Rev. Sci. Tech. Off. Int. Epiz., 2016, **35** (1), 67-81

Next-generation sequencing workflows in veterinary infection biology: towards validation and quality assurance

S. Van Borm [(1)*], J. Wang [(2)], F. Granberg [(3)] & A. Colling [(2)]

(1) Operational Directorate of Viral Diseases, Veterinary and Agrochemical Research Center (CODA-CERVA), Groeselenberg 99, B1180 Ukkel, Belgium
(2) Australian Animal Health Laboratory, CSIRO, Private Bag 24, Geelong, VIC 3220, Australia
(3) Swedish University of Agricultural Sciences (SLU), Dept. Biomedical Sciences and Veterinary Public Health, Box 7028, 750 07 Uppsala, Sweden
*Corresponding author: steven.vanborm@coda-cerva.be

Summary
Recent advancements in DNA sequencing methodologies and sequence data analysis have revolutionised research in many areas of biology and medicine, including veterinary infection biology. New technology is poised to bridge the gap between the research and diagnostic laboratory. This paper defines the potential diagnostic value and purposes of next-generation sequencing (NGS) applications in veterinary infection biology and explores their compatibility with the existing validation principles and methods of the World Organisation for Animal Health. Critical parameters for validation and quality control (quality metrics) are suggested, with reference to established validation and quality assurance guidelines for NGS-based methods of diagnosing human heritable diseases. Although most currently described NGS applications in veterinary infection biology are not primary diagnostic tests that directly result in control measures, this critical reflection on the advantages and remaining challenges of NGS technology should stimulate discussion on its diagnostic value and on the potential to validate NGS methods and monitor their diagnostic performance.

Keywords
Molecular diagnostics – Next-generation sequencing – Quality assurance – Validation.

Introduction

Genetic characterisation of infectious agents plays a central role in the diagnosis, monitoring and control of infectious diseases. The World Organisation for Animal Health (OIE) requires sequence information as part of pathogen characterisation. Rapid DNA sequencing methods based on the selective incorporation of chain-terminating dideoxynucleotides, termed Sanger or first-generation sequencing (1), and DNA amplification using the polymerase chain reaction (PCR) (2, 3) are now routinely used in the characterisation of infectious agents and provide valuable information for decision-makers. More recently, technologies for next-generation sequencing (NGS), also referred to as high-throughput sequencing or second-generation sequencing, have been developed, mostly in response to the development of an increasing number of large-scale genome sequencing projects, which are very demanding in terms of both time and resources. As a result of the development of desktop NGS platforms (4) and their availability in numerous laboratories that provide sequencing services, these technologies are now accessible to smaller research and diagnostic laboratories. Moreover, in recent years, there have been significant improvements in read length and error rate, as well as a reduction in the input DNA requirement and in protocol complexity, for all major NGS platforms. Major advantages of NGS technologies include:

– the dramatic increase in the amount of sequence information generated from a single sample

– the low cost per information unit (nucleotide)

– the ability to perform unbiased sequencing (i.e. without prior knowledge of the DNA content in a sample) or targeted sequencing.

doi: http://dx.doi.org/10.20506/rst.35.1.2418

A number of NGS platforms are currently available, each having their own strengths and limitations (reviewed in [5, 6, 7]). Second-generation sequencing platforms typically require the preparation of a DNA library from the sample, followed by clonal amplification of individual DNA molecules in the library and massive parallel sequencing of clonal template DNA. They currently offer the most cost-effective solution to producing large sequence datasets from a sample. Third-generation sequencing technologies, also known as single-molecule sequencing, (reviewed in [7, 8]), do not require clonal amplification of DNA molecules. Although they have additional advantages, such as a long read length and lack of clonal amplification errors, third-generation sequencing platforms have a higher sequencing error rate and higher operational cost at present.

In human diagnostic science, a number of NGS applications are currently transitioning from a research environment into diagnostic settings, where they result in clinically actionable decisions. Recently, guidelines were published for validating the NGS methods used in the diagnosis of heritable disorders (9) and in clinical oncology (10). Applications to infectious disease in the clinical setting include monitoring resistance mutations in RNA virus therapy (reviewed in [11]) and molecular typing of bacteria (reviewed in [12]). However, the revolutionising impact of NGS on genomic science is not limited to human health diagnostics. Published applications within veterinary infection biology include pathogen discovery and metagenomics, pathogen typing and characterisation, fine-scale molecular epidemiology, pathogenesis, and investigations of pathogen evolutionary responses to prophylaxis and treatment (reviewed in [13]). The increasing use of genomics in both human and veterinary medicine increases the availability of genomic information, allowing improvements in the diagnostic tools currently available in the clinic, e.g. by designing PCR assays for detecting microbial infectious agents, virulence factors and antibiotic-resistance determinants, and even developing optimised culture media for 'uncultivable' infectious agents (reviewed in [12]). For example, the detection of a novel orthobunyavirus in cattle in Germany (Schmallenberg virus) demonstrated the power of a metagenomic approach to discovering emerging pathogens, because specific and sensitive quantitative reverse-transcription PCRs (RT-qPCRs) could be developed quickly and used to analyse herd infections (14).

The purpose of the present paper is to investigate the potential for the validation of tests that apply NGS data to the diagnosis and characterisation of veterinary infectious disease. In Chapter 1.1.5 of the *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* (*Terrestrial Manual*) – 'Principles and methods of validation of diagnostic assays for infectious diseases' – the OIE underlines the importance of determining that diagnostic tests are fit for purpose, so that their results can be interpreted with confidence by decision-makers during disease outbreaks or used to support international trade (15). For example, specific guidelines are available for the development and validation of fundamentally different assays such as antibody, antigen and nucleic acid detection tests (15, 16, 17, 18). However, new and unique diagnostic reagents are continually being developed, along with novel assay platforms and protocols such as NGS methods, and there are ongoing discussions on how to validate these assays properly. To what degree can existing guidelines be used to validate methods based on NGS data? The authors identify current challenges associated with NGS technology and propose solutions for characterising the analytical test properties and monitoring analytical performance. Analytical and clinical validation is considered from a fitness-for-purpose perspective, with a focus on diagnostic relevance. The authors hope that this initial attempt to identify the critical parameters of NGS performance will contribute to discussions on the potential validation of these methods and ultimately lead to the formulation of guidelines by more formal workgroups.

Existing international guidelines and recommendations for the use of next-generation sequencing

Most existing guidelines stem from exhaustive workgroup efforts in the field of diagnosis of human heritable disease. For example, the Nex-StoCT workgroup, a group which is coordinated by the Centers for Disease Control and Prevention (9), has already developed the first FDA-approved NGS platform and diagnostic approach (19). European consortia are also publishing guidelines for the diagnostic use of NGS (20). Although most documents are not specifically tailored to the validation of NGS approaches for detecting or characterising infectious agents, several guidelines and recommendations for validation and quality assurance are generally applicable to any diagnostic use of NGS. Readers interested in establishing high-quality standardised NGS workflows may investigate details in the guidelines comprehensively reviewed by Bennett and Farah (10). Moreover, it is important to consider existing regulatory guidelines on diagnostic methods in general, such as Chapter 1.1.5 of the OIE *Terrestrial Manual* (15), and guidelines for the development, validation and quality monitoring of fundamentally differing assays, such as those for antibody, antigen and nucleic acid detection (16, 17, 18). The following sections describe some of the critical aspects of NGS applications in veterinary infection diagnostics. For any diagnostic or characterisation method using NGS technology, a critical reflection on the pathways and parameters for assay development, validation and

quality control, in accordance with the *Terrestrial Manual*, is crucial.

