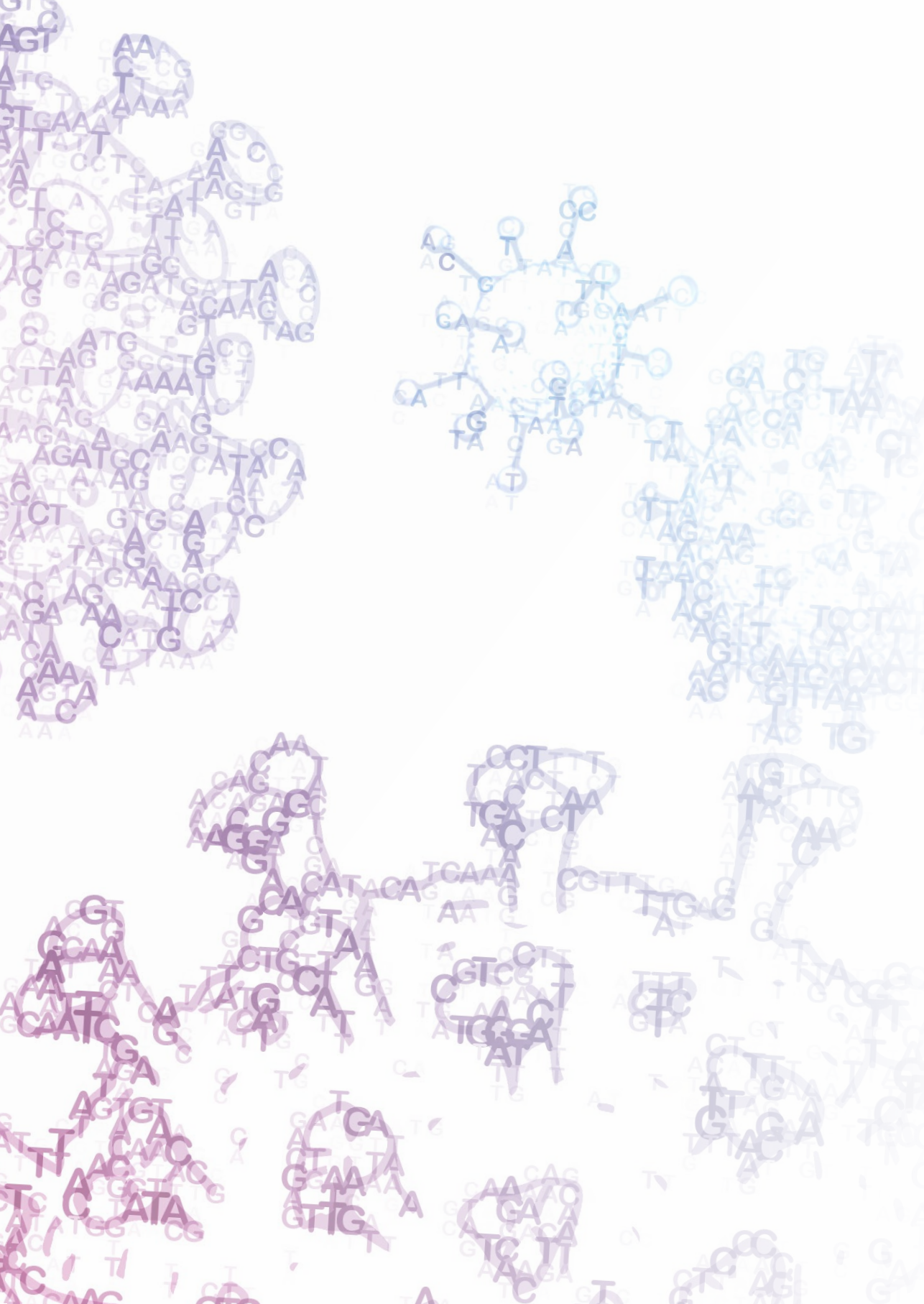
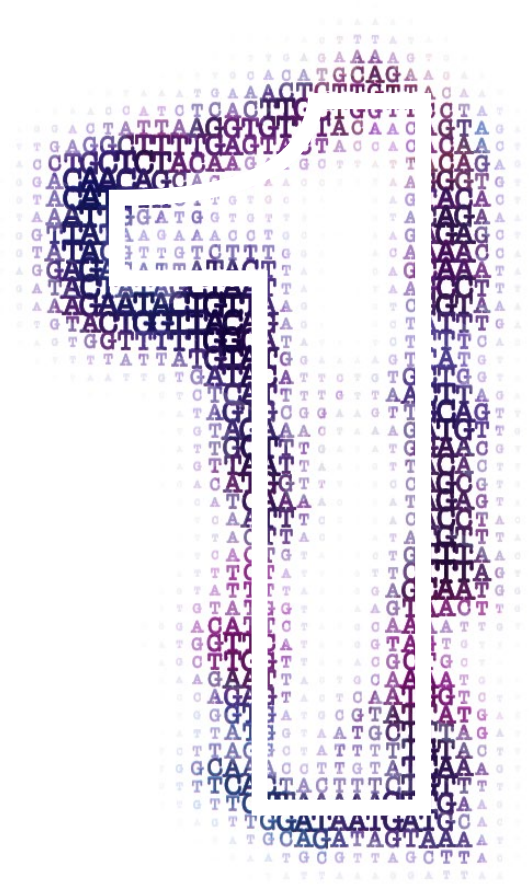
Metagenomic sequencing in clinical virology: advances in pathogen detection and future prospects

Carbo, E.C.





Chapter 1 **Introduction**

Molecular viral diagnostics is based on sequential tests identifying up to a handful of viruses at a time using polymerase chain reaction (PCR). However, with over 1,000 virus species known to infect humans [1], it is impossible to test all these viruses at once in a patient sample using this PCR method. And with a total presence of 10^{31} virus particles on earth [2] and novel or zoonotic viruses potentially infecting humans [3], there is a great expectation for a one-test-catches-all method. Viral metagenomic next-generation sequencing is such a method, though some early studies have shown a lower sensitivity compared to PCR [4,5]. The work performed in this thesis investigates the application and possibilities of viral metagenomic sequencing, further enhancing the accuracy of this test to make it suitable for application in a clinical setting.

This current chapter will first illustrate the relevance of infectious diseases as a societal burden and provide a summary of the history of infectious diseases and sequencing. Subsequently, several applications of sequencing will be explained with a focus on their utilization in viral metagenomic sequencing.

Background

Infections and pandemics

Infectious diseases have troubled humankind since recorded history. Tuberculosis (TB) is perhaps the longest known [6], and the causative pathogen may have already existed 15,000-20,000 years ago. Additionally, several types of infections have been described by both Hippocrates and the ancient Egyptians [7], and one of the first plagues described was the Athenian plague of 430 BC [8]. More recent pandemics of the last century with a large effect on society were the influenza pandemic between 1918-1920 resulting in approximate 50 million fatalities [9], and the outbreak of HIV/AIDs in the '80s and up until present day leading to over 38 million deaths [10]. Currently, infectious diseases and pandemics are still a threat, with SARS-CoV-2 as the causative agent of the latest large known outbreak.

All eukaryote cells harbour small pieces of bacterial DNA in their mitochondria, demonstrating that infections as encounters between different cells have a very long history [11]. The human genome is composed of 8% viruses due to integration of human endogenous retroviruses in our DNA [12]. Additionally, the human microbiome consists of about 10-100 trillion microbial cells that are permanent residents of the human gastrointestinal tract [13]. These symbiotic interactions with microorganisms are often ignored until the moment we get afflicted by an infection that will give us specific symptoms. Infectious diseases are part of our existence; we exist because they exist, and they exist since we exist, as many of them use the Homo sapiens species as a host. In addition, pathogen infection and replication are dependent on population density, with pathogens becoming or remaining endemic in a highly dense population [14,15]. This was already known in ancient history with the Harappan civilization, in the present day Pakistan, building a brick sewer system before 2000 BC, that probably was intended for proper sanitation [16]. With the current increasing urbanization of the world's population and ever more travel movements we are more at risk for infectious outbreaks [17,18].

There are over 1,000 virus species known to infect humans [1], and there are currently 10^{31} viral particles on Earth [2], of which a small number might be identified as infectious for humans in the future, and of which some might be zoonotic viral infections that can potentially infect humans [3]. Estimates indicate that up to 60% of human infectious diseases are from zoonotic origin [19], exemplified by the recent outbreak of SARS-CoV-2 [20]. Not only infectious diseases can directly affect

an individual's health, but they can also have broader consequences. What is often overlooked is that infectious diseases have also shaped economic, political and social aspects of our society. Black Death (Yersinia pestis in different forms) is an example of an infectious disease that impacted society greatly. It was the largest catastrophe to have ever happened to mankind, resulting in the death of one third of Europe's population around the 1300s [21,22]. Large outbreaks did not only have deleterious effects but also resulted in improvements in health care, and pushed the need for epidemiologic insight in prevention, immunity and antimicrobial treatment [8]. The SARS-CoV-2 outbreak also necessitated an urgent update of our surveillance of the infection in all details, and therefore viral DNA sequencing needs [23-25].

Infectious disease burden

Worldwide, the leading cause of death is thought to be ischemic heart disease with a crude death rate (CDR) of 115.3 per 100,000 individuals in 2019 [26] (Figure 1). However, when taking infectious and parasitic diseases into account and all lower respiratory infections, the combined death rate is 100 per 100,000 individuals, illustrating the impact of infectious diseases on human health. Underestimations of infectious diseases have been common due to absent or inconsistent surveillance, identification, and registration by healthcare organizations [27]. This problem can be larger in underdeveloped countries where infectious diseases are more frequent than ischemic heart diseases. The latest disease burden statistics presented by the WHO date from 2019, and are excluding an estimate of over six million deaths worldwide due to the outbreak of SARS-CoV-2 [28]. SARS-CoV-2 epidemiological data published in 2021 showed a crude death rate of 81.7 per 100,000 individuals in the Netherlands and 180.9 per 100,000 individuals in Belgium, further increasing the infectious disease burden [29]. Infectious disease are also part of the cause for diarrhoeal diseases and can cause cancer, with estimates of approximately one in six cancers having an infectious origin [30]. With cancer having a combined total CDR of 120.6 [26], another 20.1 deaths per 100,000 would be linked to infectious disease for a total of approximately 120 deaths per 100,000 – surpassing ischemic heart disease, and illustrating the impact of infectious diseases have on human health and the importance of studying this topic.

