REPORT OF THE VIRUS DISCOVERY WORKSHOP
Held at the International Livestock Research Institute, Nairobi
From 5th to 8th November 2012

1.0 Organisers

1. Steve Kemp - ILRI
2. George Michuki - ILRI
3. Sarah Nyongesa - ILRI
4. Anne Fischer – ICIPE/ILRI

2.0 Introduction

The virus discovery workshop was held at the International Livestock Research Institute (ILRI) Nairobi from 5th to 8th November 2012. The workshop brought together scientists conducting research in the area of virus/pathogen discovery. The workshop aimed at enabling ILRI make informed decisions about which technologies to invest in to support the activities in the CGIAR Research Program on Agriculture for Improved Nutrition and Health (A4NH). The workshop was supported by A4NH with partial input from Roche Applied Sciences.

3.0 Participating institutions

Attending participants were drawn from the Commonwealth Scientific and Industrial Research Organisation (CSIRO); Australian Animal Health Laboratory; Roche; Natural History Museum of Denmark; Los Alamos National Laboratory; University of Chicago; University of Colorado; Uganda Virus Research Institute; Centre for Disease Control and Prevention (CDC) Kenya; JICA/Nagasaki University, KEMRI, Nairobi; International Centre of Insect Physiology and Ecology *(*ICIPE*)*; Onderstepoort Veterinary Institute (OVI); the United States Department of Homeland Security (DHS); the United States Defense Threat Reduction Agency (DTRA) and the International Livestock Research Institute.

4.0 Discussions

Discussion topics revolved in the areas of field sampling strategies, virus samples handling/shipment and preservation, viral genome detection, sequencing, isolation or enrichment from mixed samples and metagenomic data analysis. The workshop program, specific discussion topics, speakers and participant's contacts and their biographies are available at http://virusdiscoveryworkshop2012.sched.org/.

ILRI was identified to be moving in the right direction in terms of its bio-repository system and the genomics platform especially the possession of the 454 genome sequencer and the acquisition of the MiSeq.

The workshop also had brainstorming sessions which resulted in a number of recommendations for each discussion topic. It was suggested that ILRI should focus on pathogen discovery rather than virus discovery especially considering a list of potentially dangerous pathogens is available and it is known which are important. In addition, pathogen discovery requires integration of biology and it was therefore, agreed to set up a virus culture

facility at ILRI. Pre-sequencing techniques such as ELISA, phenotypic and genotypic characterization were suggested. Projects should be question driven more than part of surveillance work. With integration of biology, biosecurity becomes very important in the lab where organisms are being cultivated. It is then necessary to know who has access to the lab, to the samples and trust these people.

Data management and achieving is currently well established at ILRI. Emphasis was made in the use of tape drives for long term storage and developing SOPs on data. The SOPs should provide guidelines when to store raw or analyzed data allowing making the decision on which is important on a case by case basis. Raw images from next generation sequencing should not be stored and no need to store sequences already submitted to gene bank. Frequently clean up data for duplicates.

To overcome the challenges in data sharing; MOUs and statements of work should be prepared upfront to clear up expectations. It is important to have data sharing policies: e.g. six months release policy. Have set sample quality standards to control what is stored. Chris Detter from Los Alamos offered to share templates if required. It is also important to find out what are the existing regional and national policies especially regarding to intellectual property (IP) rights.

To ensure the data and information is secure a combination of physical security of storage facility and use of RAID is well implemented at ILRI. It was further recommended that metadata be separated from sequence data. SOPs should be developed to classify metadata in relation to what needs to be linked to sequences or not. E.g. organisms (sequences) identified as threats might be better not linked to metadata. Access control to data on a need to know basis would be useful too.

Rapid identification of PCR probes for subsets of species for diagnostics can be achieved by use of universal primers first, then sequencing and design of species-specific probes

To discover novel species and pathogens a number of approaches were suggested. They comprise:

1. 16S rRNA is a good strategy for a first screen of bacterial species present in a sample. A lot of bacteria will be new and the question which arises is then: which are the important ones
2. If a phylogeny is done from the identified sequences this should allow determination of whether any sequences cluster with known families of viruses/bacteria. This will help decide how to react depending on what the outcomes are.
3. Protein motifs can be checked in order to see whether any are associated with virulence factors
4. Annotation is a very useful tool
5. Use controls if you are not able to identify the pathogen

In metagenomic approaches the most efficient techniques to use for comparing the nucleotide composition are dependent on available sample metadata. In addition, a pre-treatment can be applied to the samples using the Clontech PCR Select kit, which will remove any RNA common between two samples by hybridization. To find sequences that are unique to samples that have pathologies with 16S/18S rRNA, you can look for OTUs that are correlated with the particular categories (e.g. Healthy/diseased). This could be done also with shotgun data

assigned to categories (e.g. KEGG pathways, etc…). For shotgun data, annotation will be needed. Whether we can naively identify differences between shotgun data without worrying about first identifying genes (by COG/KEGG, etc…) and subtracting all common and very low abundance data was discussed as a possibility.

The workshop also brought the developers of MGRAST and QIIME software. These developers shared first hand experiences in designing the pipelines and also received first hand feedback from the users on usability issues and possible areas of improvement on the pipelines.

5.0 Conclusion

The above discussion outcomes were very useful and for that reason the workshop met its objective in which the participants were able to come up with consolidated tools, techniques and strategies appropriate for pathogen discovery. In addition, the workshop opened up new opportunities for potential collaboration and funding opportunity from DTRA.

6.0 Follow up action points

ILRI EID team is currently developing a document to initiate collaboration with CSIRO and Los Alamos Laboratory with the aim of applying for DTRA funding.

ILRI genomics team is in the process of setting up a virus culture facility.