Definition of the intended purpose(s) of a next-generation sequencing application

Definition of the purpose(s) of the test is the first step of the assay development pathway and all subsequent decisions will depend on it (Fig. 1). Assays based on NGS have been used to study multifactorial diseases, co-infections, and new and emerging diseases caused by viruses (virus discovery) and other pathogens, and for testing for adventitious agents in vaccine seeds. Clinical and diagnostic applications for use in disease diagnosis and surveillance are in development.

The major applications of NGS assays are related to their specific methodologies, e.g. unbiased sequencing for pathogen discovery and targeted sequencing for further characterisation. If the assay is used for detecting previously unidentified microorganisms, such as during an outbreak investigation, then the primary purpose is diagnostic. If the assay is used to further characterise a previously identified pathogen or describe the development of quasispecies, then its general purpose is as an adjunct (characterisation) test.

Pathway	*Stage*	**Experiment**	**Parameter**
Assay development pathway		• Definition of intended purpose • Species, specimen and target analyte • Design & optimisation (including generation and analysis of sequence data) • Select platform, controls and reference materials • Protocol needs to be fixed before entering next stage	Development & optimisation of protocol
Assay validation pathway	*Stage 1 Analytical*	• Assay performance: ❖ Test performance ❖ Platform performance ❖ Data analysis pipelines • Suitable reference materials • Evaluation of ❖ Data analysis pipeline results ❖ NGS quality metrics (control samples)	Analytical Se & Sp Repeatability
	*Stage 2 Diagnostic**	• Comparison with standard test method • (Samples from infected & non-infected animals)*	Relative Se & Sp (Diagnostic Se & Sp)*
Validation status retention	*Stage 3 Precision*	• Daily & periodical inhouse QC • Monitoring and maintenance of validation criteria, e.g. ❖ Data analysis pipeline results (controls) ❖ NGS quality metrics of generated sequence data	Extended Repeatability &
	Stage 4 Implementation & monitoring	• Proficiency testing/Interlaboratory comparison	Reproducibility

*Only if used as a primary diagnostic test
NGS: next-generation sequencing
QC: quality control
Se: sensitivity
Sp: specificity

Fig. 1
Pathways, stages, experiments and parameters for validation of next-generation sequencing methods
Adapted from Chapter 1.1.5 of the OIE *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*, 'Principles and methods of validation of diagnostic assays for infectious diseases' (15). The focus is on the use of NGS methods as an adjunct test, requiring only analytical validation. If used as a primary diagnostic test, diagnostic validation parameters also need to be assessed; for example, testing samples from infected and non-infected animals for estimates of diagnostic sensitivity and specificity

Methodologically, an NGS application can rely on the unbiased generation of sequence information or on targeted sequencing:

– **Unbiased sequencing** or random sequencing: nucleic acids are extracted, DNA or complementary DNA is sequenced and the results are bioinformatically deduced. The applications for this type of sequencing include identifying unknown targets in metagenomics studies (characterising the entire nucleic acid content of a sample) and studying multifactorial disease syndromes. For example, a recent study investigated the association of bovine respiratory disease (BRD) with the presence of viral sequences; it identified several known and novel viruses, occurring either alone or in combination, in animals with BRD. It showed significant associations of BRD with bovine adenovirus 3, bovine rhinitis A virus and bovine influenza D virus (21). NGS assays may also be applied to known pathogens; for example, whole-genome sequencing without target amplification enables sequence information to be obtained from unexpected variants or multiple infections. In one study, the random generation of sequence information from multiple paramyxovirus isolates from wild ducks allowed the characterisation of avian paramyxoviruses 4 and 6, either alone or as a dual infection; in contrast, classic haemagglutination inhibition tests failed to detect the dual infection (22). Another study using randomly generated NGS data from pigeon paramyxovirus isolates identified co-infection with pigeon circoviruses (23). As sequence information is generated randomly, competition with other microbial and host nucleic acids is a challenge. In some applications, such as viral discovery metagenomics, a fraction of the sample is enriched (using ultracentrifugation, filtration, nuclease treatments, selective hybridisation, etc.) to increase the analytical sensitivity for a particular group (e.g. viral) of sequences (reviewed in [24, 25]). Another interesting strategy is the use of customised capture panels to enrich the viral sequence content in a metagenomics approach (26, 27).

– **Targeted sequencing** increases the analytical sensitivity and enables sequence information on a known (by primary diagnosis) infectious agent or on a genomic target region of a pathogen, to be directly obtained from the sample. It also allows cost-effective whole-genome sequencing of large numbers of samples. Sequencing is performed on nucleic acid targets that have been selectively enriched using amplification and/or alternative enrichment strategies. For example, high-throughput sequencing of PCR amplicons that target specific genomic regions permitted the direct analysis of adenovirus diversity in water samples (28) and multilocus sequence typing of *Salmonella* strains [28, 29]. In addition, alternative hybrid capture enrichment strategies (reviewed in [30]) have been applied to whole-genome sequencing of cell-associated human herpes virus genomes (30, 31). There is potential to generate deep-sequence coverage to analyse minor variants in pathogen populations (e.g. characterisation of RNA virus quasispecies to investigate emerging variants in a population).

Potential fields of application for NGS data range from extraction of known genetic marker profiles to identification of unknown infectious agents. The added value of the technology in research applications is well documented (13). Regarding the potential diagnostic use of NGS, the first step is a clear definition of the precise scope and purpose of a methodology using NGS.

Use of next-generation sequencing as a primary diagnostic method

Primary diagnostic methods are validated tests which provide information (e.g. about the infection or vaccination status of individuals or herds) that leads to actionable decisions (treatment, containment, biosecurity measures, etc.). In population-based veterinary medicine, such methods have many important applications, including research, prescribed surveillance activities, certification of freedom from disease, estimation of prevalence and epidemiological studies. Furthermore, information derived from tests often plays an important role in quantitative risk assessment and modelling of infectious diseases. In human diagnostics (e.g. heritable disease and cancer detection and pathogen typing), direct NGS-based detection of known molecular markers in patient samples is reaching the clinic. Numerous NGS applications in veterinary infectious disease biology are currently in the research phase and have the potential to provide considerable added diagnostic value. Some examples are given below. However, several factors complicate the validation of NGS tests as primary veterinary diagnostics, including:

– the cost and effort associated with analytical and diagnostic validation (OIE validation Chapter 1.1.5)

– the operational cost of the technology

– the workload associated with implementation and validation of a data analysis workflow

– the need for investment in hardware and expertise

– the time to result (currently days, compared with hours for specific molecular diagnostics such as real-time PCR).

As a result, there is potential for the technology to be used outside quality-assured laboratories or as adjunct characterisation tests. One potential application of NGS as a primary diagnostic method is in the detection, identification and characterisation of previously unidentified microorganisms. Although an important first step, the identification and genetic characterisation of candidate infectious agents is not enough to establish causal relationships or to understand how the agents may

be associated with disease. It is therefore necessary to use a synergistic approach that combines molecular diagnostic tools, such as NGS-based metagenomics and follow-up PCR-based assays targeting detected pathogen sequences, with more conventional diagnostic methods, including isolation and characterisation, electron microscopy, (histo)pathology and epidemiological findings. This synergistic approach has been used to characterise a novel orthomyxovirus from farmed ducks (32), novel rhabdoviruses from mosquitoes (33) and a novel phlebovirus from ticks (34). In an investigation of BRD, specific PCR assays targeting the most frequent viral NGS signatures in BRD samples demonstrated an association with BRD in a larger sample panel (21). Additional potential applications of NGS in veterinary primary diagnosis mirror the applications currently being developed for human diagnostics, mainly focusing on direct extraction of predetermined evidence-based molecular marker profiles from clinical samples. Diagnostic validation parameters are difficult to obtain, therefore the first objectives for these applications should be:

– assuring proper analytical validation

– establishing NGS-specific quality metrics to enhance validation and quality monitoring

– establishing a system to monitor quality.

