

High-throughput sequencing in veterinary infection biology and diagnostics

S. Belák^{(1, 2, 3)*}, O.E. Karlsson^(2, 3, 4), M. Leijon^(1, 3)
& F. Granberg^(2, 3)

This paper (No. 04122013-00020-EN) has been peer-reviewed, accepted, edited, and corrected by authors. It has not yet been formatted for printing. It will be published in December 2013 in issue 32-3 of the *Scientific and Technical Review*.

(1) Department of Virology, Immunobiology and Parasitology (VIP), National Veterinary Institute (SVA), Ulls väg 2B, SE-751 89 Uppsala, Sweden

(2) Department of Biomedical Sciences and Veterinary Public Health (BVF), Swedish University of Agricultural Sciences (SLU), Ulls väg 2B, SE-751 89 Uppsala, Sweden

(3) World Organisation for Animal Health (OIE) Collaborating Centre for Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine, Ulls väg 2B, SE-751 89 Uppsala, Sweden

(4) SLU Global Bioinformatics Center, Department of Animal Breeding and Genetics (HGEN), SLU, Gerda Nilssons väg 2, SE-756 51 Uppsala, Sweden.

*Corresponding author: sandor.belak@slu.se

Summary

Sequencing methods have improved rapidly since the first versions of the Sanger techniques, facilitating the development of very powerful tools for detecting and identifying various pathogens, such as viruses, bacteria and other microbes. The ongoing development of high-throughput sequencing (HTS; also known as next-generation sequencing) technologies has resulted in a dramatic reduction in DNA sequencing costs, making the technology more accessible to the average laboratory. In this White Paper of the World Organisation for Animal Health (OIE) Collaborating Centre for the Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine

(Uppsala, Sweden), several approaches and examples of HTS are summarised, and their diagnostic applicability is briefly discussed. Selected future aspects of HTS are outlined, including the need for bioinformatic resources, with a focus on improving the diagnosis and control of infectious diseases in veterinary medicine.

Keywords

Aetiology – Detection of viruses and bacteria – Diagnosis – Emerging viruses and bacteria – High-throughput sequencing – Metagenomics – Next-generation sequencing.

Introduction

During recent decades, the rapid development and wide application of novel molecular techniques have resulted in a steady advance in diagnostics and in studies of the infection biology of animal diseases. The capability to detect and identify infectious agents, based on the presence of their nucleic acid molecules in clinical samples, has been exploited by various methods in molecular diagnostics, including nucleic acid hybridisation, amplification methods and nucleotide sequencing.

Several examples of rapid developments in biotechnology-based diagnostics, with a special focus on the various methods of nucleic acid hybridisation and amplification (1, 2, 3, 4), have been reported and summarised in previous reviews from the World Organisation for Animal Health (OIE) Collaborating Centre for Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine in Uppsala, Sweden. The Collaborating Centre is a partnership between the institutes of the authors of the present paper, i.e. the Swedish National Veterinary Institute (SVA) and the Swedish University for Agricultural Sciences (SLU). The introduction of various DNA sequencing technologies has allowed the direct detection, investigation and characterisation of the genome sequences of various infectious agents, including elucidation of their role, relatedness and evolutionary aspects related to various scenarios of infection biology and disease development. Furthermore, insights from comparative

sequencing studies on the genomes, exomes or transcriptomes of healthy and diseased cells and organisms are now providing more powerful diagnostic opportunities, as well as improved classification, forecasting and therapy selection for many infectious illnesses.

The importance of nucleotide sequencing was fully realised after the establishment of the Sanger method (5). For the past 35 years, this technology has been the dominant approach for DNA sequencing. It enabled the first viral genomes to be sequenced (6), and also laid the foundation for the development of polymerase chain reaction (PCR), which is the best-known and most successfully implemented diagnostic molecular technology to date (7, 8).

Since the first versions of the Sanger techniques, sequencing methods have rapidly improved, generating very powerful tools for detecting and identifying various pathogens, such as viruses, bacteria and other microbes. An important step in this development was the introduction of automated multicapillary-based instruments, using fluorophore labelling with multi-spectral imaging, later referred to as ‘first-generation sequencing’ platforms (9). This type of Sanger sequencing is still extensively used in laboratories around the world and is a front-line diagnostic tool for virus characterisation, including pathotyping and phylogenetic analysis.

The OIE has actively supported the early adoption and use of these molecular techniques in the field of veterinary medicine through the OIE Reference Laboratories and by the establishment of OIE Collaborating Centres, with a strong focus on biotechnology-based diagnosis in veterinary medicine.

The ‘first-generation sequencing’ approaches have opened up new pathways for the detection and identification of various pathogens, host–pathogen interactions and the evolution of infectious agents. However, attempts to sequence larger genomes, such as the whole genomes of various animal species, using multicapillary sequencing, have encountered considerable bottlenecks in throughput, scalability, speed and resolution. This has spurred the development of new sequencing technologies (10, 11).

The subsequent major technological advances in cyclic-array sequencing gave rise to what is known as ‘second-generation sequencing’ (SGS) or ‘next-generation sequencing’ (NGS). These technologies involve iterative cycles during which the sequences of DNA features, which have been immobilised to constant locations on a solid substrate, are determined, one base position at a time, using enzymatic manipulation and imaging-based data collection (12, 13). Today, there are several different NGS platforms with tailored protocols and approaches to template preparation and sequencing that determine the type of data produced.

The 454 FLX pyrosequencing platform (13) was established as the high-throughput method of choice for discovery and *de novo* assembly of novel microorganisms (14, 15, 16). In addition to having the longest read lengths, which simplifies *de novo* assembly, this was also the first NGS platform on the market, becoming available in 2005. In 2007, the 454 FLX system was followed by the Genome Analyser developed by Illumina/Solexa (17).

The field has since developed rapidly, as a result of the continuous improvement and refinement of existing systems and the release of completely new platforms, such as the Ion Torrent Personal Genome Machine (PGM). As a consequence, the efficiency and throughput of DNA sequencing are now increasing at a rate even faster than that projected by Moore’s law for computing power (a doubling every two years) (18, 19). This has also resulted in a dramatic reduction in DNA sequencing costs, making the technology more accessible to the average laboratory (Fig. 1).

Recently, a new generation of single-molecule sequencing technologies, termed ‘third-generation sequencing’ (TGS), has emerged, as summarised, for example, in the review article of Schadt *et al.* (20).

In this White Paper of the OIE, several approaches and examples of high-throughput sequencing (HTS) are summarised, and their diagnostic applicability is discussed briefly. Selected future aspects of

HTS are also outlined, with a focus on improved diagnosis and control of infectious diseases in veterinary medicine.

High-throughput sequencing as a diagnostic tool

As demonstrated by several peer-reviewed articles, HTS has shown great potential in the detection and discovery of novel pathogens (see ‘Examples of the application of high-throughput sequencing in veterinary diagnostic microbiology’, below). In this regard, it is common to distinguish between the spread of known infections to new areas and/or the emergence of completely novel, ‘unknown’ pathogens. Contrary to earlier techniques, HTS is unbiased and reports all nucleotide sequences present in the original sample. However, as with earlier techniques, the lower limit for detection is still ultimately determined by the abundance of pathogens in relation to host background material. By enabling deeper sequencing to be performed more quickly and cheaply, the continuing development of HTS techniques is also continuing to improve the likelihood of detecting low copy-number pathogens (21). In addition, sampling, sample preparation and enrichment protocols have all been demonstrated to have dramatic effects on the outcome of HTS-based diagnostics (22), and thus should each be considered as integral steps in the overall detection scheme (Fig. 2). Since proper sampling of diagnostic specimens is a key issue for all laboratory investigations of an animal disease, the subject has already been extensively covered in the *OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* and *Manual of Diagnostic Tests for Aquatic Animals* (23, 24). Besides the general rule that the samples collected should be representative of the condition being investigated, downstream analysis and interpretation rely on the inclusion of contextual metadata. As argued by the Genomic Standards Consortium (www.gensc.org), with sequencing costs for complete microbial and viral genomes steadily decreasing, the associated contextual information is increasingly valuable (25).

Sample preparation

Most sample preparation and enrichment protocols are composed of several individual steps, including homogenisation, filtration,

ultracentrifugation and nuclease treatment, as well as nucleic acid extraction and purification followed by amplification. These steps are outlined briefly below.

Depending on the nature of the starting material, i.e. type of tissue, blood or other body fluids, it may first be necessary to homogenise, fractionise or dilute a representative amount of the original sample. This can be performed according to existing protocols for other purposes or applications, and it is important to remember that the choice of procedure can affect the final outcome, due to differences in host cell disruption and preservation of intact pathogens (22). It is therefore advisable to find a homogenisation procedure optimised for each sample matrix rather than a one-fits-all solution (26). Regardless of the initial procedure, centrifugation of the homogenates and fluids followed by microfiltration is an effective method to separate free and released pathogens from larger particles, such as cellular debris.

To purify and concentrate virus particles, ultracentrifugation has proven to be a reliable method, especially in combination with density gradients (27, 28). Even though a standard, full-sized ultracentrifuge is not readily available in all laboratories, smaller facilities might still consider a benchtop version, if dealing with large quantities of small-volume samples.

Nuclease treatment is a widely used enrichment method for viral nucleic acid, and can be applied directly to filtered material or as a complement to ultracentrifugation. Deoxyribonuclease (DNase) and ribonuclease (RNase) are used in combination or alone to remove host contaminants by exploiting the protection from digestion afforded by the virus capsid (29, 30). Even though this is a powerful method to remove the majority of host-derived sequence contaminants, the material extracted after nuclease treatment is often present in such small quantities that amplification is required before sequencing. Various sequence-independent amplification strategies have been employed, and the two most widespread are phi29 polymerase-based multiple displacement amplification (31) and random PCR, using modified versions of sequence-independent, single-primer

amplification (SISPA) (30, 32). Although a combined approach with nuclease treatment and amplification has been demonstrated to be very effective, care should be taken since nuclease treatment might destroy free microbial and viral nucleic acid if the samples are old or not well preserved. There is also a possibility that the amplification method can introduce a bias (33), which is of less concern for pathogen detection but may influence the recovery of whole-genome sequences.

For non-culturable bacteria, phi29 polymerase-based amplification has been used successfully to obtain sufficient quantities of DNA for sequencing (34). However, 16S ribosomal RNA (rRNA) gene deep sequencing has proven more useful than random amplification for metagenomics studies of microbial communities. In this approach, primers designed for conserved regions in the bacterial 16S rRNA gene are fused with adaptors and barcodes and used to generate PCR products. With the incorporation of adaptors and barcodes, the products already constitute a sequencing library (see ‘Library construction and sequencing’, below), and can be directly used for sequencing (35). The HTS of PCR products for specific loci is commonly referred to as amplicon sequencing and can also be used as a powerful tool for the detection and identification of viruses and other microorganisms (36).

Library construction and sequencing

Most existing HTS platforms require the viral genomic sequences (directly as DNA or reverse-transcribed RNA) in the samples to be converted into sequencing libraries suitable for subsequent cluster generation and sequencing. This process usually consists of four main steps:

- fragmentation of DNA, performed by mechanical or enzymatic shearing
- end-repair, modification and ligation of adaptors, which enable amplification of the sheared DNA by adapter-specific primers

- size selection of DNA molecules with a certain optimal length for the current application or instrument
- enrichment of adapter-ligated DNA by PCR (37).

The last step is only necessary if the amount of starting material is small and has been observed to introduce amplification-based artefacts (38). In addition, by adding specific short sequence tags (indexes) during the construction of multiple libraries, it is possible to pool libraries for increased throughput and reduced costs. The necessary reagents for preparing a sequencing library can often be purchased as platform-specific commercial kits.

The prepared sequencing library fragments are then usually immobilised on a solid support for clonal amplification in a platform-dependent manner to generate separate clusters of DNA copies. The distinctive sequencing strategies of the main commercial HTS platforms currently available, as well as their application areas, are described briefly below and are also summarised in Table I, together with other properties of interest, such as running time and sequence data output.

454 pyrosequencing (Roche)

The 454 pyrosequencing technology is based on emulsion PCR (emPCR), for clonal amplification of single library fragments on beads inside aqueous reaction bubbles, in combination with a sequencing-by-synthesis approach. The DNA-containing beads are loaded into individual wells on a PicoTiterPlate, which is subjected to a cyclic flow of sequentially added nucleotide reagents. When a polymerase-mediated incorporation event occurs, a chemiluminescent enzyme generates an observable light signal that is recorded by a charge-coupled device (CCD) camera (13). There are currently two platforms on the market using this technology, the GS FLX system and the GS Junior system. While the latest GS FLX system can now generate around 700 megabases (Mb) of sequence data with read lengths of up to 1,000 base pairs (bp) in a day, it is still associated with high running costs and is most commonly used by larger

sequencing facilities. The GS Junior, on the other hand, is essentially a smaller benchtop version with lower set-up and running costs aimed at research laboratories. However, although 454 pyrosequencing has been the most commonly used technology for HTS to date, it has a high error rate in homopolymer regions (i.e. three or more consecutive, identical DNA bases), caused by accumulated light intensity variance (13).

