Metagenomic Sequencing for Surveillance of Food- and Waterborne Viral Diseases

David F. Nieuwenhuijse and Marion P. G. Koopmans*

Department of Viroscience, Erasmus Medical Center, Rotterdam, Netherlands

A plethora of viruses can be transmitted by the food- and waterborne route. However, their recognition is challenging because of the variety of viruses, heterogeneity of symptoms, the lack of awareness of clinicians, and limited surveillance efforts. Classical food- and waterborne viral disease outbreaks are mainly caused by caliciviruses, but the source of the virus is often not known and the foodborne mode of transmission is difficult to discriminate from human-to-human transmission. Atypical food- and waterborne viral disease can be caused by viruses such as hepatitis A and hepatitis E. In addition, a source of novel emerging viruses with a potential to spread via the food- and waterborne route is the repeated interaction of humans with wildlife. Wildlife-to-human adaptation may give rise to self- limiting outbreaks in some cases, but when fully adjusted to the human host can be devastating. Metagenomic sequencing has been investigated as a promising solution for surveillance purposes as it detects all viruses in a single protocol, delivers additional genomic information for outbreak tracing, and detects novel unknown viruses. Nevertheless, several issues must be addressed to apply metagenomic sequencing in surveillance. First, sample preparation is difficult since the genomic material of viruses is generally overshadowed by host- and bacterial genomes. Second, several data analysis issues hamper the efficient, robust, and automated processing of metagenomic data. Third, interpretation of metagenomic data is hard, because of the lack of general knowledge of the virome in the food chain and the environment. Further developments in virus-specific nucleic acid extraction methods, bioinformatic data processing applications, and unifying data visualization tools are needed to gain insightful surveillance knowledge from suspect food samples.

Keywords: viral pathogens, metagenomics, high-throughput sequencing, surveillance, foodborne

INTRODUCTION

Transmission via the food- and waterborne route is a common mode of spread of a wide range of viruses. Many commonly recognized food- and waterborne infections are caused by viruses that are transmitted by the fecal-oral route. Particularly caliciviruses (norovirus, sapovirus) can cause diarrhea and vomiting and less commonly astroviruses, rotaviruses, and adenoviruses (Newell et al., 2010). Other viruses cause symptoms resulting from extra-intestinal spread, like hepatitis A (HAV), and hepatitis E (HEV). High levels of viral shedding through stool and vomit lead to dispersal in the environment. Moreover, the stability of many food- and waterborne viruses allows

for prolonged persistence in the environment. Food- and water associated transmission is also suspected to enhance the spread and emergence of zoonotic viruses (e.g., Middle East Respiratory Syndrome-coronavirus and Nipah virus) and facilitates the occurrence of zoonotic events though the handling of bushmeat (Ebola virus) (Wolfe et al., 2005; European Food Safety Authority, 2014; Mann et al., 2015).

Challenges of detecting viruses transmitted by the food- and waterborne route are their diversity and the frequent secondary person-to-person transmissions, which may mask an initial food- or waterborne introduction. In addition, there is a lack of awareness among clinicians (Beersma et al., 2012), as the symptoms caused by foodborne viruses are not specific to the viruses causing the illness. Furthermore, there is limited coverage in surveillance of food- and waterborne viral disease, hampering detecting and tracing (Ahmed et al., 2014; Verhoef et al., 2015).

In the past years, high-throughput sequencing technologies have increased the ability to measure genomic material from diverse samples tremendously. These methods will most likely continue to improve in the future (Aarestrup et al., 2012). Specifically, metagenomic analysis using untargeted sequencing has received a lot of attention, because the high throughput of current sequencing technologies has made it possible to obtain multiple high coverage genomes from highly complex samples (Cotten et al., 2014; Smits et al., 2015). Even though it is still a developing field, metagenomics is starting to become mature enough for applications outside of the research environment.

With the development of multiplex real-time polymerase chain reaction (RT-PCR) protocols came the realization that unraveling etiologies of main disease syndromes is more complex than previously recognized. This led to questions about the detection of viruses for which the role as causes of illness remains to be evaluated, the importance of co-infections and recognition of less common disease etiologies (Binnicker, 2015). Similarly, high throughput metagenomic sequencing broadens the scope of detectable viruses, which, apart from making it more complex, make us further understand the role of viruses in health and disease. The biggest promise, however, is that of routine application of metagenomic sequencing in diagnostic context, facilitating viral detection and offering huge potential for tracing of viruses in (foodborne) outbreaks.

RECOGNIZING FOOD- AND WATERBORNE VIRAL DISEASE

Given the number of different viral pathogens potentially associated with food- and waterborne transmission their detection has not been straightforward. Partly because many of these pathogens lack cell culture systems that are sensitive and robust enough for application in routine settings (Amar et al., 2007). The entry point for disease-based surveillance of viruses spreading by food and water is the reporting of patients presenting to a clinician. However, patients only present themselves in case of a severe symptomatic infection, or in case self-help is not sufficient. Mild symptoms are therefore generally not registered creating a bias in surveillance. This phenomenon

is captured in the surveillance pyramid (**Figure 1**), and the full extent of disease can only be captured through epidemiological studies addressing incidence and etiology at community level coupled with severity of a range of enteric pathogens (Sethi et al., 1999; de Wit et al., 2001; Tam et al., 2012). Additionally, it is challenging to distinguish between foodborne outbreaks and outbreaks caused by direct contact between humans. Classic clinical symptoms of foodborne disease vary, ranging from diarrhea and vomiting to abdominal cramps and general malaise, which makes it hard for clinicians to pinpoint the exact causative agent. This leads to misdiagnosis if the diagnostic workup is selective, and if there are no obvious signs of food-related exposure (Beersma et al., 2012). Moreover, heterogeneity in clinical interpretation can be caused by host factors, such as differences in the expression of histo-blood-group antigens that are receptors for rota- and noroviruses (Payne et al., 2015; de Graaf et al., 2016). Susceptibility to fecal-orally transmitted viruses may also be influenced by the established microbiome and virome in the host population, of which the prior is shown to differ between different locations and age groups (Yatsunenko et al., 2012). It is reasonable to think that the differences in the gut environment are more pronounced between countries with larger social and economic differences such as first and third world countries, which often differ in their resident pathogens (Ott et al., 2012; Yatsunenko et al., 2012; Hay et al., 2013). The role of the gut virome, in addition to the gut microbiome, is a relatively new concept and has been described as potentially

FIGURE 1 | Schematic representation of the phenomenon known as the "surveillance pyramid." Layers represent different categories of infected individuals. Width of the layers represents the estimated number of individuals in that category. As indicated, individuals reported by surveillance programs generally originate from the hospitalized category.
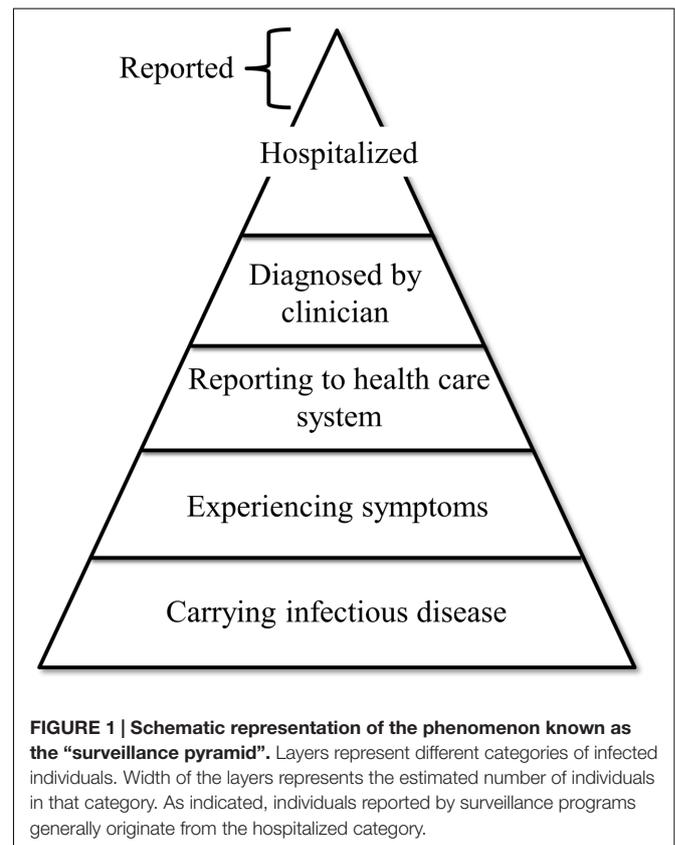
having influence on gut health and therefore expression of disease (Cadwell, 2015). Because of under and miss-diagnoses, clinical surveillance likely only captures the tip of the iceberg of food- and waterborne viral disease cases.

