GAO

United States Government Accountability Office

Science, Technology Assessment, and Analytics

Report to Congressional Requesters

TECHNOLOGY ASSESSMENT

January 2023

Pandemic Origins

Technologies and Challenges for Biological Investigations

Accessible Version

GAO-23-105406

The cover image displays examples of possible pandemic origin scenarios. These scenarios include natural origin—such as the accidental infection of one or more individuals by a pathogen transmitted from animals, including via insects or other sources such as the environment. Scenarios also include laboratory origin that begins with either the infection of individuals by a pathogen in a laboratory setting, or infections outside the laboratory caused by an accidental or intentional release of the pathogen from a laboratory.

Cover source: GAO. | GAO-23-105406

Highlights of GAO-23-105406, a report to congressional requesters

January 2023

TECHNOLOGY ASSESSMENT

Pandemic Origins

Technologies and Challenges for Biological Investigations

Why GAO did this study

Pandemics are global disease outbreaks that can greatly increase morbidity and mortality and cause significant economic and social disruptions. According to the scientific literature, most pandemics where the origin is known were caused by the natural transmission of a virus through animal-to-human contact; however, there is potential for a pandemic to originate from laboratory research.

GAO was asked to conduct a technology assessment on pandemic origins. This report describes: (1) key technologies available for pandemic origin investigations, (2) strengths and limitations of these tools and how researchers use them to investigate pandemic origins, and (3) cross-cutting challenges researchers face in trying to determine a pandemic's origin.

GAO reviewed peer-reviewed scientific literature and other documents, including reports from the Centers for Disease Control and Prevention, Office of the Director of National Intelligence, the Johns Hopkins Center for Health Security, World Health Organization, and select national laboratories; interviewed government, industry, and academic representatives; and convened a meeting of 27 experts in March 2022 with assistance from the National Academies of Sciences, Engineering, and Medicine.

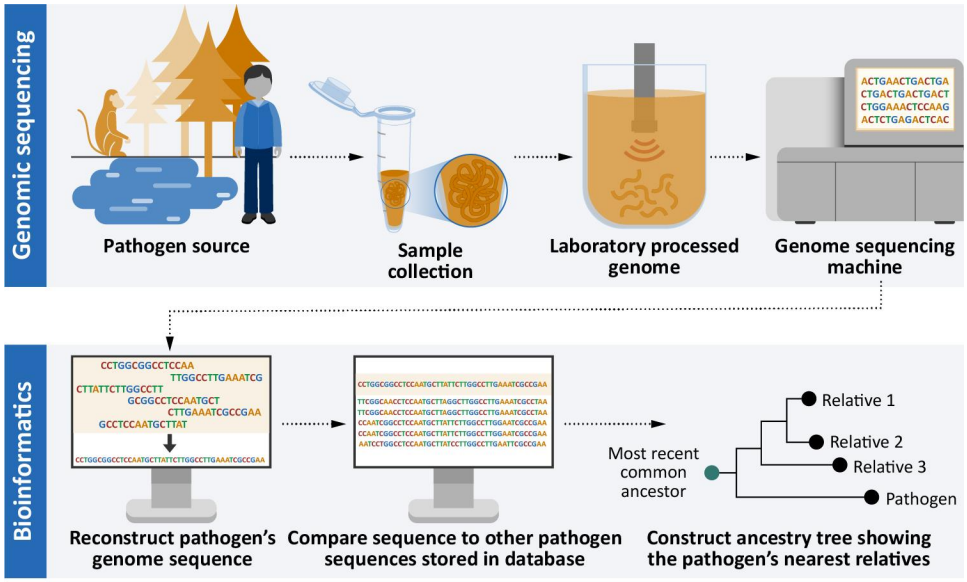
GAO is identifying policy options in this report.

View GAO-23-105406. For more information, contact Karen L. Howard at (202) 512-6888, howardk@gao.gov.

What GAO found

Determining the likely origin of pandemics is challenging. Researchers may use several technologies to investigate a pandemic's origin. For example, researchers use technologies such as genomic sequencing, bioinformatics analysis, and genetic databases to generate, analyze, and compare a pathogen's genetic makeup against that of other pathogens. A key limitation of these technologies is that some laboratory-based genetic modifications may be indistinguishable from natural variations. Access to samples is critical for conducting genetic sequence analysis, which allows researchers to generate and analyze the data needed to support the likely origin of a pandemic.

Examples of technologies for pandemic origin investigations



Source: GAO. | GAO-23-105406

Researchers also use technologies such as serology (i.e., blood analysis) and epidemiological surveillance—tracking a disease as it moves through a population—to monitor pathogen infection and disease occurrence in human and animal populations. The resulting data can support pandemic origin investigations. However, for these technologies to be effective in determining a pandemic's likely origin, investigators need access to samples and data from infected or exposed individuals from early in an outbreak to reliably trace the disease back to the first human infection(s). Further, researchers may conduct laboratory-based pathogen studies to generate data to support known natural patterns or unusual patterns of spread indicative of a possible laboratory-related origin. However, some pathogens cannot be easily cultured in a laboratory setting, and some pathogens may require enhanced biosafety-level facilities.

However, experts told GAO that technologies are not the limiting factor for determining the likely origin of a pandemic. GAO identified three cross-cutting challenges that hinder pandemic origin investigations. These include a lack of sufficient access to samples and genetic sequence data; a lack of standardized processes for submitting, accessing, and using genetic sequence data stored in databases around the world; and a lack of a sufficient and skilled interdisciplinary workforce.

GAO identified five policy options that may help address the cross-cutting challenges. These policy options represent possible actions that policymakers—who may include Congress, federal agencies, state and local governments, academia, industry, and international organizations—could consider taking. See below for a summary of the policy options and relevant opportunities and considerations.

Policy Options to Address Three Cross-Cutting Challenges in Pandemic Origin Investigations

Policy Option	Opportunities	Considerations
Establish multilateral agreements for accessing and sharing samples and genetic sequence data (report p. 21) Federal policymakers and others could encourage international preparedness in advance of future outbreaks by establishing multilateral agreements for accessing and sharing samples and genetic sequence data.	• Ensuring timely access to genetic information and samples in the critical beginning stages of a pandemic as well as throughout an origin investigation may help in the determination of a pandemic's origin. • Establishing standing agreements between nations before a pandemic occurs could assist in the determination of a pandemic's origin.	• Countries may be unwilling to participate in multilateral, international agreements because of concerns related to national sovereignty, among other reasons. • Identifying an appropriate responsible entity to determine and monitor whether countries are following agreed-upon standard processes may be time consuming and challenging.
Develop standardized processes for genetic sequence database use (report p. 22) Federal policymakers and others could empower or establish a working group to develop standardized processes for database use to support pandemic origin investigations.	• Developing standardized processes for database use could help ensure consistency of submitted data and metadata across multiple databases, improve researchers' access, and help researchers comprehensively compare genetic sequences. • Implementing leading practices for genetic data integrity and associated metadata could help improve the quality of data in genetic sequence databases.	• Standardized processes may be difficult to develop as there are risk-benefit trade-offs. For example, access to certain novel pathogen sequences should be limited to trusted and credentialed individuals with a need to access those sequences. • It may be challenging for multiple stakeholders to agree on what data are important.
Improve current, or develop new, genetic sequence database tools (report p. 23) Policymakers could encourage the improvement of current, or development of new, genetic sequence database tools.	• Improved or new database interfaces could streamline researchers' data submission, access, and use as well as improve data quality. • Improved or new database interfaces could help address the projected future growth in genetic sequence data.	• Building new, or retooling current, database interfaces could be time- and labor-intensive. • It may be challenging for groups of users to agree on what database interface features are important.
Encourage the development, retention, and growth of a workforce with the critical skills needed for pandemic origin investigations (report p. 24) Policymakers could encourage mechanisms to provide training, workforce development, and capacity-building, including in areas considered hot spots of emerging infectious disease.	• Encouraging development of expertise in geographic areas where novel pathogens are likely to emerge could increase the overall global supply of skilled workers and help to ensure the workforce is not concentrated in any geographic region. • A trained workforce skilled in origin investigations could contribute to other areas such as public health, or other types of related activities.	• Pandemic origin investigations tend to be episodic. As a result, it may be difficult to adequately plan for and consistently fund staffing in science fields related to pandemic origin investigations. • Researchers may experience unwanted attention or pressure because of their involvement in pandemic origin investigations and leave the field or refuse to participate.
Augment or develop a national strategy to better coordinate and collaborate domestically and internationally on pandemic origin investigations (report p. 25) Federal policymakers could better coordinate and collaborate with domestic and international partners by augmenting or developing a national strategy for pandemic origin investigations. This could be a standalone strategy or a component of existing strategies such as the National Biodefense Strategy.	• A national strategy could help address the challenges that hinder pandemic origin investigations. • Federal coordination and collaboration leadership, guided by a national strategy, could increase preparedness for future pandemic origin investigations. • Understanding pandemic origins could help mitigate health and economic costs associated with pandemics by, for example, facilitating surveillance that could identify future pandemics more quickly.	• Allocating resources and defining how federal agencies and others will collaborate may be challenging because of the number and types of entities with relevant expertise. • During nonpandemic periods, other priorities and needs may arise and make it challenging to provide sustained resources and support needed for maintaining a national strategy.

Source: GAO. | GAO-23-105406

Table of Contents

Figures

Abbreviations

API	application programming interface
CDC	Centers for Disease Control and Prevention
MERS	Middle East respiratory syndrome
MERS-CoV	MERS coronavirus
SARS	severe acute respiratory syndrome
SARS-CoV	SARS-associated coronavirus

U.S. GOVERNMENT ACCOUNTABILITY OFFICE
441 G St. N.W.
Washington, DC 20548

January 27, 2023

Congressional Requesters

Pandemics and epidemics—such as plague, cholera, influenza, severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and COVID-19—have afflicted humanity throughout history, causing millions of deaths and costing trillions of dollars.[1] For example, prior to a successful vaccination campaign that eradicated smallpox in 1980, the disease killed approximately 300 million people globally between 1900 and 1980.[2]

The COVID-19 pandemic has highlighted how infectious diseases continue to have a devastating impact. As of the week ending January 7, 2023, the U.S. had about 1,090,000 reported deaths attributed to COVID-19.[3] A recent assessment estimated the human and economic cost of the COVID-19 pandemic to the U.S. totaled more than $10 trillion.[4]

Given the magnitude of the health and economic costs of pandemics, policymakers—which include Congress, federal agencies, state and local governments, academic and research institutions, industry, and international organizations—have a need to better understand how and where they originate.[5] This understanding could help inform preparation and response to future epidemics and pandemics. However, determining the origin of a pathogen—a bacterium, virus, or other microorganism that can cause disease—requires evidence that may, in some cases, take decades of research to acquire. The accumulated data from these investigations may lay the foundation for future pandemic origin-tracing. For example, it took approximately 13 years to determine the origin of the SARS-associated coronavirus (SARS-CoV) pathogen that

[1]The Centers for Disease Control and Prevention (CDC) describes a pandemic as an epidemic that has spread over several countries or continents; an epidemic as an increase in the number of cases of a disease above what is normally expected in an area; and an outbreak as an epidemic, but in a more limited geographic area. However, these terms are not always consistently used for every disease. For example, while some researchers describe MERS as a pandemic, others describe it as an epidemic or outbreak.