SOLiD (Life Technologies)

The sequencing by oligonucleotide ligation and detection (SOLiD) technology also uses emPCR to generate clonal DNA fragments on beads. The enriched beads are then deposited into separate ‘pico-wells’ on a glass slide to allow sequencing by ligation. This involves iterative rounds of oligonucleotide ligation extension, during which every base is scored at least twice by using fluorescently labelled di- or tri-base probes, with data translated into ‘colour space’ rather than conventional base space (12). While this approach provides very high accuracy, the maximum read length is relatively short (75 bp). The SOLiD platforms have therefore been used mainly for applications not requiring *de novo* assembly of reads, such as transcriptomics, epigenomics and resequencing of large mammalian genomes. There are currently two SOLiD platforms available, the 5500 and the 5500xl. In addition, the Wildfire upgrade, with improved library construction, was recently released.

Illumina (Illumina Inc.)

Illumina uses a process known as bridge-PCR to generate clusters of amplified sequencing library fragments directly on the surface of a glass flowcell. Immobilised amplification primers are used in the process (39). The unique feature of the sequencing-by-synthesis approach applied by Illumina is the use of a proprietary reversible terminator technology that enables the detection of single bases as they are incorporated into growing DNA strands. Briefly, a fluorescently labelled terminator is imaged as each deoxyribonucleotide is incorporated and then cleaved to allow addition of the next. The two most recent HTS platforms introduced

by Illumina are the HiSeq 2500, which is available as either a new instrument or an upgrade for the HiSeq 2000, and the MiSeq. While the design and capacity of the HiSeq 2500 make it most suitable for sequencing centres, the MiSeq benchtop, equipped with a smaller flow cell, reduced imaging time and faster microfluidics, is aimed at a wider range of laboratories and the clinical diagnostic market. Illumina is recognised to be the dominant HTS technology and also claims to generate the industry's most accurate data (40).

Ion Torrent (Life Technologies)

Contrary to other sequencing-by-synthesis methods that also use emPCR, Ion Torrent does not rely on the detection of fluorescence or chemiluminescence signals. Instead, it uses ion-sensitive, field-effect transistor-based sensors to measure hydrogen ions released during polymerase-mediated incorporation events in individual microwells (41). As the sequencing chips are produced in the same way as semiconductors, it has been possible to increase capacity by constructing chips with a higher density of sensors and microwells. There are currently three different chips available for the Personal Genome Machine (PGM), the first platform for Ion Torrent sequencing, which range in capacity from 10 Mb to 1 gigabase (Gb). With a short run time and flexible capacity, the PGM represents an affordable and rapid benchtop system designed for small projects, such as identifying and sequencing microbial pathogens.

To enable even higher throughput with the Ion Torrent technology, the Proton platform was recently released with two different chips, aimed at sequencing exomes and large eukaryotic genomes.

PacBio RS (Pacific Biosciences)

The PacBio RS is the first commercially available TGS platform that is able to sequence single DNA molecules in real time without the need for clonal amplification. This is achieved by using a nanophotonic confinement structure, referred to as a zero-mode waveguide (ZMW) (42), as a window to monitor the activity of a single DNA polymerase molecule attached to the bottom surface.

During complementary strand synthesis, each incorporation event is recorded as it causes the release of a nucleotide-specific fluorescent dye (43). Because a single-molecule real-time sequencing chip contains thousands of ZMWs, the PacBIO RS can handle a large number of sequences simultaneously. Even though its throughput is lower than most HTS platforms on the market, the PacBIO RS still has several advantages that make it attractive for clinical laboratories and microbiology research. Sample preparation is fast, there is no introduction of amplification artefacts, run times are relatively short (finished within 1 h or 2 h), and read length is the greatest currently available (around 1,500 bp). However, in a comparison with other sequencing platforms, the PacBio RS has been reported to have the highest raw error rate (7% to 13%) (40).

Future sequencing technologies

An interesting emerging technology which holds the promise of taking HTS to the next level is offered by Oxford Nanopore Technologies Ltd. By successfully feeding a DNA strand through an engineered nanopore protein immobilised on a solid surface, and thereby obtaining its sequence (44, 45, 46), this approach finally realises a concept first pioneered during the technology race to complete the human genome (47). A small-scale implementation of the technology, termed the MinION, for single-molecule sequencing, is currently being developed. This sequencing device is aimed at pen-side applications, requiring only a standard computer. A considerably larger technology platform is also in development for large-scale studies (e.g. plant genomes, cooperative genomics and metagenomics), called the GridION (48). Both systems are scheduled for early access in spring 2013 and commercial release during late 2013. However, even though the Oxford Nanopore systems are, it seems, those closest to commercial release, a number of competitors have also developed new sequencing technologies. These include companies such as Genia, with a biological nanopore system, and the Cambridge-based GnuBio, which is developing a microfluidics-based sequencing system.

Bioinformatics

Bioinformatics, the research field focusing on the study of methods for retrieving, analysing and storing biological data, is an integral part of all HTS applications (49). For most molecular methods, a normal benchtop computer can handle the data reasonably well. However, when the basic data are in the gigabyte range and the analytical steps can produce several hundred gigabytes of information, the capacity of standard computers (in terms of memory, storage and processor power) is insufficient (50).

The process of handling the flow of data within a software structure, i.e. the chain of separate elements run in cohesion, is usually referred to as a pipeline. The bioinformatics pipelines within typical HTS applications begin with the first crucial algorithms structuring and interpreting the data from the technological platform into machine- or human-readable form. At present, the most common form is the FastQ format (51), which specifies a combination of sequence and quality for each base within a sequence.

Following the acquisition of data, a researcher normally runs a quality control check of the data set, establishing that the experiment or sample was processed according to specifications (52). This is usually followed by screening for irrelevant sequence information, such as contamination, host material or tag sequences (53). Depending on the project, the data set can then be treated in various ways, the two most common being:

- mapping towards the host genome (54), where the short reads are aligned towards a previously known homologue
- *de novo* assembly (55), where reads are processed through an algorithm to form larger sets of continuous reads, termed ‘contigs’, which are then mapped towards the closest relative homologue to form a draft genome.

Metagenomics, combined with an HTS methodology, has proven to be an even greater challenge to bioinformatics (56). Not only is the

amount of reads staggering due to little or no filtering, but also the few variables from normal genomics are removed and in their place is a shifting number of possible genomes to map the reads against. A metagenomic analysis is therefore computationally demanding, often requiring the processing of several million reads through different sequence classification algorithms, such as the basic local alignment search tool (BLAST) for sequence similarity searches, to determine possible species in the sample (57, 58). This can be experimentally circumvented by targeting specific parts of the metagenome, either by sample preparation methods or by molecular means (using amplicon-based methods to target specific gene groups) (59). This simplifies the bioinformatics and reduces the time needed for analysis.

The targeted sequencing of the total, or selected, RNA content of a sample (RNA-seq), introduces the possibility of not only mapping available messenger RNA (mRNA) sequences but also non-coding RNAs: microRNA (miRNA), small interfering RNA (siRNA), long non-coding RNA (lncRNA), ribosomal RNA (rRNA) and riboswitches (60). As the number of RNA molecule types that can be targeted grows, so does the need for efficient mapping and tools for visualising RNA content within a sample. This also introduces a number of possible problems with previous technologies, so quality control, mapping and assembly tools need to be adapted to these new targets (61).

The structure of data storage needs serious consideration when deriving bioinformatics solutions for HTS applications. The amount of information is massive, as stated above, and, even though the final amount of data can be much smaller, the requirement during processing might be measured in terabytes. It is possible, of course, to store this amount of data on small servers or powerful desktop machines, but it quickly becomes inconvenient when larger computing resources are needed or when the information is shared among several partners who need to work with and analyse the data. Thus, a solid infrastructure for data storage, analysis and dissemination is needed (62, 63). Previously, most of this need has been filled by governmental organisations, such as the European Bioinformatics

Institute or the National Center for Biotechnology Information (64). With the current influx of new data, old paradigms need to change. For example, the Sequence Read Archive (SRA) offers a storage solution for raw data without further analysis but in a way that is publicly accessible (65, 66). As seen in Figure 3, the SRA is growing at very high speed, illustrating the paradigm shift in sequencing technology that is being brought about by HTS.

However, the very large growth of sequencing data is not sustainable indefinitely and has been called one of the greatest challenges faced by computational biology. The European Union (EU) project ELIXIR (European Life Sciences Infrastructure for Biological Information) calculated that data growth is doubling every nine months, while storage capacity, due to technical limitations, only doubles every 18 months (62, 67). The total data increase in GenBank over the last 20 years is illustrated in Figure 4.

These observations, combined with the prospect of even more powerful sequencing technologies, indicate that future developments (including the use of sequencing in diagnostics) will require new approaches to providing the infrastructure needed for collecting, curating, storing and archiving biomolecular data.

Examples of the application of high-throughput sequencing in veterinary diagnostic microbiology

Given the limited space in this White Paper, only a few representative examples are listed below, showing the application of HTS in selected scenarios. The focus is placed on known and emerging new infections and diseases occurring in four groups of farm animals: swine, cattle, sheep and poultry. To provide a practical structure for presentation, the examples are summarised according to host species. However, it should be noted that the majority of the infectious agents being discussed, such as various influenza viruses, bluetongue virus (BTV) or foot and mouth disease virus (FMDV), are not restricted to a single host, but are able to establish infections in multiple animal species. Thus, the structure used below serves practical purposes and is not

intended to limit the complex infection biology of many of the viruses and viral diseases described.

Infections and diseases of swine

The global swine industry is of very great importance as a supplier of protein to the increasing human population. It has been estimated that the global yearly meat demand is more than 200 million tonnes, of which pork – with 40% – has the largest share. In addition, it has been forecast that the demand for pork will rise by 31% between 1997 and 2020, second only to the predicted increase in demand for poultry meat (40%) (68). Thus, infectious diseases in the pig population can have a profound impact and lead to large economic losses for the swine industry. In addition, pigs are physiologically similar to humans and can act as a source of zoonotic transmission, as evidenced in the cases of influenza A and hepatitis E. For these reasons, there is great interest in monitoring the infectious disease status of pig populations, and HTS methodologies have been employed by many investigators to this end.

Detection and identification of various infectious agents involved in complex diseases, such as post-weaning multisystemic wasting syndrome

Clarifying the possibilities of the combined pathobiology behind various disease complexes, such as post-weaning multisystemic wasting syndrome (PMWS), and investigating the presence of various infectious agents in such complicated disease scenarios are of particular interest. Although porcine circovirus type 2 (PCV-2) has been strongly implicated as the main aetiological agent, it was always suspected from the wide spectrum of clinical presentations that PMWS is a multifactorial disease. The HTS method is well suited to attack this kind of problem, since samples can be indiscriminately interrogated for all nucleic acid content without prior knowledge of the sequence of any suspected infectious agent. At the Collaborating Centre, lymph nodes were collected from pigs with confirmed PMWS and 454-sequencing was employed to investigate the presence of additional infectious agents in the background of the PCV-2 infection,

by Blomström *et al.* (69). Genome sequences of Torque teno virus, as well as of a novel boca-like parvovirus, were found. Subsequently, it was confirmed that this boca-like parvovirus was correlated with PMWS when 34 pigs affected with PMWS and 24 symptomless pigs were screened (70). Continued research is now required to investigate the possible pathobiological roles of the newly detected viruses in the complex PMWS scenario, with special regard to the pathogenicity of the individual viruses and the synergistic effects of the various infectious agents.

Detection of infectious agents in mixed infections in enteric disease complexes and characterisation of the porcine microbiota

The enteric disease complexes of young piglets are another example of the recent application of deep sequencing. In this case, it is used to investigate the aetiology when a herd exhibits symptoms that may be caused by many different infectious agents, such as in diarrhoea, where the presence of various agents is very common. As mentioned above, an added advantage of HTS is that more complex multifactorial disease scenarios can be defined by broad detection of the whole virome; for instance, in pig faeces. By comparing the viromes of healthy and diarrhoeic pigs, an indication of the aetiological agents may be obtained. In two studies, the faecal viromes of diarrhoeic and healthy pigs were compared (71, 72). Although these studies investigated herds on different continents, several common observations were made. The two investigations, performed in the United States (USA) and Europe, both detected kobuviruses, enteroviruses, sapoviruses, teschoviruses, astroviruses and bocaviruses in the faeces of young piglets with enteric problems. It is worth noting that, in the reviewed metagenomic studies, rotaviruses were detected only in pigs from Europe (72) and coronaviruses only in the pigs from the United States (71). Incidentally, these discrepant viruses were also those that appeared to be correlated with disease. It is likely that this is just the beginning of such investigations and that further extended metagenomic studies will similarly reveal the presence of both viruses in swine populations on the two continents. However, we can see even

at this stage that the reported studies clearly show the usefulness of HTS in acquiring more advanced knowledge by analysing the whole faecal virome, which can aid greatly in identifying the aetiological agents. However, they also highlight the importance of detailed knowledge of the commensal virome in drawing appropriate conclusions. Kobuviruses (73) and astroviruses (74) were also identified in wild boar, as well as enteroviruses and teschoviruses (73). Thus, basically the same viruses were identified in wild boar as in the domestic pig population (71, 72). The authors suggested that wild boar could be an important host and reservoir for picornaviruses and astroviruses. The increased need for livestock production and the exploitation of wildlife habitats for animal agriculture make it prudent to monitor the pathogenic status of wildlife with increased vigilance. In sera collected from African bush pigs, Torque teno sus virus (TTSuV); porcine parvovirus 4 (PPV-4); porcine endogenous retrovirus; a GB hepatitis C-like virus; and a sclerotinia hypovirulence-associated-like virus were detected (75). This was the first detection of PPV-4 in wild African suids, and novel variants of TTSuV-1 and 2 were also identified in bush pigs. This study illustrates the power of HTS as a tool for virus discovery. As mentioned above, the zoonotic aspect of such diseases in domestic pigs is highly relevant, due to pigs' close contact with humans and their similar physiology.