DETECTION OF FOOD- AND WATERBORNE VIRAL DISEASE OUTBREAKS

In cases where a cluster of patients with similar symptoms presents itself, there can be an investigation to look for epidemiological clues of the link between the cases. Additional information is garnered from the use of viral genome sequencing, making it possible to track origins of outbreaks, and to estimate how much of the observed human disease is attributable to foodborne infection by computerized linking of epidemiologic data to aligned viral genomic sequences (Verhoef et al., 2011). However, often the original source or evidence of it being food- or waterborne cannot be found, which means that outbreaks often are merely registered. Of the 941 viral disease outbreaks reported as foodborne in the joint ECDC-EFSA surveillance report of 2015, only 9.1% had robust evidence of food- or waterborne transmission (Eurosurveillance editorial team, 2015). Routine application of genotyping of HAV in newly diagnosed cases quadrupled the number of cases in which food was the most likely source of infection a 3 year enhanced surveillance study in The Netherlands, but this is not commonly done (Petrignani et al., 2014). In an investigation of 1794 food- and waterborne outbreaks in Korea, roughly 75% of the outbreaks reported in schools and public restaurants were attributed to an unknown origin (Moon et al., 2014). Availability and costs of molecular testing combined with sequencing, additional to the limited success of virus detection in food products, are likely further limiting their use in food and water surveillance. This is demonstrated by the fact that formal confirmation of a viral outbreak associated with food- and waterborne transmission still requires extensive epidemiological analysis or confirmation of a virus in the infected individual, or both (ESFA, 2016). However, due to the increase of genomic information of viruses, sequence data is increasingly used to support and strengthen outbreak investigations. Nevertheless, the surveillance programs for these viruses in the human food chain is limited, in contrast with the American CDC[1] and the European ECDC[2] surveillance programs for bacteria and parasitic pathogens causing food- and waterborne diseases (Deng et al., 2016) and does not have widespread coverage. As an example, to comply with European food safety regulations, shellfish, a well-known source of foodborne pathogens, need to be tested for enteric bacteria. However, it has been well documented that shellfish that pass quality control based on bacterial counts may still contain human pathogenic viruses (Rodriguez-Manzano et al., 2014). To be able to recognize food- waterborne viral disease outbreaks and

stop underestimation of its disease burden there should be innovations in the current foodborne surveillance system.

CLASSICAL VIRUSES ASSOCIATED WITH FOOD- AND WATERBORNE DISEASES

Although the list of viruses causing acute gastroenteritis is long, norovirus ranks among the top causes of diarrheal disease (Ahmed et al., 2014). Reporting of outbreaks suggests that the food- and waterborne disease transmission route is relatively rare, but provides an underestimate, bearing in mind that it may be hard to recognize a food- and waterborne transmission route in community-acquired diarrheal disease. To quantify the burden of all diarrheal disease attributable to foodborne transmission, the World Health Organization commissioned a study that combined data from surveillance and exhaustive literature reviews with a systematic approach to calculation of the fraction of disease attributable to food contamination (Havelaar et al., 2015). This ranked the burden of norovirus illness among the top causes of foodborne disease, along with *Campylobacter*, and listed HAV associated disease among other significant causes of foodborne disease, along with *Salmonella* and *Taenia solium*.

For bacterial foodborne pathogens, the analysis of systematically collected surveillance data has been used as the basis of attribution analysis (Pires et al., 2009). A popular approach has been to quantify the proportion of foodborne disease of humans to their likely origin, by comparing diversity of strains found in human disease outbreaks with that found in animal and environmental reservoirs (Hald et al., 2007). While this model does not allow estimating the foodborne disease where food is a vehicle for person-to-person transmission, which is common for noroviruses, it has been used with some success to quantify the contribution of foodborne viral disease stemming from environmentally contaminated food (e.g., associated with shellfish; (Verhoef et al., 2015)). This builds from the observation that there is a large discrepancy between the norovirus variants in clinical settings and environmental samples (Tao et al., 2015; Kazama et al., 2016). Norovirus GII.4, found in clinical setting, is generally related to person-to-person transmission, however, several other norovirus genotypes and genogroups were found in environmental samples in the same area. However, food associated acute gastroenteritis is not limited to norovirus infections. In a large retrospective study of oyster-related acute gastroenteritis outbreaks in Osaka City in Japan 30.7% of the cases were attributed to other pathogens such aichivirus, astrovirus, sapovirus rotavirus A, and enteroviruses (Iritani et al., 2014). Furthermore, outbreaks can be caused by a mixture of these viruses and viral variants (Wang et al., 2015).

OTHER VIRUSES TRANSMITTED VIA THE FOOD- AND WATERBORNE ROUTE

Apart from viruses causing gastroenteritis, there are viruses causing food- and waterborne diseases that are associated with

---

[1] http://ecdc.europa.eu/en/healthtopics/food_and_waterborne_disease/surveillance/Pages/index.aspx

[2] http://www.cdc.gov/ncezid/dfwed/keyprograms/surveillance.html

a variety of other syndromes. The second most common disease syndrome is hepatitis, caused by HAV, a fecal-orally transmitted virus (Havelaar et al., 2015). By decreasing natural exposure in regions with low endemicity, the susceptibility of the population for outbreaks of HAV disease in these regions is increasing (Newell et al., 2010). Because of increased globalization, contamination of food products by viruses prevalent in food producing regions can increase the risk of outbreaks in these regions. Several outbreaks of HAV infection have been reported in recent years both in the USA and Europe (Gossner et al., 2014). Most of these outbreaks could be identified as foodborne infections after intense investigations (Bruni et al., 2016). Especially fresh (imported) food products (e.g., fresh frozen berries, pomegranate seeds, and sun-dried tomatoes) have been identified as sources of the virus (Gossner et al., 2014; Tavoschi et al., 2015). Tracking the foodborne source of infection is challenging for HAV, because of an underestimation of the contribution of food as a source of infection due to the long incubation period in infected individuals (Petrignani et al., 2014).

Another foodborne virus gaining increased attention is zoonotic HEV, associated with genotype 3 and 4 HEV. HEV is widespread in commercially held pigs, as well as in wild pigs, and deer (Guillois et al., 2016). Human disease with genotype 3 HEV is increasingly recognized, but in the large majority of the cases the source of the virus is unknown (Lewis et al., 2010). There is clear evidence that food can be a source of zoonotic HEV infections. Outbreaks that have been confirmed to be caused by foodborne transmission of the virus by consumption of wild meat from boar, deer, and rabbit (Tei et al., 2003; Izopet et al., 2012; Guillois et al., 2016). Several studies have shown the zoonotic potential of HEV from pigs (Teixeira et al., 2016), HEV can also be readily detected in pork products such as dried meats and liver sausages (Di Bartolo et al., 2015). A large proportion of food-related HEV infections, however, does not lead to hospitalization of the patient, leading to under-reporting and unrecognized risk and burden of the disease (Guillois et al., 2016).