Use of next-generation sequencing as an adjunct diagnostic method

Next-generation sequencing can be used as an adjunct diagnostic method as well as a primary diagnostic method. Chapter 1.1.5 of the *Terrestrial Manual* introduces the concept of adjunct diagnostic methods and describes them as '…secondary tests or procedures that are applied to an analyte that has been detected in a primary assay. The purpose of such analytical tools is to further characterise the analyte detected in the primary assay. Examples of such adjunct tests include virus neutralisation to type an isolated virus, and molecular sequencing' (15). According to the OIE, such tests need to be validated only for their analytical performance (stage 1: analytical sensitivity/specificity, repeatability, initial reproducibility) and not to the full diagnostic extent (stage 2: diagnostic sensitivity/specificity) (Fig. 1). If an NGS-based test is used for further characterisation after initial diagnosis it would fall under this category.

Adjunct tests for the molecular characterisation of veterinary infectious agents may include whole-genome sequencing, targeted resequencing and variant detection. Typically, the pathogen (e.g. avian influenza virus, bluetongue virus) will have been identified in a primary test. The adjunct molecular characterisation approach can target an isolated pathogen, the original clinical material or purified nucleic acids. One specific application is the direct extraction of molecular marker profiles from NGS data to provide information on pathotyping or resistance typing. Future tests may, for example, provide information on subtypes of avian influenza neuraminidase and haemagglutinin (e.g. H5N1) or a specific serotype within the bluetongue virus group. Examples in bacteriology include the determination of antibiotic resistance profiles, pathogenicity or genotype/serotype/serovar profiles (applications reviewed in [12]). In virology, the possibilities include extraction of known molecular markers for host adaptation, antiviral resistance or pathogenicity. An example in human virology is the use of NGS to document subtype co-infections and emergence of resistance mutations against antiviral drugs in circulating influenza A viruses (35). A second application is the determination of complete or partial genome sequences for molecular epidemiology purposes and molecular characterisation. Provided that there is sufficient depth of coverage of the targeted genomic regions, the analysis can include the detection of minor genetic variants in the pathogen population, allowing quasispecies resolution in molecular epidemiological or evolutionary dynamics studies. NGS approaches have allowed quasispecies characterisation of populations of foot and mouth disease virus in different lesions in a single animal (36) and during transmission experiments (37). The increased resolution in genetic characterisation that allows the detection of minor genetic variants in addition to the consensus genome has allowed detailed molecular epidemiological investigations of the emergence of a highly pathogenic avian influenza virus from a low pathogenic precursor (38).

Numerous additional applications of sequencing technologies in veterinary diagnostic science are reviewed in other chapters in the present volume.

Analytical validation of next-generation sequencing applications

Once the developmental pathway is deemed satisfactory and the protocol is established, the assay can proceed to the analytical validation stage (stage 1) (Fig. 1).

Most current NGS applications in veterinary diagnostics can be categorised as adjunct characterisation tests or as exploratory (primary) diagnostics. The analytical performance of a test can be assessed using well-described samples with known concentrations of target and non-target analytes and matrix components. Defining the analytical sensitivity, specificity and repeatability of the tests is the first, critical step in confidently interpreting the results. Analysis of statistically relevant numbers of samples from field-infected and non-infected animals to determine diagnostic sensitivity and specificity (Fig. 1, stage 2) may be problematic because of factors such as sample size and

representativeness, the nature and performance of the test, and the purpose and confidence level required for clinical use (e.g. what are the chances that a sick animal would be detected at a prevalence of 5% at 95% confidence interval) (39). Since diagnostic validation is not required for adjunct characterisation tests for designated purposes, this aspect is not discussed further.

Analytical sensitivity

The OIE defines analytical sensitivity as: 'synonymous with "limit of detection", smallest detectable amount of analyte that can be measured with a defined certainty; analyte may include antibodies, antigens, nucleic acids or live organisms' (40). In the NGS context, it can be defined as the likelihood that the assay will detect the targeted sequence variations, if present, at a given probability (e.g. 95% confidence), or formally:

$$Se = \frac{Number\ of\ positive\ results\ from\ samples\ tested}{Number\ of\ positive\ samples} \times 100$$

The target region may be a complete genome, a target gene, a marker region or a known (set of) variant(s). An exact definition thus depends on the application or purpose. For unbiased sequencing (pathogen discovery), sensitivity is the ability to detect a pathogen genome (including unknown targets) within the complete nucleic acid content (including host and other background) of the sample. For targeted resequencing (characterisation), sensitivity is the ability to generate sufficient sequence depth over the target region to allow predictions about the phenotype. For unbiased sequence data, the analytical sensitivity of NGS approaches is not easy to evaluate and is more critically influenced by matrix properties than in specific molecular diagnostic methods. The ability to eliminate host nucleic acid depends on the amount, its physical state and its association with particular proteins; however, its persistence may lead to false-negative sequencing results. Analytical sensitivity is also critically dependent on the depth of sequencing and the length of the target (genome or target region). Particularly in unbiased sequencing approaches (e.g. metagenomics), a longer genome (or target region) provides a greater number of potential read targets from a single genome copy, which confers an advantage for the detection of larger genomes. For targeted sequencing, analytical sensitivity should be assessed at a given coverage threshold across the genomic regions targeted for analysis.

Analytical specificity

The OIE defines analytical specificity as the 'degree to which the assay distinguishes between the target analyte and other components in the sample matrix; the higher the analytical specificity, the lower the level of false-positives' (40). Specificity can be further described in terms of selectivity, interference and carry-over. Selectivity refers to the extent to which a method can accurately quantify the targeted analyte in the presence of:

– interference from matrix components (e.g. enzyme inhibitors)

– degradants (e.g. nucleases)

– non-specific binding of reagents.

Interference may cause falsely reduced or elevated responses in the assay that negatively affect its analytical specificity. Careful design of the assay, such as targeted primer design to avoid non-specific amplification and the use of suitable control samples, assists in the monitoring of selectivity issues. In the NGS context, analytical specificity can be defined as the probability that an assay will not detect sequence variations when none are present at a given probability (95% confidence), or formally:

$$Sp = \frac{Number\ of\ negative\ results\ from\ samples\ tested}{Number\ of\ negative\ samples} \times 100$$

A better term for analytical specificity in NGS assays is the false-positive rate; this is affected by the sequence coverage, type of sequence variation and sequence context. Specificity should be assessed for the full workflow at a given coverage threshold across the genomic regions targeted for analysis.

Repeatability

Repeatability is the degree of agreement in the results for sample replicates both within and between runs for the same test method in a given laboratory. In some guidelines it is also referred to as 'precision'. The availability of reference materials to determine these parameters may be limited in the case of NGS. Because of the larger size of genomic regions analysed and the associated costs, only a limited number of samples can be sequenced and compared. Related parameters such as measurements of the distribution of coverage across the targeted region are useful for more detailed documentation of repeatability and reproducibility. Mean values and 95% confidence intervals are frequently used parameters of precision. Reproducibility is defined as the ability of a test method to provide consistent results when applied to aliquots of the same sample tested by the same method in different laboratories.

Accuracy

Accuracy, or the degree of agreement between the nucleic acid sequences derived from the assay and a reference sequence, depends on sequence coverage. The OIE defines accuracy as the nearness of a test value to a reference standard reagent of known activity or titre (40). The quality score is the standard method of reporting the quality of a base call. It should be noted that quality scores are generated

by platform-specific algorithms and are thus not directly comparable between platforms.

Comparison with standard methods or alternative tests

During validation, the results obtained with an NGS-based assay should be compared with those obtained using alternative methods, including genetic tests (e.g. Sanger sequencing, restriction fragment length polymorphism typing, genotyping PCR, etc.) and phenotyping methods. Results of one or more different tests obtained from a panel of samples from infected and non-infected animals can be expressed as the relative diagnostic sensitivity and specificity of the candidate tests.