Viral metagenomics, using 454-sequencing of sera collected from domestic pigs from various locations in Uganda, has recently demonstrated that these animals are potential reservoirs for Ndumu virus (NDUV) (76). This is an alphavirus and belongs to the same genus as, for example, the serious human pathogen chikungunya virus. Antibodies have indeed been detected against NDUV in human patients in several African countries, indicating both zoonotic effects and a considerable geographic distribution of this virus on the African continent (76). As with other observations, these data indicate convincingly that HTS can serve as an important tool for the surveillance of potentially zoonotic pathogens in various hosts.

Pathogen–host interactions

High-throughput sequencing has the capacity not only to discover novel pathogens and provide a detailed picture of the virome, but also to contribute to understanding of the host responses to various viral infections, thereby achieving a greater knowledge of infection pathology. In this respect, HTS can substitute for high-density arrays. Porcine reproductive and respiratory syndrome (PRRS) is economically one of the most important diseases affecting the swine industry. Xiao and colleagues (77) carried out a detailed investigation of the host response to PRRS virus using the Illumina Digital Gene Expression system, which employs Illumina deep-sequencing technology. This proved to be a powerful approach, as it verified many of the features and expression patterns that were already known, while also revealing several new ones. Thus, when used to study altered gene expression in viral disease, HTS could potentially provide a better understanding of pathogenesis and aid in the development of antiviral therapies and the identification of genetic markers for resistance (77).

Infections and diseases of cattle and sheep

The large-scale cattle and sheep industries have a very important role in feeding the rapidly increasing human population on our planet. There are currently approximately 1.5 billion cattle populating the Earth. Their pastures constitute a major land use, taking up nearly 27% of the land mass of the planet, and they consume enough grain to feed hundreds of millions of people. Accordingly, cattle and sheep health are major targets of the animal health organisations, and they have a crucial importance in OIE networks. Cattle and sheep diseases are classified into various main categories, such as those that affect the respiratory, enteric, reproductive or metabolic systems and those that affect the skin, eyes, feet and udder, etc. Diseases reportable to the OIE and various other zoonotic diseases have special importance in the wide range of health problems experienced by cattle and sheep (www.oie.int; www.thecattlesite.com/diseaseinfo/). In the following sections, several examples are discussed in which HTS has opened up

new possibilities for virus detection and identification, leading to a better understanding of infection biology and improved control of infectious diseases.

Characterisation and phylogeny of new variants of bluetongue virus

Bluetongue virus has been detected in many parts of the world, including Africa, Europe, the Middle East, Australia, the South Pacific, North and South America and parts of Asia. The disease has an emerging character, since the vectors and infections are spreading to new areas of the globe, such as (relatively recently) Northern Europe. As reported in a recent article from Boyle *et al.* (78), ten BTV serotypes have been isolated in Australia so far, apparently without associated disease. High-throughput sequencing of the ten genome segments of all Australian BTV prototype serotypes, followed by comparative phylogenetic analysis, reinforced the Western and Eastern topotypes previously characterised, and also revealed unique features of several Australian BTVs. Many of the Australian BTV genome segments (Seg-) are closely related, clustering together within the Eastern topotypes. A novel Australian topotype for Seg-5 (NS1) was identified in these experiments, with taxa spread across several serotypes and over time. These studies support the conclusion that BTV-2 entered Australia recently. The Australian BTV_15_AUS_1982 prototype was revealed to be unusual among the Australian BTV isolates, with Seg-3 and Seg-8 distantly related to other BTV sequences from all serotypes. See further details in Boyle's article (78).

Determination of host viral population diversity

An interesting article has been published recently by Wright *et al.*, describing the viral population diversity of FMDV, using HTS approaches (79). The authors of this report state:

'The diverse sequences of viral populations within individual hosts are the starting material for selection and subsequent evolution of RNA viruses such as foot and mouth disease virus (FMDV). Using next-

generation sequencing (NGS) performed on a Genome Analyser platform (Illumina), this study compared the viral populations within two bovine epithelial samples (foot lesions) from a single animal with the inoculum used to initiate experimental infection. Genomic sequences were determined in duplicate sequencing runs, and the consensus sequence of the inoculum determined by NGS was identical to that previously determined using the Sanger method. However, NGS revealed the fine polymorphic substructure of the viral population, from nucleotide variants present at just below 50% frequency to those present at fractions of 1%. Some of the higher-frequency polymorphisms identified encoded changes within codons associated with heparin sulphate binding and were present in both foot lesions, revealing intermediate stages in the evolution of a tissue culture-adapted virus replicating within a mammalian host.’

The researchers identified considerable numbers of polymorphisms in the inoculum and in the two foot lesions: most of the substitutions occurred in only a small fraction of the population and represented the progeny from recent cellular replication events before the onset of any selective pressures. The upper limit for the genome-wide mutation rate of the virus within a cell was 7.8×10^{-4} per nucleotide. The authors stated that: ‘the greater depth of detection achieved by NGS demonstrates that this method is a powerful and valuable tool for the dissection of FMDV populations within hosts’. See the detailed results in the article of Wright *et al.* (79).

Considering that the majority of viral infections are caused not by a single homologous viral variant, but by mixed populations of variants that can seriously influence the outcome of the infection and the disease, HTS-based studies are very important for a better understanding of the infection biology of FMD, which is a highly important infectious disease of global significance.

Detection of unknown and emerging new pathogens: the rapid detection and identification of Schmallerberg virus

In 2011, a novel orthobunyavirus, of the Simbu serogroup, was identified in congenital malformations of bovines and ovines in

Germany and other West European countries, such as the Netherlands. This virus was named Schmallenberg virus (SBV). Later, the presence of the virus was confirmed using real-time PCR and by HTS in several locations in Europe. As with related viruses, such as Akabane virus, SBV appears to be transmitted by biting midges. Metagenomic approaches, including HTS, played an important role in the detection and rapid identification of this emerging, very important virus (80, 81).

As described in the article of Hoffmann *et al.* (81) for metagenomic analysis, four sequencing libraries were prepared and sequenced, using the 454 Genome Sequencer FLX (81). Two libraries each were generated from the DNA and RNA isolated from plasma samples. By using a combination of BLAST and sequence mapping with the 454-reference mapper application, reads were classified into the three domains of bacteria, archaea and eukaryotes. In addition to the anticipated high number of host sequences, the authors detected in some samples a considerable proportion of reads representing diverse bacterial species. It was supposed that these bacteria most likely grew in the samples during the prolonged storage before extraction of the nucleic acids used to prepare the sequencing libraries. Seven orthobunyavirus sequences were detected in the library prepared from pooled RNA from three animals at a farm. Repeated sequencing of this library resulted in 22 additional reads of orthobunyavirus-specific sequences. A few sequence gaps were filled by Sanger sequencing and by NGS of the cell culture isolate. The sequence comparison analyses and phylogenetic investigations revealed that SBV clusters closely with Shamonda viruses within the Simbu serogroup, which suggests that the novel virus is a Shamonda-like virus within the genus *Orthobunyavirus*. A further interesting event in the SBV history occurred when Rosseel *et al.* (82), in the absence of specific sequencing protocols for SBV, confirmed its presence in field samples testing positive by reverse transcription-quantitative PCR (RT-qPCR), using DNase SISPA-HTS, a virus discovery method based on random amplification and NGS. Two field samples of aborted lambs allowed the unambiguous identification of SBV. This was summarised as follows:

‘One sample yielded 192 SBV reads covering about 81% of the L segment, 56% of the M segment and 13% of the S segment. The other sample resulted in eight reads distributed over the L and M segments. Three weak positive field samples (one from an aborted calf, two from aborted lambs) containing virus quantities equivalent to 4.27 to 4.89 log₁₀ RNA copies per µl did not allow identification using DNase SISPA-HTS. This partial sequence information was compared with the whole-genome sequence of SBV isolated from bovines in Germany, identifying several sequence differences.’

The authors stated that the applied viral discovery method allowed the confirmation of SBV in RT-qPCR-positive brain samples. However, they believed that the failure to confirm SBV in weak PCR-positive samples illustrated the importance of the selection of properly targeted and fresh field samples in any virus discovery method.

The partial sequences derived from the field samples showed several differences compared to the sequences from bovines in Germany, indicating sequence divergence within the epidemic. These data show that DNase SISPA-HTS viral discovery technology can be used on samples of limited amounts of field tissue to identify emerging infectious agents. However, the sensitivity of the method seems to be rather limited. Consequently, when applying this methodology to a cluster of cases of an undiagnosed disease, it is important to select properly targeted and fresh samples, as well as to test multiple diseased animals to allow correct identification of an associated virus, as recommended by Rosseel *et al.* (82).

Infections and diseases of poultry and other avian species

Migrating birds can cover thousands of kilometres in their annual travels, which gives them a unique role as very potent biological and mechanical vectors, transporting microorganisms between countries and continents (83). Avian influenza virus (AIV) and avian paramyxovirus (APMV) are two of the most prominent examples of microorganisms transported in this way. Both of these viruses have spread to local domestic birds (84, 85), which has resulted in significant economic impacts on poultry production in many

countries. To avoid negative effects on productivity, different vaccine strategies have therefore been pursued. However, these efforts have not diminished the need to monitor the circulation of viruses in wild bird populations, since they may carry both novel versions and completely unknown viruses that, in the event of transmission, could have similar or even more serious consequences. Early detection and identification are also of paramount importance once an outbreak occurs, to allow rapid containment and to minimise its impact.

Characterisation and phylogeny of new variants of various viruses

The RNA viruses are well known to exhibit high mutation rates and frequent genetic reassortment, in addition to having high yields and short replication times. These characteristics have also been demonstrated to be true for AIV type A which, in addition to birds, has managed to establish itself in various mammals, including humans and pigs (86). Since there are several different influenza virus variants circulating, detailed information at the genomic level can be crucial to decide what action is required in an outbreak situation. To this end, protocols have been established to use HTS for rapid whole-genome sequencing of AIV (87, 88). Full genome sequences allow the properties of new isolates to be assessed, and HTS has therefore proved to be a valuable tool for rapid characterisation. One example is the phylogenetic study of influenza viruses isolated from gulls and shorebirds in Belgium, where gulls were found to carry viruses that had undergone intercontinental reassortment as a consequence of mixed infections (89). High-throughput sequencing has also been employed to investigate the emergence of escape mutants during serial *in-vitro* passaging of AIV under immune pressure (90). However, to really understand the underlying evolutionary processes, it is also important to consider the quasi-species nature of AIV and other RNA viruses. The feasibility of using high-resolution or ultra-deep sequencing to reveal the genomic diversity of AIV quasi-species has already been demonstrated (91).

Taken together, these studies suggest that HTS has great potential for use in routine diagnostics to monitor the genomic diversity of avian influenza viruses and the early emergence of potentially pathogenic variants. This is further supported by two field monitoring studies using this approach, one on AIV in sentinel birds (92) and the other on transmission from waterfowl to domestic poultry (93).

In addition to the AIV studies, HTS has also been employed to characterise the complete genomes of an avian group A rotavirus containing a novel viral protein (VP) 4 gene (94) and an African APMV type 4 strain detected in a wild duck (95). Similar studies have also been performed on relatively larger DNA viruses. For example, HTS has been used to conduct a comparative full genome analysis of infectious gallid herpesvirus 1 (GaHV-1), the causative agent of infectious laryngotracheitis (96), as well as sequence determination of a mildly virulent strain of gallid herpesvirus 2 (GaHV-2) (97).

Detection of unknown and 'unexpected' pathogens

The unbiased HTS approach has proven to be particularly useful when an outbreak occurs and both pathogen-specific tests and more advanced multiplex assays fail to identify the causative agent. It has also been used to find the likely causative agents of well-known but hitherto unexplained syndromes. Examples of the latter include proventricular dilatation syndrome (PDS) in birds, characterised by gastrointestinal dysfunction and neurologic signs, and turkey viral hepatitis (TVH), an acute, highly contagious disease of turkeys. Even though a viral cause had been suspected for both, it was not until HTS enabled a broader screening approach that plausible candidates could be identified. In the case of PDS, sequencing of complementary DNA from the brains of infected parrots revealed two strains of a novel bornavirus, which were confirmed by real-time PCR to be present only in the affected birds (98). The same approach identified a novel picornavirus in RNA material from the liver of turkey poults with TVH, and, in addition to real-time PCR, *in situ* hybridisation and immunohistological staining were performed to verify further that the virus was present only in the livers of diseased poults (99).