Beside viruses circulating in livestock, wildlife has the potential to be a large reservoir of unknown zoonotic viruses. Hunting, trading, preparing, and consuming so-called "bushmeat" is one of the routes by which novel viruses can be introduced into the human population (Karesh and Noble, 2009). It may be difficult to disentangle foodborne infection from direct zoonotic exposure, but it is important to consider local practices before ruling out food as a source of human infection. A special example are the occasional introductions of Nipah viruses from bats into humans through contamination of date palm sap which is collected in open containers to which bats that harbor these viruses have access (Rahman et al., 2012). Not proven but certainly interesting is the practice of drinking unprocessed camel urine which may contain MERS coronavirus, a practice that came to light during the investigations into sources of MERS coronavirus infection in humans (Funk et al., 2016). Even if limited in scale, small foodborne infections, originating from human-wildlife interaction, constitute as many incidents potentially pushing wildlife viruses to become human-to-human transmissible (Wolfe et al., 2005; Islam et al., 2016). In the cases of Monkeypox and Nipah this only led to small epidemics,

but when the virus is well adapted to spread from human to human this can lead to larger outbreaks, as seen during the Ebola crisis in 2015 (Wolfe et al., 2005; Mann et al., 2015). Continuing deforestation, increasing population and continued trade of bushmeat brings more humans in contact with wildlife and increases the risk of zoonosis (Karesh and Noble, 2009). Urbanization and globalization of travel and trade provides ample and increasing opportunity for further spread. Therefore, even anecdotal zoonotic introductions may constitute a public health risk, and ideally should be investigated in conjunction with the animals these humans were exposed to. As the ability to spread between humans is a key property for successful further spread, enhancing the capacity to investigate clusters of disease (in humans and animals) is important (McCloskey et al., 2014).

UNKNOWN FECAL-ORAL PASSENGERS

Bacteriophages, although not directly pathogenic to humans, could play a role in human health and disease by influencing the gut microbiome. Sequencing data from human gut samples presents a large diversity of bacteriophages in the human gut (Reyes et al., 2012). In addition to bacteriophages, untargeted sequencing of sewage samples has shown the presence of large quantities of different plant viruses (Zhang et al., 2006). Because of the presence of numerous infectious plant virus particles in human fecal waste there is ongoing research on the effect of these viruses in human health and disease (Colson et al., 2010). Similarly, there is ongoing research into the impact of bacteriophages on human health through their modulating effect on the gut microbiome (Reyes et al., 2012), and thereby, gut immunity (Honda and Littman, 2016). In what way bacteriophages protect or expose the human gut to bacterial or viral pathogens has yet to be further investigated. However, using metagenomic sequencing it will at least be possible to recognize the presence of unknown fecal-oral passengers.

METAGENOMICS FOR FOOD- AND WATERBORNE VIRAL DISEASE SURVEILLANCE

Metagenomics is a term used for experiments in which all nucleic acids in a certain sample are sequenced. For bacteria, historically, the diversity of a sample used to be expressed by performing phylogenetic analyses based on 16S ribosomal RNA (Handelsman, 2004). However, since viruses lack such a universally conserved motif, viral metagenomics refers to the attempt to recover full and partial genomes of all viruses present in the sample. Viral metagenomic analysis protocols generally start with procedures to remove host and bacterial cells followed by nuclease treatment to remove free nucleic acids. Often, the remaining nucleic acids are amplified using randomly primed (RT-) PCR and finally sequenced using high-throughput sequencing technology. Viral metagenomics has great potential in surveillance of viruses in the global food chain because of its

sensitivity, broad detection range, and detailed information of the detected virus (Aarestrup et al., 2012).

Environmental Surveillance

Metagenomic sequencing has already been used in the sampling of the world's oceans to estimate the global viral diversity (Hingamp et al., 2013). Similarly, metagenomics can be used in environments associated with viruses spread via the food- and waterborne route (**Figure 2A**), which gives an overview of all these viruses and circumvents the mentioned sampling biases. The potential of such an approach for food-related purposes was exemplified by Hellmér et al. (2014) who conducted a multi-species viral surveillance study and, albeit not metagenomic sequencing based, were able to detect several food- and waterborne viruses in sewage. Interestingly, norovirus and HAV, detected in sewage, could be related to hospitalized patients diagnosed with the viral infection in the catchment area of the sewage system. Moreover, they detected a peak in the level of norovirus several weeks before the outbreak was reported in the hospital in that area (Hellmér et al., 2014). This demonstrates the potential power of shifting the scope of surveillance of food- and waterborne viruses from the hospital to the environment. Untargeted metagenomic sequencing has been shown to be able to capture a multitude of viruses in sewage samples in several studies. Moreover, comparison between sewage viromes from Nigeria, Nepal, Bangkok, and California, four geographically distant locations, showed distinct differences in the subsets of detected human viruses (Ng et al., 2012). Interestingly, the average sequence similarity between

the reference sequences stored on the NCBI GenBank and the human viruses detected in the samples from California was higher than those from the other locations. This may indicate a bias towards American viruses in view of human virus diversity in this database (Ng et al., 2012). A study that looked at viruses from sewage capable to infect human epithelial cells was able to detect a large number of bacteriophages and several different species of the *Polyomaviridae*, *Picornaviridae*, and *Papillomaviridae* viral families.(Aw et al., 2014). Another more recent evaluation of untargeted metagenomic sequencing for surveillance purposes retrieved full genomes of Adeno-associated virus-2 as the most prominent mammalian virus in the sample. This virus is generally not associated with any pathology and cannot be grown in cell cultures, possibly underestimating its role in diarrheal disease (Furtak et al., 2016). A striking fact of these studies is the number of sequencing reads that are found that share no sequence similarity with current reference databases. Percentages of unmapped sequences range from 37 to 66% (Cantalupo et al., 2011; Ng et al., 2012). Whether these sequences represent novel viruses that can be transmitted via the food- and waterborne route remains to be determined. Nevertheless, these preliminary studies show the potential of untargeted metagenomic sequencing to detect novel and known human pathogens. Sampling a larger variety of locations, performing longitudinal studies of the same environment and deeper sequencing will provide more information on what environmental metagenomic sequencing can contribute to the monitoring of viral trends and viral diversity.
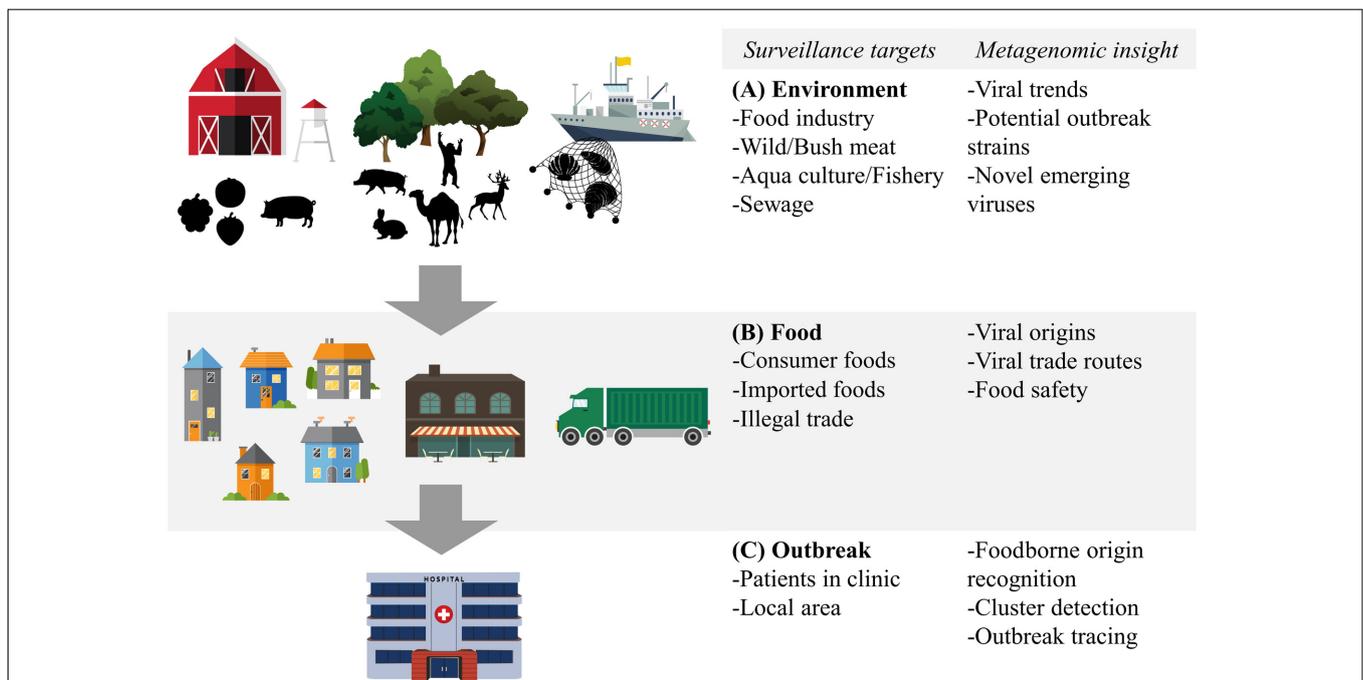
FIGURE 2 | Food- and waterborne viral surveillance targets for metagenomic sequencing approaches. (A) Environmental surveillance of food industry, wild meat and bushmeat habitat, and aquaculture and fishery environment. **(B)** Food surveillance of consumer and imported foods, including illegally imported foods. **(C)** Surveillance of food- and waterborne outbreaks, in clinic and locally. Potential of metagenomic sequencing based surveillance is listed next to each category.

