple introductions into populations. Genomic epidemiology studies provided critical early evidence of asymptomatic and presymptomatic transmission of SARS-CoV-2, ultimately influencing the selection and implementation of important mitigation strategies [59].

As the pandemic progressed, genomic tools were rapidly adapted to track emerging variants and investigate outbreaks across a wider range of local settings. These findings highlight the importance of integrating genomics with traditional epidemiological data to strengthen infectious disease surveillance and inform effective local-level public health interventions. Notably, investments in sequencing infrastructure and analytic capacity accelerated the broader application of these methods globally [53, 54].

Our study also revealed several key factors that enabled the early implementation of genomic epidemiology in local SARS-CoV-2 public health response, specifically:

1. Open-access genomic databases and standardized genotyping schemes (e.g. Pangolin and Nextclade)
2. Community-vetted sequencing protocols (e.g. ARTIC)
3. Open-source analysis tools (e.g. Nextstrain)
4. Established outbreak investigation protocols and data collection tools
5. Strong research collaborations and emergency response funding

This systematic review provides a global synthesis of how SARS-CoV-2 genomic epidemiology was applied to local outbreak investigations during the first year of the pandemic. This body of work demonstrates how genomic methods contributed to understanding transmission dynamics and evaluating control measures. Moving forward, continued integration of genomic data into surveillance systems and genomic epidemiology practices in local public health institutions will strengthen global preparedness to respond rapidly and effectively to future infectious disease threats.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0950268826101381>.

Data availability statement. All data to support the results are included in supplemental materials, and the extracted data table will be made publicly available on figshare (<https://figshare.com/>) upon publication.

Acknowledgements. We thank the librarians at the George Washington University Himmelfarb Health Sciences Library for their assistance with selecting databases, refining search terms, and training on Covidence. We thank the authors of these studies for publishing their work in peer-reviewed journals and for their contributions to the COVID-19 pandemic response. NotebookLM (Gemini 2.0) was used to identify and code common themes from manually extracted results and/or conclusions from included studies. ChatGPT 4o was used for copyediting and reducing word count in the final manuscript to meet the journal requirements. The corresponding author reviewed and modified the output and takes full responsibility for the content presented in the manuscript. Thank you to Dr. Kathleen Creppage and Dr. Shayne Gallaway for their thoughtful comments and feedback on the manuscript drafts. The opinions expressed in this article are the author's own and do not reflect the view of the U.S. Public Health Service, the Department of Defense, the Department of Health and Human Services, or the United States Government

Author contribution. Conceptualization: L.C.M., B.M.F., and A.A.R.; Data Curation and Screening: K.K., Z.T., A.R., L.H., C.S., B.M.F., L.C.M., and A.A.R.; Formal Analysis: L.C.M.; Project Administration: L.C.M.; Supervision: A.A.R. and B.M.F.; Visualization: L.C.M.; Writing – Original Draft: L.C.M.; Writing – Review and Editing: L.C.M., A.A.R., B.M.F., Z.T., L.H., A.R., C.S.

Funding statement. No specific grant or other funding was used for this project. George Washington University Himmelfarb Health Sciences Library provided access to Covidence software and peer-reviewed article databases. Article processing fees and open access were covered through the George Washington University agreement with Cambridge University Press. L.C.M. was partially supported through the George Washington University Graduate Teaching Assistantship Program. L.C.M. was a mentee of C.S. and worked on early versions

of the project as part of the Next Generation Global Health Security Network mentorship programme.

Competing interests. The authors declare none.

Ethical standard. The systematic review did not involve human subjects research and did not require approval from an institutional review board.