The HTS approach has also proven useful for finding novel or unexpected pathogens in mixed infections, as concluded in a study of two unidentified haemagglutinating agents found during a wildlife screening programme in Belgium. Using random amplification and only a moderate and cost-effective sequencing effort, the presence of three APMVs, two different APMVs type 4 and one APMV type 6, were demonstrated (100).

Investigations of microbiota: characterising bacterial and viral populations and communities in various organs

Several enteric disorders affect poultry and cause illness, mortality and economic losses. The digestive tract of an animal contains very large numbers of microorganisms and highly diverse viruses. Traditional characterisation attempts, which rely on the ability to culture each microorganism clonally, have not been able to capture the complete picture. The introduction of culture-independent HTS made a powerful diagnostic technology available, with the capacity to determine the full genomic repertoire present in complex samples. This type of characterisation of complete microbial and viral communities is more commonly known as metagenomics. Studies using HTS-based metagenomic approaches to gain new insights into the causes of enteric diseases in poultry include an examination of the turkey gut RNA virus community and a comparison of microbial communities between those from healthy chickens and those from chickens suffering from runting-stunting syndrome. Analysis of the RNA virus metagenome demonstrated the presence of members from several viral families, such as *Picornaviridae*, *Caliciviridae* and *Picobirnaviridae*, many of which had not been described in turkeys before (101). The comparative analysis of metagenomes from infected and non-infected chickens resulted in the identification of several microorganisms considered to be ‘pathogen candidates’ (102).

Among the factors that might affect the microbial and viral communities in poultry, the external environment has been of special interest. The influence of the cage system on the microbial gut flora of laying hens was investigated to determine whether it was related to

colonisation by *Salmonella* Enteritidis. Although it was found that cage type had some influence on the composition of the microbiota, no clear differences in the colonisation and excretion patterns of *Salmonella* were identified (103). The microbial community found in the immediate external environment has also been scrutinised using HTS. A study aimed at explaining recurring outbreaks of gangrenous dermatitis investigated the impact of the poultry house environment on litter bacterial community composition and found that the environment appeared to have a considerable effect and may play a key role in the emergence of foodborne pathogens (104). The lessons learned from these initial studies indicate that it would be of interest not only routinely to investigate the intestinal microbiota of poultry, but also of the poultry farm environment.

Pathogen–host interactions, studied by high-throughput sequencing in poultry

With the introduction of microarrays, it became possible to follow the host transcriptional response to infection at different time points by simultaneously measuring the expression levels of thousands of mRNAs. This approach was refined and expanded by using HTS, which allows a more comprehensive characterisation of transcriptomic events. Sequencing-based transcriptomics is commonly called RNA-seq and has, for example, been applied to analyse the transcriptomic response in chicken spleen to avian pathogenic *Escherichia coli* (APEC) infection and the integrated expression of miRNAs and mRNAs in the lungs of AIV-infected broilers. In the APEC study, candidate genes for host response were found within crucial cellular pathways, such as the T-cell receptor (TCR) signalling pathway (105). Similarly, comprehensive analysis of AIV-infected lung tissue suggested that several miRNAs and mRNAs, including those specifying MX1, IL-8, IRF-7 and TNFRS19, were likely to be involved in regulating the host response (106). The information gained by using RNA-seq to obtain an overview of pathogen-induced host responses, or to follow temporal expression changes during the infectious cycle, can be exploited to design better control strategies, as well as to identify genetic markers for resistance. However, further

specific knock-down and protein-interaction assays are generally warranted to clarify the underlying mechanisms.

Vaccine development and quality control

Vaccines are widely applied in poultry husbandry and the poultry industry to prevent and control contagious diseases. The inherited advantages of HTS as a screening and characterisation tool make it ideal for use in poultry vaccine development and quality control. Although the examples given below concern vaccines for birds, it must be emphasised that similar HTS approaches can also be applied to vaccines used in many other animal species, including humans. Using HTS, it is now possible to better understand the processes by which attenuating mutations arise during the serial passage of a virulent isolate. This is illustrated by an experiment with GaHV-2, which resulted in a wealth of information on the genetic changes that occurred during the attenuation process and also indicated that serial passage results in the generation of mixed populations (107). High-throughput sequencing can also be used to discriminate among vaccine strains, as demonstrated by a genome-level comparison of two United States live attenuated infectious laryngotracheitis virus (ILTV) vaccines with an Australian strain (108). Another interesting application is the investigation of relationships between emergent field strains and vaccine strains. A recent study of infectious ILTV indicated that independent recombination events between distinct attenuated vaccine strains resulted in virulent recombinant viruses that became the dominant strains responsible for widespread disease in commercial Australian poultry flocks (109). This illustrates the risk of using multiple different attenuated vaccines, or vectors, in the same populations.

Multi-host pathogens

Rabies virus and related viruses: complete genome sequence of the Ikoma lyssavirus, determined by high-throughput sequencing

As mentioned above, the majority of the infectious agents discussed in this paper are not restricted to a single host species, but were introduced on a host-animal-based grouping for practical reasons. In the following paragraph, a newly detected lyssavirus is introduced. Since lyssaviruses have a very broad range of hosts, this new virus is presented as a multi-host pathogen.

The diseases caused by lyssaviruses constitute one of the most important groups of viral zoonoses globally. In addition to rabies virus, the *Lyssavirus* genus in the *Rhabdoviridae* family encompasses a range of other important viruses, such as Aravan virus, Australian bat lyssavirus, Duvenhage virus, European bat lyssavirus 1 (EBLV-1), EBLV-2, Irkut virus, Khujand virus, Lagos bat virus, Mokola virus, West Caucasian bat virus and Shimoni bat virus. In general, the currently available vaccines are sufficiently protective against the predominantly circulating lyssavirus species. However, it is very important to investigate the emergence of new variants, which may differ considerably from the rabies virus and from related viruses in the genus. Using HTS technologies, the whole-genome sequence for a novel lyssavirus, Ikoma lyssavirus (IKOV), was described recently (110, 111). This virus was isolated from an African civet in Tanzania displaying clinical signs of rabies (111). Genetic investigations revealed that this virus is the most divergent within the *Lyssavirus* genus. Thus, characterisation of the entire viral genome is extremely important, with special regard to genetic and antigenic differences, compared to rabies virus and other species in the genus. Such studies will help to improve the understanding of lyssavirus diversity and enable investigations into vaccine-induced immunity and protection. The results of IKOV HTS are summarised in a recent article by Martson *et al.* (110), demonstrating the importance of the emergence of new variants.

These data must be taken into serious consideration in vaccinology, in order to provide full protection against the rabies virus and other related viruses in the *Lyssavirus* genus.

Discussion

In previous reviews from the Collaborating Centre, the various methods of nucleic acid hybridisation and amplification were reported, as powerful tools in the application of biotechnology-based molecular diagnostic virology in veterinary medicine (1, 2, 4). In the present review, further molecular approaches for the rapid detection and identification of viral genomes are summarised, with a special focus on the novel techniques of HTS and viral metagenomics. As briefly outlined in a recent book chapter presented by the Collaborating Centre (3), and illustrated by several examples in this White Paper, HTS technologies have opened up a revolutionary new area in the pathobiology of various infections and infectious diseases in both veterinary and human medicine. The examples discussed show that both known and emerging infectious agents can be detected and characterised rapidly, co-infections caused by various infectious agents can be observed, and viral populations and subpopulations mapped. Differences between ‘healthy’ and ‘sick’ animals are more easily investigated, and complex microbiological pictures can be obtained, coupled to disease scenarios and aetiology.

It should be noted that the division of subjects in this White Paper according to host-species-based groupings is not optimal for presenting HTS topics. The authors are aware that many viruses are able to cross species barriers and establish infections in various hosts simultaneously. A good example is provided by influenza viruses, which infect a wide range of hosts and cause severe disease in various host species, including humans. Nonetheless, to provide a practical perspective for OIE experts, the issues are discussed by keeping the various host species at the centre. By following this principle, influenza viruses are discussed in the section on ‘Poultry and other avian species’. However, the section on ‘Rabies virus and related viruses’ is an exception, taking into consideration that these viruses

have a very wide range of hosts. Thus, these viruses are discussed in a separate part of the paper.

In providing a brief summary of recent applications of HTS and metagenomic technologies, it is important to note that the OIE has recognised the importance and challenges of these powerful diagnostic and research tools at the appropriate time. Thus, the OIE centres of excellence (Collaborating Centres and Reference Laboratories) have a strong and timely input in these fields, detecting and characterising various scenarios of infections, co-infections and the complex infection biology features of many diseases in veterinary medicine. A good example is the input of the OIE Collaborating Centre for Zoonoses in Europe, at the Friedrich Loeffler Institute in Germany, which played a pioneering role in the detection and rapid identification of the emerging Schmallenberg virus. The authors' Collaborating Centre has also reached interesting conclusions by detecting a wide range of known and emerging new infectious agents in various disease scenarios and hosts, such as bocaviruses, Torque teno viruses, astroviruses, rotaviruses, kobuviruses and other infectious agents in several disease complexes of pigs (69, 70). The Collaborating Centre is at present studying the infection biology of various viruses in farm and non-farm animals, and HTS and metagenomic studies have been extended recently to wildlife, including aquatic diseases and diseases of insects, such as honey bees. By employing an unbiased metagenomic approach, which allows the detection of both unexpected and previously unknown infectious agents, three viruses were detected in honey bees from Spain: aphid lethal paralysis virus, Israel acute paralysis virus and Lake Sinai virus. Furthermore, this study also revealed that these bees carried a plant virus, turnip ringspot virus, and thus could potentially serve as important vector organisms. Taken together, these results demonstrate the new possibilities opened up by HTS and metagenomic analysis to study emerging new diseases in domestic and wild animal populations, including honey bees (112).

The examples provided demonstrate the important role of HTS and viral metagenomics as novel tools for the direct detection and rapid

identification of various known and emerging new, 'unknown' viruses and bacteria. Since the early detection and identification of known pathogens and emerging new agents are of great importance in veterinary medicine, the examples listed provide various aspects of this subject which should be considered by the OIE centres of excellence, both the Collaborating Centres and the Reference Laboratories.

The Schmallenberg virus story serves as a clear example of the power and potential of HTS techniques in combination with metagenomics and strong bioinformatics, in instances where emerging, 'unknown' viruses appear and must be detected as rapidly as possible. The European institutes involved have performed excellent work through their fast and effective application of HTS, metagenomics and bioinformatic analyses, and this has resulted in the very rapid detection and identification of SBV.

It is very important to note that this detection and identification of SBV was achieved by a synergistic combination and parallel application of molecular diagnostic approaches, such as metagenomics, PCR, and the tools of classical virology, including virus isolation and characterisation. This combination of diverse skills and expertise in a strong multidisciplinary approach is extremely important if we are to enhance and extend our ability to isolate viruses, properly understand their infection biology, and develop effective and accurate tools of diagnosis and disease control, such as vaccines. Unfortunately, the skills and scholarship of classical virology are today in a state of rapid decline with, for example, expertise in the techniques of virus isolation dying out in practical terms. This process is harmful to national capabilities and aspirations, and should be stopped and corrected before it becomes too late to reconstruct the skills and facilities involved. The detection and characterisation of SBV clearly indicate the need to maintain a range of complex diagnostic skills, including virus isolation, and this fact is greatly appreciated by the scientific community, including the OIE.

The parallel efforts of molecular and classical virology do not only result in the detection of viruses and other infectious agents in general, but also of virus isolates or strains that become available shortly after outbreaks, or other infection scenarios, have occurred. The proper combination of molecular and classical virology is leading towards the prompt development of novel diagnostic tools and the immediate application of specific control measures. This flexible combination of various approaches is exactly what is needed in this world of rapid globalisation, with relaxed border controls and highly increased human and animal mobility, climatic changes, and many other factors, all of which today support the rapid spread of emerging new pathogens on a global scale.

Conclusions, prognostics and recommendations

From this briefly summarised analysis of recent applications of HTS, the authors conclude that these techniques have opened up new possibilities:

- for the rapid and powerful detection and identification of both known and ‘unknown’ emerging, new, infectious agents
- to determine the complicated infection biology of disease complexes, such as PMWS or enteric disease complexes in young animals
- to obtain information about the entire genomic structures of infectious agents, which can provide insights into functions such as genomic recombination, genomic stability, viral evolution, viral phylogeny, pathogenicity markers and other properties of viral genomes
- to study and elucidate the processes of pathogen–host interactions
- to study the populations and population dynamics of various viruses and to support predictive calculations and forecasts in viral infection biology and control

- to facilitate the development of specific diagnostic tools, such as new PCR panels based on the results of HTS
- to support vaccine development by studying viral populations, population dynamics, escape from vaccine pressure and the emergence of new viral variants, which must be included in the new vaccines
- to facilitate the effective control of infectious diseases by supporting the development of approaches to differentiate infected from vaccinated animals (DIVA) in vaccinology and by supporting the control and predictive measures taken by veterinary epizootiology and epidemiology.