Microorganisms and metagenomics

In 1677, Dutch merchant Antonie van Leeuwenhoek wrote in his letters to the Royal Society about 'concerning little animals' he observed in several materials using one of his own custom-built microscopes [31], which must have been the first visualisation of individual bacteria. After this first known recognition of microorganisms,

it took two more centuries before Robert Koch developed new methods to grow microorganisms. Koch isolated and pinpointed a bacterium to cause tuberculosis: Mycobacterium tuberculosis [32,33] and formulated criteria for establishing the causality of a microbe (Koch's postulates) [34]. Shortly afterwards, in 1898, Martinus Beijerink, who received his doctorate at the University of Leiden, discovered the tobacco mosaic pathogen. Martinus Beijerink called it a 'virus' and he is now considered as one of the founders of virology [35].

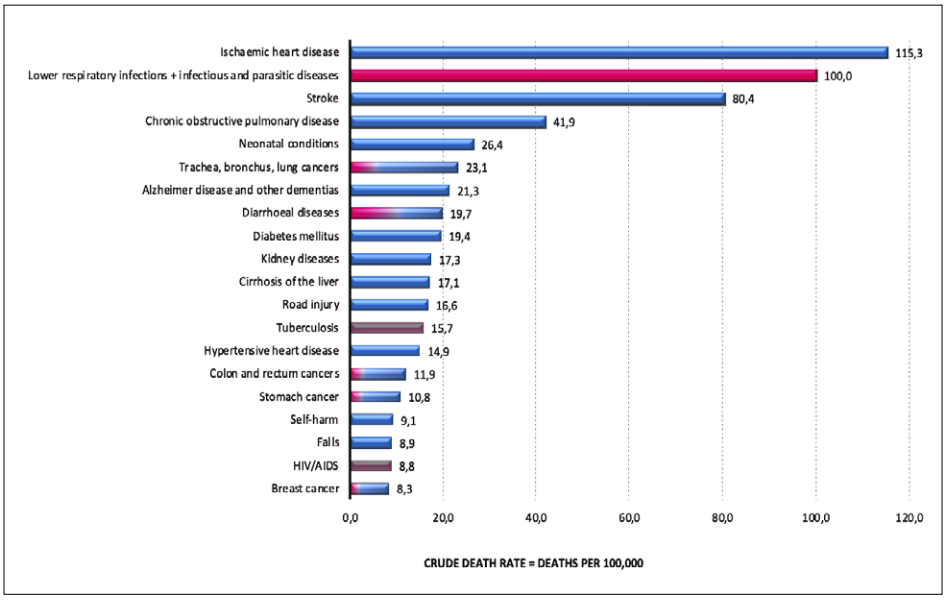
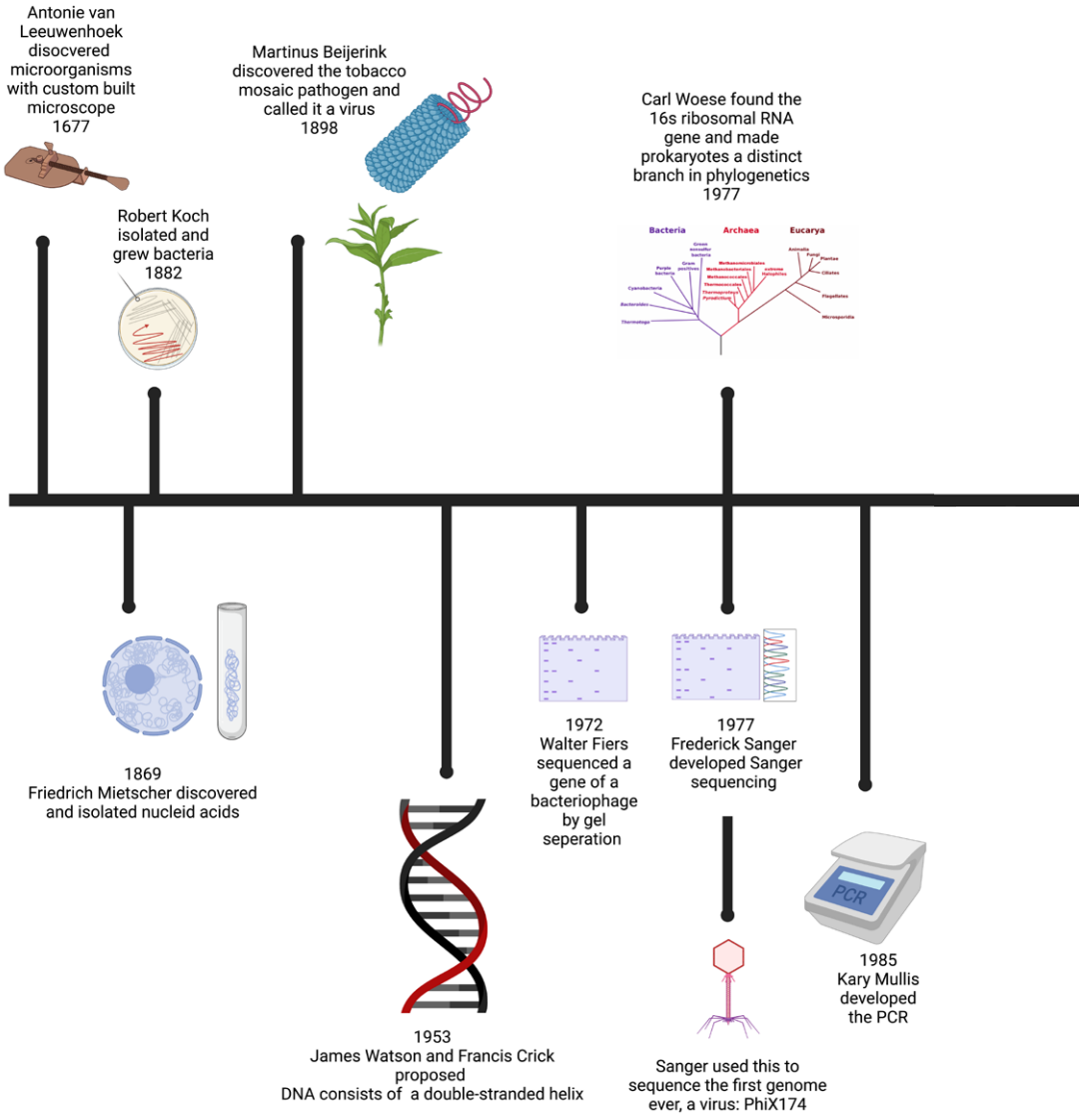


Figure 1. Disease burden worldwide: Top 20 causes of death worldwide.

Data adapted from World Health Organization, last presented for the year 2019 [26], therefore death cause data excludes a majority of COVID-19 deaths. Crude death rate (CDR) associated with infectious diseases are shown in red. CDR of lower respiratory infections and infectious and parasitic diseases is the total number including the CDR of Tuberculosis and HIV/AIDS, therefore the CDR of these individual diseases are masked.

Another century later Carl Woese distinguished the different biological kingdoms using the common 16S ribosomal RNA gene present in prokaryotes. Because this region is quite well preserved amongst species, it is a good basis for phylogenetics [36]. From this moment the prokaryotes were distinct in the tree of life and later 16S amplicon sequencing was marked as the start of microbial sequencing. This was achieved first by means of Sanger sequencing, simply by looking at short limited sequences at once and later using next-generation sequencing looking at a high number of sequences at once. In 1996 (Stein et al. 1996) sequenced


Antonie van Leeuwenhoek disocvered microorganisms with custom built microscope 1677
Martinus Beijerink discovered the tobacco mosaic pathogen and called it a virus 1898
Carl Woese found the 16s ribosomal RNA gene and made prokaryotes a distinct branch in phylogenetics 1977
Robert Koch isolated and grew bacteria 1882
Bacteria
Archaea
Eucarya
1869 Friedrich Mietscher discovered and isolated nucleid acids
1953 James Watson and Francis Crick proposed DNA consists of a double-stranded helix
1972 Walter Fiers sequenced a gene of a bacteriophage by gel seperation
1977 Frederick Sanger developed Sanger sequencing
Sanger used this to sequence the first genome ever, a virus: PhiX174
PCR
1985 Kary Mullis developed the PCR