Food Surveillance

Analogous to the environment in which it has been produced, food itself can benefit from metagenomic surveillance. Food contamination in combination with international trade, changing eating habits and food processing practices all contribute to the spread of food- and waterborne viruses and making food itself a valuable target of metagenomic surveillance (**Figure 2B**). Sentinel screening of imported foods, especially risk foods such as fresh fruits and vegetables, dried meats and seafood, could prevent foodborne viral outbreaks such as the international HAV outbreak in Europe from 2012 to 2013 (Severi et al., 2015). Successful application of metagenomic sequencing of viruses has been shown in a study isolating viruses in the family of *Reoviridae* and *Picobirnaviridae* from field-grown lettuce (Aw et al., 2014).

Apart from legal trade, illegal import of food products, such as bushmeat, could also be screened. Untargeted metagenomic sequencing is especially suited for these types of screenings, as the origin and the potential viral content of these samples are often completely unknown. In one example, metagenomic sequencing was performed on bushmeat seized by the customs officers of a French airport. Although no viruses with a potential threat to human health could be detected (Temmam et al., 2016), these initial attempts should be looked at as potentially interesting surveillance approaches, given that relatively large quantities of raw bushmeat are estimated to enter Europe and the Americas annually (Mann et al., 2015).

Another source of known and potentially unknown foodborne disease-causing viruses are shellfish. Mainly the consumption of oysters is associated with foodborne outbreaks (Bellou et al., 2013). However, oysters, cockles, and clams have been shown to accumulate norovirus, sapovirus, and HAV (Benabbes et al., 2013). To our knowledge, there are no published studies performing untargeted virome sequencing of these shellfish. Surveillance by metagenomic sequencing can be beneficial for aquaculture, also for monitoring seafood health, as in aquaculture, large numbers of animals are kept in a confined environment for an extended period, increasing opportunities for the spread of infections. Cultivated fish and other sources of seafood can be infected with a wide variety of viruses (Alavandi and Poornima, 2012).

Outbreak Surveillance

One of the main promises of surveillance using metagenomic sequencing is that of concomitant clinical application (diagnosis of patients) and public health application (typing and cluster analysis to trace of food- and waterborne outbreaks) (**Figure 2C**). Using metagenomic sequencing, the effort of detecting and genotyping of a virus can be combined to trace an outbreak, regardless of prior knowledge of the virus, provided the data is analyzed in combination with relevant metadata. The use of this integrative approach has been demonstrated in an investigation of a hospital outbreak of human parainfluenza virus, which was investigated using high-throughput metagenomic sequencing (Greninger et al., 2016). Both the detection of the virus, the diagnosis of the disease and the establishment of viral clusters and transmission routes could be derived from the metagenomic sequencing data. A similar approach should enable investigation of viruses related to food- or waterborne diseases and distinguishing between a food- and waterborne and a person-to-person transmission route. In such investigations, speed is of the utmost importance, therefore on-site sequencing strategies, enabled by novel portable sequencing platforms such as the Oxford Nanopore MinION (Hoenen et al., 2016), have potential in fast local outbreak detection and disease monitoring (Arias et al., 2016). Recent reports have shown potential in metagenomic detection of hepatitis C, chikungunya, Ebola and Zika virus in hospital settings (Greninger et al., 2015; Sardi et al., 2016). The development of on-site sequencing technology is still in its infancy, however, and it remains to be investigated if food-related viral outbreaks will be traceable and can deliver whole-genome based viral dynamics analysis analogous to the investigation of the Ebola outbreak of 2014 (Gire et al., 2014; Quick et al., 2016). However, the same on-site technology has been shown to be beneficial in tracing foodborne salmonella (Quick et al., 2015). Aspects of current on-site sequencing technologies that need to be improved for viral metagenomic sequencing are the limited throughput and sequence quality, which limit the detection of low-level viral genomes and minor variants. Nevertheless, the use of near-real-time sequencing of Ebola and Zika during the recent outbreaks has received a lot of attention and has shown that the technology works.

CHALLENGES IN SAMPLE PREPARATION AND SEQUENCING

The routine application of metagenomic sequencing for clinical diagnosis and surveillance is dependent costs versus performance criteria such as speed, reliability, and comparability of results with those of reference methods. Improvements are necessary in the standardization and speedup of sample preparation, sequencing and data analysis for clinical and public health application. A recent study has shown the potential of fast whole-genome sequence based epidemiological tracing in the recent Ebola outbreak (Arias et al., 2016). However, specific primers were used to target the Ebola genome, which is different from a metagenomic sequencing approach.

The developments and different choices of sequencing technology make it difficult to decide how to standardize routine diagnostics and surveillance protocols. Studies that directly compare platforms help in this decision making process. Two studies compared the Illumina MiSeq, Roche-454 titanium, Ion Torrent PGM, and PacBio RS platforms (Quail et al., 2012; Frey et al., 2014). For viral metagenomics application, the Illumina and the Ion Torrent platform seem to outperform the other two platforms based on their relatively low cost per giga base output. Between these two systems, the main tradeoff is the sequencing time versus the sequencing read output. The high volume of sequencing reads produced by the Illumina platform, in a longer timeframe, increases the chance that a lowly abundant viral genome is sufficiently covered, which makes it more suitable for metagenomics of complex samples. The Ion Torrent platform delivers a smaller number of reads in a smaller timeframe, which

is beneficial when a timely result is necessary, and low level presence of viruses is disregarded, for instance in diagnostic settings.

Novel approaches, such as the MinION nanopore sequencer, increase speed and depth of coverage at the cost of sequence error rate. A comparison of a metagenomics approach using the MinION nanopore or the Illumina MiSeq sequencer reports a sample-to-result time of 6 h for a MinION nanopore setup compared to 20 h using an Illumina MiSeq setup (Greninger et al., 2015). Despite their reported successes in identification of viruses, the reported error rate of 10 to 60 percent impedes high resolution sequence classification at low genome coverage, or the use of sequence data for reliable source-tracking. It does allow very rapid virus classification in cases where low coverage suffices, or at high viral titers (Hoenen et al., 2016; Quick et al., 2016). However, performance of the MinION sequencing platform remains to be tested at lower virus titers and with more complex samples which are generally encountered in surveillance and clinical settings (food, feces, sewage).