Quality metrics

Because of the larger amount of DNA analysed (larger part of the genome, larger sequencing depth in a sample) and the associated costs, only a limited number of samples can be sequenced and compared for the formal determination of analytical test properties. Fortunately, the large amount of sequence data allows the use of NGS-specific parameters for the detailed documentation of repeatability and reproducibility. These NGS quality metrics (9) evaluate whether sufficient and reproducible sequence coverage and quality has been obtained for the target regions, and are directly related to analytical sensitivity and specificity. As the scope of NGS applications differs widely, it is difficult to establish generic guidelines for acceptable values of quality metrics. Suitable control measures might include running positive, negative and no-template controls in replicates of the test, and using a quality scoring system (39). The user should determine the normal and acceptable ranges of these metrics during assay validation and then monitor the observed values during sample analysis. Procedures should be in place to outline the actions required when a control limit is exceeded. For example, what are the likely consequences in terms of false-positive or false-negative results, or retesting of inconclusive samples? Quality metrics available for evaluating the analytical performance of NGS tests include:

– *Depth of coverage*. This refers to the number of sequence reads providing information about a given nucleotide. The characteristic depth of coverage of the target region(s) should be determined during validation. Sufficient coverage (minimum cut-off) that ensures adequate sensitivity and specificity should be defined. When ongoing quality monitoring shows that the depth of coverage at a given nucleotide is below the validated minimum coverage, confirmation should be provided using alternative methods (e.g. Sanger sequencing) or additional sequencing.

– *Uniformity of coverage*. This describes how the depth of coverage is distributed over the target region(s) of the test. Acceptable values should be described during validation. Deviations in uniformity of coverage from the validated range potentially indicate errors in the testing process.

– *GC bias*. The GC content (relative abundance of G and C nucleotides) of a target region affects the efficiency of sequencing reactions and will affect the uniformity of coverage. Where possible, the degree of GC bias in the target region(s) should be determined during validation and monitored to evaluate test performance.

– *Base-call quality scores*. These scores are platform dependent; they reflect the signal-to-noise ratio and thus the probability that the base call was correct. An acceptable threshold of raw base-call quality should be established during validation and incorporated in bioinformatics filters to eliminate poor quality data during analysis. Scores of base-call quality should be monitored to assess the quality of produced sequencing data.

– *Decline in signal intensity or read length*. Depending on the exact application, NGS platform and chemistry, the resulting sequence reads have a typical distribution of read length and signal intensity. The expected signal intensity across reads (or read length distribution) should be established during validation and monitored for each run. Deviations in the distribution of read lengths may indicate problematical datasets.

– *Mapping quality*. This measures the probability of a read being correctly mapped to a genomic position within the target region. Acceptable values (e.g. proportion of reads mapping to the target) should be established during validation of bioinformatics workflows and the proportion of reads not mapping to the target should be monitored during each run.

– *Internal controls*. Most platforms offer the possibility of spiking an internal control at very low frequency during the sequencing run. The quality metrics of those reads should be compared with reported quality metrics (e.g. the Illumina *PhiX* control: http://res.illumina.com/documents/products/technotes/technote_phixcontrolv3.pdf).

Quality assurance of next-generation sequencing applications

Following proper analytical validation of an NGS-based method, its performance needs to be monitored over time. To evaluate the analytical performance of successive sequencing runs, the first step is to compare the NGS quality metrics with acceptable ranges established during assay validation. Assay performance monitoring and validation requires that NGS

quality metrics should be obtained for all runs and controls and reference materials should be used. Every sequencing test ('run') should aim to include positive and negative controls appropriate to the intended purpose of the assay. Ideally, these controls monitor the entire assay, including sample preparation, NGS and the data analysis pipeline (Fig. 1, stage 3). Appropriate controls and NGS quality metrics must be used to verify each aspect of the procedure, including nucleic acid quality control, library preparation, monitoring cross-contamination (including for multiplexing), and determining the sensitivity, repeatability and, if the technology is transferred to another laboratory, reproducibility. Reference ranges for NGS quality metrics need to be established and used as indicators for accepting or rejecting a test run and procedures need to be in place for follow-up. However, the nature and current cost of the technology may limit the use of physical control samples. In such cases, NGS quality metrics to account for naturally occurring endogenous sequences within a sample (e.g. the presence of bacteriophage sequences to validate a viral metagenomics approach) or exogenous control sequences (such as platform-specific spiking, e.g. *PhiX* for the Illumina platform) may be useful. Consideration must be given to the steps in the process that are verified by a given control strategy, the choice being driven by the purpose of the NGS test.

Depending on the part of the workflow that needs to be validated, differing types of reference material may be suitable:

– *Biological sample*. Most similar to a diagnostic sample; can be selected to have known variant(s). Sometimes only a limited amount is available and this type of control is not necessarily renewable. For example, Li and colleagues (41) used a biological reagent consisting of 25 different human RNA and DNA viral pathogens to assess the efficiency of various viral metagenomics protocols. Collaborative efforts between laboratories may be useful to exchange and supply sample material.

– *Nucleic acid from a sample*. Unlike the biological sample, this does not take sample preparation (extraction, enrichment) into account.

– *Nucleic acid from an isolate or culture*. This is renewable and in large supply but may include variants resulting from passaging/replication of the pathogen.

– *Synthetic nucleic acid*. A broad range of sequences can be synthesised, and genetic variations can be spiked in. It is possible to make sequence templates with complex regions, including deletions and duplications, and large amounts can be manufactured easily. Disadvantages are that these synthetic controls may not recapitulate biological sample composition, may not cover the whole genome and may include variants resulting from errors in synthesis.

– *Electronic reference data files*. These can be engineered with any characteristics; they provide a good means of assessing software performance but do not take into account sample preparation and sequencing steps. Raw data files must mimic output data from technologies that evolve continuously as platforms and reagents are updated, and a dataset is needed for each platform (inter-platform comparison is difficult).

Apart from monitoring the quality metrics of ongoing analyses and reference material in comparison with those assessed during validation, several **critical parameters affecting test performance** need to be acknowledged and where possible monitored.

– Selection of the most appropriate type of **sample** (e.g. swab, tissue, blood, isolate) is key to the success of the NGS approach. The scope of the application, i.e. the animal species and sample matrices that can be interrogated using the method, must be defined.

– **Sample handling**, i.e. sample collection, handling and storage, must ensure sample integrity (42). Contamination prevention measures should be in place, taking the extreme sensitivity of the technology into account.

– Because of the costs associated with and complexity of NGS runs, it is imperative to include **quality assessment of intermediate steps**. The quantity and quality of the extracted nucleic acids should be assessed and compared with an acceptable range before proceeding to library preparation. Precise quantification and characterisation of the prepared sequencing library is required before proceeding to sequencing. General metrics of the raw output data should be evaluated before accepting the dataset and proceeding to full data analysis.

– Laboratories may choose to establish in-house NGS facilities or use (commercial) sequencing services, or otherwise ensure access to NGS. In all cases, **platform validation** is necessary to ensure the system is capable of providing reliable sequence analysis across the genomic regions targeted by the test. NGS quality metrics provide suitable parameters for the validation and monitoring of platform performance. Most platforms permit controls to be spiked into reagents; QC metrics of controls can then be used to monitor the performance of both platforms and reagents. Additional technology-specific performance metrics can be used to monitor platform performance (e.g. Illumina cluster density) and to identify aberrant sequencing runs.

– Typical NGS methods are complex. To ensure repeatability and interlaboratory reproducibility, it is therefore important to optimise the laboratory and bioinformatics components of each assay and to describe these in validated

standard operating procedures. All procedures, including sample processing (nucleic acid extraction, library preparation, barcoding, target enrichment), sequencing, bioinformatics and reporting, should be documented in standard operating procedures before validation can begin (Fig. 1, Assay development pathway).