It is very important to note that HTS is becoming more and more affordable. Many HTS-based studies performed within the field of veterinary medicine were unimaginable a few years ago. Development has been rapid and, with the recent introduction of affordable and fast benchtop sequencers, HTS is more accessible than ever. Their variety of features makes it likely that multiple HTS platforms will continue to coexist in the marketplace, with some having clear advantages over others in particular applications. As a consequence, it is unlikely that the concept of ‘a general HTS’ will be validated as a diagnostic tool; it is more plausible that separate platforms will be applied for various specific purposes. However, it is not only the sequencing platform that is important, since the whole chain of processes from sample collection to final results after bioinformatics analysis should be considered. We should also be aware of, and not underestimate, the need for a structured storage system before introducing HTS-based methods into routine diagnostics operations. The HTS field also currently lacks clear diagnostic standards for analysis and the comparison of results (113).

Bioinformatics presents the community with a new and exciting research field. These days, bioinformaticians are a standard resource in human medicine and have proven their worth by developing a multitude of tools: www.allbioinformatics.eu/doku.php; www.seqahead.eu/; www.elixir-europe.org/. Even though the EU, for example, has identified the need to disseminate these tools to the

broader life sciences field, and has sponsored projects aimed at just this purpose, the need for the OIE to coordinate the incorporation of bioinformatics into the field of animal health is becoming ever more apparent.

The following developments are foreseen in the near future:

- HTS will be relatively affordable for many institutes in veterinary medicine;
- very rapid development will continue to take place in sequencing technologies and instrumentation, reducing prices and making such services more easily affordable;
- further advancement of HTS and bioinformatics techniques will provide new possibilities for improved detection of known and emerging, new, ‘unknown’ infectious agents, and lead to better diagnosis of infectious diseases;
- future pen-side versions, e.g. the MinION from Oxford Nanopore Technologies, will bring HTS to simply equipped laboratories or to veterinarians in the field for use in rapid and accurate diagnosis. Point-of-care diagnostic application and technical decentralisation are expected. However, the supporting role of central laboratories will be maintained, with special regard to bioinformatics and evaluating sequencing results;
- HTS will strongly support research on complex scenarios of co-infections, various disease complexes, and viral–viral and viral–bacterial synergetic effects;
- a more adequate determination of ‘healthy’ and ‘sick’ status will be greatly facilitated by HTS and by comparative analysis of sequence information;
- a better understanding of the differences between ‘healthy’ and ‘sick’ status will shed light on infectious agents that are commonly present, while their role has not yet been fully understood;

- vaccine development and control will be strongly supported by HTS, with special regard to emerging pathogens and escaping mutants under vaccine pressure, etc. More specific and powerful vaccines will be constructed and tailored, based on sequence information provided by HTS;
- epizootiology and epidemiology will be further strengthened by the sequence information provided by HTS. Control measures for infectious diseases will be intensified and more effective policies will be developed and applied;
- energetic progress in bioinformatics is crucial, to properly analyse and understand the huge amount of sequence information;
- OIE centres of excellence, primarily selected OIE Collaborating Centres with molecular diagnostic profiles, should be designated to establish strong links with bioinformatic centres of excellence. The purpose is to develop and strengthen links with the field of bioinformatics, coupled to biotechnology-based diagnosis of infectious diseases in veterinary medicine. As a step in this direction, the OIE Collaborating Centre for the Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine has established a relationship, and is working in close collaboration, with the Swedish University of Agricultural Sciences (SLU) Global Bioinformatics Centre in Uppsala, with joint projects on the detection of known and unknown viruses, the analysis of viral populations and subpopulations, evolutionary aspects, viral stability, escape from vaccine pressure and other aspects of infection biology.

To facilitate the issues discussed above, the authors suggest that the OIE consider and discuss several important issues, including:

- How will HTS and sequence data be further used to address animal health issues?
- How will HTS technologies be harmonised and their evaluation standardised in the OIE network?

- What are the consequences for clinical data standards and security, particularly when setting policies?
- How will the existence and use of these data be regulated at the national and state levels?
- How will the existence of very large-scale sequence data interact with intellectual property laws? What legal framework will maximise the clinical application of discoveries?

The OIE Collaborating Centre for the Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine, together with its fellow Collaborating Centres and Reference Laboratories, is ready to support the OIE in all these discussions and decision-making. This will further strengthen and secure the application of HTS technologies in veterinary medicine, in light of the ‘One World, One Health, One Virology’ principles.

Acknowledgements

The continuing support of the OIE is greatly appreciated, and has directly contributed to these achievements. This work was supported by the Award of Excellence (Excellensbidrag) provided to S. Belák by the Swedish University of Agricultural Sciences, by the EpiSEQ, RAPIDIA-FIELD and AniBioThreat EU projects, by the FORMAS BioBridges strong research environments project No. 2011-1692, by the National Veterinary Institute (SVA), and by the Department of Biomedical Sciences and Veterinary Public Health, BVF, Swedish University of Agricultural Sciences.

References

1. Belák S. (2007). – Molecular diagnosis of viral diseases, present trends and future aspects. A view from the OIE Collaborating Centre for the Application of Polymerase Chain Reaction Methods for Diagnosis of Viral Diseases in Veterinary Medicine. *Vaccine*, **25** (30), 5444–5452.

2. Belák S. (2007). – Experiences of an OIE Collaborating Centre in molecular diagnosis of transboundary animal diseases: a review. *Dev. Biol. (Basel)*, **128**, 103–112.

3. Belák S. & Liu L. (2013). – Recent advances in veterinary diagnostic virology: report from a Collaborating Centre of the World Organisation for Animal Health (OIE). *In* Advanced techniques in diagnostic microbiology (Y.-W. Tang & C.W. Stratton, eds). Springer, New York, 661–678.

4. Belák S., Thorén P., LeBlanc N. & Viljoen G. (2009). – Advances in viral disease diagnostic and molecular epidemiological technologies. *Expert Rev. mol. Diagn.*, **9** (4), 367–381.

5. Sanger F., Nicklen S. & Coulson A.R. (1977). – DNA sequencing with chain-terminating inhibitors. *Proc. natl Acad. Sci. USA*, **74** (12), 5463–5467.

6. Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R., Fiddes C.A., Hutchison C.A., Slocombe P.M. & Smith M. (1977). – Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, **265** (5596), 687–695.

7. Mullis K., Faloona F., Scharf S., Saiki R., Horn G. & Erlich H. (1986). – Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. quant. Biol.*, **51** (Pt 1), 263–273.

8. Bartlett J.M. & Stirling D. (2003). – A short history of the polymerase chain reaction. *Meth. molec. Biol.*, **226**, 3–6.

9. Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Heiner C., Kent S.B. & Hood L.E. (1986). – Fluorescence detection in automated DNA sequence analysis. *Nature*, **321** (6071), 674–679.

10. Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., Gocayne J.D., Amanatides P., Ballew R.M., Huson D.H., Wortman

J.R., Zhang Q., Kodira C.D., Zheng X.H., Chen L., Skupski M., Subramanian G., Thomas P.D., Zhang J., Gabor Miklos G.L., Nelson C., Broder S., Clark A.G., Nadeau J., McKusick V.A., Zinder N., Levine A.J., Roberts R.J., Simon M., Slayman C., Hunkapiller M., Bolanos R., Delcher A., Dew I., Fasulo D., Flanigan M., Florea L., Halpern A., Hannenhalli S., Kravitz S., Levy S., Mobarry C., Reinert K., Remington K., Abu-Threideh J., Beasley E., Biddick K., Bonazzi V., Brandon R., Cargill M., Chandramouliswaran I., Charlab R., Chaturvedi K., Deng Z., Di Francesco V., Dunn P., Eilbeck K., Evangelista C., Gabrielian A.E., Gan W., Ge W., Gong F., Gu Z., Guan P., Heiman T.J., Higgins M.E., Ji R.R., Ke Z., Ketchum K.A., Lai Z., Lei Y., Li Z., Li J., Liang Y., Lin X., Lu F., Merkulov G.V., Milshina N., Moore H.M., Naik A.K., Narayan V.A., Neelam B., Nusskern D., Rusch D.B., Salzberg S., Shao W., Shue B., Sun J., Wang Z., Wang A., Wang X., Wang J., Wei M., Wides R., Xiao C., Yan C., Yao A., Ye J., Zhan M., Zhang W., Zhang H., Zhao Q., Zheng L., Zhong F., Zhong W., Zhu S., Zhao S., Gilbert D., Baumhueter S., Spier G., Carter C., Cravchik A., Woodage T., Ali F., An H., Awe A., Baldwin D., Baden H., Barnstead M., Barrow I., Beeson K., Busam D., Carver A., Center A., Cheng M.L., Curry L., Danaher S., Davenport L., Desilets R., Dietz S., Dodson K., Doup L., Ferriera S., Garg N., Gluecksmann A., Hart B., Haynes J., Haynes C., Heiner C., Hladun S., Hostin D., Houck J., Howland T., Ibegwam C., Johnson J., Kalush F., Kline L., Koduru S., Love A., Mann F., May D., McCawley S., McIntosh T., McMullen I., Moy M., Moy L., Murphy B., Nelson K., Pfannkoch C., Pratts E., Puri V., Qureshi H., Reardon M., Rodríguez R., Rogers Y.H., Romblad D., Ruhfel B., Scott R., Sitter C., Smallwood M., Stewart E., Strong R., Suh E., Thomas R., Tint N.N., Tse S., Vech C., Wang G., Wetter J., Williams S., Williams M., Windsor S., Winn-Deen E., Wolfe K., Zaveri J., Zaveri K., Abril J.F., Guigo R., Campbell M.J., Sjolander K.V., Karlak B., Kejariwal A., Mi H., Lazareva B., Hatton T., Narechania A., Diemer K., Muruganujan A., Guo N., Sato S., Bafna V., Istrail S., Lippert R., Schwartz R., Walenz B., Yooseph S., Allen D., Basu A., Baxendale J., Blick L., Caminha M., Carnes-Stine J., Caulk P., Chiang Y.H., Coyne M., Dahlke C., Mays A., Dombroski M., Donnelly M., Ely D.,

Esparham S., Fosler C., Gire H., Glanowski S., Glasser K., Glodek A., Gorokhov M., Graham K., Gropman B., Harris M., Heil J., Henderson S., Hoover J., Jennings D., Jordan C., Jordan J., Kasha J., Kagan L., Kraft C., Levitsky A., Lewis M., Liu X., López J., Ma D., Majoros W., McDaniel J., Murphy S., Newman M., Nguyen T., Nguyen N., Nodell M., Pan S., Peck J., Peterson M., Rowe W., Sanders R., Scott J., Simpson M., Smith T., Sprague A., Stockwell T., Turner R., Venter E., Wang M., Wen M., Wu D., Wu M., Xia A., Zandieh A. & Zhu X. (2001). – The sequence of the human genome. *Science*, **291** (5507), 1304–1351.

11. Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J., Mesirov J.P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-Thomann N., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J.C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R.H., Wilson R.K., Hillier L.W., McPherson J.D., Marra M.A., Mardis E.R., Fulton L.A., Chinwalla A.T., Pepin K.H., Gish W.R., Chissoe S.L., Wendl M.C., Delehaanty K.D., Miner T.L., Delehaanty A., Kramer J.B., Cook L.L., Fulton R.S., Johnson D.L., Minx P.J., Clifton S.W., Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N., Cheng J.F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M., Gibbs R.A., Muzny D.M., Scherer S.E., Bouck J.B., Sodergren E.J., Worley K.C., Rives C.M., Gorrell J.H., Metzker M.L., Naylor S.L., Kucherlapati R.S., Nelson D.L., Weinstock G.M., Sakaki Y., Fujiyama A., Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T., Weissenbach J., Heilig R., Saurin W., Artiguenave F., Brottier P., Bruls T., Pelletier E., Robert C., Wincker P., Smith D.R., Doucette-Stamm L., Rubenfield M.,

Weinstock K., Lee H.M., Dubois J., Rosenthal A., Platzer M., Nyakatura G., Taudien S., Rump A., Yang H., Yu J., Wang J., Huang G., Gu J., Hood L., Rowen L., Madan A., Qin S., Davis R.W., Federspiel N.A., Abola A.P., Proctor M.J., Myers R.M., Schmutz J., Dickson M., Grimwood J., Cox D.R., Olson M.V., Kaul R., Raymond C., Shimizu N., Kawasaki K., Minoshima S., Evans G.A., Athanasiou M., Schultz R., Roe B.A., Chen F., Pan H., Ramser J., Lehrach H., Reinhardt R., McCombie W.R., de la Bastide M., Dedhia N., Blocker H., Hornischer K., Nordsiek G., Agarwala R., Aravind L., Bailey J.A., Bateman A., Batzoglu S., Birney E., Bork P., Brown D.G., Burge C.B., Cerutti L., Chen H.C., Church D., Clamp M., Copley R.R., Doerks T., Eddy S.R., Eichler E.E., Furey T.S., Galagan J., Gilbert J.G., Harmon C., Hayashizaki Y., Haussler D., Hermjakob H., Hokamp K., Jang W., Johnson L.S., Jones T.A., Kasif S., Kasprzyk A., Kennedy S., Kent W.J., Kitts P., Koonin E.V., Korf I., Kulp D., Lancet D., Lowe T.M., McLysaght A., Mikkelsen T., Moran J.V., Mulder N., Pollara V.J., Ponting C.P., Schuler G., Schultz J., Slater G., Smit A.F., Stupka E., Szustakowski J., Thierry-Mieg D., Thierry-Mieg J., Wagner L., Wallis J., Wheeler R., Williams A., Wolf Y.I., Wolfe K.H., Yang S.P., Yeh R.F., Collins F., Guyer M.S., Peterson J., Felsenfeld A., Wetterstrand K.A., Patrinos A., Morgan M.J., de Jong P., Catanese J.J., Osoegawa K., Shizuya H., Choi S., Chen Y.J. & International Human Genome Sequencing Consortium (2001). – Initial sequencing and analysis of the human genome. *Nature*, **409** (6822), 860–921.