To increase the viral specific output of metagenomics sequencing approaches sample preparation methods can be used to reduce non-viral genomic material or specifically select for viruses. Approaches that are being investigated, range from different extraction protocols (Cotten et al., 2014; Conceição-Neto et al., 2015) to using a virome specific capturing chip (Briese et al., 2015) or blood-derived antibodies to capture viral particles (Oude Munnink et al., 2013). Paradoxically, however, the sequencing capacity of high throughput metagenomic sequencing is sensitive enough to pick up contaminants from the lab reagents, or from previous experiments (Gruber, 2015). These pose a challenge to the interpretation of metagenomic data. To limit contamination, laboratories in which samples are processed are often separated from those in which nucleic acids are amplified and equipment is UV treated and cleaned with bleach. Additionally, alternating the sample-specific DNA barcodes in multiplex sequencing experiments reduces contamination from previous runs. Nevertheless, it is recommended to include both negative control samples, which have been processed similarly, but are believed to contain no viruses, and positive control samples, that contain known quantities of a variety of viruses (Lusk, 2014). Alternatively, bioinformatics tools such as DeconSeq (Schmieder and Edwards, 2011), have been developed to check for signals of regularly found lab contamination in the sequencing data. In conclusion, as contamination of samples and equipment may not be avoidable, its likelihood should be taken in consideration when using metagenomic sequencing technology for food-related surveillance applications.

DIFFICULTIES OF METAGENOMIC DATA ANALYSIS

Aside from lab-based technical difficulties, there are several challenges concerning data analysis of metagenomic sequencing experiments. First, due to the high and increasing read output of sequencing machines, data analysis of high throughput sequencing projects generally requires strong computational infrastructure, which, can require large investments and technological expertise (Spjuth et al., 2016). However, metagenomic data analysis tools have been improving, optimizing the ratio between computing resources needed and their speed and accuracy. Sequence annotation tools based on k-mer lookup tables such as UBLAST (Edgar, 2010), Kraken (Wood et al., 2014), Kaiju (Menzel et al., 2016), and Diamond (Buchfink et al., 2014) have increased the speed of sequence assignment to reference database with several orders of magnitude, while requiring relatively modest processing power.

Second, the assembly of millions of genomic fragments into 1000s of different individual genomes is a daunting task. Historically, short-read assemblers were developed and optimized to assemble a single genome out of a set of sequencing reads. These assemblers are therefore not suited for the reconstruction of metagenomes and are prone to creating synthetic chimeric genomes (Vázquez-Castellanos et al., 2014). Various assemblers have since been developed specifically aimed at metagenome assembly, like MetaSPAdes (Nurk et al., 2016), Ray-Meta (Boisvert et al., 2012), MetAMOS (Treangen et al., 2013), MetaVelvet (Afiahayati et al., 2015), and IDBA-UD (Peng et al., 2012). Nevertheless, metagenome assembly is still a challenging task, often requiring manual editing to resolve miss-assemblies.

Third, assigning all assembled genomes to a reference genome is hampered by miss-annotations and incomplete reference databases. One example is "non-A, non-B hepatitis virus", a sequence present in the NCBI GenBank, which was miss-annotated and the sequence was shown to belong to a bacteriophage (Cantalupo et al., 2011). The volume of sequencing databases is increasing rapidly, however, sequence annotations and metadata are of varying levels of quality and the speed of analysis decreases with increasing reference datasets. Therefore, there is a tradeoff between the rate of success of annotation of a sequence against a smaller curated reference dataset, and reliability of annotation using a large reference database with less-well curated annotation data.

Sequence homology of multiple reference genomes can lead to the spurious assignment of sequencing reads to one of these genomes. An example of the impact of spurious read annotation was the alleged detection of genomic material of *Yersinia pestis* in the New York subway system. Further inspection showed that the reads mapping to *Yersinia pestis* could have mapped with similar likelihood to other bacterial species (Afshinnekoo et al., 2015). Such miss-annotations of metagenomic sequences need to be anticipated and carefully addressed before using metagenomics in surveillance and diagnostic applications.

METAGENOMIC DATA INTERPRETATION

The final challenge of metagenomic sequencing based surveillance is the interpretation of the annotated sequences. There is still little knowledge of the presence and dynamics of viruses in the environment and the food chain, which is of influence on the interpretation of food- and waterborne viral surveillance samples. Various factors are expected to influence

the virome, and without knowledge of the typical viral content of a sample, the relevance of the detection of a virus is hard to determine. An example of this is a study showing a large discrepancy between the levels of HAV genotypes detected in sewage samples compared to the genotype infecting patients in the clinic in the same time (La Rosa et al., 2014). A potential sampling bias and asymptomatic shedding of one of the variants was proposed as an explanation of the discrepancy. However, this shows that a lack of knowledge of viral diversity in a population under surveillance could potentially lead to wrong conclusions in environmental surveillance studies.

Detection of a virus by molecular methods relies on intact genomic material of a virus being present in the sample. However, the relationship between the infectivity of a detected virus and the detection of a fragment of its genome is not unambiguous. Apart from intrinsic virus characteristics, infectivity and detection of a virus depends on its stability in the sample matrix (Cook and Rze, 2004) and during sample preprocessing steps (Conceição-Neto et al., 2015). Similarly, the detection of a virus using untargeted metagenomic sequencing does not confirm its infectivity. Cell culture based infectivity essays are the golden standard to determine virus infectivity, these methods are, however, not scalable and many viruses cannot be cultured *in vitro* (Hamza et al., 2011). High genome coverage combined with close sequence identity to a viral reference genome with a known pathogenic phenotype are currently the strongest links between metagenomic sequencing data and disease etiology. Nevertheless, currently employed PCR based methods, which are based on genome fragment detection, suffer from the same limitations (D'Agostino et al., 2011).

It is becoming increasingly clear that integration of different data sources and experimental results is crucial for the interpretation of metagenomic sequencing experiments. Therefore, browsing of these data and visualization of relationships between genome datasets and metadata should be facilitated. In the recent years, interactive web-based data browsing and visualization tools have increased in popularity to facilitate the interaction with and the browsing through highly complicated data in a user-friendly manner. Further development of tools that facilitate interaction with and visualization of metagenomic sequencing results, such as Kronatools (Ondov et al., 2011) and Taxonomer (Flygare et al., 2016), and frameworks for so-called data analysis "dashboards"[3,4,5], should make the interpretation of metagenomics experiments easier in the future.

[3] http://shiny.rstudio.com/

[4] https://plot.ly/

[5] http://jupyter.org/

CONCLUSION

In our current society, there is much attention for the diseases that are causing occasional outbreaks. However, there are multiple strong signs that there are viruses hiding below the radar, due to a focus on viruses with direct clinical impact. As such, the disease burden of food- and waterborne viral infections is mainly recorded in outbreaks, signified by severe symptoms and hospitalization. However, it is estimated that the large abundance of viral infections causing mild symptoms, and thus not being recorded, carry a large portion of the global food- and waterborne disease burden. Moreover, this disease burden is expanded by the consequential infections and outbreaks of these viruses in susceptible populations. Global food trading, diversification of food sources and interactions with animals and other reservoirs of food- and waterborne disease related viruses complicate the capability of investigators to detect the original source and to determine the transmission pattern of viruses causing foodborne outbreaks. Therefore, surveillance efforts should look to metagenomic sequencing technologies, bioinformatics analysis tools and data sharing initiatives to get a more realistic insight in the global burden of food- and waterborne viral disease, and to make informed decisions on how to reduce this burden.

AUTHOR CONTRIBUTIONS

DN and MK designed the focus of the review, DN did the literature search and drafted the manuscript, MK reviewed and revised.