– The eventual purpose of the NGS application will determine the nature of the data (complexity, full genome, depth required). The **bioinformatics workflow** comprises the processes designed to yield useful information from the raw data. Validation of data output from software pipelines can be achieved using datasets of reference sequences or control samples.

– Given the size and complexity of datasets associated with NGS applications ('big data'), laboratories need to ensure that proper **expertise in data handling** and data management is available. Information technology infrastructure and expertise should ensure safe data storage and management.

– The use of NGS currently requires a high level of **technical expertise**, including specialist knowledge of wet laboratory techniques, bioinformatics and information technology. Laboratories envisaging the use of NGS-based methods should therefore ensure that their staff are appropriately trained and certified.

Guidelines for NGS-based diagnostic genetic testing stress the importance of assessing interlaboratory reproducibility (Fig. 1, stage 4) to ensure the independent assessment of test performance. Proficiency testing or interlaboratory comparisons should focus on the analysis of sequence variations across the genomic regions targeted by the assay. The composition of the panel (species, number and nature of samples, replicates, etc.) defines the scope of the proficiency testing exercise (analytical sensitivity/specificity, repeatability, etc.). Procedures for ensuring homogeneity and stability of samples and analysis of results (z-score, Youden plot, etc.) need to be in place. OIE validation guideline 3.6.6 – 'Selection and use of reference samples and panels' describes a range of purposes and panels for proficiency testing (43). Testing programmes take into account assay differences in the recipient laboratory to allow proper comparison of interlaboratory performance and strive for harmonisation. Electronic sequence files may permit comparison of bioinformatics pipelines across laboratories but differences in sequencing platforms should be considered. At present, no proficiency testing is available for the evaluation of NGS applications in testing for veterinary infectious diseases. Although the considerable cost associated with organising and participating in an NGS-based proficiency testing exercise may currently still be prohibitive for laboratories, the organisation of such exercises in the future will be a huge step forward in method standardisation.

Challenges, and random and systematic error

The advantages of NGS are numerous, but several processes in the NGS workflow from sample to data interpretation are potentially vulnerable to error and/or introduction of bias (Fig. 2) (reviewed in [13]). Challenges and errors can be associated with sampling, sample processing, pre-amplification (targeted sequencing), library preparation, sequencing chemistry and data analysis.

Sampling

Systematic error can occur when the amount of genetic material for a particular pathogen or genetic variant in a sample from an individual or epidemiological unit is very small, as it may not be detected. This may compromise analytical sensitivity where the current cost of the method only allows a small sample size in the epidemiological unit or where unbiased sequencing results in competition with host and contaminating nucleic acids.

Assay chemistry and enzymatic reactions

Random and systematic errors can occur during pre-amplification, library preparation or sequencing and may have an impact on analytical sensitivity (i.e. decrease the chance of detecting a genetic marker present in the sample, with risk of false-negative results) or specificity (i.e. generate variation that is not present in the sample, thus increasing the chances of false-positive results). Pre-amplification for targeted sequencing may introduce point mutations and insertion/deletion errors that arise, for example, during RT and PCR amplification of the target region (depending on the error rates of the enzymes used). Amplification bias can affect the detected relative frequency of sequence variants present in the sample. Additional errors are associated with the sequencing chemistry and library construction, as well as with the clonal amplification of DNA to obtain clonal template sequences for second-generation sequencing. For example, the loss of synchronicity (dephasing) in a percentage of the clonally amplified DNA template (44) results in increased noise and sequencing errors (45). Every NGS platform available has its own distinct characteristics in terms of read and error profiles: the Illumina platform is often regarded as having the lowest error rate but other platforms can produce longer reads (4, 46).

Carry-over

As with any diagnostic method, laboratory or reagent contamination is likely to cause false-positive results. This

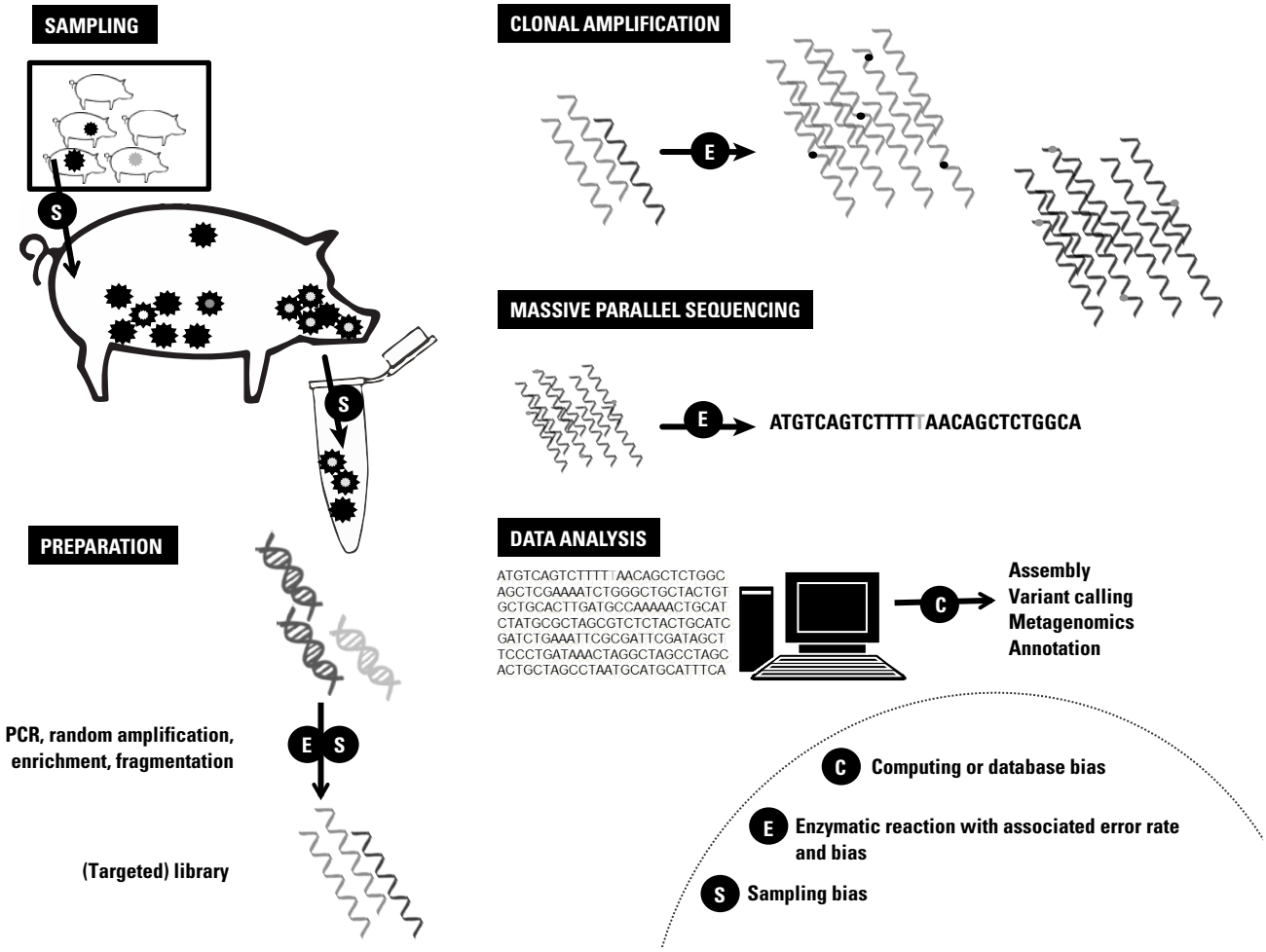

Fig. 2
Essential steps in the workflow of an NGS-based assay from sample to result
The figure indicates areas of potential error and bias introduction
Source: Van Borm *et al.* (see 13). © Springer Science+Business Media, New York, 2015

is particularly relevant in metagenomics strategies, where it is often difficult to exclude the possibility of laboratory or reagent contamination (47, 48). It is therefore important to adopt proper quality control measures that include the use of control samples. Sequencing laboratories sporadically report carry-over contamination between successive sequencing experiments (49), or crosstalk between sequencing libraries in the case of multiplexed runs (50).