[2]K.K. Thomas, "40 Years in a Post-Smallpox World," Johns Hopkins Bloomberg School of Public Health, May 8, 2020 (https://publichealth.jhu.edu/2020/40-years-in-a-post-smallpox-world).

[3]CDC's National Center for Health Statistics COVID-19 death counts in the U.S. are based on provisional counts from death certificate data, which do not distinguish between laboratory-confirmed and probable COVID-19 deaths. Provisional counts are incomplete because of an average delay of 2 weeks (a range of 1–8 weeks or longer) for death certificate processing. See CDC, National Center for Health Statistics, "Provisional Death Counts for Coronavirus Disease 2019 (COVID-19)," accessed January 10, 2023, https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm.

[4]R. Bruns and N. Teran, "Weighing the Cost of the Pandemic," Institute for Progress, April 21, 2022: 1-7 (https://progress.institute/weighing-the-cost-of-the-pandemic/).

[5]Determination of a pandemic's origin has some level of inherent scientific uncertainty. For this report, we use the term "origin" to mean "likely origin," acknowledging this uncertainty.

caused the 2002-2003 SARS pandemic.[6] However, the knowledge gained from those investigations helped researchers more quickly determine the origin of the MERS outbreak of 2012, according to literature we reviewed.

You asked us to conduct a technology assessment to understand how the U.S. can be better prepared to predict, prevent, detect, assess, and effectively respond to future pandemics, with a focus on determining the origins of pandemics. In this technology assessment, we describe

- key technologies available for pandemic origin investigations;
- strengths and limitations of these tools and how researchers use them to investigate pandemic origins;
- cross-cutting challenges researchers face in trying to determine a pandemic's origin; and
- policy options that may help address the cross-cutting challenges of using these key technologies to determine the likely origin of a pandemic.[7]

To address our objectives, we conducted literature searches and reviewed selected scholarly articles and other documents, including reports from the Centers for Disease Control and Prevention (CDC), Office of the Director of National Intelligence, the Johns Hopkins Center for Health Security, World Health Organization, and select national laboratories, describing technologies for pandemic pathogen characterization. Additionally, we interviewed stakeholders and experts with a diverse set of perspectives on the science and application of these technologies. This included holding an expert meeting with assistance from the National Academies of Sciences, Engineering, and Medicine. See appendix I for more information on our scope and methodology and appendix II for a list of participants in our expert meeting.

We conducted our work from August 2021 through January 2023 in accordance with all sections of GAO's Quality Assurance Framework relevant to technology assessments. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations to our work. We believe the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

[6]Initial evidences showed the civet cat to be the primary animal origin of the SARS-associated coronavirus (SARS-CoV). Later studies suggested that Chinese horseshoe bats were natural reservoirs—locations where the pathogen circulates among people and animals between outbreaks—and that the civet cat most likely served as an intermediate host. However, the study identifying the closest ancestor to SARS-CoV in a single bat colony in the Kunming, Yunnan Province in China was not published until December 2015.

[7]For the purposes of this report, the term "technologies" includes the instruments, techniques, skills, methods, and processes used in pathogen characterization.

1 Background

Pandemics are global infectious disease outbreaks that can greatly increase morbidity and mortality in people, and cause significant economic and social disruptions. According to the scientific literature, most pandemics where the origin is known were caused by the natural transmission of a virus through animal-to-human contact. Outbreaks have also been reported as a result of laboratory accidents, and research suggests the 1977-1978 H1N1 influenza pandemic may have been the result of a laboratory accident or other cause (see fig. 1).[8] Determining the likely origin of pandemics is challenging and requires information gathered from established methods for the investigation of disease outbreaks.

1.1 Natural origin

A pandemic with a natural origin scenario could initiate with the accidental infection of one or more individuals by a pathogen transmitted from animals, including via insects or other sources such as the environment. Pandemics are often the result of zoonotic pathogens being naturally transmitted between animals and humans.[9]

Zoonotic diseases can have several potential outcomes:

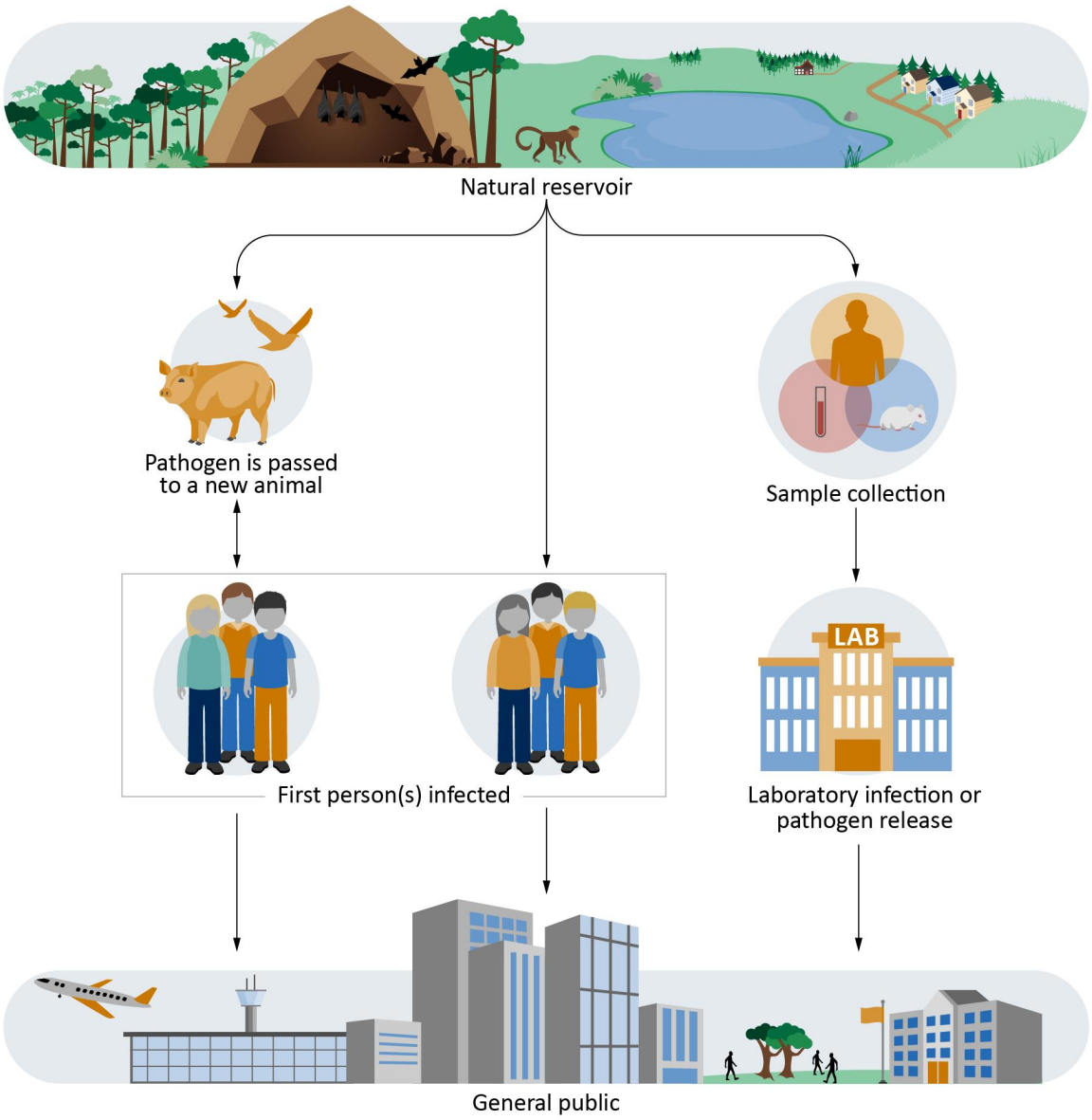
- the pathogen infects animals or humans, where it may or may not cause disease;
- the pathogen adapts so that it can be transmitted to humans without sustained human-to-human transmission, resulting in only small outbreaks among people; or
- the pathogen adapts for sustained transmission among humans, resulting in outbreaks, epidemics, pandemics, or becoming endemic in the human population.[10]

[8]Examples of known laboratory accidents involving pathogens include the unintended release of smallpox virus from a laboratory in the United Kingdom in 1978, which resulted in one death and over 300 vaccinations and surveillance of the researcher's close contacts; the accidental self-injection of the Ebola virus by a Russian scientist in 2004 that resulted in her death; and the unintended release of *Brucella* bacteria from a vaccine facility in China that began in 2019, continued in 2020, and caused over 10,000 infections. Other causes suggested for the 1977-1978 H1N1 influenza pandemic include deliberate release of the virus or a vaccine trial mishap. See M. Rozo and G.K. Gronvall, "The Reemergent 1977 H1N1 Strain and the Gain-of-Function Debate," *mBio*, vol. 6 (2015):e01013-15.

[9]Zoonotic "spillover" refers to the transmission of a pathogen from animals to humans. Zoonotic "spillback" refers to the transmission of a pathogen from humans to animals and is sometimes referred to as "reverse zoonosis."

[10]CDC describes endemic as the constant presence or the usual prevalence of a disease or infectious agent in a population within a geographic area. Adaptation of a pathogen to a new host is not an absolute requirement for transmissibility among humans.

Figure 1: Examples of pandemic origin scenarios



Source: GAO. | GAO-23-105406

Note: These pandemic origin scenarios are not meant to be exhaustive. Other scenarios may be possible. For example, researchers could be accidently infected from the environment during sample collection or during sample packaging or shipment. In the laboratory origin scenario depicted in the right column, the "first person(s) infected" may occur during sample collection, in the laboratory, or in the general public.

We identified three main factors that affect the risk of zoonotic transmission: the animals that harbor the pathogen, the nature of human interaction with those animals, and the frequency of those interactions. Scientific literature suggests that the likelihood of zoonotic disease spillover has increased in recent decades likely because of factors such

as increases in human-animal interactions through farming practices, wildlife trade, habitat loss, and climate change. These interactions facilitate the repeated exchange of pathogens between animals and humans.[11] However, most pathogens that infect humans through zoonotic transmission do not result in significant human-to-human transmission.