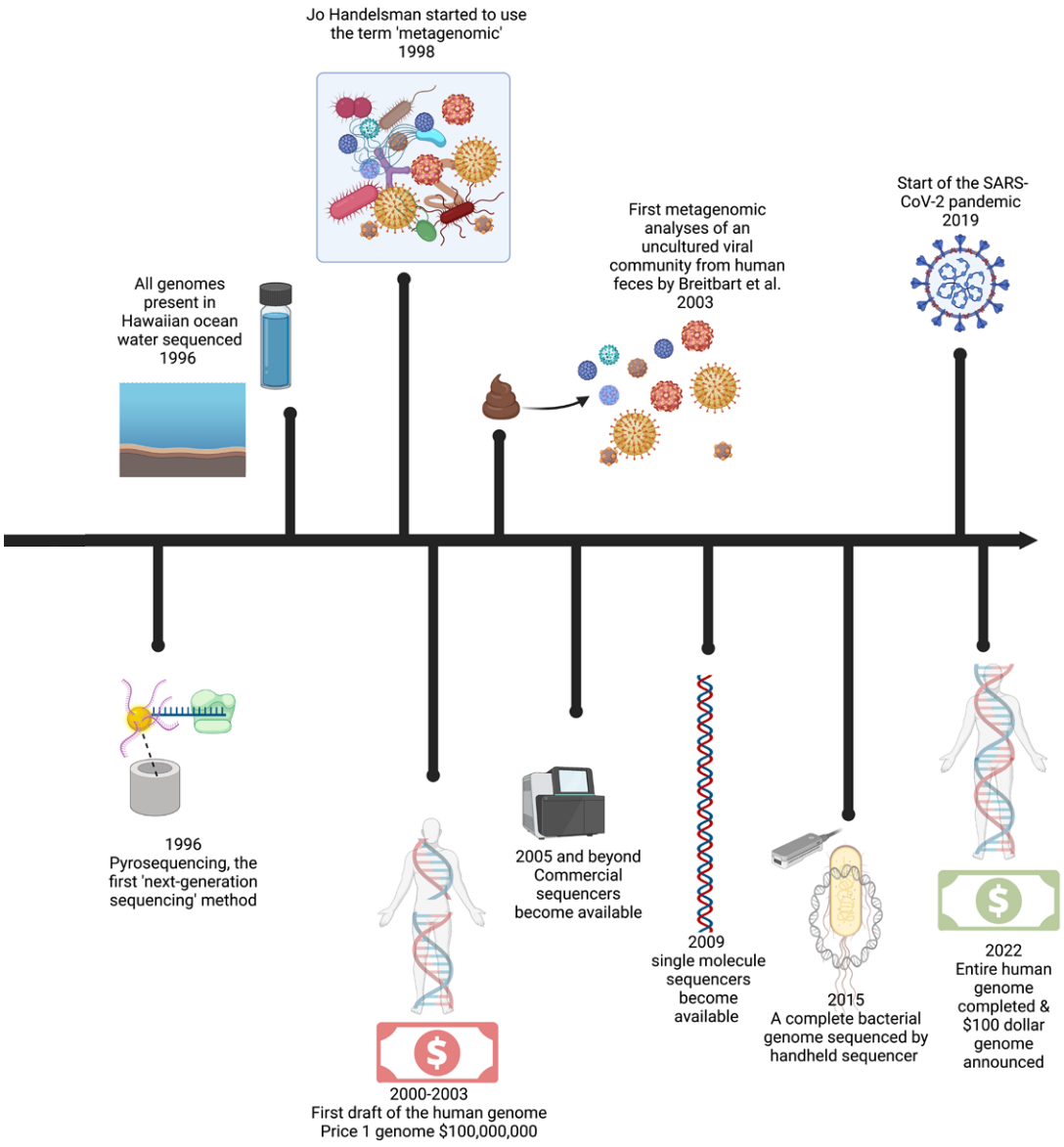


Figure 2. Timeline with milestones in microbiology and sequencing.

Above: important events over time in microbiology, virology and metagenomics, and below: milestones in molecular genetics and sequencing. Adapted from the timeline of the study of Escobar-Zepeda et al. [40] Created using Biorender.com.

Hawaiian ocean water to look at all the genomes in a sample [37], pioneering the field towards metagenomics, though the name metagenomic sequencing was coined in 1998 when Jo Handelsman used the term metagenomics in one of her studies [38], meaning sequencing all metagenomes present in a sample (see Figure 2). The first metagenomic analysis of the uncultured viral community present in human feces was studied in Breitbart et al. in 2003 [39].

History of sequencing

Just before Robert Koch isolated and pinpointed a specific disease-causing micro-organism, Friedrich Mietscher was able to discover and isolate DNA from cells for the first time in 1869. He first observed there was another structure with different kind of characteristics [41]. James Watson and Francis Crick proposed that the DNA structure consisted of a right-handed helix composed of two anti-parallel DNA strands [42]. Rosalind Franklin in her fellow research on the molecular structure of RNA/DNA and her X-Ray diffraction work provided insight in this structure that enabled them to publish this work [43]. Many attempts were undertaken to sequence DNA to determine the order of nucleotides, and Robert Holley in 1965 sequenced the first transfer RNA extracted from yeast [44]. Shortly after, in 1968, Crick published the genetic code behind all the different amino acids [45].The first gene of which the complete nucleotide code was deciphered was that of a virus, as in 1972 Walter Fiers from Belgium sequenced the first gene of bacteriophage MS2. The first sequencing protocol consisted of digesting the virus RNA into small pieces and then separating them on a gel [46]. Frederick Sanger in 1977 developed Sanger sequencing based on a chain termination method that generates partially fragmented DNA, each fragment with a radiolabelled termination, enabling the sequence to be constructed [47] Sanger also used this method to sequence and determine the nucleotide order of the first genome ever, a virus, bacteriophage PhiX174 [48]. Up to this day, PhiX174 is a virus that haunts genetic research in every sector, since one of the large sequence companies, Illumina, advises to sequence this virus in every run for a quality control check; indirectly leading to contaminated assembled genomes being uploaded to public databases, including parts of this PhiX174 genome [49]. It was also in the group of Sanger where there was the first need for computer processing of DNA data [50], whereas computer processing of biological data, bioinformatics, was previously mainly still applied in the protein field with Margaret Dayhoff probably being the first bioinformatician [50].

In 1985 Kary Mullis developed the PCR, a method we are still currently using [51]. Merely one year later Leroy Hood and Michael Hunkapiller automated the Sanger

sequencing method making sequencing possible in a quicker way [52]. In order to expand sequence capacity, pyrosequencing was introduced in 1996 [53] as the first high throughput or next-generation sequence method.

Next-generation sequencing (NGS)

The development of massively parallel sequencing contributed to the first drafts of the human genome in 2000 and 2003 [54-56]. Sanger sequencing was subsequently used to sequence Craig Venter's genome for 100 million dollars, while resequencing James Watson's genome was less than 1 million dollars using NGS [54,57]. Around that time several other companies also started to use a similar method for NGS: Solexa in 1998 (later acquired by Illumina); 454 in 2005 (acquired by Roche in 2007); SOLiD in 2007; and the IonTorrent system of Life Technologies in 2001 [58].

While NGS methods first aimed at sequencing short sequence fragments, the new era of NGS sequencers focuses on single molecule sequencing, first described in 2009 [59]. With this technique performed now by Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT), a complete genome can be sequenced in one single run. These longer sequence reads greatly improve assembling novel genomes [60]. In 2015 a complete bacterial genome was assembled using the ONT method [61]. And only in this year, 2022, the complete human genome has been finally completely sequenced by means of several different sequencing methods including Illumina, PacBio and ONT filling in the last blanks of the human genome that still currently existed [62]. Whereas sequencing Watson's genome first had a price of $1 million, in 2014 the 1000-dollar human genome became available [63]. Sequencing costs are still declining, and in 2020 the author of this thesis paid €300 for privately sequencing her whole genome and since this year, 2022, the $100 human genome is available [64].

The need for informatics in biology: bioinformatics

With the growing number of sequence reads that need interpretation, bioinformatics becomes of vital importance. Bioinformatics started with protein analysis [50], as Margaret Dayhoff, 'the mother and father of bioinformatics', used punched-card business machines to establish molecular energies of organic molecules as these calculations could not be handled on regular calculators [65]. Together with Robert S. Ledley she developed a way to use computational resources in order to establish protein structure. The software 'COMPROTEIN' running on a IBM7090 needed for this was written in FORTRAN on punch cards (see Figure 3), and it bears resemblance to current *de novo* sequence assembly methods [66]. She also developed the one-letter amino acid codes that we still use today [67].

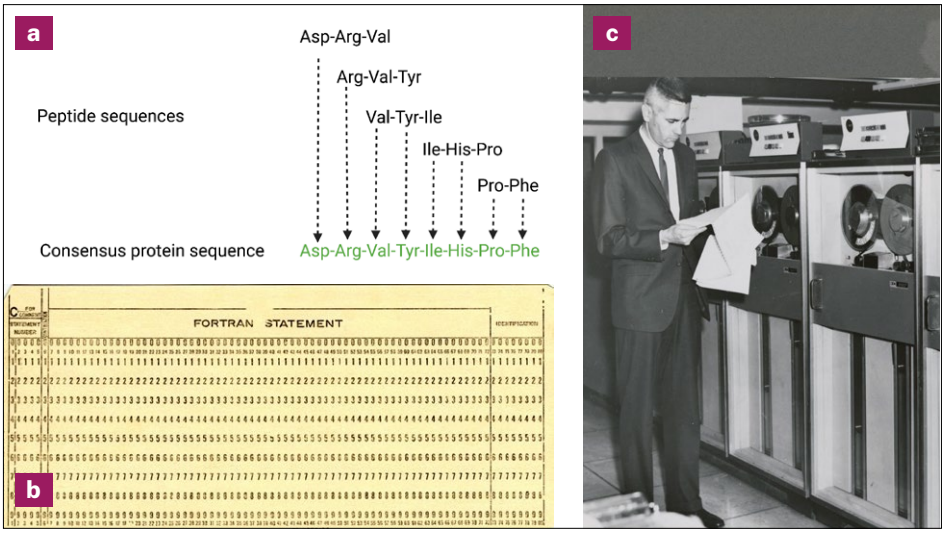


Figure 3. The first bioinformatics software.

(a) Overview of COMPROTEIN. COMPROTEIN used peptide sequences as input and gave a consensus protein sequence as output. Created using Biorender.com. **(b)** FORTRAN punch card. COMPROTEIN was written in the language FORTRAN, only one line of code could be punched per punch card. **(c)** An IBM 7090 mainframe that could run COMPROTEIN.

Two decades later came the need to use computational methods to aid the analysis of nucleotides for comparisons, calculations and matching patterns [68]. Roger Staden wrote the first software to analyse DNA data from Sanger sequencing. His software looked for overlap between gel reads, to join reads into contigs, and to annotate sequence files [69]. He also extended the DNA alphabet with codes for when base calling could quantitatively not be correct, and this is now recorded in the official nomenclature for unclear bases in sequences [70], Roger Staden's software can currently still be downloaded [71].

Next-generation sequencing in clinical settings

Nowadays, utilizing sequencing within a clinical setting is common practice within human genetics and pathology, but its use in microbiology still lags behind. In clinical genetics, the identification of the first disease causing mutations were explored mid-20th century. Linkage analysis was suggested to make a connection between a genome locus and a disease [72], for instance the Huntington disease gene in 1983 [73]. In 1986 the first gene CYBB was linked to a chronic granulomatous disease without exactly knowing what the function of the gene was [74]. In 2009, an autosomal dominant mutation was found to be causal for Freeman-Sheldon

syndrome [75]. Shortly after, whole exome sequencing became common to find causes for autosomal recessive diseases [76]. Whole genome sequencing, having a larger horizontal genome coverage of the entire human genome, is since 2010 also used for finding disease causes [77].

In microbiology, the first attempt for metagenomic sequencing in Hawaiian ocean water was already undertaken in 1996 [37] and whole viral genomes were sequenced in 2004 [78]. Sequencing was not popular in a clinical setting, due the wide availability of fast traditional methods to identify pathogens. In clinical virology, viral culture, PCR, or serology testing were used to detect 'known' viral pathogens, and later, multiplex PCR reactions were used to detect several viruses at once. Viral metagenomics, investigating all viral nucleotides from an often uncultured sample, made it possible to identify novel and unexpected but previously identified viruses [79,80]. Viral metagenomics, testing in an unbiased and agnostic way, has been suggested as a powerful tool for virus discovery in a clinical healthcare setting [81]. Viral identification and discovery are in great need, with a constant threat of zoonotic virus transfer and an estimate of at least 320,000 mammalian viruses that still need to be discovered [82]. In the beginning of the recent pandemic, it was a viral metagenomic technique that identified the disease-causing viral pathogen directly from patients' material, and established the genome at once [23,24].