Data analysis pipeline

Certain biases and errors can be introduced during the analysis of NGS datasets, because of the limitations of the algorithms or reference data(bases) used (51, 52). For example, genomic repeat regions present well-known problems for sequence assembly algorithms (53). The bioinformatics pipeline is key to managing some of the errors and biases introduced into the overall workflow: this includes sequence filtering (removing low-quality sequence information) and alignment, followed by variant calling and error correction. Discriminating true biological variants from those resulting from experimental noise is an important issue when trying to identify low-frequency variants in a population, for example, when studying viral quasispecies or in metagenomic analyses. The large size of the datasets requires significant computational and data-storage capacity, as well as expertise in information technology and bioinformatics. An interesting approach is the development of bioinformatics pipelines that are deployable on both standalone servers and publicly available cloud-based platforms (54). To evaluate the potential of NGS for a wider group of scientists and diagnosticians, there is a real need to develop flexible and practical bioinformatics workflows that can provide user-friendly tools for the analysis of massive datasets that become publicly available. Database exhaustiveness can be a determinant of the analytical sensitivity of a data analysis workflow. For example, a laboratory workflow may generate the appropriate raw sequence information (e.g. coverage of target region, sequence reads of an unknown pathogen) but the analysis

workflow, including comparison with a custom or public database, may fail to identify this sequence information if the corresponding reference information is missing from the databank. Given the exponential growth of databases, partly fuelled by the NGS revolution, this limitation will decrease over time. Database requirements are not the same for pathogen discovery and diagnostic applications. In the latter case, it may be more relevant to screen a genetic dataset against a curated database of pathogens or molecular marker profiles with documented relevance for diagnostic purposes.

Software advancements combined with platform and chemistry developments are expected to further reduce operational costs, error rates and the input DNA requirement, which will allow more careful sampling strategies to be adopted and reduce experimentally introduced error and bias in the future. These can be minimised using a proper sampling strategy, documenting wet laboratory and bioinformatics workflow to minimise error introduction, documenting assay uncertainties (from the wet laboratory and bioinformatics workflows), establishing criteria for data interpretation, and performing follow-up/QC of the critical parameters of assay performance.

Depending on the precise application of NGS data (scope), these systematic or random errors may or may not have an impact on the analytical properties of the test.

Conclusions and prospects

While recent (r)evolutions in sequencing technologies promise to dramatically change veterinary infection biology research, a critical reflection on the potential diagnostic applications of NGS technologies is necessary. Recent advancements in human diagnostics indicate that the technology is ready to be transferred from research to diagnostic laboratories and have resulted in the production of several guidelines on the validation of NGS-based diagnostics. The technology certainly has revolutionising strengths but, as discussed in the present article, there are still some weaknesses to overcome. The authors have documented potential sources of error and bias introduction and provided some preliminary guidelines for the analytical validation and quality monitoring of NGS-based methods used in the diagnosis of veterinary infectious diseases; however, these guidelines are based on personal experience and interpretation of the available literature and current guidelines about NGS and its validation. A more formal approach towards standardised and generally accepted guidelines for method validation is needed. It should be stressed that the possible diagnostic applications of NGS methodology are very diverse; therefore, users should very carefully define the scope of the NGS method under development or validation, and define values for validation parameters that are suitable for that particular method. The authors hope that this present manuscript will stimulate further discussion between expert NGS users, experts in virology, bacteriology and parasitology, and experts in the validation of diagnostic methods, with the aim of producing formally accepted guidelines for the diagnostic application of NGS technologies in veterinary infection biology.

Acknowledgements

This work was supported by Epi-SEQ, a research project supported under the 2nd joint call for transnational research projects by EMIDA ERA-NET (FP7 project no. 219235).

Vers la validation et l'assurance qualité des flux de données de séquençage de nouvelle génération utilisés en biologie des maladies animales infectieuses

S. Van Borm, J. Wang, F. Granberg & A. Colling

Résumé

Les avancées récentes enregistrées en matière de séquençage de l'ADN et d'analyse des données de séquences ont révolutionné la recherche dans de nombreux domaines de la biologie et de la médecine, notamment la biologie des maladies animales infectieuses. Ces nouvelles technologies vont permettre de combler le fossé qui séparait la recherche fondamentale du laboratoire de diagnostic. Après avoir défini l'intérêt diagnostique des applications du séquençage de nouvelle génération (SNG) ainsi que leurs finalités dans le domaine de la biologie des maladies animales infectieuses, les auteurs examinent

leur compatibilité avec les méthodes et les principes actuels de validation recommandés par l'Organisation mondiale de la santé animale. Ils proposent quelques paramètres critiques de validation et de contrôle qualité (mesure de la qualité), en se référant aux lignes directrices de validation et d'assurance qualité des techniques diagnostiques basées sur le séquençage de nouvelle génération visant à détecter les maladies humaines héréditaires. Certes, la plupart des applications actuelles des méthodes de séquençage de nouvelle génération en biologie des maladies animales infectieuses ne constituent pas des tests de diagnostic primaire (dont dépendent directement les décisions de contrôle sanitaire) ; toutefois, l'analyse critique proposée par les auteurs sur les avantages de cette technologie et sur les difficultés restant à résoudre devrait ouvrir la voie à des discussions sur l'intérêt diagnostique des méthodes recourant au séquençage de nouvelle génération ainsi que sur les perspectives de validation et de contrôle de leurs performances diagnostiques.

Mots-clés
Assurance qualité – Méthodes de diagnostic moléculaire – Séquençage de nouvelle génération – Validation.

Hacia la validación y la garantía de calidad de los procedimientos de trabajo de la secuenciación de próxima generación en biología de las infecciones veterinarias

S. Van Borm, J. Wang, F. Granberg & A. Colling

Resumen
Los recientes avances en los métodos de secuenciación del ADN y el análisis de los datos de secuencias han revolucionado la investigación en muchos ámbitos de la biología y la medicina, entre ellos la biología de las infecciones veterinarias. Las nuevas técnicas encierran la promesa de reducir la distancia entre el mundo de la investigación y los laboratorios de diagnóstico. Tras explicar el interés que pueden revestir las aplicaciones de la secuenciación de próxima generación y su posible uso con fines de diagnóstico de infecciones veterinarias, los autores examinan su compatibilidad con los principios y métodos de validación que tiene definidos la Organización Mundial de Sanidad Animal. Asimismo, proponen parámetros básicos para su validación y control de calidad (medición de la calidad), haciendo referencia a las pautas ya establecidas de validación y garantía de calidad de métodos de diagnóstico de enfermedades humanas hereditarias que reposan en técnicas de secuenciación de próxima generación. Aunque la mayoría de las aplicaciones de estas técnicas actualmente descritas en biología de las infecciones veterinarias no constituyen pruebas primarias de diagnóstico, esto es, cuyos resultados puedan inducir directamente medidas de control, esta crucial reflexión sobre las ventajas que entraña la secuenciación de próxima generación y los problemas que aún plantea debería alentar un debate sobre su interés para labores de diagnóstico y sobre la posibilidad de validar métodos basados en estas técnicas y de hacer un seguimiento de la eficacia diagnóstica que ofrezcan.

Palabras clave
Diagnóstico molecular – Garantía de calidad – Secuenciación de próxima generación – Validación.