12. Shendure J., Porreca G.J., Reppas N.B., Lin X., McCutcheon J.P., Rosenbaum A.M., Wang M.D., Zhang K., Mitra R.D. & Church G.M. (2005). – Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, **309** (5741), 1728–1732.

13. Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X.V., Godwin B.C., He W., Helgesen S., Ho C.H., Irzyk G.P., Jando S.C., Alenquer M.L., Jarvie T.P., Jirage K.B., Kim J.B., Knight J.R., Lanza J.R., Leamon J.H., Lefkowitz S.M., Lei M., Li J., Lohman K.L., Lu H., Makhijani V.B.,

McDade K.E., McKenna M.P., Myers E.W., Nickerson E., Nobile J.R., Plant R., Puc B.P., Ronan M.T., Roth G.T., Sarkis G.J., Simons J.F., Simpson J.W., Srinivasan M., Tartaro K.R., Tomasz A., Vogt K.A., Volkmer G.A., Wang S.H., Wang Y., Weiner M.P., Yu P., Begley R.F. & Rothberg J.M. (2005). – Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437** (7057), 376–380.

14. Palacios G., Druce J., Du L., Tran T., Birch C., Briese T., Conlan S., Quan P.L., Hui J., Marshall J., Simons J.F., Egholm M., Paddock C.D., Shieh W.J., Goldsmith C.S., Zaki S.R., Catton M. & Lipkin W.I. (2008). – A new arenavirus in a cluster of fatal transplant-associated diseases. *N. Engl. J. Med.*, **358** (10), 991–998.

15. Briese T., Paweska J.T., McMullan L.K., Hutchison S.K., Street C., Palacios G., Khristova M.L., Weyer J., Swanepoel R., Egholm M., Nichol S.T. & Lipkin W.I. (2009). – Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.*, **5** (5), e1000455.

16. Cox-Foster D.L., Conlan S., Holmes E.C., Palacios G., Evans J.D., Moran N.A., Quan P.L., Briese T., Hornig M., Geiser D.M., Martinson V., vanEngelsdorp D., Kalkstein A.L., Drysdale A., Hui J., Zhai J., Cui L., Hutchison S.K., Simons J.F., Egholm M., Pettis J.S. & Lipkin W.I. (2007). – A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, **318** (5848), 283–287.

17. Bennett S. (2004). – Solexa Ltd. *Pharmacogenomics*, **5** (4), 433–438.

18. Moore G.E. (1998). – Cramming more components onto integrated circuits [Reprinted from *Electronics*, 19 April 1965, **38** (8), 114–117]. In *Proc. IEEE*, **86** (1), 82–85. doi:10.1109/JPROC.1998.658762.

19. Sboner A., Mu X.J., Greenbaum D., Auerbach R.K. & Gerstein M.B. (2011). – The real cost of sequencing: higher than you think! *Genome Biol.*, **12** (8), 125.

20. Schadt E.E., Turner S. & Kasarskis A. (2010). – A window into third-generation sequencing. *Hum. molec. Genet.*, **19**, R227–R240.

21. Cheval J., Sauvage V., Frangeul L., Dacheux L., Guigon G., Dumey N., Pariente K., Rousseaux C., Dorange F., Berthet N., Brisse S., Moszer I., Bourhy H., Manuguerra C.J., Lecuit M., Burguiere A., Caro V. & Eloit M. (2011). – Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J. clin. Microbiol.*, **49** (9), 3268–3275.

22. Daly G.M., Bexfield N., Heaney J., Stubbs S., Mayer A.P., Palser A., Kellam P., Drou N., Caccamo M., Tiley L., Alexander G.J., Bernal W. & Heaney J.L. (2011). – A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. *PLoS ONE*, **6** (12), e28879.

23. World Organisation for Animal Health (OIE) (2009). – Manual of Diagnostic Tests for Aquatic Animals, 6th Ed. OIE, Paris. Available at: www.oie.int/international-standard-setting/aquatic-manual/access-online/ (accessed on 20 October 2013).

24. World Organisation for Animal Health (OIE) (2012). – Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, 7th Ed. OIE, Paris. Available at: www.oie.int/international-standard-setting/terrestrial-manual/access-online/ (accessed on 20 October 2013).

25. Field D., Amaral-Zettler L., Cochrane G., Cole J.R., Dawyndt P., Garrity G.M., Gilbert J., Glöckner F.O., Hirschman L., Karsch-Mizrachi I., Klenk H.-P., Knight R., Kottmann R., Kyrpides N., Meyer F., San Gil I., Sansone S.-A., Schriml L.M., Sterk P., Tatusova T., Ussery D.W., White O. & Wooley J. (2011). – The Genomic Standards Consortium. *PLoS Biol.*, **9** (6), e1001088.

26. Thurber R.V., Haynes M., Breitbart M., Wegley L. & Rohwer F. (2009). – Laboratory procedures to generate viral metagenomes. *Nat. Protoc.*, **4** (4), 470–483.

27. Tan le V., Van Doorn H.R., Van der Hoek L., Minh Hien V., Jebbink M.F., Quang Ha D., Farrar J., Van Vinh Chau N. & de Jong M.D. (2011). – Random PCR and ultracentrifugation increases sensitivity and throughput of VIDISCA for screening of pathogens in clinical specimens. *J. Infect. dev. Ctries*, **5** (2), 142–148.

28. Lawrence J.E. & Steward G.F. (2010). – Purification of viruses by centrifugation. *In* Manual of aquatic viral ecology (S.W. Wilhelm, M.G. Weinbauer & C.A. Suttle, eds). Association for the Sciences of Limnology and Oceanography (ASLO), Waco, Texas, 166–181. doi:10.4319/mave.2010.978-0-9845591-0-7.

29. Djikeng A. & Spiro D. (2009). – Advancing full length genome sequencing for human RNA viral pathogens. *Future Virology*, **4** (1), 47–53.

30. Allander T., Emerson S.U., Engle R.E., Purcell R.H. & Bukh J. (2001). – A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc. natl Acad. Sci. USA*, **98** (20), 11609–11614.

31. Berthet N., Reinhardt A.K., Leclercq I., van Ooyen S., Batejat C., Dickinson P., Stamboliyska R., Old I.G., Kong K.A., Dacheux L., Bourhy H., Kennedy G.C., Korfhage C., Cole S.T. & Manuguerra J.C. (2008). – Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. *BMC Molec. Biol.*, **9**, 77. doi:10.1186/1471-2199-9-77.

32. Reyes G.R. & Kim J.P. (1991). – Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Molec. cell. Probes*, **5** (6), 473–481.

33. Kim K.H. & Bae J.W. (2011). – Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl. environ. Microbiol.*, **77** (21), 7663–7668.

34. Hongoh Y. & Toyoda A. (2011). – Whole-genome sequencing of unculturable bacterium using whole-genome amplification. *Meth. molec. Biol.*, **733**, 25–33.

35. Sergeant M.J., Constantinidou C., Cogan T., Penn C.W. & Pallen M.J. (2012). – High-throughput sequencing of 16S rRNA gene amplicons: effects of extraction procedure, primer length and annealing temperature. *PLoS ONE*, **7** (5), e38094.

36. Bybee S.M., Bracken-Grissom H., Haynes B.D., Hermansen R.A., Byers R.L., Clement M.J., Udall J.A., Wilcox E.R. & Crandall K.A. (2011). – Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biol. Evol.*, **3**, 1312–1323. doi:10.1093/gbe/evr106

37. Linnarsson S. (2010). – Recent advances in DNA sequencing methods: general principles of sample preparation. *Experim. Cell Res.*, **316** (8), 1339–1343.

38. Aird D., Ross M.G., Chen W.S., Danielsson M., Fennell T., Russ C., Jaffe D.B., Nusbaum C. & Gnirke A. (2011). – Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.*, **12** (2), R18. doi:10.1186/gb-2011-12-2-r18.

39. Fedurco M., Romieu A., Williams S., Lawrence I. & Turcatti G. (2006). – BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.*, **34** (3), e22.

40. Quail M.A., Smith M., Coupland P., Otto T.D., Harris S.R., Connor T.R., Bertoni A., Swerdlow H.P. & Gu Y. (2012). – A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

41. Rothberg J.M., Hinz W., Rearick T.M., Schultz J., Mileski W., Davey M., Leamon J.H., Johnson K., Milgrew M.J., Edwards M., Hoon J., Simons J.F., Marran D., Myers J.W., Davidson J.F., Branting A., Nobile J.R., Puc B.P., Light D., Clark T.A., Huber M., Branciforte

J.T., Stoner I.B., Cawley S.E., Lyons M., Fu Y., Homer N., Sedova M., Miao X., Reed B., Sabina J., Feierstein E., Schorn M., Alanjary M., Dimalanta E., Dressman D., Kasinskas R., Sokolsky T., Fidanza J.A., Namsaraev E., McKernan K.J., Williams A., Roth G.T. & Bustillo J. (2011). – An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, **475** (7356), 348–352.

42. Levene M.J., Korlach J., Turner S.W., Foquet M., Craighead H.G. & Webb W.W. (2003). – Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, **299** (5607), 682–686.

43. Eid J., Fehr A., Gray J., Luong K., Lyle J., Otto G., Peluso P., Rank D., Baybayan P., Bettman B., Bibillo A., Bjornson K., Chaudhuri B., Christians F., Cicero R., Clark S., Dalal R., Dewinter A., Dixon J., Foquet M., Gaertner A., Hardenbol P., Heiner C., Hester K., Holden D., Kearns G., Kong X., Kuse R., Lacroix Y., Lin S., Lundquist P., Ma C., Marks P., Maxham M., Murphy D., Park I., Pham T., Phillips M., Roy J., Sebra R., Shen G., Sorenson J., Tomaney A., Travers K., Trulson M., Vieceli J., Wegener J., Wu D., Yang A., Zaccarin D., Zhao P., Zhong F., Korlach J. & Turner S. (2009). – Real-time DNA sequencing from single polymerase molecules. *Science*, **323** (5910), 133–138.

44. Dekker C. (2007). – Solid-state nanopores. *Nat. Nanotechnol.*, **2** (4), 209–215.

45. Howorka S., Cheley S. & Bayley H. (2001). – Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat. Biotechnol.*, **19** (7), 636–639.

46. Benner S., Chen R.J., Wilson N.A., Abu-Shumays R., Hurt N., Lieberman K.R., Deamer D.W., Dunbar W.B. & Akeson M. (2007). – Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat. Nanotechnol.*, **2** (11), 718–724.

47. Kasianowicz J.J., Brandin E., Branton D. & Deamer D.W. (1996). – Characterization of individual polynucleotide molecules using a membrane channel. *Proc. natl Acad. Sci. USA*, **93** (24), 13770–13773.

48. Branton D., Deamer D.W., Marziali A., Bayley H., Benner S.A., Butler T., Di Ventra M., Garaj S., Hibbs A., Huang X., Jovanovich S.B., Krstic P.S., Lindsay S., Ling X.S., Mastrangelo C.H., Meller A., Oliver J.S., Pershin Y.V., Ramsey J.M., Riehn R., Soni G.V., Tabard-Cossa V., Wanunu M., Wiggin M. & Schloss J.A. (2008). – The potential and challenges of nanopore sequencing. *Nat. Biotechnol.*, **26** (10), 1146–1153.

49. Metzker M.L. (2010). – Sequencing technologies: the next generation. *Nat. Rev. Genet.*, **11** (1), 31–46.

50. Shendure J. & Ji H. (2008). – Next-generation DNA sequencing. *Nat. Biotechnol.*, **26** (10), 1135–1345.

51. Cock P.J., Fields C.J., Goto N., Heuer M.L. & Rice P.M. (2010). – The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.*, **38** (6), 1767–1771.