The exact processes by which some pathogens adapt to infect humans and then maintain long-term human-to-human transmission are not well understood, limiting our ability to quickly or definitively establish the origin of a pandemic. For example, the origins of the Ebola virus and SARS-CoV-2, which causes COVID-19, remain inconclusive. Even established, well-understood pathogens may adapt to expand beyond their typical disease geography, become more transmissible, or cause more severe disease. Although most pathogens could evolve or be manipulated in ways that may cause a human pandemic, viruses—especially RNA viruses—are the most likely to have this ability.[12]

Further, the location of the first reported human disease case—also known as the index case—might differ from where the pathogen naturally resides, making it difficult for researchers to identify a pandemic's actual origin.

1.2 Laboratory origin

A pandemic with a laboratory origin scenario could initiate with either the accidental infection of an individual or individuals by a pathogen in a laboratory setting, or infections caused by an accidental or intentional release of the pathogen from a laboratory. For example, such an infection could occur when a researcher collects a sample containing a pathogen and transfers it to a laboratory.[13] During the course of handling the pathogen, the researcher may accidently be exposed to the pathogen and become infected. Alternatively, laboratory containment may break down, resulting in the accidental release of the pathogen into the surrounding environment and infection of individuals outside the laboratory.[14] Further, some infections with a laboratory origin could involve the intentional modifications of

[11]The repeated exchange of pathogens between animals and humans is also known as "viral chatter." The frequency of viral chatter is high on farms where wild and domesticated animals are housed and bred together as well as in live animal and wet markets. Live animal and wet markets sell perishable items—such as fresh meat and produce—and sometimes live animals which are often slaughtered on-site.

[12]A. Adalja et al., "Characteristics of Microbes Most Likely to Cause Pandemics and Global Catastrophes," *Current Topics in Microbiology and Immunol*ogy, vol. 424 (2019):1-20.

[13]A sample may be obtained from human or animal sources (e.g., blood, saliva, or tissues), the environment (e.g., water, soil, or air), food, or other sources. The sample may contain the pathogen or markers—such as antibodies—indicating pathogen exposure or infection.

[14]For example, in 1979, anthrax spores were accidentally released from a facility in the Soviet city of Sverdlovsk. The cloud of spores produced a 50-kilometer trail of disease and death in animals and humans—at least 66 people died. J.W. Sahl et al., "A *Bacillus anthracis* genome sequence from the Sverdlovsk 1979 autopsy specimens" *mBio* (2016) 7(5): e01501-16.

pathogens created using techniques such as genetic engineering or serial passaging.[15]

1.3 Investigating pandemic origin

Several key technologies and approaches can help inform a pandemic's origin. Researchers typically rely on samples and data obtained from infected people, animals, and the environment. For example, researchers may collect clinical samples from infected individuals or samples from animals in or around outbreak areas such as farms or live animal or wet markets. Researchers may also collect environmental samples—such as water, soil, or insects—in or around outbreak areas. Data may consist of information about the infected individuals collected during case investigation activities—including travel history and prior contacts with other infected people—to help determine disease spread. Data may also include pathogen genetic sequence information and how the pathogen infects or transmits between hosts.[16]

Chapters 2 and 3 of this report describe the key technologies—including their strengths and limitations—used to characterize pathogens and assist in pandemic origin investigations. Chapter 4 discusses the cross-cutting challenges researchers face when investigating the origin of a pandemic.

Chapter 5 presents five policy options that may help address these challenges and improve the ability of researchers to respond more quickly and effectively to future pandemics.

[15]Genetic engineering uses laboratory-based technologies to alter the genetic makeup of a pathogen. For example, genetic engineering may involve adding a gene from one species to an organism from a different species to produce a desired trait. Serial passaging involves iteratively growing a pathogen in animals or cell cultures in a laboratory. Over time, the pathogen could acquire mutations similar to those that arise in natural environments. Cell culture involves isolating and growing animal or plant cells in a laboratory environment. Some pathogens, such as viruses, infect and replicate inside the cells. GAO has work underway examining the Department of Health and Human Services' oversight of research involving enhanced potential pandemic pathogens.

[16]A pathogen's genetic sequence—also known as the genome—comprises the order of the chemical "letters" of a pathogen's genetic material—DNA or RNA (genomes of some viruses only contain RNA). DNA and RNA contain all of the pathogen's genetic information. For the purposes of this report, the term "sequence" refers to "genetic sequence," and the term "genetic databases" refers to "genetic sequence databases."

2 Technologies for Investigating Pandemic Origin

Several key technologies can help inform a pandemic's origin. Drawing on information from experts, stakeholders, and scientific literature, we identified the following categories of such technologies:

- genetic sequence analysis;
- pathogen exposure monitoring and disease tracking; and
- laboratory-based pathogen studies.

2.1 Genetic sequence analysis

Genetic sequence analysis involves the combination of pathogen genomic sequencing, bioinformatics analysis, and genetic databases. These technologies allow researchers to generate, analyze, and compare a pathogen's genetic makeup—its sequence—against other pathogen sequences (see fig. 2).[17]

After generating the sequence of the pathogen, researchers use different bioinformatics tools to piece together and analyze the compiled sequences. While many analyses compare the sequences against those in genetic databases, other analyses can be performed independent of the databases.

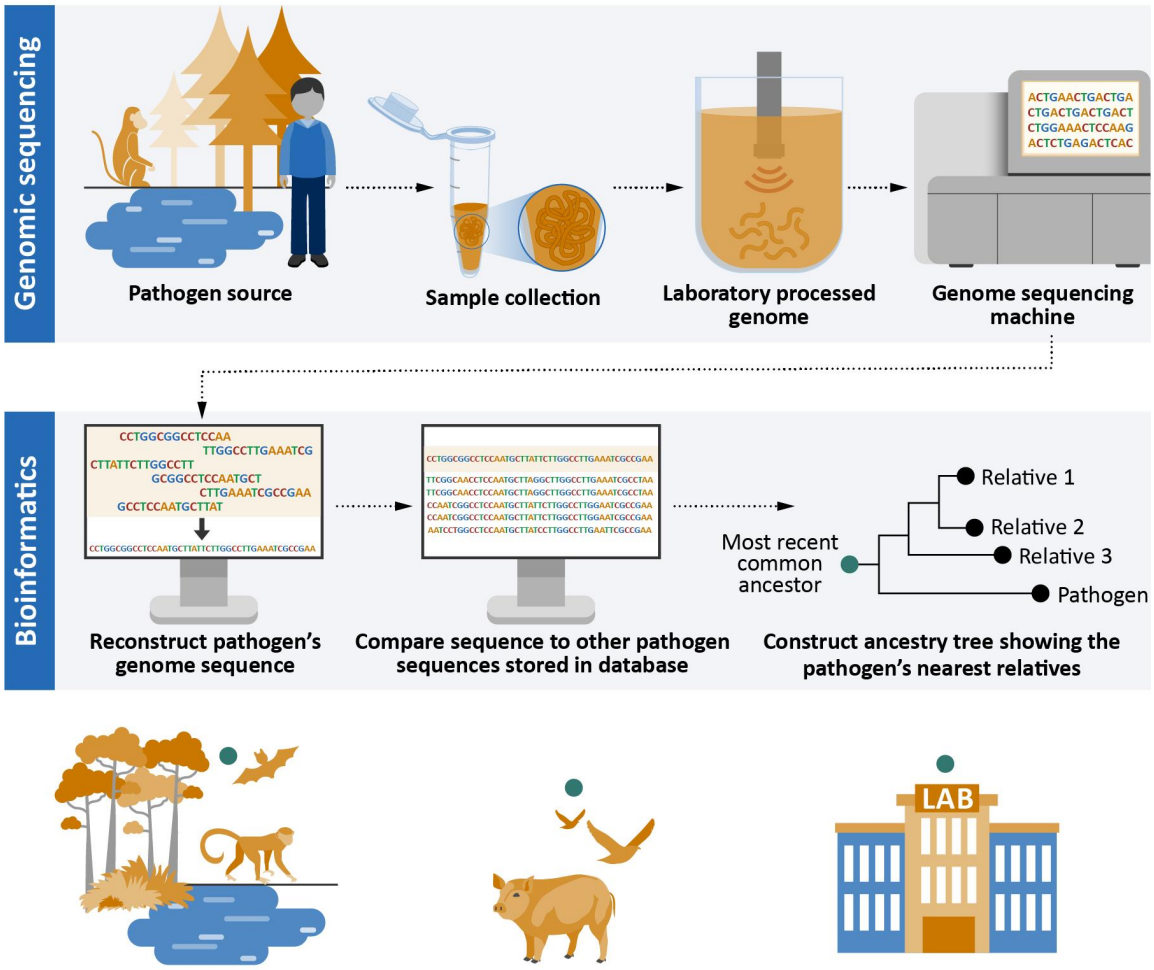
Genomic sequencing. Genomic sequencing identifies the order—or sequence—of the chemical "letters" of a pathogen's genetic material.[18] One traditional sequencing method—Sanger sequencing—copies specific segments of the pathogen's genetic material repeatedly, marks the copies with fluorescent molecules, sorts them, and then reads the individual letters.[19] Sanger sequencing produces accurate data. Reconstructing complete pathogen genomes, which are thousands to millions of letters in length, letter by letter is slow and expensive.

[17]Bioinformatics is an interdisciplinary field that uses computational algorithms for the analysis of biological data—in this case, genetic sequences.

[18]Each of the four letters—A, C, G, and T (or U in the case of RNA)—represents a chemical unit of DNA or RNA called a base.

[19]See GAO, *Science & Tech Spotlight: Genomic Sequencing of Infectious Pathogens,* GAO-21-426sp (Washington, D.C.: Mar. 30, 2021).

Figure 2: Genetic sequence analysis for pandemic origin investigations



The most recent common ancestor—denoted by the teal circle—could exist in the environment (e.g., animal, insect, water, soil), an intermediate host, or a laboratory

Source: GAO. | GAO-23-105406

Next-generation sequencing technologies can process hundreds of genomes simultaneously, enabling researchers to generate large amounts of pathogen sequence data more quickly than Sanger sequencing. Most next-generation sequencing technologies use a "massively parallel" approach to generate many short sequences of letters from different parts of the pathogen's genome at the same time. Assembling the short sequences then produces the entire sequence of the pathogen's genome.