Clinical viral metagenomics

Sequencing all genomes in a sample at once

Metagenomics enables detection of all the genetic material of organisms present ina sample, making it a pathogen-agnostic approach for detecting common and rare or novel pathogens that are not included in conventional testing. Beforehand, a clinician does not need to have a hypothesis of what pathogen is expected, unlike traditional PCR testing. Another benefit is that this technique enables investigation of multiple species at once (Figure 4). The first case reports on the identification of viral pathogens in patients by means of metagenomics focused mainly on encephalitis patients [83–95]. Immunocompromised patients are of specific interest as they are at increased risk of infections by unexpected and novel viruses and bacteria without having regular symptoms [83,96]. In 2019, two prospective clinical utility studies were performed and published where viral metagenomics was used in parallel with conventional diagnostics [97,98]. These findings demonstrated that viral enrichment was beneficial for virus detection improving the potential for diagnostics [97,99-101].

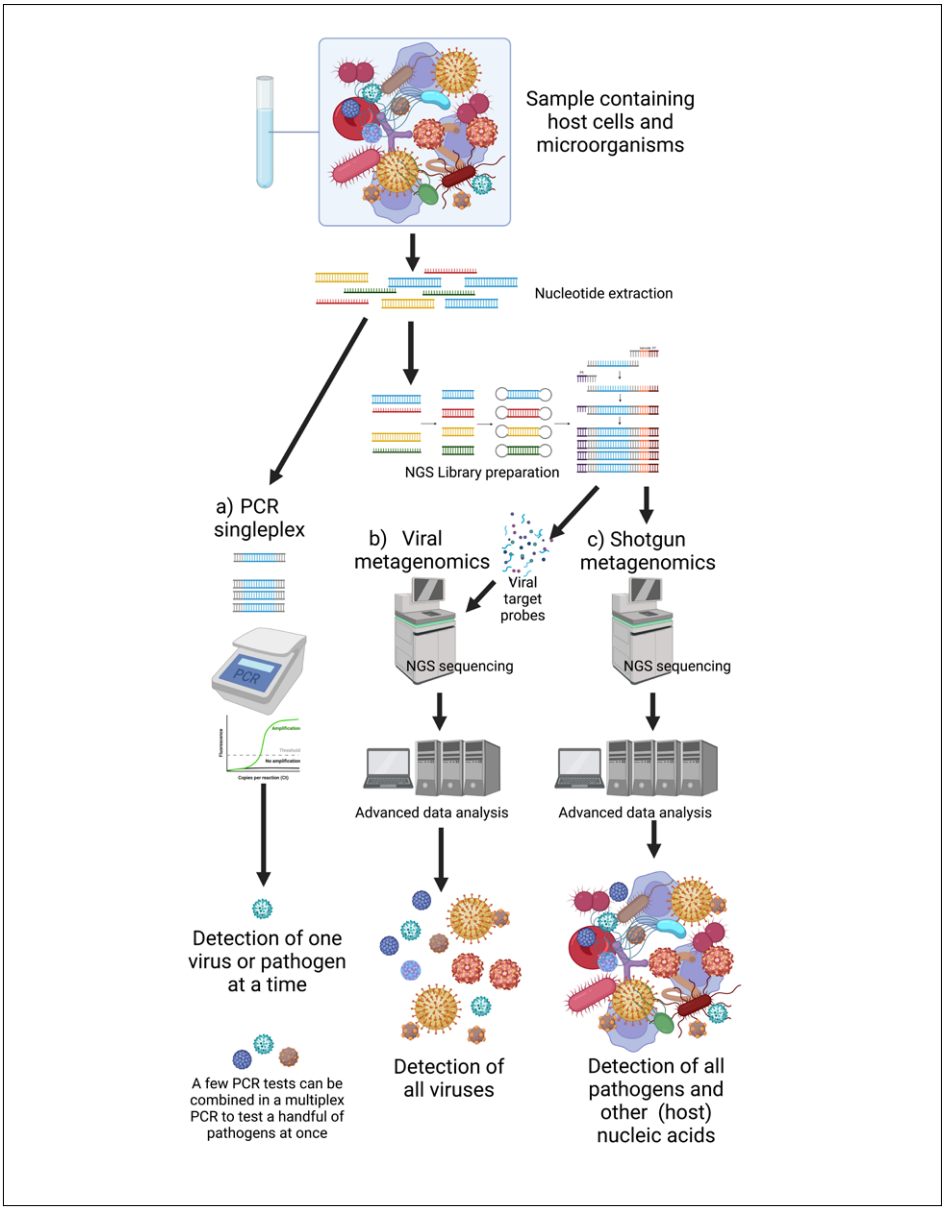


Figure 4. Overview of PCR and (targeted) metagenomic sequencing methods.

An overview of three methods for viral pathogen detection in a sample containing host cells and microorganisms. a) Conventional singleplex PCR testing of just one virus target at a time, though in a multiplex test several PCR tests can be combined to test multiple viruses at once b) Targeted viral metagenomics, using viral probes to capture only certain viral sequences after library preparation, c) Sequencing all genetic material in a sample using shotgun metagenomics, so all pathogens/species/genetic host material is available after data analysis. Created using Biorender.com.

While the genomes of viruses are sequenced, information about the type/subtype is often additionally available as well as information about resistance mutations.Another benefit of metagenomics is that a virus will not be missed due to mismatching primer pair sequences in cases of virus mutations in the primer regions. Recently, metagenomics has shown to be useful for the discovery and classification of the SARS-Cov-2 virus directly from patient material [24,25].

Increase in diagnostic yield

The additional diagnostic yield of metagenomic sequencing reported in literature is 28.73% (CI 19.80-37.63), with diagnostic yield defined as finding a potentially causative pathogen by means of metagenomic sequencing after initial/diagnostic standard tests were negative for a pathogen. The numbers are based on a systematic review and meta-analysis (Figure 5) performed by the author of this thesis and the metagenomics group. Forty-seven studies were included for systematic review, of which 27 studies were selected that targeted potential causative pathogens by means of metagenomic sequencing after the initial or diagnostic testing was negative in patients with a wide range of clinical syndromes [97,98,102-126].

Viral metagenomics laboratory protocol

For viral metagenomic sequencing, nucleotides are first extracted from a sample of RNA, DNA or both. Then in most cases RNA is converted into DNA by means of cDNA synthesis. In the beginning of the library preparation, genetic material is fragmented, after which the sequences are end-repaired and ligated with adapters needed for sequencing (Figure 6). The nucleotide sequences are ligated with barcodes to differentiate samples after sequencing. Usually, library preparation protocols involve amplification of the prepared libraries [83,97,98,128-132]. When no special filtering or additional target probes are used, this is called shotgun metagenomics (Figure 4). If required, steps can be undertaken to filter out more human cells, or ribosomal RNA before the library prep either via centrifugation or additional prep kits [133-137]. After the library preparation it is also possible to use a targeted probe kit designed to capture specifically viral sequences [100,103,133,138].

The highly sensitive PCR procedure is the gold standard to establish whether a virus is present in a given sample, and currently it is a challenge to establish a similar sensitivity using metagenomics. Due to highly abundant host material and microbiome, viral pathogen sequences are like the proverbial needle in a haystack and sensitivity remains to be improved. Additionally, it is common to detect contaminating microbial genetic material from reagents specifically used in test kits, the

'kitome' [139,140]. The current protocols are expensive and time-consuming [141] and, due to the complexity, not every lab can perform this method of sequencing [140,142].

Research was needed to further improve the viral metagenomic test sensitivity and reduce the amount of background (host) sequences and contaminating sequences. Due to the limited number of studies presenting data on the diagnostic yield of the test, additional patient cohorts with specific clinical syndromes needed to be tested to investigate the yield in different populations. In addition, the potential for virus discovery straight from clinical samples, while utilizing viral metagenomics, had to be tested for accuracy and applicability.

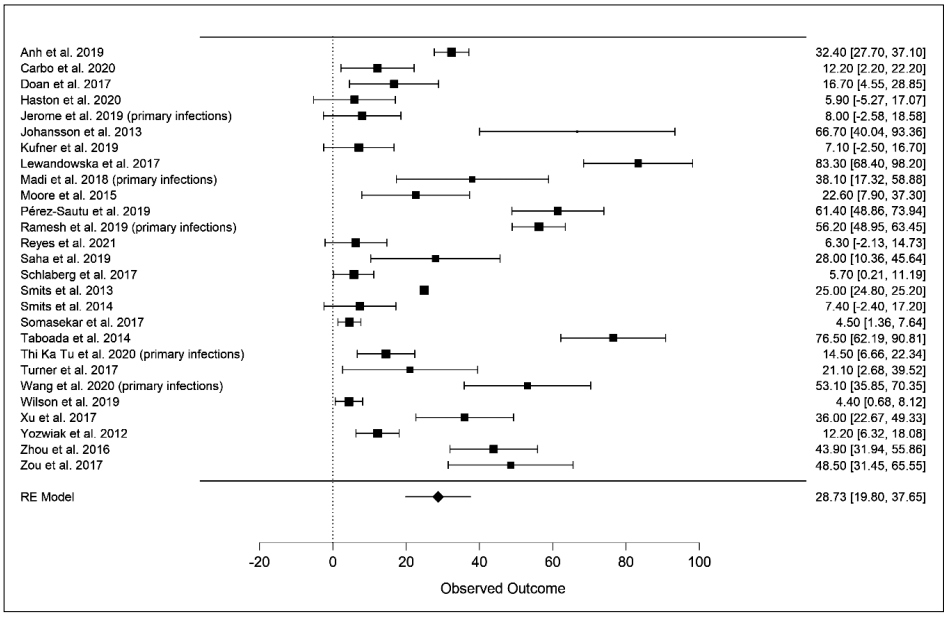


Figure 5. Meta-analysis of studies using metagenomics as a diagnostic tool to detect infectious diseases, in patients with a wide range of clinical syndromes.

Derived from author's unpublished data of a systematic review on viral metagenomics and diagnostic yield. A search in PubMed was conducted and reference lists were searched in December 2020, of which 644 studies were obtained. Of the 644 studies, microbiome/virome studies, technical validation studies, reviews, opinion papers, studies in other languages than English, case reports, and studies with sample number <7 were excluded. Forty-seven remained for systematic review, 27 identified potentially causative pathogen by means of metagenomic sequencing after the initial or diagnostic testing was negative. A random effects meta-analysis was performed given the heterogeneity of the 27 papers using JASP [127].