52. Schmieder R. & Edwards R. (2011). – Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27** (6), 863–864.

53. Schmieder R., Lim Y.W., Rohwer F. & Edwards R. (2010). – TagCleaner: identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics*, **11**, 341.

54. Li H. & Homer N. (2010). – A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinform.*, **11** (5), 473–483.

55. Pop M. (2009). – Genome assembly reborn: recent computational challenges. *Brief. Bioinform.*, **10** (4), 354–366.

56. Scholz M.B., Lo C.C. & Chain P.S. (2012). – Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr. Opin. Biotechnol.*, **23** (1), 9–15.

57. Simon C. & Daniel R. (2011). – Metagenomic analyses: past and future trends. *Appl. environ. Microbiol.*, **77** (4), 1153–1161.

58. Chen K. & Pachter L. (2005). – Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS comput. Biol.*, **1** (2), 106–112.

59. Edwards R.A. & Rohwer F. (2005). – Viral metagenomics. *Nat. Rev. Microbiol.*, **3** (6), 504–510.

60. Morozova O., Hirst M. & Marra M.A. (2009). – Applications of new sequencing technologies for transcriptome analysis. *Annu. Rev. Genomics Hum. Genet.*, **10**, 135–151.

61. Pepke S., Wold B. & Mortazavi A. (2009). – Computation for ChIP-seq and RNA-seq studies. *Nat. Methods*, **6** (11 Suppl.), S22–S32.

62. Editorial (2012). – The data deluge. *Nat. Cell Biol.*, **14** (11), 1231.

63. Editorial (2009). – Access denied? *Nature*, **462** (7271), 252. doi:10.1038/462252a

64. Brooksbank C., Cameron G. & Thornton J. (2010). – The European Bioinformatics Institute's data resources. *Nucleic Acids Res.*, **38** (Database issue), D17–D25.

65. Leinonen R., Sugawara H., Shumway M. & International Nucleotide Sequence Database Collaboration (2011). – The Sequence Read Archive. *Nucleic Acids Res.*, **39** (Database issue), D19–D21.

66. Kodama Y., Shumway M., Leinonen R. & International Nucleotide Sequence Database Collaboration (2012). – The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40** (Database issue), D54–D56.

67. European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL–EBI) (2012). – How fast is life science data growing? European Life Sciences Infrastructure for Biological Information (ELIXIR). ELIXIR, Hinxton, Cambridge, United Kingdom. Available at: www.elixir-europe.org/news/how-fast-life-science-data-growing (accessed on 25 November 2012).

68. Rosegrant M.W., Paisner M.S., Meijer S. & Witcover J. (2001). – 2020 Global food outlook: trends, alternatives, and choices. International Food Policy Research Institute, Washington, DC.

69. Blomström A.L., Belák S., Fossum C., McKillen J., Allan G., Wallgren P. & Berg M. (2009). – Detection of a novel porcine boca-like virus in the background of porcine circovirus type 2 induced postweaning multisystemic wasting syndrome. *Virus Res.*, **146** (1–2), 125–129.

70. Blomström A.L., Belák S., Fossum C., Fuxler L., Wallgren P. & Berg M. (2010). – Studies of porcine circovirus type 2, porcine boca-like virus and Torque teno virus indicate the presence of multiple viral infections in postweaning multisystemic wasting syndrome pigs. *Virus Res.*, **152** (1–2), 59–64.

71. Shan T., Li L., Simmonds P., Wang C., Moeser A. & Delwart E. (2011). – The fecal virome of pigs on a high-density farm. *J. Virol.*, **85** (22), 11697–11708.

72. Sachsenröder J., Twardziok S., Hammerl J.A., Janczyk P., Wrede P., Hertwig S. & Johne R. (2012). – Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS ONE*, **7** (4), e34631.

73. Reuter G., Nemes C., Boros A., Kapusinszky B., Delwart E. & Pankovics P. (2013). – Porcine kobuvirus in wild boars (*Sus scrofa*). *Arch. Virol.*, **158** (1), 281–282. doi:10.1007/s00705-012-1456-y.

74. Reuter G., Nemes C., Boros A., Kapusinszky B., Delwart E. & Pankovics P. (2012). – Astrovirus in wild boars (*Sus scrofa*) in Hungary. *Arch. Virol.*, **157** (6), 1143–1147.

75. Blomström A.L., Ståhl K., Masembe C., Okoth E., Okurut A.R., Atmnedi P., Kemp S., Bishop R., Belák S. & Berg M. (2012). – Viral metagenomic analysis of bushpigs (*Potamochoerus larvatus*) in Uganda identifies novel variants of porcine parvovirus 4 and Torque teno sus virus 1 and 2. *Virol. J.*, **9**, 192.

76. Masembe C., Michuki G., Onyango M., Rumberia C., Norling M., Bishop R.P., Djikeng A., Kemp S.J., Orth A., Skilton R.A., Ståhl K. & Fischer A. (2012). – Viral metagenomics demonstrates that domestic pigs are a potential reservoir for Ndumu virus. *Virol. J.*, **9** (1), 218.

77. Xiao S., Jia J., Mo D., Wang Q., Qin L., He Z., Zhao X., Huang Y., Li A., Yu J., Niu Y., Liu X. & Chen Y. (2010). – Understanding PRRSV infection in porcine lung based on genome-wide transcriptome response identified by deep sequencing. *PLoS ONE*, **5** (6), e11377.

78. Boyle D.B., Bulach D.M., Amos-Ritchie R., Adams M.M., Walker P.J. & Weir R. (2012). – Genomic sequences of Australian bluetongue virus prototype serotypes reveal global relationships and possible routes of entry into Australia. *J. Virol.*, **86** (12), 6724–6731.

79. Wright C.F., Morelli M.J., Thebaud G., Knowles N.J., Herzyk P., Paton D.J., Haydon D.T. & King D.P. (2011). – Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *J. Virol.*, **85** (5), 2266–2275.

80. Beer M., Conraths F.J. & van der Poel W.H. (2012). – ‘Schmallenberg virus’: a novel orthobunyavirus emerging in Europe. *Epidemiol. Infect.*, **141** (1), 1–8. doi:10.1017/S0950268812002245.

81. Hoffmann B., Scheuch M., Hoper D., Jungblut R., Holsteg M., Schirrmeier H., Eschbaumer M., Goller K.V., Wernike K., Fischer

M., Breithaupt A., Mettenleiter T.C. & Beer M. (2012). – Novel orthobunyavirus in cattle, Europe, 2011. *Emerg. infect. Dis.*, **18** (3), 469–472.

82. Rosseel T., Scheuch M., Hoper D., De Regge N., Caij A.B., Vandenbussche F. & Van Borm S. (2012). – DNase SISPA-next generation sequencing confirms Schmallenberg virus in Belgian field samples and identifies genetic variation in Europe. *PLoS ONE*, **7** (7), e41967.

83. Hubalek Z. (2004). – An annotated checklist of pathogenic microorganisms associated with migratory birds. *J. Wildl. Dis.*, **40** (4), 639–659.

84. Alexander D.J. (2000). – A review of avian influenza in different bird species. *Vet. Microbiol.*, **74** (1–2), 3–13.

85. Gilchrist P. (2005). – Involvement of free-flying wild birds in the spread of the viruses of avian influenza, Newcastle disease and infectious bursal disease from poultry products to commercial poultry. *World Poult. Sci. J.*, **61** (2), 198–214.

86. Cardona C.J., Xing Z., Sandrock C.E. & Davis C.E. (2009). – Avian influenza in birds and mammals. *Comp. Immunol. Microbiol. infect. Dis.*, **32** (4), 255–273.

87. Hoper D., Hoffmann B. & Beer M. (2009). – Simple, sensitive, and swift sequencing of complete H5N1 avian influenza virus genomes. *J. clin. Microbiol.*, **47** (3), 674–679.

88. Kampmann M.L., Fordyce S.L., Avila-Arcos M.C., Rasmussen M., Willerslev E., Nielsen L.P. & Gilbert M.T. (2011). – A simple method for the parallel deep sequencing of full influenza A genomes. *J. virol. Meth.*, **178** (1–2), 243–248.

89. Van Borm S., Rosseel T., Vangeluwe D., Vandenbussche F., van den Berg T. & Lambrecht B. (2012). – Phylogeographic analysis of avian influenza viruses isolated from Charadriiformes in

Belgium confirms intercontinental reassortment in gulls. *Arch. Virol.*, **157** (8), 1509–1522.

90. Hoper D., Kalthoff D., Hoffmann B. & Beer M. (2012). – Highly pathogenic avian influenza virus subtype H5N1 escaping neutralization: more than HA variation. *J. Virol.*, **86** (3), 1394–1404.

91. Ramakrishnan M.A., Tu Z.J., Singh S., Chockalingam A.K., Gramer M.R., Wang P., Goyal S.M., Yang M., Halvorson D.A. & Sreevatsan S. (2009). – The feasibility of using high resolution genome sequencing of influenza A viruses to detect mixed infections and quasispecies. *PLoS ONE*, **4** (9), e7105. doi:10.1371/journal.pone.0007105.

92. Dugan V.G., Dunham E.J., Jin G., Sheng Z.M., Kaser E., Nolting J.M., Alexander H.L. Jr, Slemons R.D. & Taubenberger J.K. (2011). – Phylogenetic analysis of low pathogenicity H5N1 and H7N3 influenza A virus isolates recovered from sentinel, free flying, wild mallards at one study site during 2006. *Virology*, **417** (1), 98–105. doi:10.1016/j.virol.2011.05.004.

93. Croville G., Soubies S.M., Barbieri J., Klopp C., Mariette J., Bouchez O., Camus-Bouclainville C. & Guerin J.L. (2012). – Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *J. clin. Microbiol.*, **50** (9), 2881–2887.

94. Trojnar E., Sachsenroder J., Twardziok S., Reetz J., Otto P.H. & Johne R. (2013). – Identification of an avian group A rotavirus containing a novel VP4 gene of close relationship to those of mammalian rotaviruses. *J. gen. Virol.*, **94** (Pt 1), 136–142. doi:10.1099/vir.0.047381-0.

95. Abolnik C., de Castro M. & Rees J. (2012). – Full genomic sequence of an African avian paramyxovirus type 4 strain isolated from a wild duck. *Virus Genes*, **45** (3), 537–541.

96. Spatz S.J., Volkening J.D., Keeler C.L., Kutish G.F., Riblet S.M., Boettger C.M., Clark K.F., Zsak L., Afonso C.L., Mundt E.S.,

Rock D.L. & Garcia M. (2012). – Comparative full genome analysis of four infectious laryngotracheitis virus (Gallid herpesvirus-1) virulent isolates from the United States. *Virus Genes*, **44** (2), 273–285. doi:10.1007/s11262-011-0696-3.

97. Spatz S.J. & Rue C.A. (2008). – Sequence determination of a mildly virulent strain (CU-2) of Gallid herpesvirus type 2 using 454 pyrosequencing. *Virus Genes*, **36** (3), 479–489.

98. Honkavuori K.S., Shivaprasad H.L., Williams B.L., Quan P.L., Hornig M., Street C., Palacios G., Hutchison S.K., Franca M., Egholm M., Briese T. & Lipkin W.I. (2008). – Novel borna virus in psittacine birds with proventricular dilatation disease. *Emerg. infect. Dis.*, **14** (12), 1883–1886.

99. Honkavuori K.S., Shivaprasad H.L., Briese T., Street C., Hirschberg D.L., Hutchison S.K. & Lipkin W.I. (2011). – Novel picornavirus in Turkey poults with hepatitis, California, USA. *Emerg. infect. Dis.*, **17** (3), 480–487.

100. Rosseel T., Lambrecht B., Vandebussche F., van den Berg T. & Van Borm S. (2011). – Identification and complete genome sequencing of paramyxoviruses in mallard ducks (*Anas platyrhynchos*) using random access amplification and next generation sequencing technologies. *Virol. J.*, **8**, 463.

101. Day J.M., Ballard L.L., Duke M.V., Scheffler B.E. & Zsak L. (2010). – Metagenomic analysis of the turkey gut RNA virus community. *Virol. J.*, **7**, 313.

102. Kim T. & Mundt E. (2011). – Metagenomic analysis of intestinal microbiomes in chickens. *Meth. molec. Biol.*, **733**, 185–194.

103. Nordentoft S., Molbak L., Bjerrum L., De Vylder J., Van Immerseel F. & Pedersen K. (2011). – The influence of the cage system and colonisation of *Salmonella* Enteritidis on the microbial gut flora of laying hens studied by T-RFLP and 454 pyrosequencing. *BMC Microbiol.*, **11**, 187.

104. Dumas M.D., Polson S.W., Ritter D., Ravel J., Gelb J. Jr, Morgan R. & Wommack K.E. (2011). – Impacts of poultry house environment on poultry litter bacterial community composition. *PLoS ONE*, **6** (9), e24785.

105. Nie Q., Sandford E.E., Zhang X., Nolan L.K. & Lamont S.J. (2012). – Deep sequencing-based transcriptome analysis of chicken spleen in response to avian pathogenic *Escherichia coli* (APEC) infection. *PLoS ONE*, **7** (7), e41645.