Another next-generation sequencing technology—nanopore sequencing—uses an electrical current to thread single DNA or RNA strands through tiny pores of a membrane. As the DNA or RNA strand passes through the pore, the electrical field varies based on the specific sequence passing through the pore. By measuring and analyzing variations in the electrical field, the technology can sequence long stretches of the DNA or RNA strand.

Bioinformatics. Researchers use many types of bioinformatics tools to analyze genomic sequences. One type assembles the stretches of DNA or RNA generated by next-generation sequencing instruments to reconstruct the pathogen's genome. A second type compares the pathogen's genetic sequence to sequences stored in genetic databases.[20] Some of these tools allow researchers to analyze the structural and functional information of a gene or protein from the sequences. These tools may also identify mutations in the sequences and potential genetically-engineered sequences. A third kind of tool analyzes genetic sequences to identify likely evolutionary relationships between pathogens and their nearest relatives. This process is known as phylogenetic analysis.

Genetic databases. Researchers use genetic databases to organize the biological information gathered from many different types of pathogens.[21] Many of these genetic databases contain millions of sequences from thousands of pathogens, allowing users to compare genetic sequences of a given pathogen against many other pathogens that were previously catalogued.[22]

2.2 Infectious disease surveillance

Other tools can help researchers understand the path of the disease. The study of the presence of antibodies in the blood in response to pathogens, serology, enables the characterization and monitoring of pathogen infections in human and animal populations.[23] Serology can help establish whether a human or animal has been infected by a pathogen, sometimes long after the initial infection. Examples of technologies used for serology include biological and chemical tests.[24]

Epidemiology—the study of disease occurrence in humans and animal populations—provides information about the timing and geographic spread of the disease. Epidemiological surveillance tracks disease in populations to try to determine when and where the disease originated, among other things.[25] For example, epidemiology may help identify the source of the pathogen, its possible spread, and possible "reservoirs"

[20]Bioinformatics tools, such as the National Center for Biotechnology Information's (NCBI) Basic Local Alignment Search Tool (BLAST), identify similarities between nucleic acid or amino acid sequences. BLAST also scores the statistical degree of similarities between the sequences. Higher scores indicate a higher degree of similarity—or likely relatedness—between sequences. For more information, see S.F. Altschul et al., "Basic Local Alignment Search Tool," *Journal of Molecular Biology,* vol. 215 (1990): 403-410.

[21]This information includes DNA, RNA, and amino acid sequences from organisms collected from the environment and research conducted in laboratories. Amino acids are the fundamental building blocks of proteins.

[22]Examples of genetic databases include GenBank®, European Nucleotide Archive (ENA), DNA Data Bank of Japan (DDBJ), and Global Initiative on Sharing All Influenza Data (GISAID).

[23]An antibody is a protein component of the immune system that circulates in the blood, recognizes foreign substances like bacteria and viruses, and neutralizes them. The percentage of individuals in a population whose blood contains antibodies to a pathogen is called seroprevalence.

[24]For example, an enzyme-linked immunosorbent assay (ELISA) detects host antibodies by binding to pathogen proteins—called antigens—coated in wells on test plates. The presence or absence of these antibody-antigen complexes can then be determined using enzymes. A chemiluminescent immunoassay (CLIA) uses chemical probes that detect and label antibodies by generating light emissions (i.e., luminescence) through a chemical reaction.

[25]According to CDC, epidemiological surveillance is the ongoing and systematic collection, analysis, and interpretation of health data in the process of describing and monitoring a health event.

where the pathogen circulates among people and animals between outbreaks.

2.3 Laboratory-based pathogen studies

Laboratory-based pathogen studies examine interactions between the pathogen and the host animal or person infected with the pathogen. Such studies can reveal how pathogens infect hosts and are transmitted from one host to another. The results of these studies help researchers understand the distribution and spread—epidemiology—of the disease caused by the pathogen. Researchers also study the degree to which a pathogen can infect and transmit between hosts using animals known as *in vivo* studies, or cell cultures known as *in vitro* studies. For example, laboratory-based pathogen studies may use animals and cell cultures to determine a pathogen's transmission rate between infected and uninfected animals and cells as well as the pathogen's infectious dose.

Experts told us that other laboratory-based technologies may enable researchers to identify modifications to nucleic acids or proteins. These technologies include proteomics, the study of host and pathogen proteins; glycomics, the study of sugar molecules occurring on proteins; and epigenetics, the study of chemical modifications to host or pathogen nucleic acids—see text box for further explanation. The information gained from these technologies could help researchers in pandemic origin investigations; however, these technologies are not fully developed for such use.

Epigenetics

Researchers use epigenetics to study how behavior and the environment may cause changes in DNA and RNA that affect genes and proteins. For example, DNA and RNA may be modified through the addition of chemical groups. Typically, these chemical groups occur at specific places on the DNA and RNA. The modifications affect the ability of enzymes to "read" the DNA and RNA and produce proteins, resulting in cellular changes.

Experts and literature note that certain pathogens can cause epigenetic changes in infected people; some ongoing research is focused on detecting whether exposure to certain biological agents can be identified by examining such epigenetic changes. Further, one expert noted that it is not yet possible to detect laboratory manipulation-based epigenetic changes, but epigenetics may offer this capability for future origins investigations.

Source: GAO review of literature and the March 2022 expert meeting. | GAO-23-105406

3 Researchers Have Used a Variety of Technologies for Pandemic Origin Investigations

Researchers have used a variety of technologies for pandemic origin investigations. For example, researchers have generated pathogen sequence data using genomic sequencing, then used bioinformatics tools to analyze and compare the sequence to reference sequences stored in genetic databases. Three outcomes can result from these comparisons:

- If a pathogen's sequence matches sequences from naturally-occurring organisms, this could provide support for a natural origin. Further, phylogenetic analyses may be conducted to identify the pathogen's closest relatives or its most recent common ancestor.[26]
- If a pathogen's sequence, or parts of its sequence, matches known, laboratory-generated sequences, this could provide support that a pathogen may have a laboratory origin.
- If a pathogen's sequence does not closely match any sequences in the genetic databases, this could indicate a novel pathogen. This could also indicate the genetic databases lack the diversity of sequences needed to accurately compare the pathogen's sequence.

Other approaches, such as serology, epidemiology, and laboratory-based pathogen studies, have also been used to support such pandemic origin investigations. However, multiple lines of evidence are often needed to establish a pandemic's likely origin. Further, experts told us technologies are not the limiting factor for investigating the likely origin of a pandemic.

3.1 Researchers used genetic sequence analysis to determine the likely origin of several pandemics

Researchers used genetic sequence analysis to help establish the likely natural origins of several pandemics and outbreaks, including the 2002-2003 SARS pandemic, the 2009 H1N1 influenza pandemic, and the initial MERS outbreak in 2012.[27] Researchers also

[26]The most recent common ancestor of any set of individuals—such as viruses—is the most recent individual virus from which all of the other individual viruses in the group are directly descended. This definition is adapted from the International Society of Genetic Genealogy.

[27]Genetic sequence analysis of samples from civet cats and a raccoon dog from a live animal market showed that the animal SARS-CoV strains were 99.8 percent identical to the SARS-CoV strains isolated from infected humans. See L.-F. Wang and B.T. Eaton, "Bats, civets and the emergence of SARS," *Current Topics in Microbiology and Immunology*, vol. 315 (2007):325-344. Genetic sequence analysis also showed that MERS-CoV strains isolated from camels were almost identical to those isolated from humans and were phylogenetically related to bat coronaviruses. See J. Cui et al., "Origin and evolution of pathogenic coronaviruses," *Nature Reviews Microbiology*, vol. 17 (2019): 181-192. Genetic sequence analysis of samples from humans and pigs established the origin of the H1N1 influenza virus in central Mexico, where it jumped from pigs to humans. See I. Mena et al., "Origins of the 2009 H1N1 influenza pandemic in swine in Mexico," *eLife* (2016) 10.7554/eLife.16777.

used phylogenetic analysis to trace the transmission of HIV-1 from Africa to Haiti, followed by its subsequent transmission from Haiti to North American populations around the 1960s. Researchers continue to use genetic sequence analysis to investigate the origin of other pandemics, including the COVID-19 pandemic caused by SARS-CoV-2.[28]

The increasing speed and accuracy and decreasing cost of genomic sequencing technologies, such as next-generation sequencing, allow researchers to simultaneously process hundreds of pathogen genomes. Researchers are thus able to quickly generate pathogen sequence data necessary for investigating potential origin**.** Experts told us that because of these strengths, they consider genomic sequencing a key technology for pandemic origin investigations.

A key limitation of genetic sequence analysis is that some laboratory-based genetic modifications may be indistinguishable from natural variations. For example:

- Some traditional genetic engineering techniques and newer genome editing tools—such as CRISPR-Cas9—may leave no detectable trace of genetic modification.[29] Some bioinformatics tools that use artificial intelligence (AI) may help researchers detect patterns indicative of genome editing.[30] However, these are currently limited by a lack of large sequence datasets on which to train the algorithms.
- One agency official described a 2011 large foodborne outbreak in Germany that was caused by a strain of *Escherichia coli* (*E. coli*) bacteria. Genetic sequence analysis showed the strain contained genetic sequences from two strains of *E. coli*. This unusual genetic makeup potentially supported a laboratory origin. However, researchers later determined, through additional research, that a natural origin was more likely.
- Sequence changes (i.e., mutations) resulting from laboratory adaptation experiments—such as serial passaging—may be more difficult to detect than genome editing because the laboratory adaptation more closely mimics aspects of natural processes of evolution. For example, some researchers argue that serial passaging may explain certain features of the SARS-CoV-2 genome, while others argue that a zoonotic origin is the more likely explanation for those features.[31]

Some phylogenetics software tools are limited in their utility for assessing pathogen origins because of technical limitations of the

[28]J.E. Pekar et al., "The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2," *Science* (2022) 10.1126/science.abp8337; M. Worobey et al., "The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic," *Science* (2022) 10.1126/science.abp8715.

[29]Clustered Regularly Interspaced Palindromic Repeats (CRISPR)-associated protein number 9 (Cas9) is one type of genome editing technology that allows scientists to precisely modify a pathogen's genome, potentially leading to changes in a pathogen's characteristics.

[30]E.C. Alley et al., "A machine learning toolkit for genetic engineering attribution to facilitate biosecurity," *Nature Communications* (2020) 10.1038/s41467-020-19612-0.

[31]K.G. Andersen et al., "The proximal origin of SARS-CoV-2," *Nature Medicine*, vol. 26 (2020): 450-455.

analysis programs and deficiencies in databases used for sequence comparisons. For example, some phylogenetic tools use a certain pattern of pathogen evolution from other organisms when comparing sequences. However, many pathogens do not follow the types of evolutionary patterns that other organisms follow. As a result, conclusions based on the use of these tools should be confirmed with other methods. More recently, network-based approaches have been used to reconstruct virus evolution more realistically.