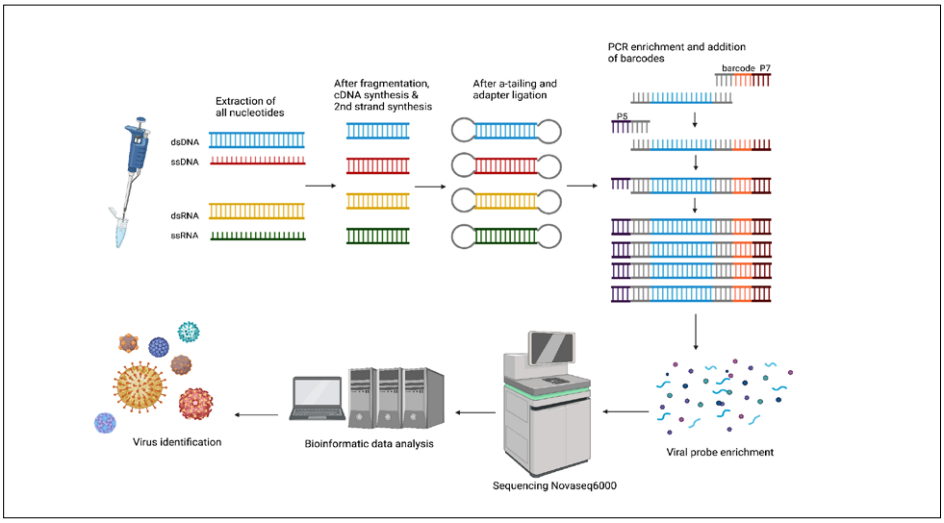


Figure 6. Steps in our current metagenomics NGS protocol applicable in a clinical setting.

Library preparation protocol as performed in this thesis starts after extracting all nucleotides of a clinical sample with enzymatic fragmentation, cDNA synthesis and 2nd strand synthesis. Next, A-tailing of the sequences is carried out and adapters are ligated to the blunt ended A-tailed overhang. A PCR enrichment step is carried out and sample specific barcodes are added to the sample.

To select viruses and minimize sequence reads of other (host) species, a viral probe enrichment can be carried out, though samples could also be sequenced for a shotgun approach without probe enrichment. Sequencing is carried out on an Illumina sequencer followed by bioinformatic data analysis using a taxonomic classifier for virus identification. Created using Biorender.com.

Viral metagenomics data analysis

The large amount of data that needs to be handled after generation of sequence reads is often a bottleneck, since specifically trained personnel, hardware (both sequencing platform and computing methods) and software is required. Data analysis of up to millions of sequence reads usually starts with removing bad quality sequence reads and sequence adapters. These adapters are DNA fragments that aid sequencing but do not provide information about the original content of the sample. It is optional to first filter out host material by means of mapping to the human genome or to first assemble reads into genome parts called contigs. Thereafter, read classification takes place by a taxonomic classifier to lay out the virome present in a sample. If one is additionally interested to zoom in at the genome of specific viruses that were detected, immediate mapping against the reference genomes can be performed, followed by variant calling analysis if more specific characterization is needed. If the user is interested in completely novel viruses, a *de novo* assembly step can be performed after the first step.

Bioinformatic sequence analyses tools are mainly orientated on human genetics and most taxonomic classification tools have originally been designed for bacteria. Therefore, analysis and optimization of sensitivity and specificity specifically for viral metagenomic testing is needed. Furthermore, the impact of filtering out host reads on accuracy needs to be investigated for viral metagenomics, and the possibilities for quantification and typing by means of NGS.

Thesis aim

The research in this thesis has several aims. Firstly, establishing the diagnostic yield of viral metagenomics in specific patient populations: patients suspected of encephalitis and travellers returning with febrile illness. Secondly, the identification, typing and quantification of viruses by means of viral metagenomics as a diagnostic tool is evaluated. Another aim is to improve sensitivity and specificity of the wet and dry (bioinformatic) lab components of viral metagenomics, in order to achieve a better performance of the method in clinical practice.

Lastly, we investigated the best methods and approaches of performing genetic analysis of just one viral genome.

Outline of this thesis

This thesis is focusing on diagnostic yield, clinical findings and enhancing technical opportunities in viral metagenomics. **Chapter 7 and 8** are devoted to **whole genome sequencing** of one specific viral genome by means of metagenomics, and a comparison of sequencing methods of SARS-CoV-2.

Chapter 2 is focused on the estimated **diagnostic yield**, the number of extra viruses that can be found using metagenomics after traditional test remain negative in cases of meningoencephalitis. In this invited review, the technical and bioinformatic advances of viral metagenomics and the remaining challenges are explained.

To further enhance sensitivity, shotgun metagenomic sequencing and sequencing with viral capture probes was compared in a cohort of **encephalitis patients** with a known virus in **chapter 3**. In this chapter, an additional cohort of adult and paediatric hematological patients without etiologic agent detected by conventional assays was assessed using metagenomic sequencing.

Chapter 4 describes a metagenomics protocol with viral capture probes that was applied on a cohort of **international travellers** with febrile illness. We focus on confirming and typing of the original positive test results, and on detection of viruses that remained undetected using traditional assays.

Almost a billion sequence reads generated for 88 respiratory samples were used to assess the performance of various bioinformatic **taxonomic classification** tools based on the original qPCR results in **chapter 5**.

In **chapter 6** a metagenomics protocol is applied to **quantify** the number of viruses in transplant patients over the course of the disease. Thus, in addition to establishing the type of virus, the number of viral particles was assessed.

In order to assess the performance of a metagenomic protocol for **virus discovery** directly in a patient sample, viral metagenomic sequencing is performed on clinical samples containing SARS-CoV, MERS-CoV and SARS-CoV-2 with viral databases from the time of original discovery. In **chapter 7,** we explain the process, and the steps taken for virus discovery in a clinical setting.

In chapter 8 it is described how **SARS-CoV-2** samples were handled by one metagenomic sequencing and four amplicon-based **WGS protocols** of three different sequence platforms to assess the performance for analyses of this one specific genome.

Chapter 9 contains the **general discussion** on the methodological breakthroughs, remaining challenges in viral metagenomics and viral sequencing. In addition, the future opportunities for metagenomic NGS in the future viral or molecular diagnostic laboratories are discussed.

References

[1] G. Lasso et al., 'A Structure-Informed Atlas of Human-Virus Interactions', Cell, vol. 178, no. 6, pp. 1526-1541.e16, Sep. 2019, doi: 10.1016/j.cell.2019.08.005.

[2] M. Breitbart and F. Rohwer, 'Here a virus, there a virus, everywhere the same virus?', Trends in Microbiology, vol. 13, no. 6, pp. 278–284, Jun. 2005, doi: 10.1016/j.tim.2005.04.003.

[3] L. H. Taylor, S. M. Latham, and M. E. J. woolhouse, 'Risk factors for human disease emergence', Phil. Trans. R. Soc. Lond. B, vol. 356, no. 1411, pp. 983–989, Jul. 2001, doi: 10.1098/rstb.2001.0888.

[4] Z. Diao, D. Han, R. Zhang, and J. Li, 'Metagenomics next-generation sequencing tests take the stage in the diagnosis of lower respiratory tract infections', Journal of Advanced Research, vol. 38, pp. 201–212, May 2022, doi: 10.1016/j.jare.2021.09.012.

[5] K. N. Govender, T. L. Street, N. D. Sanderson, and D. W. Eyre, 'Metagenomic Sequencing as a Pathogen-Agnostic Clinical Diagnostic Tool for Infectious Diseases: a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies', J Clin Microbiol, vol. 59, no. 9, pp. e02916-20, Aug. 2021, doi: 10.1128/JCM.02916-20.

[6] I. Barberis, N. L. Bragazzi, L. Galluzzo, and M. Martini, 'The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus', J Prev Med Hyg, vol. 58, no. 1, pp. E9–E12, Mar. 2017.

[7] M. Mohammadpour, M. Abrishami, A. Masoumi, and H. Hashemi, 'Trachoma: Past, present and future', Journal of Current Ophthalmology, vol. 28, no. 4, pp. 165–169, Dec. 2016, doi: 10.1016/j.joco.2016.08.011.

[8] D. Huremović, 'Brief History of Pandemics (Pandemics Throughout History)', in Psychiatry of Pandemics, D. Huremović, Ed. Cham: Springer International Publishing, 2019, pp. 7–35. doi: 10.1007/978-3-030-15346-5_2.

[9] J. K. Taubenberger and D. M. Morens, '1918 Influenza: the Mother of All Pandemics', Emerg. Infect. Dis., vol. 12, no. 1, pp. 15–22, Jan. 2006, doi: 10.3201/eid1209.05-0979.

[10] 'UNAIDS statistics', Accessed: Aug. 03, 2022. [Online]. Available: https://www.unaids.org/en/resources/fact-sheet

[11] C. G. Kurland and S. G. E. Andersson, 'Origin and Evolution of the Mitochondrial Proteome', Microbiol Mol Biol Rev, vol. 64, no. 4, pp. 786–820, Dec. 2000, doi: 10.1128/MMBR.64.4.786-820.2000.

[12] M. Horie et al., 'Endogenous non-retroviral RNA virus elements in mammalian genomes', Nature, vol. 463, no. 7277, pp. 84–87, Jan. 2010, doi: 10.1038/nature08695.

[13] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon, 'The Human Microbiome Project', Nature, vol. 449, no. 7164, pp. 804–810, Oct. 2007, doi: 10.1038/nature06244.

[14] R. M. May and R. M. Anderson, 'Endemic infections in growing populations', Mathematical Biosciences, vol. 77, no. 1–2, pp. 141–156, Dec. 1985, doi: 10.1016/0025-5564(85)90093-8.

[15] R. M. Anderson and R. M. May, Eds., Population Biology of Infectious Diseases: Report of the Dahlem Workshop on Population Biology of Infectious Disease Agents Berlin 1982, March 14 – 19. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982. doi: 10.1007/978-3-642-68635-1.

[16] 'Harappan civilization building a brick sewer system', Accessed: Aug. 03, 2022. [Online]. Available: https://www.jstor.org/stable/26426051

[17] J. F. Lindahl and D. Grace, 'The consequences of human actions on risks for infectious diseases: a review', Infection Ecology & Epidemiology, vol. 5, no. 1, p. 30048, Jan. 2015, doi: 10.3402/iee.v5.30048.