References

1. Sanger F., Nicklen S. & Coulson A.R. (1977). – DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74** (12), 5463–5467. doi:10.1073/pnas.74.12.5463.

2. 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** (1), 263–273. doi:10.1101/SQB.1986.051.01.032.

3. Bartlett J.M. & Stirling D. (2003). – A short history of the polymerase chain reaction. *Meth. Molec. Biol.*, **226**, 3–6. doi:10.1385/1-59259-384-4:3.

4. Loman N.J., Misra R.V., Dallman T.J., Constantinidou C., Gharbia S.E., Wain J. & Pallen M.J. (2012). – Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.*, **30** (5), 434–439. doi:10.1038/nbt.2198.

5. Glenn T.C. (2011). – Field guide to next-generation DNA sequencers. *Molec. Ecol. Resour.*, **11** (5), 759–769. doi:10.1111/j.1755-0998.2011.03024.x.

6. Radford A.D., Chapman D., Dixon L., Chantrey J., Darby A.C. & Hall N. (2012). – Application of next-generation sequencing technologies in virology. *J. Gen. Virol.*, **93** (Pt 9), 1853–1868. doi:10.1099/vir.0.043182-0.

7. Pareek C.S., Smoczynski R. & Tretyn A. (2011). – Sequencing technologies and genome sequencing. *J. Appl. Genet.*, **52** (4), 413–435. doi:10.1007/s13353-011-0057-x.

8. Schadt E.E., Turner S. & Kasarskis A. (2010). – A window into third-generation sequencing. *Hum. Molec. Genet.*, **19** (R2), R227–R240. doi:10.1093/hmg/ddq416.

9. Gargis A.S., Kalman L., Berry M.W., Bick D.P., Dimmock D.P., Hambuch T., Lu F., Lyon E., Voelkerding K.V., Zehnbauer B.A., Agarwala R., Bennett S.F., Chen B., Chin E.L., Compton J.G., Das S., Farkas D.H., Ferber M.J., Funke B.H., Furtado M.R., Ganova-Raeva L.M., Geigenmuller U., Gunselman S.J., Hegde M.R., Johnson P.L., Kasarskis A., Kulkarni S., Lenk T., Liu C.S., Manion M., Manolio T.A., Mardis E.R., Merker J.D., Rajeevan M.S., Reese M.G., Rehm H.L., Simen B.B., Yeakley J.M., Zook J.M. & Lubin I.M. (2012). – Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.*, **30** (11),1033–1036. doi:10.1038/nbt.2403.

10. Bennett N.C. & Farah C.S. (2014). – Next-generation sequencing in clinical oncology: next steps towards clinical validation. *Cancers (Basel)*, **6** (4), 2296–2312. doi:10.3390/cancers6042296.

11. Quinones-Mateu M.E., Avila S., Reyes-Teran G. & Martinez M.A. (2014). – Deep sequencing: becoming a critical tool in clinical virology. *J. Clin. Virol.*, **61** (1), 9–19. doi:10.1016/j.jcv.2014.06.013.

12. Fournier P.E., Dubourg G. & Raoult D. (2014). – Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Med.*, **6** (11), 114. doi:10.1186/s13073-014-0114-2.

13. Van Borm S., Belák S., Freimanis G., Fusaro A., Granberg F., Hoper D., King D.P., Monne I., Orton R. & Rosseel T. (2015). – Next-generation sequencing in veterinary medicine: how can the massive amount of information arising from high-throughput technologies improve diagnosis, control, and management of infectious diseases? *In* Veterinary infection biology: molecular diagnostics and high-throughput strategies. *Meth. Molec. Biol.*, **1247**, 415–436. doi:10.1007/978-1-4939-2004-4_30.

14. 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. doi:10.3201/eid1803.111905.

15. World Organisation for Animal Health (OIE) (2013). – Chapter 1.1.5. Principles and methods of validation of diagnostic assays for infectious diseases. *In* Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.05_VALIDATION.pdf (accessed on 17 December 2015).

16. World Organisation for Animal Health (OIE) (2014). – Validation guideline 3.6.1. Development and optimisation of antibody detection assays. *In* Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/GUIDELINE_3.6.1_ANTIBODY_DETECT.pdf (accessed on 17 December 2015).

17. World Organisation for Animal Health (OIE) (2014). – Validation guideline 3.6.2. Development and optimisation of antigen detection assays. *In* Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/GUIDELINE_3.6.2_ANTIGEN_DETECT.pdf (accessed on 17 December 2015).

18. World Organisation for Animal Health (OIE) (2014). – Validation guideline 3.6.3. Development and optimisation of nucleic acid detection assays. *In* Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/GUIDELINE_3.6.3_NAD_ASSAYS.pdf (accessed on 17 December 2015).

19. Sheridan C. (2014). – Milestone approval lifts Illumina's NGS from research into clinic. *Nat. Biotechnol.*, **32** (2), 111–112. doi:10.1038/nbt0214-111.

20. Vrijenhoek T., Kraaijeveld K., Elferink M., de Ligt J., Kranendonk E., Santen G., Nijman I.J., Butler D., Claes G., Costessi A., Dorlijn W., van Eyndhoven W., Halley D.J., van den Hout M.C., van Hove S., Johansson L.F., Jongbloed J.D., Kamps R., Kockx C.E., de Koning B., Kriek M., Lekanne dit Deprez R., Lunstroo H., Mannens M., Nelen M., Ploem C., Rijnen M., Saris J.J., Sinke R., Sistermans E., van Slegtenhorst M., Sleutels F., van der Stoep N., van Tienhoven M., Vermaat M., Vogel M., Waisfisz Q., Marjan Weiss J., van den Wijngaard A., van Workum W., Ijntema H., van der Zwaag B., van IJken W.F.J., den Dunnen J., Veltman J.A., Hennekam R. & Cuppen E. (2015). – Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects. *Eur. J. Hum. Genet.,* **23** (9), 1270. doi:10.1038/ejhg.2014.279 (Corrigendum: 10.1038/ejhg.2015.44).

21. Ng T.F., Kondov N.O., Deng X., van Eenennaam A., Neibergs H.L. & Delwart E. (2015). – A metagenomics and case-control study to identify viruses associated with bovine respiratory disease. *J. Virol.*, **89** (10), 5340–5349. doi:10.1128/JVI.00064-15.

22. Rosseel T., Lambrecht B., Vandenbussche 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. doi:10.1186/1743-422X-8-463.

23. Van Borm S., Rosseel T., Steensels M., van den Berg T. & Lambrecht B. (2013). – What's in a strain? Viral metagenomics identifies genetic variation and contaminating circoviruses in laboratory isolates of pigeon paramyxovirus type 1. *Virus Res.*, **171** (1), 186–193. doi:10.1016/j.virusres.2012.11.017.

24. Belák S., Karlsson O.E., Blomström A.L., Berg M. & Granberg F. (2013). – New viruses in veterinary medicine, detected by metagenomic approaches. *Vet. Microbiol.*, **165** (1–2), 95–101. doi:10.1016/j.vetmic.2013.01.022.

25. Granberg F., Karlsson O.E. & Belák S. (2015). – Metagenomic approaches to disclose disease-associated pathogens: detection of viral pathogens in honeybees. *Meth. Molec. Biol.*, **1247**, 491–511. doi:10.1007/978-1-4939-2004-4_33.

26. Briese T., Kapoor A., Mishra N., Jain K., Kumar A., Jabado O.J. & Lipkin W.I. (2015). – Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio*, **6** (5), e01491-15. doi:10.1128/mbio.01491-15.

27. Wylie T.N., Wylie K.M., Herter B.N. & Storch G.A. (2015). – Enhanced virome sequencing through solution-based capture enrichment. *Genome Res.*, **25**, 1910–1920. doi:10.1101/gr.191049.115.