Additionally, some phylogenetic tools are not capable of analyzing the millions of sequences currently being generated. For example, one expert told us that the volume and complexity of SARS-CoV-2 data crashed a commonly used phylogenetics program. The lack of reference sequences and metadata in databases also impacts researchers' ability to conduct meaningful phylogenetic analyses.[32]

Further, multiple experts told us that it can be problematic when databases have sequences overrepresented by specific countries. For example, the SARS-CoV-2 sequences in the Global Initiative on Sharing All Influenza Data (GISAID) database are dominated by data from the U.S. and U.K., whereas data from relevant locations elsewhere in the world are scarcer. This underrepresentation negatively affects the ability to determine where a pathogen may have originated.

3.2 Researchers used serology and epidemiological surveillance for pandemic origin investigations

Researchers have also used serology and epidemiological surveillance to monitor pathogen infection and disease occurrence in human and animal populations to support pandemic origin investigations. Serology and epidemiological surveillance can provide information regarding the timing and geographic spread of the pathogen and disease. For example, if serology studies detect antibodies in animal populations near a suspected disease outbreak in humans where the disease is not normally present or expected, this could lend support to a natural origin. Further, epidemiological surveillance can be used to generate models to predict how a pathogen spreads. These models can also be run in reverse to trace the spread of the disease back to the early stages of a pandemic. However, for serology and epidemiological surveillance to be effective in determining a pandemic's origin, investigators need access to samples and data from infected or exposed individuals from early in an outbreak and as close to index cases as possible to reliably trace the disease back to the first human infection(s).

Serology surveillance in people and camels provided two key pieces of information that contributed to the determination that camels were direct sources of human infection with MERS-CoV. First, researchers detected MERS-CoV antibodies from archived camel blood samples dating back to 1983. Second,

[32]In this report, we refer to information about genetic sequences, such as when and where a sample was collected, as metadata.

serology surveillance showed a higher prevalence of MERS-CoV antibodies in humans exposed to camels relative to the general population. Together with other studies, this information led researchers to conclude that MERS-CoV was likely transmitted to people from camels.

Epidemiological studies of the first SARS cases in Guangdong Province, China in 2002-2003 suggested a zoonotic origin of the virus. For example, several of the early cases were associated with occupations that involved contact with wildlife, including handling, killing, and selling wild animals as well as preparing and serving wildlife animal meat in restaurants. Subsequent serology surveillance found a higher than normal seroprevalence of SARS-CoV antibodies among wild animal traders as compared to vegetable traders from the same Guangdong market. Further, serology surveillance of animal traders in three different live animal markets found that 13 percent had SARS-CoV antibodies, whereas 72 percent of traders of civet cats had SARS-CoV antibodies.[33]

Researchers also used epidemiological data, among other types of data, to investigate the hypothesis that the COVID-19 epidemic in Wuhan began at the Huanan market. Based on the geographic and timing patterns of reported cases within the city and the specific locations of cases within the Huanan market, recent studies assessed that this market was "an early and major epicenter" of COVID-19 emergence.[34] However, researchers and agency analysts reported that uncertainty still exists about where the first SARS-CoV-2 infections occurred because of a lack of clinical samples available for serological and genetic analyses as well as a lack of epidemiological data from the earliest cases.[35]

Serology and epidemiological surveillance may be limited by the ability to collect and analyze samples from infected humans and animal populations. For example, certain countries may refuse or limit researchers' access to field sites, facilities, data, or people. Further, researchers conducting field-based sample collections may encounter logistical and operational barriers to accessing remote field sites, including personal protective equipment constraints.[36] Sensitive and specific serology tests may also take time to develop and validate.

Researchers may also face technical challenges for collecting, preserving, and transporting samples. For example, many viruses, such as SARS-CoV-2, only contain RNA, which is less chemically stable than DNA, and may require specialized preservatives. Samples may also require cold storage and shipment—known as cold chains—to maintain their integrity. In remote parts of the world, cold chain infrastructure may be lacking. Further, samples collected

[33]L.-F. Wang and B.T. Eaton, "Bats, Civets and the Emergence of SARS," *Current Topics in Microbiology and Immunology,* vol. 315 (2007): 325–344.

[34]E.C. Holmes et al., "The Origins of SARS-CoV-2: A Critical Review," *Cell,* vol. 184 (2021): ep. 1-9.

[35]Office of the Director of National Intelligence, National Intelligence Council, "Updated Assessment on COVID-19 Origins" (2021): ep. 1-18.

[36]Collecting animal samples can be dangerous both to the individual researchers collecting the samples as well as the public. To collect samples, researchers typically need to make personal contact with animals. One expert told us about a project that uses drones or robots to collect guano samples from bat caves, mitigating the possibility of researchers contracting viruses by eliminating the need to enter the caves themselves.

from humans or animals have high amounts of host genetic material, making it difficult or more time-consuming to extract, isolate, and analyze a pathogen's genome.

Finally, even comprehensive field-based sampling aimed at investigating the origins of pandemic pathogens may be inconclusive. For example, researchers recently reported a sampling effort in China aimed at tracing the origin of two pandemic pathogens, SARS-CoV and SARS-CoV-2.[37] Despite generating a database of over 17,500 animal samples, researchers did not find any closely related coronaviruses.

3.3 Researchers used laboratory-based pathogen studies for pandemic origin investigations

Laboratory-based pathogen studies using cell cultures or animals have generated information about a pathogen's ability to infect, mutate, adapt to, and spread between hosts. Results from these laboratory studies provided evidence supporting known natural patterns of spread or unusual patterns of spread indicative of a possible laboratory-related origin. For example, researchers studying pandemic H1N1 influenza virus in ferrets identified the viral genes, proteins of transmission, and host receptor sites that drive different routes of transmission.[38] The results of these studies supported the conclusion that this virus likely originated from animal-to-human transmission.

Several cell culture and animal studies have also been used for studying SARS-CoV-2 infection and spread. For example, researchers used cell cultures to isolate and study the virus samples from some of the first COVID-19 patients and to identify host factors required for SARS-CoV-2 replication. Researchers also used cell cultures to study genetic changes in the virus during serial passaging, including confirming the ability of the virus to adapt quickly to the host. Further, researchers used different animal studies to determine the ability of the virus to transfer to and infect healthy animals, which may provide evidence for the virus reservoir and intermediate hosts.

Laboratory-based pathogen studies are useful for studying pathogen biology under highly controlled conditions. Cell culture studies and animal studies each have strengths. Cell culture studies comply with the ethical desire for reducing the use of animals, and they are less expensive, faster, and allow for the study of specific pathogen-host targets, which could not be assessed in humans or animals. Animal studies help researchers better understand pathogen infection and transmission, and they have the potential to elucidate the natural history of the disease.

Key limitations of laboratory-based pathogen studies are that some pathogens cannot be easily cultured in a laboratory setting, and some pathogens require enhanced biosafety-level facilities. Results from controlled laboratory transmission studies also may not accurately represent the natural environment, making it difficult for

[37]Z. Wu et al., "A Comprehensive Survey of Bat Sarbecoviruses across China for the Origin Tracing of SARS-CoV and SARS-CoV-2," *Research Square* (2021): ep. 1-37.

[38]J.S. Long et al., "Host and viral determinants of influenza A virus species specificity," *Nature Reviews Microbiology* (2019) 10.1038/s41579-018-0115-z.

researchers to clearly distinguish between natural versus laboratory-controlled transmission patterns. For example, cell culture studies do not resemble the complexity of a human or animal host, and translating cell culture-generated data to animal models can be particularly challenging. Further, animal studies are costly and raise ethical concerns.

4 Researchers Face Three Key Challenges When Investigating Pandemic Origin

In addition to the specific technology limitations discussed earlier, researchers also encounter three challenges at various stages in the pandemic origin investigation process, according to experts. Specifically,

- Lack of sufficient access to samples and genetic sequence data,
- Lack of standardized processes for submitting, accessing, and using genetic sequence data stored in databases around the world, and
- Lack of a sufficient and skilled interdisciplinary workforce.[39]

4.1 Researchers lack sufficient access to critical samples and data

We found that access to samples from index cases and other primary and secondary cases or genetic sequence data derived from those samples may be restricted in two broad ways.

- Local concerns may limit access to samples and data. For example, primary care physicians may not collaborate with public health officials. Therefore, data from medical testing and patient care may not be available for pathogen surveillance. Privacy concerns, general mistrust, perceived infringements on a country's sovereignty, or fear of negative consequences can also result in restricted access.
- Even if researchers have access to samples and data, their ability to extract suitable information may be limited by a lack of standardized processes. For example, health officials may collect samples for a purpose other than pathogen surveillance or store and process the data obtained from the samples in a way that precludes investigations into the origin of the pandemic. Further, no one entity is responsible for determining and enforcing standardized processes.

Experts told us that multilateral agreements on sample and data sharing are necessary because pandemics can originate from anywhere and rapidly spread internationally. They also said that negotiating or modifying agreements each time a pandemic occurs is not effective.

[39]Sufficient and prompt access to initial outbreak samples enables actions to prevent current disease spread (e.g., via travel restrictions, testing programs, vaccine development). However, for pandemic origin investigations, which may occur months or years after the initial outbreak, sufficient and timely access to such samples is important to maximize the chances of a reliable result.

4.2 Lack of standardized processes for genetic sequence databases prevents researchers from analyzing data effectively

Some genetic sequence databases used by researchers may lack standardized processes for data submission, access, and use. To investigate the origin of a pandemic, researchers need access to genetic sequence data, which may be stored in multiple databases, such as the National Center for Biotechnology Information's (NCBI) GenBank®, GISAID, and the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI).[40] Experts cited three main challenges to working across multiple databases:

- Each genetic sequence database may have different processes for submitting, accessing, and using the data. GenBank, which is one of the most widely used databases, is open access, places no restrictions on the distribution of data, and provides multiple submission tools depending on the type of sequence data to be submitted. GISAID, on the other hand, requires personal access credentials, prohibits any re-distribution of data, and provides a web portal for submissions. As a result, gathering all of the data necessary to investigate the origin of a pandemic can be challenging.
- Genetic sequence databases generally lack standardized user interfaces for data submission and access, and some existing user interfaces can be cumbersome. For example, experts told us that submission processes for some major genetic sequences databases are not user-friendly, and previous submissions can be difficult to edit.[41] Similarly, interfaces for accessing data are not standardized. For example, some major databases lack application programming interfaces (API) that would provide access to the data from other applications.[42] Because researchers lack standardized submission and access interfaces, they may have to use different procedures to submit and retrieve needed data from relevant databases, which can be time-consuming and inefficient.
- Metadata are crucial for investigating the origin of a pathogen, but their availability and quality may vary. For example, GenBank's submission process allows researchers to submit information in distinct metadata fields with few constraints on content. One record that we examined lists "Japan" as the country where the sample was collected and "2020-07" as the collection date. Another

[40]GenBank is part of the International Nucleotide Sequence Database Collaboration, which includes the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank. These three databases exchange data on a daily basis.