REFERENCES

Aarestrup, F. M., Brown, E. W., Detter, C., Gerner-Smidt, P., Gilmour, M. W., Harmsen, D., et al. (2012). Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg. Infect. Dis.* 18:e1. doi: 10.3201/eid/1811.120453

Afiahayati, R., Sato, K., Sakakibara, Y., Robertson, D. L., Prosperi, M., Afiahayati, K., et al. (2015). MetaVelvet-SL: an extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. *DNA Res.* 22, 69–77. doi: 10.1093/dnares/dsu041

Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., et al. (2015). Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Syst.* 1, 72–87. doi: 10.1016/j.cels.2015.01.001

Ahmed, S. M., Hall, A. J., Robinson, A. E., Verhoef, L., Premkumar, P., Parashar, U. D., et al. (2014). Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet. Infect. Dis.* 14, 725–730. doi: 10.1016/S1473-3099(14)70767-4

Alavandi, S. V., and Poornima, M. (2012). Viral metagenomics: a tool for virus discovery and diversity in aquaculture. *Indian J. Virol.* 23, 88–98. doi: 10.1007/s13337-012-0075-2

Amar, C. F. L., East, C. L., Gray, J., Iturriza-Gomara, M., Maclure, E. A., and McLauchlin, J. (2007). Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control infectious intestinal disease study (1993–1996). *Eur. J. Clin. Microbiol. Infect. Dis.* 26, 311–323. doi: 10.1007/s10096-007-0290-8

Arias, A., Watson, S. J., Asogun, D., Tobin, E. A., Lu, J., Phan, M. V. T., et al. (2016). Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evol.* 2:vew016. doi: 10.1093/ve/vew016

Aw, T. G., Howe, A., and Rose, J. B. (2014). Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *J. Virol. Methods* 210, 15–21. doi: 10.1016/j.jviromet.2014.09.017

Beersma, M. F. C., Sukhrie, F. H. A., Bogerman, J., Verhoef, L., Melo, M. M., Vonk, A. G., et al. (2012). Unrecognized norovirus infections in health care institutions and their clinical impact. *J. Clin. Microbiol.* 50, 3040–3045. doi: 10.1128/JCM.00908-12

Bellou, M., Kokkinos, P., and Vantarakis, A. (2013). Shellfish-borne viral outbreaks: a systematic review. *Food Environ. Virol.* 5, 13–23. doi: 10.1007/s12560-012-9097-6

Benabbes, L., Ollivier, J., Schaeffer, J., Parnaudeau, S., Rhaissi, H., Nourlil, J., et al. (2013). *Norovirus* and other human enteric viruses in moroccan shellfish. *Food Environ. Virol.* 5, 35–40. doi: 10.1007/s12560-012-9095-8

Binnicker, M. J. (2015). Multiplex molecular panels for diagnosis of gastrointestinal infection: performance, result interpretation, and cost-effectiveness. *J. Clin. Microbiol.* 53, 3723–3728. doi: 10.1128/JCM.02103-15

Boisvert, S., Raymond, F., Godzaridis, É, Laviolette, F., Corbeil, J., Wold, B., et al. (2012). Ray meta: scalable de novo metagenome assembly and profiling. *Genome Biol.* 13:R122. doi: 10.1186/gb-2012-13-12-r122

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O. J., et al. (2015). Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio* 6:e1491-15. doi: 10.1128/mBio.01491-15

Bruni, R., Taffon, S., Equestre, M., Chionne, P., Madonna, E., Rizzo, C., et al. (2016). Key role of sequencing to trace hepatitis a viruses circulating in Italy during a large multi-country European foodborne outbreak in 2013. *PLoS ONE* 11:e0149642. doi: 10.1371/journal.pone.0149642

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Cadwell, K. (2015). The virome in host health and disease. *Immunity* 42, 805–813. doi: 10.1016/j.immuni.2015.05.003

Cantalupo, P. G., Calgua, B., Zhao, G., Hundesa, A., Wier, A. D., Katz, J. P., et al. (2011). Raw sewage harbors diverse viral populations. *MBio* 2:e180-11. doi: 10.1128/mBio.00180-11

Colson, P., Richet, H., Desnues, C., Balique, F., Moal, V., Grob, J.-J., et al. (2010). Pepper mild mottle virus, a plant virus associated with specific immune responses, fever, abdominal pains, and pruritus in humans. *PLoS ONE* 5:e10041. doi: 10.1371/journal.pone.0010041

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., et al. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* 5:16532. doi: 10.1038/srep16532

Cook, N., and Rze, A. (2004). Survival of human enteric viruses in the environment and food. *FEMS Microbiol. Rev.* 28, 441–453. doi: 10.1016/j.femsre.2004.02.001

Cotten, M., Oude Munnink, B., Canuti, M., Deijs, M., Watson, S. J., Kellam, P., et al. (2014). Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS ONE* 9:e93269. doi: 10.1371/journal.pone.0093269

D'Agostino, M. D., Cook, N., Rodriguez-Lazaro, D., and Rutjes, S. (2011). Nucleic acid amplification-based methods for detection of enteric viruses: definition of controls and interpretation of results. *Food Environ. Virol.* 3, 55–60.

de Graaf, M., van Beek, J., and Koopmans, M. P. (2016). Human norovirus transmission and evolution in a changing world. *Nat. Rev. Microbiol.* 14, 421–433. doi: 10.1038/nrmicro.2016.48

de Wit, M. A., Kortbeek, L. M., Koopmans, M. P., de Jager, C. J., Wannet, W. J., Bartelds, A. I., et al. (2001). A comparison of gastroenteritis in a general

practice-based study and a community-based study. *Epidemiol. Infect.* 127, 389–397.

Deng, X., den Bakker, H. C., and Hendriksen, R. S. (2016). Genomic epidemiology: whole-genome-sequencing–powered surveillance and outbreak investigation of foodborne bacterial pathogens. *Annu. Rev. Food Sci. Technol.* 7, 353–374. doi: 10.1146/annurev-food-041715-033259

Di Bartolo, I., Angeloni, G., Ponterio, E., Ostanello, F., and Ruggeri, F. M. (2015). Detection of hepatitis E virus in pork liver sausages. *Int. J. Food Microbiol.* 193, 29–33. doi: 10.1016/j.ijfoodmicro.2014.10.005

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

ESFA (2016). Manual for reporting on food−borne outbreaks in accordance with directive 2003/99/EC for information deriving from the year 2015. *EFSA Support Publ.* 13:989E. doi: 10.2903/SP.EFSA.2016.EN-989

European Food Safety Authority (2014). An update on the risk of transmission of Ebola virus (EBOV) via the food chain. *EFSA J.* 12:3884. doi: 10.2903/j.efsa.2014.3884

Eurosurveillance editorial team (2015). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2011 has been published. *Euro Surveill.* 18:20449. doi: 10.2903/j.efsa.2015.3991

Flygare, S., Simmon, K., Miller, C., Qiao, Y., Kennedy, B., Di Sera, T., et al. (2016). Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol.* 17:111. doi: 10.1186/s13059-016-0969-1

Frey, K. G., Herrera-Galeano, J., Redden, C. L., Luu, T. V., Servetas, S. L., Mateczun, A. J., et al. (2014). Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics* 15:96. doi: 10.1186/1471-2164-15-96

Funk, A. L., Goutard, F. L., Miguel, E., Bourgarel, M., Chevalier, V., Faye, B., et al. (2016). MERS-CoV at the animal–human interface: inputs on exposure pathways from an expert-opinion elicitation. *Front. Vet. Sci.* 3:88. doi: 10.3389/fvets.2016.00088

Furtak, V., Roivainen, M., Mirochnichenko, O., Zagorodnyaya, T., Laassri, M., Zaidi, S. Z., et al. (2016). Environmental surveillance of viruses by tangential flow filtration and metagenomic reconstruction. *Euro Surveill.* 21:30193. doi: 10.2807/1560-7917.ES.2016.21.15.30193

Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S. G., Park, D. J., Kanneh, L., et al. (2014). Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345, 1369–1372. doi: 10.1126/science.1259657