106. Wang Y., Brahmakshatriya V., Lupiani B., Reddy S.M., Soibam B., Benham A.L., Gunaratne P., Liu H.C., Trakooljul N., Ing N., Okimoto R. & Zhou H. (2012). – Integrated analysis of microRNA expression and mRNA transcriptome in lungs of avian influenza virus infected broilers. *BMC Genomics*, **13**, 278.

107. Spatz S.J. (2010). – Accumulation of attenuating mutations in varying proportions within a high passage very virulent plus strain of Gallid herpesvirus type 2. *Virus Res.*, **149** (2), 135–142.

108. Chandra Y.G., Lee J. & Kong B.W. (2012). – Genome sequence comparison of two United States live attenuated vaccines of infectious laryngotracheitis virus (ILTV). *Virus Genes*, **44** (3), 470–474.

109. Lee S.W., Markham P.F., Coppo M.J., Legione A.R., Markham J.F., Noormohammadi A.H., Browning G.F., Ficorilli N., Hartley C.A. & Devlin J.M. (2012). – Attenuated vaccines can recombine to form virulent field viruses. *Science*, **337** (6091), 188.

110. Marston D.A., Ellis R.J., Horton D.L., Kuzmin I.V., Wise E.L., McElhinney L.M., Banyard A.C., Ngeleja C., Keyyu J., Cleaveland S., Lembo T., Rupprecht C.E. & Fooks A.R. (2012). – Complete genome sequence of Ikoma lyssavirus. *J. Virol.*, **86** (18), 10242–10243.

111. Marston D.A., Horton D.L., Ngeleja C., Hampson K., McElhinney L.M., Banyard A.C., Haydon D., Cleaveland S., Rupprecht C.E., Bigambo M., Fooks A.R. & Lembo T. (2012). –

Ikoma lyssavirus, highly divergent novel lyssavirus in an African civet. *Emerg. infect. Dis.*, **18** (4), 664–667.

112. Granberg F., Vicente-Rubiano M., Rubio-Guerri C., Karlsson O.E., Kukielka D., Belák S. & Sanchez-Vizcaino J.M. (2013). – Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by aphid lethal paralysis, Israel acute paralysis and Lake Sinai viruses. *PLoS ONE*, **8** (2), e57459.

113. Dunne W.M. Jr, Westblade L.F. & Ford B. (2012). – Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. *Eur. J. clin. Microbiol. infect. Dis.*, **31** (8), 1719–1726.

Table I
Current high-throughput sequencing platforms and their characteristics

Platform	Library preparation time	Run time	Average read length (bp)	Throughput per run	Instrument cost	Run cost	Main biological applications
<i>Roche/454</i>							
GS FLX Titanium XL+	2 h to 6 h	23 h	Up to 1,000	700 Mb	~\$500k	~\$6k	<i>De novo</i> genome sequencing and resequencing, targeted amplicon sequencing, genotyping, transcriptomics, metagenomics
GS Junior	2 h to 6 h	10 h	~400	35 Mb	\$125k	~\$1k	
<i>Illumina</i>							
HiSeq 2500 (rapid)	3 h to 5 h	27 h	2 × 100	180 Gb	\$740k	~\$23k	Genome resequencing, targeted amplicon sequencing, genotyping, transcriptomics, metagenomics
HiSeq 2500 (max)	3 h to 5 h	11 days	2 × 100	600 Gb	\$740k	~\$23k	
MiSeq	90 min	~24 hours	2 × 250	7.5–8.5 Gb	\$125k	\$965	
<i>SOLID</i>							
5500 xl	~8 h	6 days	75	95 Gb	\$595k	~\$10k	Genome resequencing, genotyping, transcriptomics
5500 xl Wildfire	2 h	10 days	75	240 Gb	\$70k upgrade	~\$5k	
<i>Ion Torrent PGM</i>							
314 Chip	4 h to 6 h	2 h to 3 h	~400	10–40 Mb	\$50k	\$350	<i>De novo</i> microbial genome sequencing and resequencing, targeted amplicon sequencing, genotyping, RNA-seq on low-complexity transcriptome, metagenomics
316 Chip		2 h to 3 h	~400	100–400 Mb		\$550	
318 Chip		2 h to 3 h	~400	~1 Gb		\$750	
<i>Ion Torrent Proton</i>							
Chip I	4 h to 6 h	2 h to 4 h	~200	~10 Gb	\$149k	~\$1k	<i>De novo</i> genome sequencing and resequencing, targeted amplicon sequencing, genotyping, transcriptomics, metagenomics
Chip II		2 h 4 h	~20	~100 Gb		~\$1k	
<i>Pacific Biosciences</i>							

PacBio RS	Not applicable	2 h	~1,500 (C1 chemistry)	100 Mb	\$700k	\$100	Microbial genome sequencing, targeted amplicon sequencing, aids full-length transcriptomics, discovery of large structural variants and haplotypes
-----------	----------------	-----	-----------------------	--------	--------	-------	----------------------------------------------------------------------------------------------------------------------------------------------------

bp: base pair
Gb: gigabase (1,000 Mb)
k: thousand
Mb: megabase (1,000,000 bp)

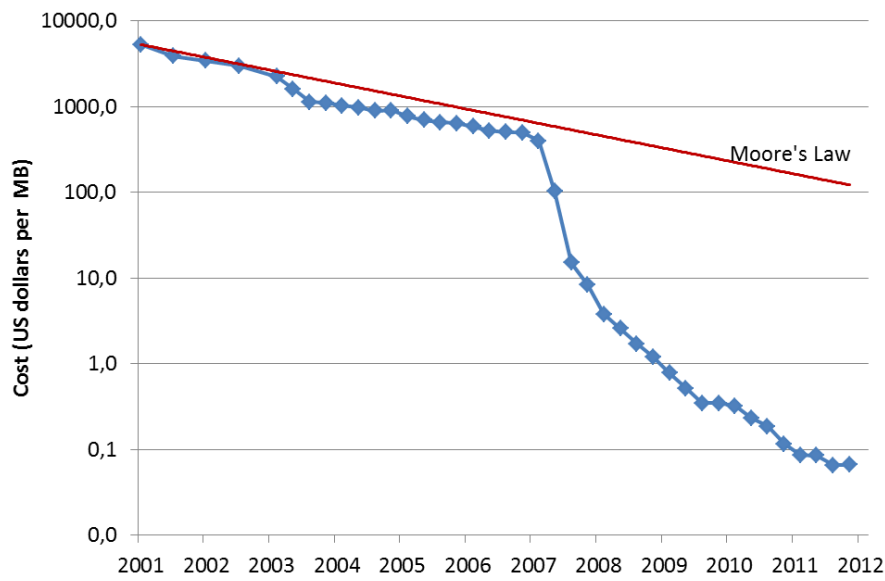


Fig. 1
The change over time for cost per raw megabase of DNA sequencing

Source: www.genome.gov/sequencingcosts/

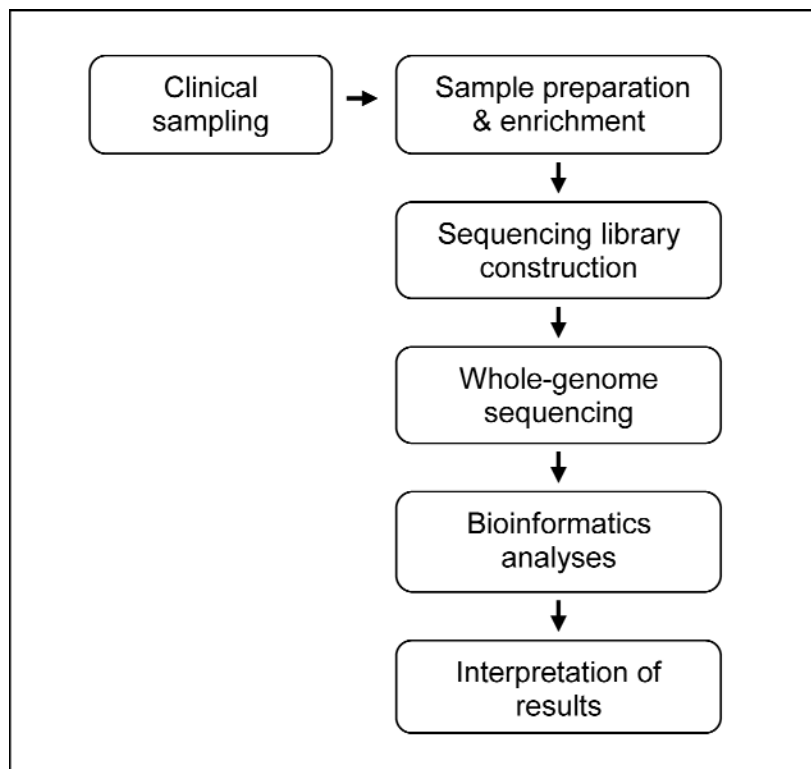


Fig. 2
Workflow for pathogen detection using high-throughput sequencing

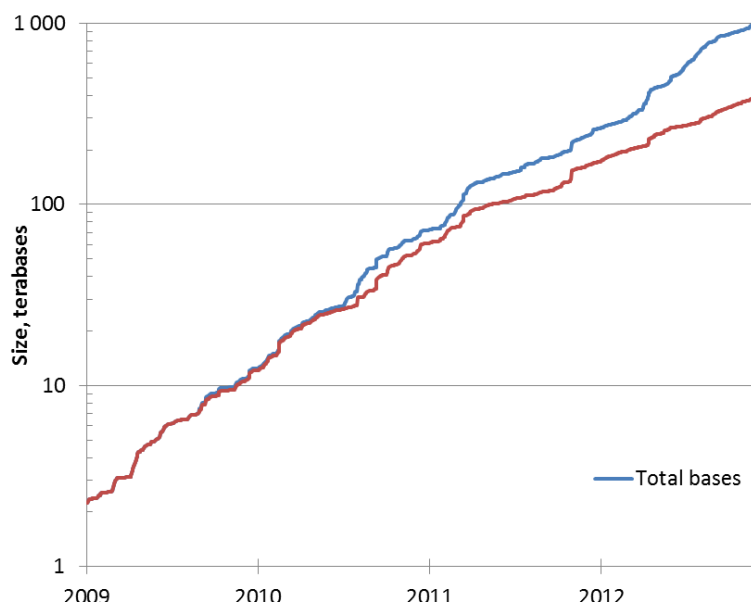


Fig. 3
The growth of information in the Sequence Read Archive, which stores raw sequencing data from next-generation sequencing platforms

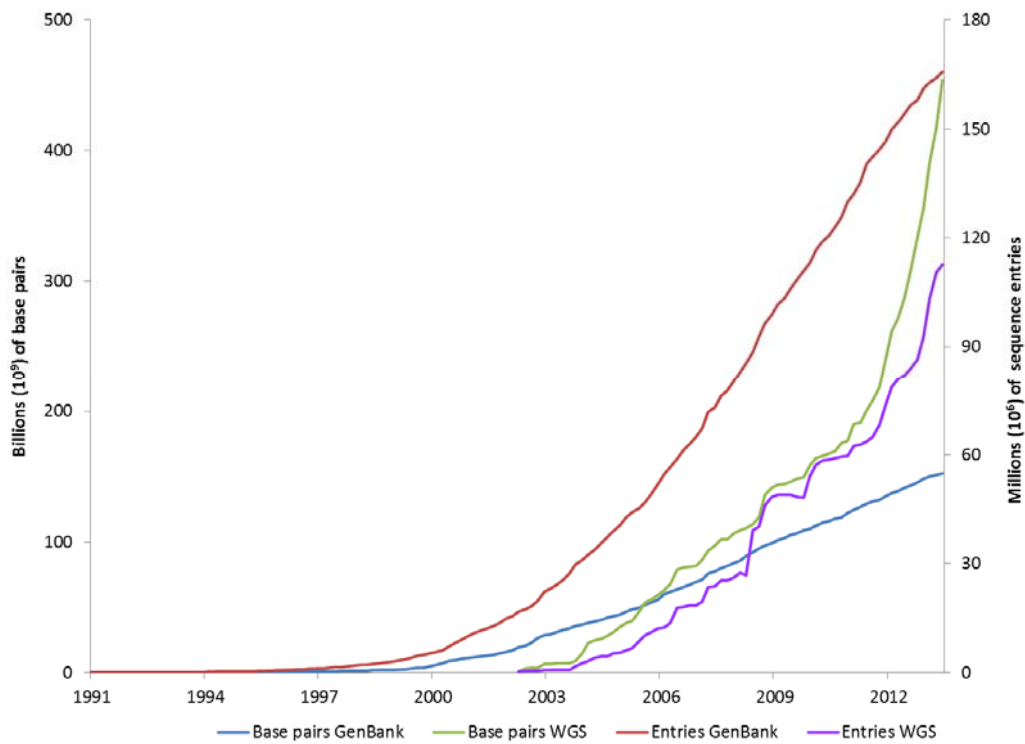


Fig. 4

The total number of base pairs and sequence entries in GenBank over the last 20 years, compared with the growth of whole-genome shotgun data from larger sequencing projects