[41]For example, experts told us that GenBank allows only the original author to edit a submission. This could be problematic if an error to the record exists and the original author is no longer active in research. In this case, the error may become permanent. However, the National Institutes of Health noted a record cannot be publicly released in GenBank until it has a valid scientific classification. Further, if an organism's valid scientific classification is revised by an international standards group, then the record can be updated accordingly without requiring a submitter request.

[42]An application programming interface (API) enables machine-to-machine communication, allowing users to obtain real-time data updates. GAO, *Open Data: Treasury Could Better Align USAspending.gov with Key Practices and Search Requirements*, GAO-19-72 (Washington, D.C.: Dec. 13, 2018).

record of a different genetic sequence lists a more specific location, "Canada: Toronto," as the country where the sample was collected, but no collection date. Although GenBank allows users to report the latitude and longitude where samples were collected, a 2017 study estimated that 99 percent of records do not include that information.[43]

These challenges may be exacerbated by the immense scale and continued growth of genetic sequence data. (See text box for a prediction on the future growth of genomic data.) As the amount of data in each database grows, and as more databases are added, standardized processes are crucial to ensure that researchers can compile, analyze, and share all the genetic sequence data necessary to investigate the origin of a pandemic. However, it is unclear if the existing infrastructure of multiple independent databases worldwide can support the growth of genomic data.

Rapid growth of big data

A 2015 study predicted that, by 2025, genomics research worldwide will generate between 2 and 40 exabytes of data annually. (For reference, 1 exabyte equals 1 billion gigabytes.) This would make genomics one of the most challenging domains of Big Data in terms of data acquisition, storage, distribution, and analysis.

Accommodating the expected growth of genomic data will require advancements in computational speed and power, as well as algorithms optimized for Big Data.

Source: GAO review of literature. | GAO-23-105406

4.3 The global research community lacks a sufficient and skilled interdisciplinary workforce

Pandemic origin investigations require a highly skilled workforce with expertise in multiple fields. We identified four main challenges to developing and retaining such a workforce based on information we gathered from experts and literature:

- Demand for workers in relevant fields tends to increase when pandemics occur and decrease when pandemics end. Likewise, funding for relevant research tends to fluctuate. This makes it challenging to keep the workforce "warm" (i.e., available and proficient) to conduct investigations promptly when pandemics occur.
- Pandemic origin investigations require expertise in multiple fields such as biology, virology, microbiology, immunology, epidemiology, ecology, genomics, bioinformatics, and computer science. However, the current workforce is siloed because of factors such as academic structures, funding priorities, and grant processes, according to experts we interviewed.[44] This makes it challenging to build and maintain the multidisciplinary workforce necessary to conduct investigations.
- The current uneven global distribution of the workforce leads to political and

[43]T. Tahsin et al., "Named Entity Linking of Geospatial and Host Metadata in GenBank for Advancing Biomedical Research," *Database* (2017): https://doi.org/10.1093/database/bax093. National Institutes of Health officials told us they have since made concerted efforts to increase collection and harmonization of sample collection location and date information. They also noted that in some cases, such data may be unavailable due to privacy or ethical concerns.

[44]The term "academic structure" is defined as the components of academic institutions and how they relate to each other. Components include academic careers, departments, plans, and subplans.

logistical challenges during a pandemic. For example, a 2021 study concluded that inadequate sequencing capacity because of limited skillsets, among other factors, hindered biosurveillance during the COVID-19 pandemic.[45]

- Some researchers told us that they faced criticism because of their involvement in investigating the origin of a pandemic, particularly when their conclusions were considered controversial. These researchers said they and others may be reluctant to participate in further investigations because of personal and professional risks.

We found that a national strategy could help to address these challenges. National strategies are "whole of nation" efforts that frequently include international components. They may be part of a structure of overlapping or supporting national strategies and typically involve sectors, organizations, entities, and resources outside the control of the federal government.[46]

[45]M. Dzobo et al., "Inadequate SARS-CoV-2 Genetic Sequencing Capacity in Zimbabwe: A Call to Urgently Address this Key Gap to Control Current and Future Waves," *IJID Regions*, vol. 1 (2021): ep. 3-4. https://doi.org/10.1016/j.ijregi.2021.09.004.

[46]See GAO, *Combating Terrorism: Evaluation of Selected Characteristics in National Strategies Related to Terrorism*, GAO-04-408T (Washington, D.C.: Feb. 3, 2004).

5 Selected Policy Options to Help Address Three Cross-Cutting Key Challenges for Investigating Pandemic Origin

Chapter 4 described three cross-cutting challenges that hinder researchers trying to investigate the origin of a pandemic:

- Lack of sufficient access to samples and genetic sequence data,
- Lack of standardized processes for genetic databases, and
- Lack of a sufficient and skilled interdisciplinary workforce.

GAO identified five policy options that may help address these challenges. These policy options are not mutually exclusive and represent possible actions that policymakers—who may include Congress, federal agencies, state and local governments, academic and research institutions, industry, and international organizations—could consider taking. Addressing the three broad challenges with these policy options could also help improve the ability of researchers to respond more quickly and effectively to potential future pandemics.

Policy Option: Federal policymakers and others could encourage international preparedness in advance of future outbreaks by supporting the development of multilateral agreements for accessing and sharing samples and genetic sequence data.

Challenge Addressed: Access to samples and genetic sequence data

Federal policymakers and others could help establish comprehensive multilateral, international agreements for accessing and sharing genetic sequence samples and data in advance of future outbreaks. These proactive agreements could include definitions of the roles and responsibilities of international investigation teams and incentives for adherence, helping ensure more timely access to critical information.

Potential implementation approaches

- Develop multilateral sample and data-sharing agreements—for example, to include expectations of timely access to samples and detailed standards for sample collection, sample storage, and metadata that countries will supply—as an objective in national pandemic origin investigation strategies.
- Work with international health organizations, such as the World Health Organization, to identify and address barriers to establishing multilateral, international agreements for ensuring access to genetic sequence samples and data, and support the development of such agreements.
- Seek agreement with stakeholders on incentives for participation, such as equitable access to vaccines and therapeutics. These incentives could also include economic assistance and assurances to mitigate stigmatization when promptly sharing samples and genetic sequence data.

Opportunities

- Ensuring timely access to genetic information and samples in the critical beginning stages of a pandemic as well as throughout an origin investigation may help in the determination of a pandemic's origin.
- Establishing standing agreements between nations before a pandemic occurs could assist in determination of a pandemic's origin.
- Incentives may help encourage reluctant countries to participate.

Considerations

- Countries may be unwilling to participate in such multilateral, international agreements because of concerns related to national sovereignty.
- Identifying an appropriate responsible entity to determine and monitor whether countries are following agreed-upon standard processes and their implementation may be time-consuming and challenging.[47]

Policy Option: Federal policymakers and others could empower or establish a working group to develop standardized processes for database use to support pandemic origin investigations.

Challenge Addressed: Lack of standardized processes for data submission, access, and use

A working group could develop standardized processes for submission of and access to data in databases such as GenBank.[48] Standardized processes could help ensure that all users submit and access the same kinds of data used for pandemic origin investigations.

Potential implementation approach

Federal policymakers and others—such as state and local policymakers, current database providers, developers, and users—could collaborate to identify and develop standardized processes for using genetic sequence databases. This could include updating documentation processes—such as clear instructions for types of sample metadata—for using GenBank and other databases and encouraging those database providers to implement these standardized processes.

[47]For example, it took 6 years for the Secretariat of the Convention on Biological Diversity's *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity* (an international agreement which aims at sharing the benefits arising from the use of genetic resources in a fair and equitable way) to develop and implement the agreement. However, the protocol still lacks a strong plan for compliance. The U.S. is not a signatory to the Nagoya Protocol or the Convention on Biological Diversity.

[48]Other databases may be operated by other countries or nongovernmental organizations.

Opportunities

- Developing standardized processes for databases could help ensure consistency of submitted data and metadata across multiple databases, improve researchers' access, and help researchers comprehensively compare genetic sequences. For example, standardized processes for recording geographic details of sample collections could help researchers who use the database examine information to better understand where a pathogen resides naturally.

- Implementing leading practices for genetic data integrity and associated metadata could help improve the quality of data in genetic sequence databases. For example, as discussed previously, we heard from researchers that some databases would only allow the researcher who entered a genetic sequence to change any of that information or to delete the sequence. Database governance practices that give database administrators a greater role in performing quality control could help ensure more data can be used to comprehensively compare genetic sequences to determine a pathogen's evolutionary ancestry and origin.

Considerations

- Standardized processes may be difficult to develop as there are risk-benefit trade-offs. For example, it is critical that access to certain novel pathogen sequences in databases be limited to trusted and credentialed individuals with a need to access those sequences. The working group would therefore need to balance the security of the databases with ensuring that researchers can access novel pathogen sequences, as needed, for critical work.
- Universities and industry researchers may have existing policies governing metadata to ensure privacy. For example, the benefits of including specific geographic information with biological samples must be weighed against any privacy concerns of the people and communities from which those samples were collected.
- It may be challenging for multiple stakeholders to agree on what data are important. For example, stakeholders may have different perspectives on what metadata should be required versus optional.

Policy Option: Policymakers could encourage the improvement of current, or development of new, genetic sequence database tools.

Challenge Addressed: Lack of standard user and application programming interfaces

Improving current genetic sequence database tools or developing new ones may help investigators determine a pandemic's origin more effectively. For example, redesigning current or creating new database user interfaces or APIs could help researchers perform genetic sequence comparisons more efficiently and aid in phylogenetic analyses.

Potential implementation approaches

- Policymakers could encourage improvements to sequence database tools—such as user interfaces or APIs—of current databases.
- Policymakers could incentivize—for example, via funding—the creation of new database user interfaces or APIs.

Opportunities

- Improved or new database user interfaces and APIs—as agreed upon by groups of end users and in conjunction with standard processes— could, for example, streamline researchers' data submission, access, and use and improve data quality.
- Improved or new database user interfaces and APIs could assist in addressing the projected future growth in genetic sequence data by, for example, enabling the analysis of large datasets stored in distributed cloud-based systems.[49]

Considerations

- Building new, or retooling current, database user interfaces and APIs could be time- and labor- intensive.
- It may be challenging for groups of users to agree on what database user interfaces and APIs features are important. For example, users may have different opinions on what is important to include in the user interfaces to make the databases more user-friendly or what applications need to communicate with the databases.