[18] C. Connolly, R. Keil, and S. H. Ali, 'Extended urbanisation and the spatialities of infectious disease: Demographic change, infrastructure and governance', Urban Studies, vol. 58, no. 2, pp. 245–263, Feb. 2021, doi: 10.1177/0042098020910873.

[19] 'Zoonotic origin', Accessed: Aug. 03, 2022. [Online]. Available: https://iris.wpro.who.int/bitstream/handle/10665.1/7819/9789290615040_eng.pdf

[20] K. Dhama et al., 'SARS-CoV-2 jumping the species barrier: Zoonotic lessons from SARS, MERS and recent advances to combat this pandemic virus', Travel Medicine and Infectious Disease, vol. 37, p. 101830, Sep. 2020, doi: 10.1016/j.tmaid.2020.101830.

[21] O. Benedictow, 'The black death: the greatest catastrophe ever.', Hist Today, p. 55(3):42–9, 2005.

[22] W. Scheidel, The great leveler: violence and the history of inequality from the Stone Age to the twenty-first century. Princeton, New Jersey: Princeton University Press, 2017.

[23] N. Zhu et al., 'A Novel Coronavirus from Patients with Pneumonia in China, 2019', N Engl J Med, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[24] C. Huang et al., 'Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China', The Lancet, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[25] P. Zhou et al., 'Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin', Microbiology, preprint, Jan. 2020. doi: 10.1101/2020.01.22.914952.

[26] 'CDR', Accessed: Aug. 03, 2022. [Online]. Available: ttps://www.who.int/data/maternal-newborn-child-adolescent-ageing/indicator-explorer-new/mca/crude-death-rate-(deaths-per-1000-population)

[27] C. L. Gibbons et al., 'Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods', BMC Public Health, vol. 14, no. 1, p. 147, Dec. 2014, doi: 10.1186/1471-2458-14-147.

[28] 'https://covid19.who.int', Accessed: Aug. 03, 2022. [Online]. Available: https://covid19.who.int

[29] S. Soneji, H. Beltrán-Sánchez, J. W. Yang, and C. Mann, 'Population-level mortality burden from novel coronavirus (COVID-19) in Europe and North America', Genus, vol. 77, no. 1, p. 7, Dec. 2021, doi: 10.1186/s41118-021-00115-9.

[30] D. M. Parkin, 'The global health burden of infection-associated cancers in the year 2002', Int. J. Cancer, vol. 118, no. 12, pp. 3030–3044, Jun. 2006, doi: 10.1002/ijc.21731.

[31] A. Leewenhoeck, 'Observations, Communicated to the Publisher by Mr. Antony van Leewenhoeck, in a Dutch Letter of the 9th of Octob. 1676. Here English'd: concerning Little Animals by Him Observed in Rain-Well-Sea. and Snow Water; as Also in Water Wherein Pepper Had Lain Infused', Philosophical Transactions, vol. Vol. 12 (1677-1678), pp. 821–831.

[32] S. R. Lakhani, 'Early clinical pathologists: Robert Koch (1843-1910).', Journal of Clinical Pathology, vol. 46, no. 7, pp. 596–598, Jul. 1993, doi: 10.1136/jcp.46.7.596.

[33] S. M. Blevins and M. S. Bronze, 'Robert Koch and the "golden age" of bacteriology', International Journal of Infectious Diseases, vol. 14, no. 9, pp. e744–e751, Sep. 2010, doi: 10.1016/j.ijid.2009.12.003.

[34] R. Koch, Xth International Congress of Medicine, vol. Berlin, 1890.

[35] A. Kammen, 'Beijerinck's contribution to the virus concept — an introduction', in 100 Years of Virology, C. H. Calisher and M. C. Horzinek, Eds. Vienna: Springer Vienna, 1999, pp. 1–8. doi: 10.1007/978-3-7091-6425-9_1.

[36] C. R. Woese and G. E. Fox, 'Phylogenetic structure of the prokaryotic domain: The primary kingdoms', Proc. Natl. Acad. Sci. U.S.A., vol. 74, no. 11, pp. 5088–5090, Nov. 1977, doi: 10.1073/pnas.74.11.5088.

[37] J. L. Stein, T. L. Marsh, K. Y. Wu, H. Shizuya, and E. F. DeLong, 'Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon', J Bacteriol, vol. 178, no. 3, pp. 591–599, Feb. 1996, doi: 10.1128/jb.178.3.591-599.1996.

[38] J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman, 'Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products', Chemistry & Biology, vol. 5, no. 10, pp. R245–R249, Oct. 1998, doi: 10.1016/S1074-5521(98)90108-9.

[39] M. Breitbart et al., 'Metagenomic Analyses of an Uncultured Viral Community from Human Feces', J Bacteriol, vol. 185, no. 20, pp. 6220–6223, Oct. 2003, doi: 10.1128/JB.185.20.6220-6223.2003.

[40] A. Escobar-Zepeda et al., 'Analysis of sequencing strategies and tools for taxonomic annotation: Defining standards for progressive metagenomics', Sci Rep, vol. 8, no. 1, p. 12034, Dec. 2018, doi: 10.1038/s41598-018-30515-5.

[41] R. Dahm, 'Friedrich Miescher and the discovery of DNA', Developmental Biology, vol. 278, no. 2, pp. 274–288, Feb. 2005, doi: 10.1016/j.ydbio.2004.11.028.

[42] J. D. Watson and F. H. C. Crick, 'Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid', Nature, vol. 171, no. 4356, pp. 737–738, Apr. 1953, doi: 10.1038/171737a0.

[43] B. Maddox, 'The double helix and the "wronged heroine"', Nature, vol. 421, no. 6921, pp. 407–408, Jan. 2003, doi: 10.1038/nature01399.

[44] R. W. Holley, G. A. Everett, J. T. Madison, and A. Zamir, 'NUCLEOTIDE SEQUENCES IN THE YEAST ALANINE TRANSFER RIBONUCLEIC ACID', J Biol Chem, vol. 240, pp. 2122–2128, May 1965.

[45] F. H. C. Crick, 'The origin of the genetic code', Journal of Molecular Biology, vol. 38, no. 3, pp. 367–379, Dec. 1968, doi: 10.1016/0022-2836(68)90392-6.

[46] W. M. Jou, G. Haegeman, M. Ysebaert, and W. Fiers, 'Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein', Nature, vol. 237, no. 5350, pp. 82–88, May 1972, doi: 10.1038/237082a0.

[47] F. Sanger, S. Nicklen, and A. R. Coulson, 'DNA sequencing with chain-terminating inhibitors', Proc. Natl. Acad. Sci. U.S.A., vol. 74, no. 12, pp. 5463–5467, Dec. 1977, doi: 10.1073/pnas.74.12.5463.

[48] F. Sanger et al., 'Nucleotide sequence of bacteriophage ΦX174 DNA', Nature, vol. 265, no. 5596, pp. 687–695, Feb. 1977, doi: 10.1038/265687a0.

[49] S. Mukherjee, M. Huntemann, N. Ivanova, N. C. Kyrpides, and A. Pati, 'Large-scale contamination of microbial isolate genomes by Illumina PhiX control', Stand in Genomic Sci, vol. 10, no. 1, p. 18, Dec. 2015, doi: 10.1186/1944-3277-10-18.

[50] 'Computer processing of DNA sequence data', Journal of Molecular Biology, vol. 116, no. 1, pp. 29–30, Oct. 1977, doi: 10.1016/0022-2836(77)90116-4.

[51] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich, 'Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction', Cold Spring Harbor Symposia on Quantitative Biology, vol. 51, no. 0, pp. 263–273, Jan. 1986, doi: 10.1101/SQB.1986.051.01.032.

[52] L. M. Smith et al., 'Fluorescence detection in automated DNA sequence analysis', Nature, vol. 321, no. 6071, pp. 674–679, Jun. 1986, doi: 10.1038/321674a0.

[53] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén, and P. Nyrén, 'Real-Time DNA Sequencing Using Detection of Pyrophosphate Release', Analytical Biochemistry, vol. 242, no. 1, pp. 84–89, Nov. 1996, doi: 10.1006/abio.1996.0432.

[54] J. C. Venter et al., 'The Sequence of the Human Genome', Science, vol. 291, no. 5507, pp. 1304–1351, Feb. 2001, doi: 10.1126/science.1058040.

[55] J. Schmutz et al., 'Quality assessment of the human genome sequence', Nature, vol. 429, no. 6990, pp. 365–368, May 2004, doi: 10.1038/nature02390.

[56] International Human Genome Sequencing Consortium et al., 'Initial sequencing and analysis of the human genome', Nature, vol. 409, no. 6822, pp. 860–921, Feb. 2001, doi: 10.1038/35057062.

[57] D. A. Wheeler et al., 'The complete genome of an individual by massively parallel DNA sequencing', Nature, vol. 452, no. 7189, pp. 872–876, Apr. 2008, doi: 10.1038/nature06884.

[58] J. M. Heather and B. Chain, 'The sequence of sequencers: The history of sequencing DNA', Genomics, vol. 107, no. 1, pp. 1–8, Jan. 2016, doi: 10.1016/j.ygeno.2015.11.003.

[59] E. Check Hayden, 'Genome sequencing: the third generation', Nature, p. news.2009.86, Feb. 2009, doi: 10.1038/news.2009.86.

[60] J. Quick, A. R. Quinlan, and N. J. Loman, 'A reference bacterial genome dataset generated on the MinIONTM portable single-molecule nanopore sequencer', GigaSci, vol. 3, no. 1, p. 22, Dec. 2014, doi: 10.1186/2047-217X-3-22.

[61] N. J. Loman, J. Quick, and J. T. Simpson, 'A complete bacterial genome assembled de novo using only nanopore sequencing data', Nat Methods, vol. 12, no. 8, pp. 733–735, Aug. 2015, doi: 10.1038/nmeth.3444.

[62] S. Nurk et al., 'The complete sequence of a human genome', Science, vol. 376, no. 6588, pp. 44–53, Apr. 2022, doi: 10.1126/science.abj6987.

[63] E. Check Hayden, 'Technology: The $1,000 genome', Nature, vol. 507, no. 7492, pp. 294–295, Mar. 2014, doi: 10.1038/507294a.

[64] G. Almogy et al., 'Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform', Genomics, preprint, May 2022. doi: 10.1101/2022.05.29.493900.

[65] G. Moody, Digital code of life: how bioinformatics is revolutionizing science, medicine, and business. Hoboken, N.J: Wiley, 2004.