Gossner, C., Gossner, C., and Severi, E. (2014). Three simultaneous, food-borne, multi-country outbreaks of hepatitis A virus infection reported in EPIS-FWD in 2013: What does it mean for the European Union? *Euro Surveill.* 19:20941. doi: 10.2807/1560-7917.ES2014.19.43.20941

Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., et al. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* 7:99. doi: 10.1186/s13073-015-0220-9

Greninger, A. L., Zerr, D. M., Qin, X., Adler, A. L., Sampoleo, R., Kuypers, J. M., et al. (2016). Rapid metagenomic next-generation sequencing during an investigation of hospital-acquired human parainfluenza virus 3 infections. *J. Clin. Microbiol.* 55, 177–182. doi: 10.1128/JCM.01881-16

Gruber, K. (2015). Here, there, and everywhere: from PCRs to next-generation sequencing technologies and sequence databases, DNA contaminants creep in from the most unlikely places. *EMBO Rep.* 16, 898–901. doi: 10.15252/embr.201540822

Guillois, Y., Abravanel, F., Miura, T., Pavio, N., Vaillant, V., Lhomme, S., et al. (2016). High proportion of asymptomatic infections in an outbreak of hepatitis E associated with a spit-roasted piglet, France, 2013. *Clin. Infect. Dis.* 62, 351–357. doi: 10.1093/cid/civ862

Hald, T., Lo Fo Wong, D. M. A., and Aarestrup, F. M. (2007). The Attribution of human infections with antimicrobial resistant *Salmonella* bacteria in Denmark to sources of animal origin. *Foodborne Pathog. Dis.* 4, 313–326. doi: 10.1089/fpd.2007.0002

Hamza, I. A., Jurzik, L., Überla, K., and Wilhelm, M. (2011). Methods to detect infectious human enteric viruses in environmental water samples. *Int. J. Hyg. Environ. Health* 214, 424–436. doi: 10.1016/j.ijheh.2011.07.014

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685. doi: 10.1128/MMBR.68.4.669-685.2004

Havelaar, A. H., Kirk, M. D., Torgerson, P. R., Gibb, H. J., Hald, T., Lake, R. J., et al. (2015). World health organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLoS Med.* 12:e1001923. doi: 10.1371/journal.pmed.1001923

Hay, S. I., Battle, K. E., Pigott, D. M., Smith, D. L., Moyes, C. L., Bhatt, S., et al. (2013). Global mapping of infectious disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20120250. doi: 10.1098/rstb.2012.0250

Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., et al. (2014). Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Appl. Environ. Microbiol.* 80, 6771–6781. doi: 10.1128/AEM.01981-14

Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., et al. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* 7, 1678–1695. doi: 10.1038/ismej.2013.59

Hoenen, T., Groseth, A., Rosenke, K., Fischer, R. J., Hoenen, A., Judson, S. D., et al. (2016). Nanopore sequencing as a rapidly deployable ebola outbreak tool. *Emerg. Infect. Dis.* 22, 331–334. doi: 10.3201/eid2202.151796

Honda, K., and Littman, D. R. (2016). The microbiota in adaptive immune homeostasis and disease. *Nature* 535, 75–84. doi: 10.1038/nature18848

Iritani, N., Kaida, A., Abe, N., Kubo, H., Sekiguchi, J.-I., Yamamoto, S. P., et al. (2014). Detection and genetic characterization of human enteric viruses in oyster-associated gastroenteritis outbreaks between 2001 and 2012 in Osaka City, Japan. *J. Med. Virol.* 86, 2019–2025. doi: 10.1002/jmv.23883

Islam, M. S., Sazzad, H. M. S., Satter, S. M., Sultana, S., Hossain, M. J., Hasan, M., et al. (2016). Nipah virus transmission from bats to humans associated with drinking traditional liquor made from date palm sap, Bangladesh, 2011–2014. *Emerg. Infect. Dis.* 22, 664–670. doi: 10.3201/eid2204.151747

Izopet, J., Dubois, M., Bertagnoli, S., Lhomme, S., Marchandeau, S., Boucher, S., et al. (2012). Hepatitis E virus strains in rabbits and evidence of a closely related strain in humans, France. *Emerg. Infect. Dis.* 18, 1274–1281. doi: 10.3201/eid1808.120057

Karesh, W. B., and Noble, E. (2009). The bushmeat trade: increased opportunities for transmission of zoonotic disease. *Mt. Sinai J. Med.* 76, 429–434. doi: 10.1002/msj.20139

Kazama, S., Masago, Y., Tohma, K., Souma, N., Imagawa, T., Suzuki, A., et al. (2016). Temporal dynamics of norovirus determined through monitoring of municipal wastewater by pyrosequencing and virological surveillance of gastroenteritis cases. *Water Res.* 92, 244–253. doi: 10.1016/j.watres.2015.10.024

La Rosa, G., Libera, S. D., Iaconelli, M., Ciccaglione, A. R., Bruni, R., Taffon, S., et al. (2014). Surveillance of hepatitis A virus in urban sewages and comparison with cases notified in the course of an outbreak, Italy 2013. *BMC Infect. Dis.* 14:419. doi: 10.1186/1471-2334-14-419

Lewis, H. C., Wichmann, O., and Duizer, E. (2010). Transmission routes and risk factors for autochthonous hepatitis E virus infection in Europe: a systematic review. *Epidemiol. Infect.* 138, 145–166. doi: 10.1017/S0950268809990847

Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS ONE* 9:e110808. doi: 10.1371/journal.pone.0110808

Mann, E., Streng, S., Bergeron, J., Kircher, A., Morvan, J., Deubel, V., et al. (2015). A review of the role of food and the food system in the transmission and spread of Ebolavirus. *PLoS Negl. Trop. Dis.* 9:e0004160. doi: 10.1371/journal.pntd.0004160

McCloskey, B., Dar, O., Zumla, A., and Heymann, D. L. (2014). Emerging infectious diseases and pandemic potential: status quo and reducing risk of global spread. *Lancet Infect. Dis.* 14, 1001–1010. doi: 10.1016/S1473-3099(14)70846-1

Menzel, P., Ng, K. L., Krogh, A., Riesenfeld, C., Schloss, P., Handelsman, J., et al. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7:11257. doi: 10.1038/ncomms11257

Moon, S., Sohn, I.-W., Hong, Y., Lee, H., Park, J.-H., Kwon, G.-Y., et al. (2014). Emerging pathogens and vehicles of food- and water-borne disease outbreaks in Korea, 2007-2012. *Osong Public Health Res. Perspect.* 5, 34–39. doi: 10.1016/j.phrp.2013.12.004

Newell, D. G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., et al. (2010). Food-borne diseases – the challenges of 20 years ago still persist

while new ones continue to emerge. *Int. J. Food Microbiol.* 139(Suppl.), S3–S15. doi: 10.1016/j.ijfoodmicro.2010.01.021

Ng, T. F. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., et al. (2012). High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J. Virol.* 86, 12161–12175. doi: 10.1128/JVI.00869-12

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2016). metaSPAdes: a new versatile de novo metagenomics assembler. arXiv:1604.03071

Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12:385. doi: 10.1186/1471-2105-12-385

Ott, J. J., Stevens, G. A., Groeger, J., and Wiersma, S. T. (2012). Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 30, 2212–2219. doi: 10.1016/j.vaccine.2011.12.116

Oude Munnink, B. B., Jazaeri Farsani, S. M., Deijs, M., Jonkers, J., Verhoeven, J. T. P., Ieven, M., et al. (2013). Autologous antibody capture to enrich immunogenic viruses for viral discovery. *PLoS ONE* 8:e78454. doi: 10.1371/journal.pone.0078454

Payne, D. C., Currier, R. L., Staat, M. A., Sahni, L. C., Selvarangan, R., Halasa, N. B., et al. (2015). Epidemiologic Association between FUT2 secretor status and severe rotavirus gastroenteritis in children in the United States. *JAMA Pediatr.* 169, 1040–1045. doi: 10.1001/jamapediatrics.2015.2002

