

NEXT GENERATION SURVEILLANCE SYSTEMS – INTEGRATING WHOLE GENOME SEQUENCING DATA INTO REAL-TIME DISEASE DETECTION AND CONTROL

Peter Durr, Jianning Wang, Frank Wong, Ivano Broz, Kerryne Graham and Sam McCullough

CSIRO-Australian Animal Health Laboratory, Private Bag 24, East Geelong, VIC. 3219, Australia
Tel.: +61-3 5227.5139; peter.durr@csiro.au

“Next-gen” (NG) sequencing technologies are radically changing the diagnostic landscape for transboundary animal diseases (TAD). Not only do they enable whole genome sequencing (WGS) at reasonable cost, but additionally provide a rapid turn-around, with assembled sequences potentially being ready within 24–48 hours of specimen collection. Already examples exist where (conventional) WGS data has informed pathways of farm-to-farm transmission for foot and mouth disease (1) and highly pathogenic avian influenza (2) outbreaks in Europe. The current challenge is how to develop systems that use comparable analyses in the face of a TAD outbreak in near real-time, and for this several specific difficulties can be identified.

The first main difficulty in developing these NG outbreak surveillance systems relates to maintaining sequence data quality, and to assure that all nucleotides detected as “variants” really are as such, and are not an error. Thus for example, in deciding if one farm might have infected another, the outbreak investigator may rely on the evidence of minimal sequence variation from the two farms. However, if the sequencing or assembly process is suboptimal and erroneous nucleotides are identified, then an incorrect conclusion that the first farm could not have infected the second might result. A second problem relates to the integration of the WGS data with covariate data relating to potential “paths of transmission” of pathogens from farm to farm. Thus in the farm outbreak scenario, potentially relevant data includes wind speed and direction, livestock movements and vehicle traffic between farms. For many countries, these data are becoming increasingly available, but often exist in disparate databases. Even if all the relevant databases can be successfully queried and integrated with the NG-WGS data, a third challenge is how to return the results to veterinarians working in disease control centres in a manner that is readily comprehensible and thus facilitates disease control and surveillance decision-making.

All three of these challenges are not unique to veterinary surveillance systems and are encountered in all disciplines where “Big Data” are being generated and used. Practical solutions to overcoming them lie in using iterative approaches to development of systems, with a priority given to maintaining data quality and early end-user engagement. Currently our laboratory is developing a web-based application, the “Molecular Analysis and Reporting System” (“MARS”), using these principles, and this will be described and demonstrated.

1. COTTAM E.M., WADSWORTH J., SHAW A.E., ROWLANDS R.J., GOATLEY L., MAAN S., *et al.* (2008). Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. *PLoS pathogens*, **4** (4): e1000050. PubMed PMID: 18421380.
2. BATAILLE A., VAN DER MEER F., STEGEMAN A. & KOCH G. (2011). Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. *PLoS pathogens*, **7** (6): e1002094. PubMed PMID: 21731491.