[66] M. O. Dayhoff and R. S. Ledley, 'Comprotein: a computer program to aid primary protein structure determination', in Proceedings of the December 4-6, 1962, fall joint computer conference on - AFIPS '62 (Fall), Philadelphia, Pennsylvania, 1962, pp. 262–274. doi: 10.1145/1461518.1461546.

[67] M. O. Dayhoff and National Biomedical Research Foundation., Atlas of protein sequence and structure. Silver Spring. Md: National Biomedical Research Foundation, 1969.

[68] J. Gauthier, A. T. Vincent, S. J. Charette, and N. Derome, 'A brief history of bioinformatics', Briefings in Bioinformatics, vol. 20, no. 6, pp. 1981–1996, Nov. 2019, doi: 10.1093/bib/bby063.

[69] R. Staden, 'A strategy of DNA sequencing employing computer programs', Nucl Acids Res, vol. 6, no. 7, pp. 2601–2610, 1979, doi: 10.1093/nar/6.7.2601.

[70] A. Cornish-Bowden, 'Nomenclature for incompletely specified bases in nucleic acid sequences: rcommendations 1984', Nucl Acids Res, vol. 13, no. 9, pp. 3021–3030, 1985, doi: 10.1093/nar/13.9.3021.

[71] 'Staden', Accessed: Aug. 03, 2022. [Online]. Available: http://staden.sourceforge.net

[72] D. Botstein, R. L. White, M. Skolnick, and R. W. Davis, 'Construction of a genetic linkage map in man using restriction fragment length polymorphisms', Am J Hum Genet, vol. 32, no. 3, pp. 314–331, May 1980.

[73] J. F. Gusella et al., 'A polymorphic DNA marker genetically linked to Huntington's disease', Nature, vol. 306, no. 5940, pp. 234–238, Nov. 1983, doi: 10.1038/306234a0.

[74] R. L. Baehner et al., 'DNA linkage analysis of X chromosome-linked chronic granulomatous disease.', Proc. Natl. Acad. Sci. U.S.A., vol. 83, no. 10, pp. 3398–3401, May 1986, doi: 10.1073/pnas.83.10.3398.

[75] S. B. Ng et al., 'Targeted capture and massively parallel sequencing of 12 human exomes', Nature, vol. 461, no. 7261, pp. 272–276, Sep. 2009, doi: 10.1038/nature08250.

[76] M. Choi et al., 'Genetic diagnosis by whole exome capture and massively parallel DNA sequencing', Proc. Natl. Acad. Sci. U.S.A., vol. 106, no. 45, pp. 19096–19101, Nov. 2009, doi: 10.1073/pnas.0910672106.

[77] J. R. Lupski et al., 'Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy', N Engl J Med, vol. 362, no. 13, pp. 1181–1191, Apr. 2010, doi: 10.1056/NEJMoa0908094.

[78] J. C. Venter et al., 'Environmental Genome Shotgun Sequencing of the Sargasso Sea', Science, vol. 304, no. 5667, pp. 66–74, Apr. 2004, doi: 10.1126/science.1093857.

[79] R. A. Edwards and F. Rohwer, 'Viral metagenomics', Nat Rev Microbiol, vol. 3, no. 6, pp. 504–510, Jun. 2005, doi: 10.1038/nrmicro1163.

[80] E. L. Delwart, 'Viral metagenomics', Rev. Med. Virol., vol. 17, no. 2, pp. 115–131, Mar. 2007, doi: 10.1002/rmv.532.

[81] S. Svraka, K. Rosario, E. Duizer, H. van der Avoort, M. Breitbart, and M. Koopmans, 'Metagenomic sequencing for virus identification in a public-health setting', Journal of General Virology, vol. 91, no. 11, pp. 2846–2856, Nov. 2010, doi: 10.1099/vir.0.024612-0.

[82] S. J. Anthony et al., 'A Strategy To Estimate Unknown Viral Diversity in Mammals', mBio, vol. 4, no. 5, pp. e00598-13, Nov. 2013, doi: 10.1128/mBio.00598-13.

[83] J. R. Brown, T. Bharucha, and J. Breuer, 'Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases', Journal of Infection, vol. 76, no. 3, pp. 225–240, Mar. 2018, doi: 10.1016/j.jinf.2017.12.014.

[84] C. Y. Chiu et al., 'Diagnosis of Fatal Human Case of St. Louis Encephalitis Virus Infection by Metagenomic Sequencing, California, 2016', Emerg. Infect. Dis., vol. 23, no. 10, pp. 1964–1968, Oct. 2017, doi: 10.3201/eid2310.161986.

[85] A. W. D. Edridge et al., 'Novel Orthobunyavirus Identified in the Cerebrospinal Fluid of a Ugandan Child With Severe Encephalopathy', Clinical Infectious Diseases, vol. 68, no. 1, pp. 139–142, Jan. 2019, doi: 10.1093/cid/ciy486.

[86] H. Fridholm et al., 'Human pegivirus detected in a patient with severe encephalitis using a metagenomic pan-virus array', Journal of Clinical Virology, vol. 77, pp. 5–8, Apr. 2016, doi: 10.1016/j.jcv.2016.01.013.

[87] B. Hoffmann et al., 'A Variegated Squirrel Bornavirus Associated with Fatal Human Encephalitis', N Engl J Med, vol. 373, no. 2, pp. 154–162, Jul. 2015, doi: 10.1056/NEJMoa1415627.

[88] D. Lipowski et al., 'A Cluster of Fatal Tick-borne Encephalitis Virus Infection in Organ Transplant Setting', The Journal of Infectious Diseases, vol. 215, no. 6, pp. 896–901, Mar. 2017, doi: 10.1093/infdis/jix040.

[89] N. T. H. Mai et al., 'Central Nervous System Infection Diagnosis by Next-Generation Sequencing: A Glimpse Into the Future?', Open Forum Infectious Diseases, vol. 4, no. 2, p. ofx046, Apr. 2017, doi: 10.1093/ofid/ofx046.

[90] J. A. Murkey et al., 'Hepatitis E Virus–Associated Meningoencephalitis in a Lung Transplant Recipient Diagnosed by Clinical Metagenomic Sequencing', Open Forum Infectious Diseases, vol. 4, no. 3, p. ofx121, Jul. 2017, doi: 10.1093/ofid/ofx121.

[91] S. N. Naccache et al., 'Diagnosis of Neuroinvasive Astrovirus Infection in an Immunocompromised Adult With Encephalitis by Unbiased Next-Generation Sequencing', Clinical Infectious Diseases, vol. 60, no. 6, pp. 919–923, Mar. 2015, doi: 10.1093/cid/ciu912.

[92] K. Perlejewski et al., 'Next-generation sequencing (NGS) in the identification of encephalitis-causing viruses: Unexpected detection of human herpesvirus 1 while searching for RNA pathogens', Journal of Virological Methods, vol. 226, pp. 1–6, Dec. 2015, doi: 10.1016/j.jviromet.2015.09.010.

[93] A. Piantadosi et al., 'Rapid Detection of Powassan Virus in a Patient With Encephalitis by Metagenomic Sequencing', Clinical Infectious Diseases, vol. 66, no. 5, pp. 789–792, Feb. 2018, doi: 10.1093/cid/cix792.

[94] F. Tschumi et al., 'Meningitis and epididymitis caused by Toscana virus infection imported to Switzerland diagnosed by metagenomic sequencing: a case report', BMC Infect Dis, vol. 19, no. 1, p. 591, Dec. 2019, doi: 10.1186/s12879-019-4231-9.

[95] M. R. Wilson et al., 'A novel cause of chronic viral meningoencephalitis: Cache Valley virus: Orthobunyavirus Encephalitis', Ann Neurol., vol. 82, no. 1, pp. 105–114, Jul. 2017, doi: 10.1002/ana.24982.

[96] D. Saylor, K. Thakur, and A. Venkatesan, 'Acute encephalitis in the immunocompromised individual', Current Opinion in Infectious Diseases, vol. 28, no. 4, pp. 330–336, Aug. 2015, doi: 10.1097/QCO.0000000000000175.

[97] M. R. Wilson et al., 'Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis', N Engl J Med, vol. 380, no. 24, pp. 2327–2340, Jun. 2019, doi: 10.1056/NEJMoa1803396.

[98] Kufner et al., 'Two Years of Viral Metagenomics in a Tertiary Diagnostics Unit: Evaluation of the First 105 Cases', Genes, vol. 10, no. 9, p. 661, Aug. 2019, doi: 10.3390/genes10090661.

[99] B. M. O'Flaherty et al., 'Comprehensive viral enrichment enables sensitive respiratory virus genomic identification and analysis by next generation sequencing', Genome Res., vol. 28, no. 6, pp. 869–877, Jun. 2018, doi: 10.1101/gr.226316.117.

[100] T. N. Wylie, K. M. Wylie, B. N. Herter, and G. A. Storch, 'Enhanced virome sequencing using targeted sequence capture', Genome Res., vol. 25, no. 12, pp. 1910–1920, Dec. 2015, doi: 10.1101/gr.191049.115.

[101] J. S. Kalpoe, A. C. M. Kroes, S. Verkerk, E. C. J. Claas, R. M. Y. Barge, and M. F. C. Beersma, 'Clinical relevance of quantitative varicella-zoster virus (VZV) DNA detection in plasma after stem cell transplantation', Bone Marrow Transplant, vol. 38, no. 1, pp. 41–46, Jul. 2006, doi: 10.1038/sj.bmt.1705397.

[102] L.-A. Le, T.-P. Nguyen-Hoang, V.-P. Huynh, T.-H. Nguyen, T.-V. Nguyen, and T.-D. Ho-Huynh, 'Microbiome dataset analysis from a shrimp pond in Ninh Thuan, Vietnam using shotgun metagenomics', Data in Brief, vol. 31, p. 105731, Aug. 2020, doi: 10.1016/j.dib.2020.105731.

[103] E. C. Carbo et al., 'Improved diagnosis of viral encephalitis in adult and pediatric hematological patients using viral metagenomics', Journal of Clinical Virology, vol. 130, p. 104566, Sep. 2020, doi: 10.1016/j.jcv.2020.104566.

[104] T. Doan et al., 'Metagenomic DNA Sequencing for the Diagnosis of Intraocular Infections', Ophthalmology, vol. 124, no. 8, pp. 1247–1248, Aug. 2017, doi: 10.1016/j.ophtha.2017.03.045.