Policy Option: Policymakers could incentivize the development, retention, and growth of a workforce with the critical skills needed to conduct or support the work of characterizing the likely origin of a pandemic.

Challenge Addressed: Lack of a sufficient and skilled interdisciplinary workforce

Incentivizing the development of the workforce could increase the availability of skilled workers by creating international partnerships, among other things, and leveraging or creating training programs to encourage workforce growth and retention.

Potential implementation approaches

- Policymakers could encourage mechanisms to provide training, workforce development, and capacity building, including in areas considered hot spots of emerging infectious disease. Focusing on recruitment and consistent investment in global as well as domestic programs may increase the available workforce by increasing the number of skilled workers and retaining those workers.

[49]Additional technological needs to address the future growth in genetic sequence data may include data centers with fast, tiered storage systems, improved algorithms, data streaming approaches, and large-scale machine learning systems.

- Policymakers could leverage or enhance existing programs to provide incentives for students and research professionals to pursue careers in fields with skills necessary for pandemic origin investigations.

Opportunities

- Encouraging development of expertise in geographic areas where novel pathogens are likely to emerge would not only increase the overall global supply of skilled workers but also help to ensure the workforce is not concentrated in any one particular geographic region.
- Increased and improved educational initiatives could foster a generation of students and professionals with the multidisciplinary qualifications and skills needed to support pandemic origin investigations. For example, the National Science Foundation currently invests in numerous graduate student educational activities through a program that provides activities and training opportunities to augment students' research assistantships with non-academic research internships. Policymakers could continue to leverage or expand these types of programs by, for example, encouraging investment in multidisciplinary scientific fields that may support pandemic origin investigations.
- A sufficient and trained workforce skilled in origin investigations could contribute to other areas such as public health, biotechnology, infectious diseases, or other types of related biological research and development.

Considerations

- Pandemic origin investigations tend to be episodic and irregular. As a result, it may be difficult to adequately plan for and consistently fund staffing in science fields related to pandemic investigations.
- The scientific community may resist any alteration to current academic structures, and it may be challenging to adapt priorities, processes, and funding in a sufficiently timely manner needed to respond to a pandemic. As a result, attracting qualified people into the necessary workforce fields may be challenging if those fields are marginalized and underfunded.
- Researchers may experience unwanted attention, pressure, harassment, or influence because of their involvement in pandemic origin investigations. As a result, increasing the size of the workforce may not lead to sustained expertise if experienced researchers leave the field or refuse to participate in pandemic origin investigations.

Policy Option: Federal policymakers could augment or develop a national strategy to better coordinate and collaborate domestically and internationally on pandemic origin investigations.

Challenges Addressed: All

The 2022 National Biodefense Strategy and Implementation Plan may assist in addressing the cross-cutting challenges we identified. For example, the 2022 Strategy includes an Early Warning priority area that encompasses targets and corresponding

actions related to determining the origin of biological events, including infectious disease outbreaks.[50] However, the 2022 Strategy does not specifically outline how the lead and support departments and agencies will coordinate and collaborate to address origin determination. Augmenting the 2022 Strategy or developing a separate strategy with these specifics could better position the nation to play a leading role in pandemic origin investigations.

Potential implementation approaches

- Federal policymakers could augment the National Biodefense Strategy to specify how lead and support departments and agencies will coordinate and collaborate with domestic and international partners to address pandemic origin investigations.
- Federal policymakers could develop a new, standalone, national strategy focused on pandemic origin investigations that describes how federal entities will coordinate and collaborate with domestic and international partners on such investigations.

Opportunities

- A national strategy could help address the challenges that hinder pandemic origin investigations.
- Federal coordination and collaboration leadership, guided by a national strategy, could increase preparedness for future pandemic origin investigations.
- Understanding pandemic origins could help mitigate health and economic costs associated with pandemics by, for example, facilitating surveillance that could identify future pandemics more quickly.
- A national strategy that includes pandemic origin investigations could help identify and quickly deploy resources needed for timely investigation of a pandemic's origin.

Considerations

- Allocating resources and defining how federal agencies and others will collaborate may be challenging because of the number and types of entities with relevant expertise that would be involved.
- During nonpandemic periods, other priorities and needs may arise and make it challenging to provide sustained resources and support

[50]This priority area includes characterizing biological material to support investigations, origin determination, and attribution as well as supporting United Nations investigations of outbreaks of unknown origin. See Office of Science and Technology Policy, *National Biodefense Strategy and Implementation Plan for Countering Biological Threats, Enhancing Pandemic Preparedness, and Achieving Global Health Security* (Washington, D.C.: October 2022).

needed for maintaining a national strategy.

- Augmenting or developing a new strategy would require careful consideration to avoid duplication, overlap, or fragmentation with existing related strategies, such as those for biodefense.
- Integrating a goal of pandemic origin investigations into existing strategies could dilute the focus and resources of the existing strategies.

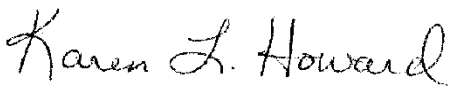
6 Agency and Expert Comments

We provided a draft of this product to the Department of State, Department of Defense, Department of Homeland Security, Department of Energy's Office of Science and National Nuclear Security Administration Laboratories, Office of the Director of National Security's Intelligence Advanced Research Projects Activity, Office of Science and Technology Policy, Department of Health and Human Services' Centers for Disease Control and Prevention and National Institutes of Health, Department of Justice's Federal Bureau of Investigation, National Institute of Standards and Technology, National Science Foundation, and United States Agency for International Development for review. Six agencies provided technical comments on the draft report, which we incorporated as appropriate.

We also invited the participants from our expert meeting to review our draft report. Of the 27 experts, 17 agreed to receive the draft for review and 10 provided technical comments. We incorporated their technical comments as appropriate.

As agreed with your offices, unless you publicly announce the contents of this report earlier, we plan no further distribution until 5 days from the report date. At that time, we will send copies of this report to the appropriate congressional committees and other interested parties. In addition, the report is available at no charge on the GAO website at https://www.gao.gov.

If you or your staff have any questions about this report, please contact me at (202) 512-6888 or howardk@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix III.



Karen L. Howard, PhD
Director
Science, Technology Assessment, and Analytics

List of Requesters

The Honorable Cathy McMorris Rodgers
Chair
Committee on Energy and Commerce
House of Representatives

The Honorable Bob Latta
Chair
Subcommittee on Communications and Technology
Committee on Energy and Commerce
House of Representatives

The Honorable Jeff Duncan
Chair
Subcommittee on Energy, Climate, and Grid Security
Committee on Energy and Commerce
House of Representatives

The Honorable Bill Johnson
Chair
Subcommittee on Environment, Manufacturing, and Critical Minerals
Committee on Energy and Commerce
House of Representatives

The Honorable Brett Guthrie
Chair
Subcommittee on Health
Committee on Energy and Commerce
House of Representatives

The Honorable Gus Bilirakis
Chair
Subcommittee on Innovation, Data, and Commerce
Committee on Energy and Commerce
House of Representatives

The Honorable H. Morgan Griffith
Chair
Subcommittee on Oversight and Investigations
Committee on Energy and Commerce
House of Representatives

The Honorable Markwayne Mullin
United States Senate

The Honorable Kelly Armstrong
House of Representatives

The Honorable Larry Bucshon, MD
House of Representatives

The Honorable Michael Burgess, MD
House of Representatives

The Honorable Earl L. "Buddy" Carter
House of Representatives

The Honorable Dan Crenshaw
House of Representatives

The Honorable John Curtis
House of Representatives

The Honorable Neal P. Dunn, MD
House of Representatives

The Honorable Richard Hudson
House of Representatives

The Honorable John Joyce, MD
House of Representatives

The Honorable Debbie Lesko
House of Representatives

The Honorable Gary Palmer
House of Representatives

The Honorable Greg Pence
House of Representatives

The Honorable Steve Scalise
House of Representatives

The Honorable Tim Walberg
House of Representatives

Appendix I: Objectives, Scope, and Methodology

Objectives

This report identifies and discusses:

1. key technologies available for pandemic origin investigations;
2. strengths and limitations of these tools and how researchers use them to investigate pandemic origins;
3. cross-cutting challenges researchers face in trying to determine a pandemic's origin; and
4. policy options that may help address the limitations and cross-cutting challenges of using these key technologies to determine the origin of a pandemic.

Scope and methodology

To address our first three objectives, we assessed available and developing technologies and approaches that are currently used in pandemic origin investigations. For all of our objectives we reviewed peer-reviewed scientific literature and other documents describing current and developing tools, including reports from the Centers for Disease Control and Prevention, Office of the Director of National Intelligence, the Johns Hopkins Center for Health Security, World Health Organization, and select national laboratories; interviewed federal agency officials and experts from government, academia, industry, and the nonprofit sector; and convened a 3-day expert meeting with assistance from the National Academies of Sciences, Engineering, and Medicine to discuss the objective topics. We also reviewed federal agency guidance on the development and deployment of these technologies for pandemic origin investigations.

Limitations to scope

The list of key technologies for pandemic origin investigations discussed in this report is not intended to be exhaustive. Based on our review of the literature and discussions with federal agency officials and other experts, we selected technologies currently in use or under development by researchers to investigate a pandemic's origin. We did not include all possible types of pathogens; we focused on those that are likely to lead to direct human-human transmission. For example, we did not include pathogens that cause foodborne outbreaks. We also did not review or include classified data or intelligence. Since pandemics pose a global threat, the policy options we identified represent possible actions U.S. policymakers and international stakeholders could take.

Literature search

In the course of our review, we worked with a GAO research librarian to conduct a literature search of key technologies for identifying and characterizing pandemic pathogens.[51] The librarian conducted literature searches with

[51]For the purposes of this report, the term "technologies" includes the instruments, techniques, skills, methods, and processes used in pathogen characterization.

Scopus using search terms including "pandemic origins," "biosurveillance," "SARS-CoV-2," and "bioinformatics," among other keywords relevant to technologies for characterizing pathogens. We conducted a broad search of materials published within the last 10 years, including scholarly articles and government reports. From these searches, we identified and selected relevant articles to include in our review. We used the results of our literature review to inform our findings as well as identify experts to interview or invite to participate in our expert meeting.