References

- [1] **Heather JM and Chain B** (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8.
- [2] **Gardly JL and Loman NJ** (2018) Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics* **19**, 9–20.
- [3] **Behrmann O and Spiegel M** (2020) COVID-19: From rapid genome sequencing to fast decisions. *The Lancet Infectious Diseases* **20**, 1218.
- [4] **Pollett S, et al.** (2020) Genomic epidemiology as a public health tool to combat mosquito-borne virus outbreaks. *The Journal of Infectious Diseases* **221**, S308–S318.
- [5] **Faria NR, et al.** (2016) Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352**, 345–349.
- [6] **Quick J, et al.** (2016) Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232.
- [7] **Davedow T, et al.** (2022) PulseNet international survey on the implementation of whole genome sequencing for foodborne disease surveillance in low and middle-income countries. *Foodborne Pathogens and Disease* **19**, 332.
- [8] **Higgins J, et al.** (eds.) (2019) *Cochrane Handbook for Systematic Reviews of Interventions*, 2nd Edn. Chichester (UK): John Wiley & Sons.
- [9] **Page MJ, et al.** (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **372**, n71.
- [10] **World Health Organization** (2020) *WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020*. Available at <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020> (accessed 5 November 2024).
- [11] **National Heart, Lung, and Blood Institute** (2021) *Study Quality Assessment Tools*. Available at <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools> (accessed 3 November 2024).
- [12] **Joanna Briggs Institute** (2020) *JBICritical Appraisal Tools*. Available at <https://jbi.global/critical-appraisal-tools> (accessed 3 November 2024).
- [13] **Field N, et al.** (2014) Strengthening the reporting of molecular epidemiology for infectious diseases (STROME-ID): An extension of the STROBE statement. *The Lancet Infectious Diseases* **14**, 341–352.
- [14] **McHugh ML** (2012) Interrater reliability: The kappa statistic. *Biochemia Medica* **22**, 276–282.
- [15] **Arons MM, et al.** (2020) Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *New England Journal of Medicine* **382**, 2081–2090.
- [16] **Karmarkar E, et al.** (2021) Timely intervention and control of a novel coronavirus (COVID-19) outbreak at a large skilled nursing facility - San Francisco, California, 2020. *Infection Control and Hospital Epidemiology* **42**, 1173–1180.
- [17] **Klompas M, et al.** (2021) A SARS-CoV-2 cluster in an acute care hospital. *Annals of Internal Medicine* **174**, 794–802.
- [18] **Addetia A, et al.** (2020) Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with a high attack rate. *Journal of Clinical Microbiology* **58**, e02107–e02120.
- [19] **Firestone MJ, et al.** (2020) COVID-19 outbreak associated with a 10-day motorcycle rally in a neighboring state - Minnesota, August–September 2020. *MMWR. Morbidity and Mortality Weekly Report* **69**, 1771–1776.
- [20] **Hershow RB, et al.** (2021) Rapid spread of SARS-CoV-2 in a state prison after introduction by newly transferred incarcerated persons - Wisconsin, August 14–October 22, 2020. *MMWR. Morbidity and Mortality Weekly Report* **70**, 478–482.
- [21] **Szablewski CM, et al.** (2021) SARS-CoV-2 transmission dynamics in a sleep-away camp. *Pediatrics* **147**, e2020046524.
- [22] **Letizia AG, et al.** (2020) SARS-CoV-2 transmission among marine recruits during quarantine. *New England Journal of Medicine* **383**, 2407–2416.
- [23] **Lovestad AH, et al.** (2021) Investigation of intra-hospital SARS-CoV-2 transmission using nanopore whole-genome sequencing. *Journal of Hospital Infection* **111**, 107–116.
- [24] **Lucey M, et al.** (2021) Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks. *Clinical Infectious Diseases* **72**, e727–e735.
- [25] **Olmos C, et al.** (2021) SARS-CoV-2 infection in asymptomatic healthcare workers at a clinic in Chile. *PLoS One* **16**, e0245913.
- [26] **Paltansing S, et al.** (2021) Transmission of SARS-CoV-2 among health-care workers and patients in a teaching hospital in the Netherlands confirmed by whole-genome sequencing. *Journal of Hospital Infection* **110**, 178–183.
- [27] **Rivett L, et al.** (2020) Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *eLife* **9**, e58728.
- [28] **Selhorst P, et al.** (2021) Symptomatic SARS-CoV-2 reinfection of a health care worker in a Belgian nosocomial outbreak despite primary neutralizing antibody response. *Clinical Infectious Diseases* **73**, e2985–e2991.
- [29] **Takenouchi T, et al.** (2021) Clinical utility of SARS-CoV-2 whole genome sequencing in deciphering source of infection. *Journal of Hospital Infection* **107**, 40–44.
- [30] **Voeten HACM, et al.** (2021) Unravelling the modes of transmission of SARS-CoV-2 during a nursing home outbreak: Looking beyond the church super-spread event. *Clinical Infectious Diseases* **73** (Suppl 2), S163–S169.
- [31] **Wang X, et al.** (2020) Nosocomial outbreak of COVID-19 pneumonia in Wuhan, China. *European Respiratory Journal* **55**, 2000544.
- [32] **Weinberger T, et al.** (2021) Prospective longitudinal serosurvey of health Care Workers in the First Wave of the SARS-CoV-2 pandemic in a quaternary Care Hospital in Munich, Germany. *Clinical Infectious Diseases* **73**, e3055–e3065.
- [33] **Pirnay J-P, et al.** (2020) Study of a SARS-CoV-2 outbreak in a Belgian military education and training Center in Maradi, Niger. *Viruses* **12**, 949.
- [34] **Taylor H, et al.** (2021) Cross-sectional investigation of a COVID-19 outbreak at a London Army barracks: Neutralising antibodies and virus isolation. *Lancet Regional Health - Europe* **2**, 100015.
- [35] **Plucinski MM, et al.** (2021) Coronavirus disease 2019 (COVID-19) in Americans aboard the diamond princess cruise ship. *Clinical Infectious Diseases* **72**, e448–e457.
- [36] **Sekizuka T, et al.** (2020) SARS-CoV-2 genome analysis of Japanese travelers in Nile River cruise. *Frontiers in Microbiology* **11**, 1316.
- [37] **Sekizuka T, et al.** (2020) Haplotype networks of SARS-CoV-2 infections in the diamond princess cruise ship outbreak. *Proceedings of the National Academy of Sciences* **117**, 20198–20201.
- [38] **Eichler N, et al.** (2021) Transmission of severe acute respiratory syndrome coronavirus 2 during border quarantine and air travel, New Zealand (Aotearoa). *Emerging Infectious Diseases* **27**, 1274–1278.
- [39] **Speake H, et al.** (2020) Flight-associated transmission of severe acute respiratory syndrome coronavirus 2 corroborated by whole-genome sequencing. *Emerging Infectious Diseases* **26**, 2872–2880.
- [40] **Muller N, et al.** (2020) Severe acute respiratory syndrome coronavirus 2 outbreak related to a nightclub, Germany, 2020. *Emerging Infectious Diseases* **27**, 645–648.
- [41] **Günther T, et al.** (2020) SARS-CoV-2 outbreak investigation in a German meat processing plant. *EMBO Molecular Medicine* **12**, e13296.
- [42] **Zuckerman NS, et al.** (2020) Comprehensive analyses of SARS-CoV-2 transmission in a public health virology laboratory. *Viruses* **12**, 854.
- [43] **Chau NVV, et al.** (2021) Superspreading event of SARS-CoV-2 infection at a Bar, Ho Chi Minh City Vietnam. *Emerging Infectious Diseases* **27**, 310–314.
- [44] **Zhang J, et al.** (2021) Phylogenomic tracing of asymptomatic transmission in a COVID-19 outbreak. *The Innovation* **2**, 100099.
- [45] **Oude Munnink BB, et al.** (2020) Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nature Medicine* **26**, 1405–1410.
- [46] **Paden CR, et al.** (2020) Rapid, sensitive, full-genome sequencing of severe acute respiratory syndrome coronavirus 2. *Emerging Infectious Diseases* **26**, 2401.