[105] J. C. Haston et al., 'Prospective Cohort Study of Next-Generation Sequencing as a Diagnostic Modality for Unexplained Encephalitis in Children', Journal of the Pediatric Infectious Diseases Society, vol. 9, no. 3, pp. 326–333, Jul. 2020, doi: 10.1093/jpids/piz032.

[106] H. Jerome et al., 'Metagenomic next-generation sequencing aids the diagnosis of viral infections in febrile returning travellers', Journal of Infection, vol. 79, no. 4, pp. 383–388, Oct. 2019, doi: 10.1016/j.jinf.2019.08.003.

[107] H. Johansson, D. Bzhalava, J. Ekström, E. Hultin, J. Dillner, and O. Forslund, 'Metagenomic sequencing of "HPV-negative" condylomas detects novel putative HPV types', Virology, vol. 440, no. 1, pp. 1–7, May 2013, doi: 10.1016/j.virol.2013.01.023.

[108] D. W. Lewandowska et al., 'Unbiased metagenomic sequencing complements specific routine diagnostic methods and increases chances to detect rare viral strains', Diagnostic Microbiology and Infectious Disease, vol. 83, no. 2, pp. 133–138, Oct. 2015, doi: 10.1016/j.diagmicrobio.2015.06.017.

[109] N. Madi, W. Al-Nakib, A. S. Mustafa, and N. Habibi, 'Metagenomic analysis of viral diversity in respiratory samples from patients with respiratory tract infections in Kuwait', J Med Virol, vol. 90, no. 3, pp. 412–420, Mar. 2018, doi: 10.1002/jmv.24984.

[110] N. E. Moore et al., 'Metagenomic Analysis of Viruses in Feces from Unsolved Outbreaks of Gastroenteritis in Humans', J Clin Microbiol, vol. 53, no. 1, pp. 15–21, Jan. 2015, doi: 10.1128/JCM.02029-14.

[111] U. Pérez-Sautu et al., 'Target-independent high-throughput sequencing methods provide evidence that already known human viral pathogens play a main role in respiratory infections with unexplained etiology', Emerging Microbes & Infections, vol. 8, no. 1, pp. 1054–1065, Jan. 2019, doi: 10.1080/22221751.2019.1640587.

[112] A. Ramesh et al., 'Metagenomic next-generation sequencing of samples from pediatric febrile illness in Tororo, Uganda', PLoS ONE, vol. 14, no. 6, p. e0218318, Jun. 2019, doi: 10.1371/journal.pone.0218318.

[113] A. Reyes et al., 'Viral metagenomic sequencing in a cohort of international travellers returning with febrile illness', Journal of Clinical Virology, vol. 143, p. 104940, Oct. 2021, doi: 10.1016/j.jcv.2021.104940.

[114] S. Saha et al., 'Unbiased Metagenomic Sequencing for Pediatric Meningitis in Bangladesh Reveals Neuroinvasive Chikungunya Virus Outbreak and Other Unrealized Pathogens', mBio, vol. 10, no. 6, pp. e02877-19, /mbio/10/6/mBio.02877-19.atom, Dec. 2019, doi: 10.1128/mBio.02877-19.

[115] R. Schlaberg et al., 'Viral Pathogen Detection by Metagenomics and Pan-Viral Group Polymerase Chain Reaction in Children With Pneumonia Lacking Identifiable Etiology', The Journal of Infectious Diseases, vol. 215, no. 9, pp. 1407–1415, May 2017, doi: 10.1093/infdis/jix148.

[116] S. L. Smits et al., 'Novel Cyclovirus in Human Cerebrospinal Fluid, Malawi, 2010–2011', Emerg. Infect. Dis., vol. 19, no. 9, Sep. 2013, doi: 10.3201/eid1909.130404.

[117] S. L. Smits et al., 'New Viruses in Idiopathic Human Diarrhea Cases, the Netherlands', Emerg. Infect. Dis., vol. 20, no. 7, Jul. 2014, doi: 10.3201/eid2007.140190.

[118] B. Taboada et al., 'Is There Still Room for Novel Viral Pathogens in Pediatric Respiratory Tract Infections?', PLoS ONE, vol. 9, no. 11, p. e113570, Nov. 2014, doi: 10.1371/journal.pone.0113570.

[119] N. Thi Kha Tu et al., 'The Virome of Acute Respiratory Diseases in Individuals at Risk of Zoonotic Infections', Viruses, vol. 12, no. 9, p. 960, Aug. 2020, doi: 10.3390/v12090960.

[120] P. Turner et al., 'The aetiologies of central nervous system infections in hospitalised Cambodian children', BMC Infect Dis, vol. 17, no. 1, p. 806, Dec. 2017, doi: 10.1186/s12879-017-2915-6.

[121] H. Wang et al., 'Clinical diagnostic application of metagenomic next-generation sequencing in children with severe nonresponding pneumonia', PLoS ONE, vol. 15, no. 6, p. e0232610, Jun. 2020, doi: 10.1371/journal.pone.0232610.

[122] L. Xu et al., 'Characterization of the nasopharyngeal viral microbiome from children with community-acquired pneumonia but negative for Luminex xTAG respiratory viral panel assay detection', J Med Virol, vol. 89, no. 12, pp. 2098–2107, Dec. 2017, doi: 10.1002/jmv.24895.

[123] N. L. Yozwiak, P. Skewes-Cox, M. D. Stenglein, A. Balmaseda, E. Harris, and J. L. DeRisi, 'Virus Identification in Unknown Tropical Febrile Illness Cases Using Deep Sequencing', PLoS Negl Trop Dis, vol. 6, no. 2, p. e1485, Feb. 2012, doi: 10.1371/journal.pntd.0001485.

[124] Y. Zhou et al., 'Metagenomics Study of Viral Pathogens in Undiagnosed Respiratory Specimens and Identification of Human Enteroviruses at a Thailand Hospital', The American Journal of Tropical Medicine and Hygiene, vol. 95, no. 3, pp. 663–669, Sep. 2016, doi: 10.4269/ajtmh.16-0062.

[125] X. Zou et al., 'Simultaneous virus identification and characterization of severe unexplained pneumonia cases using a metagenomics sequencing technique', Sci. China Life Sci., vol. 60, no. 3, pp. 279–286, Mar. 2017, doi: 10.1007/s11427-016-0244-8.

[126] S. I. Sardi et al., 'Coinfections of Zika and Chikungunya Viruses in Bahia, Brazil, Identified by Metagenomic Next-Generation Sequencing', J Clin Microbiol, vol. 54, no. 9, pp. 2348–2353, Sep. 2016, doi: 10.1128/JCM.00877-16.

[127] 'JASP', Accessed: Mar. 23, 2021. [Online]. Available: https://jasp-stats.org

[128] C. Rodriguez et al., 'Fatal Encephalitis Caused by Cristoli Virus, an Emerging Orthobunyavirus, France', Emerg Infect Dis, vol. 26, no. 6, pp. 1287–1290, Jun. 2020, doi: 10.3201/eid2606.191431.

[129] C. Rodriguez et al., 'Viral genomic, metagenomic and human transcriptomic characterization and prediction of the clinical forms of COVID-19', PLoS Pathog, vol. 17, no. 3, p. e1009416, Mar. 2021, doi: 10.1371/journal.ppat.1009416.

[130] J. R. Brown et al., 'Astrovirus VA1/HMO-C: An Increasingly Recognized Neurotropic Pathogen in Immunocompromised Patients', Clinical Infectious Diseases, vol. 60, no. 6, pp. 881–888, Mar. 2015, doi: 10.1093/cid/ciu940.

[131] M. Alawi et al., 'DAMIAN: an open source bioinformatics tool for fast, systematic and cohort based analysis of microorganisms in diagnostic samples', Sci Rep, vol. 9, no. 1, p. 16841, Nov. 2019, doi: 10.1038/s41598-019-52881-4.

[132] M. Christopeit et al., 'Suspected encephalitis with Candida tropicalis and Fusarium detected by unbiased RNA sequencing', Ann Hematol, vol. 95, no. 11, pp. 1919–1921, Nov. 2016, doi: 10.1007/s00277-016-2770-3.

[133] F. X. López-Labrador et al., 'Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure', Journal of Clinical Virology, vol. 134, p. 104691, Jan. 2021, doi: 10.1016/j.jcv.2020.104691.

[134] C. P. Oechslin et al., 'Limited Correlation of Shotgun Metagenomics Following Host Depletion and Routine Diagnostics for Viruses and Bacteria in Low Concentrated Surrogate and Clinical Samples', Front. Cell. Infect. Microbiol., vol. 8, p. 375, Oct. 2018, doi: 10.3389/fcimb.2018.00375.

[135] A. L. van Rijn et al., 'The respiratory virome and exacerbations in patients with chronic obstructive pulmonary disease', PLoS ONE, vol. 14, no. 10, p. e0223952, Oct. 2019, doi: 10.1371/journal.pone.0223952.

[136] S. van Boheemen et al., 'Retrospective Validation of a Metagenomic Sequencing Protocol for Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric Patients', The Journal of Molecular Diagnostics, vol. 22, no. 2, pp. 196–207, Feb. 2020, doi: 10.1016/j.jmoldx.2019.10.007.

[137] J. R. Brown, T. Bharucha, and J. Breuer, 'Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases', Journal of Infection, vol. 76, no. 3, pp. 225–240, Mar. 2018, doi: 10.1016/j.jinf.2017.12.014.

[138] T. Briese et al., 'Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis', mBio, vol. 6, no. 5, pp. e01491-15, Sep. 2015, doi: 10.1128/mBio.01491-15.

[139] M. Asplund et al., 'Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries', Clinical Microbiology and Infection, vol. 25, no. 10, pp. 1277–1285, Oct. 2019, doi: 10.1016/j.cmi.2019.04.028.

[140] J. C. Wooley and Y. Ye, 'Metagenomics: Facts and Artifacts, and Computational Challenges', J. Comput. Sci. Technol., vol. 25, no. 1, pp. 71–81, Jan. 2010, doi: 10.1007/s11390-010-9306-4.

[141] A. L. Greninger, 'The challenge of diagnostic metagenomics', Expert Review of Molecular Diagnostics, vol. 18, no. 7, pp. 605–615, Jul. 2018, doi: 10.1080/14737159.2018.1487292.

[142] T. Li et al., 'Metagenomic Next-Generation Sequencing of the 2014 Ebola Virus Disease Outbreak in the Democratic Republic of the Congo', J Clin Microbiol, vol. 57, no. 9, pp. e00827-19, Sep. 2019, doi: 10.1128/JCM.00827-19.