28. Ogorzaly L., Walczak C., Galloux M., Etienne S., Gassilloud B. & Cauchie H.M. (2015). – Human adenovirus diversity in water samples using a next-generation amplicon sequencing approach. *Food Environ. Virol.*, **7** (2), 112–121. doi:10.1007/s12560-015-9194-4.

29. Singh P., Foley S.L., Nayak R. & Kwon Y.M. (2012). – Multilocus sequence typing of *Salmonella* strains by high-throughput sequencing of selectively amplified target genes. *J. Microbiol. Meth.*, **88** (1), 127–133. doi:10.1016/j.mimet.2011.11.004.

30. Mertes F., Elsharawy A., Sauer S., van Helvoort J.M., van der Zaag P.J., Franke A., Nilsson M., Lehrach H. & Brookes A.J. (2011). – Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief. Funct. Genomic.*, **10** (6), 374–386. doi:10.1093/bfgp/elr033.

31. Depledge D.P., Palser A.L., Watson S.J., Lai I.Y., Gray E.R., Grant P., Kanda R.K., Leproust E., Kellam P. & Breuer J. (2011). – Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One,* **6** (11), e27805. doi:10.1371/journal.pone.0027805.

32. Kessell A., Hyatt A., Lehmann D., Shan S., Crameri S., Holmes C., Marsh G., Williams C., Tachedjian M., Yu M., Bingham J., Payne J., Lowther S., Wang J., Wang L.F. & Smith I. (2012). – Cygnet River virus, a novel orthomyxovirus from ducks, Australia. *Emerg. Infect. Dis.*, **18** (12), 2044–2046. doi:10.3201/eid1812.120500.

33. Quan P.L., Williams D.T., Johansen C.A., Jain K., Petrosov A., Diviney S.M., Tashmukhamedova A., Hutchison S.K., Tesh R.B., Mackenzie J.S., Briese T. & Lipkin W.I. (2011). – Genetic characterization of K13965, a strain of Oak Vale virus from Western Australia. *Virus Res.*, **160** (1–2), 206–213. doi:10.1016/j.virusres.2011.06.021.

34. Wang J., Selleck P., Yu M., Ha W., Rootes C., Gales R., Wise T., Crameri S., Chen H., Broz I., Hyatt A., Woods R., Meehan B., McCullough S. & Wang L.F. (2014). – Novel phlebovirus with zoonotic potential isolated from ticks, Australia. *Emerg. Infect. Dis.*, **20** (6), 1040–1043. doi:10.3201/eid2006.140003.

35. Ghedin E., Laplante J., DePasse J., Wentworth D.E., Santos R.P., Lepow M.L., Porter J., Stellrecht K., Lin X., Operario D., Griesemer S., Fitch A., Halpin R.A., Stockwell T.B., Spiro D.J., Holmes E.C. & St George K. (2011). – Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *J. Infect. Dis.*, **203** (2), 168–174. doi:10.1093/infdis/jiq040.

36. Wright C.F., Morelli M.J., Thébaud 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. doi:10.1128/JVI.01396-10.

37. Morelli M.J., Wright C.F., Knowles N.J., Juleff N., Paton D.J., King D.P. & Haydon D.T. (2013). – Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. *Vet. Res.*, **44** (1), 12. doi:10.1186/1297-9716-44-12.

38. Monne I., Fusaro A., Nelson M.I., Bonfanti L., Mulatti P., Hughes J., Murcia P.R., Schivo A., Valastro V., Moreno A., Holmes E.C. & Cattoli G. (2014). – Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. *J. Virol.,* **88** (8), 4375–4388. doi:10.1128/JVI.03181-13.

39. Mattocks C.J., Morris M.A., Matthijs G., Swinnen E., Corveleyn A., Dequeker E., Muller C.R., Pratt V. & Wallace A. (2010). – A standardized framework for the validation and verification of clinical molecular genetic tests. *Eur. J. Hum. Genet.*, **18** (12), 1276–1288. doi:10.1038/ejhg.2010.101.

40. World Organisation for Animal Health (2012). – Glossary of terms, Volume 1. *In* Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris, xvii–xxi. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/0.04_GLOSSARY.pdf (accessed on 13 April 2016).

41. Li L., Deng X., Mee E.T., Collot-Teixeira S., Anderson R., Schepelmann S., Minor P.D. & Delwart E. (2015). – Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J. Virol. Meth.*, **213**, 139–146. doi:10.1016/j.jviromet.2014.12.002.

42. World Organisation for Animal Health (OIE) (2013). – Chapter 1.1.1. Collection, submission and storage of diagnostic specimens. *In* Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.01_COLLECTION_DIAG_SPECIMENS.pdf (accessed on 17 December 2015).

43. World Organisation for Animal Health (OIE) (2014). – Validation guideline 3.6.6. Selection and use of reference samples and panels. *In* Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris. Available at: www.oie.int/fileadmin/Home/eng/Health_standards/tahm/GUIDELINE_3.6.6_REFERENCE_SAMPLES.pdf (accessed on 17 December 2015).

44. Whiteford N., Skelly T., Curtis C., Ritchie M.E., Lohr A., Zaranek A.W., Abnizova I. & Brown C. (2009). – Swift: primary data analysis for the Illumina Solexa sequencing platform. *Bioinformatics*, **25** (17), 2194–2199. doi:10.1093/bioinformatics/btp383.

45. Metzker M.L. (2010). – Sequencing technologies: the next generation. *Nat. Rev. Genet.*, **11** (1), 31–46. doi:10.1038/nrg2626.

46. 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. doi:10.1186/1471-2164-13-341.

47. Rosseel T., Pardon B., De Clercq K., Ozhelvaci O. & Van Borm S. (2014). – False-positive results in metagenomic virus discovery: a strong case for follow-up diagnosis. *Transbound. Emerg. Dis.*, **61** (4), 293–299. doi:10.1111/tbed.12251.

48. Laurence M., Hatzis C. & Brash D.E. (2014). – Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One*, **9** (5), e97876. doi:10.1371/journal.pone.0097876.

49. Quail M.A., Smith M., Jackson D., Leonard S., Skelly T., Swerdlow H.P., Gu Y. & Ellis P. (2014). – SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics*, **15**, 110. doi:10.1186/1471-2164-15-110.

50. Buschmann T., Zhang R., Brash D.E. & Bystrykh L.V. (2014). – Enhancing the detection of barcoded reads in high throughput DNA sequencing data by controlling the false discovery rate. *BMC Bioinformatics*, **15**, 264. doi:10.1186/1471-2105-15-264.

51. Archer J., Rambaut A., Taillon B.E., Harrigan P.R., Lewis M. & Robertson D.L. (2010). – The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time: an ultra-deep approach. *PLoS Comput. Biol.*, **6** (12), e1001022. doi:10.1371/journal.pcbi.1001022.

52. Degner J.F., Marioni J.C., Pai A.A., Pickrell J.K., Nkadori E., Gilad Y. & Pritchard J.K. (2009). – Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, **25** (24), 3207–3212. doi:10.1093/bioinformatics/btp579.

53. Treangen T.J. & Salzberg S.L. (2012). – Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.*, **13** (1), 36–46. doi:10.1038/nrg3164.

54. Naccache S.N., Federman S., Veeraraghavan N., Zaharia M., Lee D., Samayoa E., Bouquet J., Greninger A.L., Luk K.C., Enge B., Wadford D.A., Messenger S.L., Genrich G.L., Pellegrino K., Grard G., Leroy E., Schneider B.S., Fair J.N., Martinez M.A., Isa P., Crump J.A., DeRisi J.L., Sittler T., Hackett J., Jr, Miller S. & Chiu C.Y. (2014). – A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.*, **24** (7), 1180–1192. doi:10.1101/gr.171934.113.