- [47] **Chen C**, et al. (2020) MINERVA: A facile strategy for SARS-CoV-2 whole-genome deep sequencing of clinical samples. *Molecular Cell* **80**, 1123–1134.e[4]
- [48] **Addetia A**, et al. (2020) Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *Journal of Clinical Virology* **129**, 104523.
- [49] **Quick J** (2020) *nCoV-2019 Sequencing Protocol*. <https://doi.org/10.17504/protocols.io.bbmuik6w>.
- [50] **Grubaugh ND**, et al. (2019) An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology* **20**, 8.
- [51] **Quick J**, et al. (2017) Multiplex PCR method for MinION and Illumina sequencing of zika and other virus genomes directly from clinical samples. *Nature Protocols* **12**, 1261–1276.
- [52] **Hadfield J**, et al. (2018) Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123.
- [53] **World Health Organization** (2022) *WHO Global Genomic Surveillance Strategy for Pathogens with Pandemic and Epidemic Potential 2022–2032*. Geneva: World Health Organization. Available at <https://www.who.int/publications/i/item/9789240046979> (accessed 8 April 2025).
- [54] **Centers for Disease Control and Prevention** (2024) *CDC Announces \$90M Funding to Support New Pathogen Genomics Centers of Excellence Network*. Available at <https://www.cdc.gov/advanced-molecular-detection/php/success-stories/pgcoe-announcement.html> (accessed 8 April 2025).
- [55] **Brito AF**, et al. (2022) Global disparities in SARS-CoV-2 genomic surveillance. *Nature Communications* **13**, 7003.
- [56] **Oxford Nanopore Technologies** (2024) *Nanopore Store: MinION Starter Pack*. Available at <https://store.nanoporetech.com/us/minion-basic.html> (accessed 19 November 2024).
- [57] **National Academies of Sciences, Engineering, and Medicine** (2020) *Genomic Epidemiology Data Infrastructure Needs for SARS-CoV-2: Modernizing Pandemic Response Strategies*. Washington, DC: The National Academies Press.
- [58] **Knyazev S**, et al. (2022) Unlocking capacities of genomics for the COVID-19 response and future pandemics. *Nature Methods* **19**, 374–380.
- [59] **Qiu X**, et al. (2021) The role of asymptomatic and pre-symptomatic infection in SARS-CoV-2 transmission—A living systematic review. *Clinical Microbiology and Infection* **27**, 511–519.