High-throughput Sequencing in Veterinary Infection Biology and Diagnostics

S. Belák1,2,3 *, O.E. Karlsson2,3, M. Leijon1,3, and F. Granberg2,3

1Department of Virology, Immunobiology and Parasitology, VIP National Veterinary Institute (SVA); 2Department of Biomedical Sciences and Veterinary Public Health, BVF, Swedish University of Agricultural Sciences (SLU); 3World Organisation for Animal Health (OIE) Collaborating Centre for the Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine, Ulls väg 2B, SE-751 89, Uppsala, Sweden

Keywords: metagenomics; sequencing; unknown viruses, virus detection, diagnosis, unknown aetiology

During the past few decades, the rapid development and wide application of novel molecular techniques have resulted in unprecedented advances in animal disease diagnostics. The possibility of detecting and identifying infectious agents brought about by the presence of their nucleic acid molecules in clinical samples has been exploited by various molecular diagnostic methods including nucleic acid hybridisation, several amplification methods and nucleotide sequencing. The OIE has actively supported the development and use of these molecular techniques in the field of veterinary medicine by OIE Reference Laboratories and through the establishment of OIE Collaborating Centres focused on biotechnology-based diagnosis in veterinary medicine.

Since the development of DNA sequencing methods in the 1970s, efforts to improve the technology led to the important step of the introduction of automated multicapillary-based sequencing instruments. Recently a major technological advancement in cyclic array sequencing gave rise to “second or next generation sequencing”, which today encompasses several different platforms.

As a result of the rapid improvement of existing systems and the release of completely new platforms, 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 (doubling every 2 years). This has also resulted in a dramatic reduction in the cost of DNA sequencing, making the technology more accessible to the average laboratory. The recent emergence of a new generation of single-molecule sequencing technologies, termed “third-generation sequencing”, has made it more suitable to collectively refer to all the novel sequencing approaches as high-throughput sequencing (HTS).

Contrary to earlier diagnostic tools, HTS is unbiased with regard to the target and it detects and reports practically on all kinds of nucleotide sequences present in the original clinical sample. This wide-range capacity allows the prompt detection of all microbes and viruses in a sample, including novel pathogens. However, the limit detection is still determined by the abundance of the pathogens in relation to the host background material. Recent advances in HTS technologies allow sequencing to be performed faster and at lower cost, and are continuously improving the capacities and the likelihood of detecting even low-copy-number pathogens. Nonetheless, sampling as well as sample preparation and enrichment protocols have been demonstrated to have a dramatic effect on the outcome of HTS-based diagnostics and should therefore also be considered as integrated and important steps in the overall detection scheme.

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 that allows large amounts of data that are being produced to be handled. For a given application, it is common to fully automate the analysis process by establishing what is known as a pipeline, i.e. a continuous data flow between different software programs. In the case of analysing HTS data, the bioinformatics pipelines can contain several elements, such as quality control, assembly of reads into longer contiguous sequences (contigs), homology searches against sequence databases, alignment against reference genomes and taxonomic classification of metagenomic sequences. However, in the long run it is the growing amounts of sequencing data from HTS that is expected to be the greatest challenge and not the analysis of individual samples.

HTS has already been applied to advance veterinary diagnostic microbiology, with extremely important results. In this presentation, a few noteworthy and representative examples within the major farm animal groups will be given.