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Petrignani, M., Verhoef, L., Vennema, H., van Hunen, R., Baas, D., van Steenbergen, J. E., et al. (2014). Underdiagnosis of foodborne hepatitis A, the Netherlands, 2008–2010. *Emerg. Infect. Dis.* 20, 596–602. doi: 10.3201/eid2004.130753

Pires, S. M., Evers, E. G., van Pelt, W., Ayers, T., Scallan, E., Angulo, F. J., et al. (2009). Attributing the human disease burden of foodborne infections to specific sources. *Foodborne Pathog. Dis.* 6, 417–424. doi: 10.1089/fpd.2008.0208

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics* 13:341. doi: 10.1186/1471-2164-13-341

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., et al. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* 16:114. doi: 10.1186/s13059-015-0677-2

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530, 228–232. doi: 10.1038/nature16996

Rahman, M. A., Hossain, M. J., Sultana, S., Homaira, N., Khan, S. U., Rahman, M., et al. (2012). Date palm sap linked to nipah virus outbreak in Bangladesh, 2008. *Vector Borne Zoonotic Dis.* 12, 65–72. doi: 10.1089/vbz.2011.0656

Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F., and Gordon, J. I. (2012). Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat. Rev. Microbiol.* 10, 607–617. doi: 10.1038/nrmicro2853

Rodriguez-Manzano, J., Hundesa, A., Calgua, B., Carratala, A., Maluquer de Motes, C., Rusiñol, M., et al. (2014). Adenovirus and norovirus contaminants in commercially distributed shellfish. *Food Environ. Virol.* 6, 31–41. doi: 10.1007/s12560-013-9133-1

Sardi, S. I., Somasekar, S., Naccache, S. N., Bandeira, A. C., Tauro, L. B., Campos, G. S., et al. (2016). Coinfections of Zika and Chikungunya viruses in Bahia, Brazil, identified by metagenomic next-generation sequencing. *J. Clin. Microbiol.* 54, 2348–2353. doi: 10.1128/JCM.00877-16

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* 6:e17288. doi: 10.1371/journal.pone.0017288

Sethi, D., Wheeler, J., Rodrigues, L. C., Fox, S., and Roderick, P. (1999). Investigation of under-ascertainment in epidemiological studies based in general practice. *Int. J. Epidemiol.* 28, 106–112.

Severi, E., Verhoef, L., Thornton, L., Guzman-Herrador, B. R., Faber, M., Sundqvist, L., et al. (2015). Large and prolonged food-borne multistate hepatitis A outbreak in Europe associated with consumption offrozen berries,

2013 to 2014. *Euro Surveill.* 20:21192. doi: 10.2807/1560-7917.ES2015.20.29.
21192

Smits, S. L., Bodewes, R., Ruiz-González, A., Baumgärtner, W., Koopmans, M. P.,
Osterhaus, A. D. M. E., et al. (2015). Recovering full-length viral genomes from
metagenomes. *Front. Microbiol.* 6:1069. doi: 10.3389/fmicb.2015.01069

Spjuth, O., Bongcam-Rudloff, E., Dahlberg, J., Dahlö, M., Kallio, A., Pireddu, L.,
et al. (2016). Recommendations on e-infrastructures for next-generation
sequencing. *Gigascience* 5:26. doi: 10.1186/s13742-016-0132-7

Tam, C. C., Rodrigues, L. C., Viviani, L., Dodds, J. P., Evans, M. R., Hunter, P. R.,
et al. (2012). Longitudinal study of infectious intestinal disease in the UK (IID2
study): incidence in the community and presenting to general practice. *Gut* 61,
69–77. doi: 10.1136/gut.2011.238386

Tao, Z., Xu, M., Lin, X., Wang, H., Song, L., Wang, S., et al. (2015). Environmental
surveillance of genogroup i and ii noroviruses in Shandong Province, China in
2013. *Sci. Rep.* 5:17444. doi: 10.1038/srep17444

Tavoschi, L., Severi, E., Niskanen, T., Boelaert, F., Rizzi, V., Liebana, E., et al.
(2015). Food-borne diseases associated with frozen berries consumption: a
historical perspective, European Union, 1983 to 2013. *Euro Surveill.* 20:21193.
doi: 10.2807/1560-7917.ES2015.20.29.21193

Tei, S., Kitajima, N., Takahashi, K., and Mishiro, S. (2003). Zoonotic transmission
of hepatitis E virus from deer to human beings. *Lancet* 362, 371–373. doi:
10.1016/S0140-6736(03)14025-1

Teixeira, J., Mesquita, J. R., Pereira, S. S., Oliveira, R. M. S., Abreu-Silva, J.,
Rodrigues, A., et al. (2016). Prevalence of hepatitis E virus antibodies in workers
occupationally exposed to swine in Portugal. *Med. Microbiol. Immunol.* 206,
77–81. doi: 10.1007/s00430-016-0484-8

Temmam, S., Davoust, B., Chaber, A.-L., Lignereux, Y., Michelle, C., Monteil-
Bouchard, S., et al. (2016). Screening for viral pathogens in african simian
bushmeat seized at a French airport. *Transb. Emerg. Dis.* doi: 10.1111/tbed.
12481 [Epub ahead of print].

Treangen, T. J., Koren, S., Sommer, D. D., Liu, B., Astrovskaya, I., Ondov, B.,
et al. (2013). MetAMOS: a modular and open source metagenomic
assembly and analysis pipeline. *Genome Biol.* 14:R2. doi: 10.1186/gb-2013-14-
1-r2

Vázquez-Castellanos, J. F., García-López, R., Pérez-Brocal, V., Pignatelli, M.,
Moya, A., Handelsman, J., et al. (2014). Comparison of different assembly and

annotation tools on analysis of simulated viral metagenomic communities in
the gut. *BMC Genomics* 15:37. doi: 10.1186/1471-2164-15-37

Verhoef, L., Hewitt, J., Barclay, L., Ahmed, S. M., Lake, R., Hall, A. J., et al.
(2015). Norovirus genotype profiles associated with foodborne transmission,
1999–2012. *Emerg. Infect. Dis.* 21, 592–599. doi: 10.3201/eid2104.14
1073

Verhoef, L., Kouyos, R. D., Vennema, H., Kroneman, A., Siebenga, J., van Pelt, W.,
et al. (2011). An integrated approach to identifying international foodborne
norovirus outbreaks. *Emerg. Infect. Dis.* 17, 412–418. doi: 10.3201/eid1703.
100979

Wang, Y., Zhang, J., and Shen, Z. (2015). The impact of calicivirus mixed infection
in an oyster-associated outbreak during a food festival. *J. Clin. Virol.* 73, 55–63.
doi: 10.1016/j.jcv.2015.10.004

Wolfe, N. D., Daszak, P., Kilpatrick, A. M., and Burke, D. S. (2005). Bushmeat
hunting, deforestation, and prediction of zoonotic disease. *Emerg. Infect. Dis.*
11, 1822–1827. doi: 10.3201/eid1112.040789

Wood, D. E., Salzberg, S. L., Venter, C., Remington, K., Heidelberg, J., Halpern, A.,
et al. (2014). Kraken: ultrafast metagenomic sequence classification using exact
alignments. *Genome Biol.* 15:R46. doi: 10.1186/gb-2014-15-3-r46

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G.,
Contreras, M., et al. (2012). Human gut microbiome viewed across age and
geography. *Nature* 486, 222–227. doi: 10.1038/nature11053

Zhang, T., Breitbart, M., Lee, W. H., Run, J.-Q., Wei, C. L., Soh, S. W. L., et al.
(2006). RNA viral community in human feces: prevalence of plant pathogenic
viruses. *PLoS Biol.* 4:e3. doi: 10.1371/journal.pbio.0040003