Interviews

We interviewed federal agency officials and researchers as well as nonfederal experts with a diverse set of perspectives on the science and application of these technologies. These experts included individuals from 11 relevant federal agencies: the Department of State, Department of Defense, Department of Homeland Security, Department of Energy's Office of Science and National Nuclear Security Administration Laboratories, Office of the Director of National Security's Intelligence Advanced Research Projects Activity, Office of Science and Technology Policy, Department of Health and Human Services' Centers for Disease Control and Prevention and National Institutes of Health, Department of Justice's Federal Bureau of Investigation, National Institute of Standards and Technology, National Science Foundation, and United States Agency for International Development. We also interviewed experts from technology companies, universities, and research institutes that use or develop genome sequencing, proteomics technologies, and laboratory characterization methods for pathogen characterization; representatives from national and international health organizations (e.g., the Association of Public Health Laboratories, Association of State and Territorial Health Officials, EcoHealth Alliance, and World Health Organization); and other individuals with expertise with technologies used for pandemic origin investigations.

Expert meeting

To address all of our objectives, we also held an expert meeting March 22-24, 2022. This meeting was held with assistance from the National Academies of Sciences, Engineering, and Medicine and was divided into six sessions: (1) genomic technologies for determining pathogen sequences; (2) genomic technologies for characterizing pathogen sequences to inform origin; (3) genomic technologies for determining analytical confidence and reproducibility; (4) non-genomic technologies for characterizing pathogens; (5) surveillance technologies that would inform pandemic pathogen origin; and (6) potential policy options that could help address technology limitations and other challenges.[52]

We selected meeting participants based on their expertise in at least one area related to our four objectives. We provided the National

[52]This meeting of experts was planned and convened with assistance from the National Academies of Sciences, Engineering, and Medicine to better ensure that a breadth of expertise was brought to bear in its preparation. However, all final decisions regarding meeting substance and expert participation are the responsibility of GAO.

Academies of Sciences, Engineering, and Medicine with descriptions of the expertise needed by expert meeting participants. From this information, the National Academies of Sciences, Engineering, and Medicine provided an initial list of potential participants for the expert meeting. We reviewed the list and provided an additional list of experts based on our review of the literature.

In addition to evaluating experts on the basis of their expertise, we evaluated them for any conflicts of interest. A conflict of interest was considered to be any current financial or other interest, such as an organizational position, that might conflict with the service of an individual because it could (1) impair objectivity or (2) create an unfair competitive advantage for any person or organization. Of the 27 experts who participated in the expert meeting, some were affiliated with companies, government, or research-funding entities. We took these affiliations into consideration as potential conflicts of interest when conducting our analysis and preparing our report. We determined that these experts' affiliations were unlikely to bias our overall reporting.

Policy options

Based on our research, we developed a series of policy options. Policy options are not formal recommendations for federal agencies, or matters for congressional consideration, but they are intended to represent possible options policymakers can take to address a policy objective. For each policy option, we discussed potential opportunities and considerations. These are not listed in any particular order, nor are they inclusive of all possible policy options. Based on the goal of improving U.S. pandemic preparedness, we decided on an objective designed to identify options that could help improve capabilities for pandemic origin investigations. We limited policy options to those that fit the objective and fell within the report scope.

To develop our policy options, we compiled a list of possible options over the course of our work based on review of the literature, interviews with experts, and our expert meeting held March 22–24, 2022. We further refined and assessed these options to ensure they were adequately supported by the evidence we collected, could be feasibly implemented, and fit into the overall scope of our work. We then analyzed the information we collected to identify potential benefits and considerations of implementing each policy option. The policy options and analyses were supported by documentary and testimonial evidence.

We conducted our work from August 2021 to January 2023 in accordance with all sections of GAO's Quality Assurance Framework that are relevant to technology assessments. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations to our work. Consistent with our quality assurance framework, we provided the relevant agencies and experts with a draft of our report and solicited their feedback, which we incorporated as appropriate. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

Appendix II: Expert Participation

We convened a 3-day meeting of 27 experts with assistance from the National Academies of Sciences, Engineering, and Medicine to inform our work on technologies for determining pandemic origin; the meeting was held virtually March 22–24, 2022. The experts who participated in this meeting are listed below. Some of these experts gave us additional assistance throughout our work, including four experts who provided additional assistance during our study by sending material for review or participating in interviews and 10 experts who reviewed our draft report for accuracy and provided technical comments.

David B. Allison, PhD

Dean, Distinguished Professor and Provost Professor

Indiana University–Bloomington School of Public Health

Jesse Bloom, PhD

Professor, Basic Sciences Division

Professor, Herbold Computational Biology Program, Public Health Sciences Division

Fred Hutchinson Cancer Research Center

Roger Brent, PhD

Professor, Basic Sciences Division

Professor, Public Health Sciences Division

Fred Hutchinson Cancer Research Center

James Diggans, PhD

Distinguished Scientist, Bioinformatics and Biosecurity

Twist Bioscience

Joshua Dunn, PhD

Head of Design

Ginkgo Bioworks, Inc.

Livia Schiavinato Eberlin, PhD

Associate Professor, Department of Surgery

Baylor College of Medicine

Patrick Fitch, PhD

Associate Director of Chemical, Earth and Life Sciences

Los Alamos National Laboratory

A. Oveta Fuller, PhD

Associate Professor of Microbiology and Immunology

Medical School at University of Michigan

Gigi Kwik Gronvall, PhD

Senior Scholar

Johns Hopkins Center for Health Security

Associate Professor, Department of Environmental Health and Engineering

Johns Hopkins Bloomberg School of Public Health

India Hook-Barnard, PhD

Executive Director

Engineering Biology Research Consortium (EBRC)

Katrina Kalantar, PhD

Computational Biology Lead, Infectious Diseases

Chan Zuckerberg Initiative

Ali S. Khan, MD, MPH, MBA

Dean, College of Public Health

University of Nebraska Medical Center (UNMC)

Former Assistant Surgeon General

U.S. Public Health Service

Andy Kilianski, PhD

Senior Director for Emerging Infectious Diseases

International AIDS Vaccine Initiative (IAVI)

Adjunct Professor, Schar School of Policy and Government

George Mason University

Sergios-Orestis Kolokotronis, PhD, MPhil, MA

Assistant Professor, Department of Epidemiology and Biostatistics. School of Public Health

The State University of New York (SUNY) Downstate Health Sciences University

Suresh Kuchipudi, BVSc, MVSc, PhD, PGCHE, FHEA, Dip. ACVM, MBA

Professor and Endowed Chair in Emerging Infectious Diseases

Pennsylvania State University

Associate Director

Penn State Animal Diagnostic Laboratory (ADL)

Jacob Lemieux, MD, DPhil

NIH-funded Physician/Scientist, Division of Infectious Disease

Massachusetts General Hospital (MGH) and Harvard Medical School (HMS)

Bronwyn MacInnis, PhD

Director of Pathogen Genomic Surveillance, Infectious Disease and Microbiome Program

Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard

Alemka Markotić, MD, PhD

Director

University Hospital for Infectious Diseases, Zagreb, Croatia

Head of Department for Research and Head of Clinical Department for Urinary Tract Infections and Full Professor

Medical School, University of Rijeka and Catholic University Zagreb

Associate Member, Croatian Academy

Jonna Mazet, DVM, MPVM, PhD

Vice Provost – Grand Challenges

University of California (UC) Davis

Chancellor's Leadership Professor of Epidemiology and Disease Ecology and Founder, One Health Institute

UC Davis School of Veterinary Medicine

Folker Meyer, PhD

Professor of Data Science

University Hospital, University of Duisburg-Essen

Tara O'Toole, MD, MPH

Senior Fellow

In-Q-Tel

Director, IQT Lab

BiologyNext

Rushika Perera, PhD

Associate Professor, Department of Anatomy

University of California (UC) San Francisco

Brian Plew

Director, Public Health Solutions

Thermo Fisher Scientific

David Relman, MD

Thomas C. and Joan M. Merigan Professor in Medicine

Professor of Microbiology & Immunology

Senior Fellow, Center for International Security and Cooperation

Stanford University

Chief of Infectious Diseases

Veterans Affairs Palo Alto Health Care System

Aaron Streets, PhD

Assistant Professor in Bioengineering

University of California (UC) Berkeley

Core Member, Biophysics Program and Center for Computational Biology

Investigator, Chan Zuckerberg Biohub

David Walt, PhD

Hansjörg Wyss Professor of Bioinspired Engineering

Harvard Medical School

Professor of Pathology

Brigham and Women's Hospital

Core Faculty Member

Wyss Institute at Harvard University

Susan Weiss, PhD

Professor and Vice Chair, Department of Microbiology and Co-Director

Penn Center for Research on Coronaviruses and Other Emerging Pathogens,

Perelman School of Medicine

University of Pennsylvania

Governor

American Academy of Microbiology

Appendix III: GAO Contact and Staff Acknowledgments

GAO contact

Karen L. Howard, PhD, Director, Science, Technology Assessment, and Analytics (STAA), at (202) 512-6888 or howardk@gao.gov

Staff acknowledgments

In addition to the contact named above, the following STAA staff made key contributions to this report:

Hayden Huang, PhD, Assistant Director and Senior Engineer

Michael Dickens, PhD, Analyst-in-Charge and Senior Biological Scientist

Calaera Powroznik, MS, Analyst

Craig Starger, PhD, Biological Scientist

Adam Wells, PhD, Data Scientist

These staff also contributed to this work:

Nora Adkins, JD, Senior Attorney
Virginia Chanley, PhD, Senior Design Methodologist
Louise Fickel, Communications Analyst
Anika McMillon, Visual Communications Analyst
Matthew Metz, Senior Electrical Engineer
Cindy Korir-Morrison, PhD, Senior Biological Scientist
Robert Rivas, MS, Senior Analyst
Sushil Sharma, PhD, DrPH, Assistant Director
Amber Sinclair, PhD, Senior Design Methodologist
Walter Vance, PhD, Assistant Director

(105406)

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's website (https://www.gao.gov). Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to https://www.gao.gov and select "E-mail Updates."

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, https://www.gao.gov/ordering.htm.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or
TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on Facebook, Flickr, Twitter, and YouTube.
Subscribe to our RSS Feeds or E-mail Updates.
Listen to our Podcasts and read The Watchblog.
Visit GAO on the web at https://www.gao.gov.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact: Website: https://www.gao.gov/fraudnet/fraudnet.htm
Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

A. Nicole Clowers, Managing Director, ClowersA@gao.gov, (202) 512-4400,
U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, YoungC1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149, Washington, DC 20548

Strategic Planning and External Liaison

Stephen Sanford, Managing Director, spel@gao.gov, (202) 512-9715
U.S. Government Accountability Office, 441 G Street NW, Room 7B37N, Washington, DC 20548