Developing disease monitoring and diagnostic tools via generation of integrated geospatial immune repertoire signature sequence data and relational databases

This proposal seeks to respond to the call on Monitoring Pests and Diseases**. Genomic signatures can serve as surrogates of a phenotype, function, or health status. We propose to generate geospatial integrated livestock immune repertoire signature database with unique information for diagnosis, disease surveillance and monitoring.** This approach is both unique and different from current diagnostic approaches that focus largely on pathogen detection without global studies on the immune response to them, an important parameter in the epidemiology of infectious diseases.

Immune repertoire profiling enables identification of gene sequences that encode antibodies with specificity to pathogen antigens. These data can be used to develop sensitive diagnostic reagents as well as *in silico* immunity map for disease surveillance and monitoring that not only detect an on-going immune response to pathogens, but also provide a historical account of previous pathogenic encounters. This information will feedback directly to the farmer for immediate decision on the immune status of the livestock and also to the veterinary services for future intervention plans. Integrating immune repertoire data from animals (~ 1×10^6 sequence data points/individual) in endemic populations with associated phenotypic and epidemiological meta-data e. g., breed, age, sex, nutritional status, burden of infection, clinical diagnoses and vaccination history, should provide important geospatial information for disease control and epidemiology. Mining such information for pathogens of pandemic, epidemic and endemic importance should provide important digital footprints of immunity (immunity map) to diseases and “invisible pathogens”, thereby providing warning signs for planning against periodic disease outbreaks and for public health interventions.

This project will aim to collect peripheral blood mononuclear cell (PBMC) from fish and livestock species within different epidemiological zones together with extensive metadata. Using special immune repertoire genomic technology that has been developed at ILRI, nucleic acid extracted from PBMC or purified B-cells that secrete antibodies will be sequenced at high quality and depth. The sequence data will be analyzed using specialized algorithms to generate a curated database of immune repertoire signatures. Machine learning approaches will be employed to link immune repertoire sequence signatures to associated meta-data and generate geospatial immunity map. The data will be stored in the cloud livestock immunogenetic repository for free global crowd-sourcing and downstream applications. A network of continuous input between the farmer, veterinary services and data center is envisaged to create a near real-time livestock immunity map for disease control, surveillance and monitoring. This data can be used to help farmers make decisions on which, when and how to vaccinate and treat their livestock thereby reducing the cost of production. Other applications include, generation of diagnostic reagents, supporting new vaccine discovery and development, and also developing new therapies.

As a proof of concept, 50 blood samples together with extensive meta-data will be collected from livestock (both infected and healthy controls) in Kenya. Sampling will be based on previously determined different epidemiological zones. Targeted immunoglobulin sequencing will be performed to obtain immune repertoire sequence signatures for generating immunity maps that will be used to create initial basic data-base. In collaboration with our partners at The Technical University of Denmark (Immunoinformatics and Machine-Learning Group) and University of Toronto (offering synthetic antibody expression technology), the database will be mined to extract key epidemiological information that will be fed back to the farmers and veterinary services for immediate action or future disease control planning. For diagnostic purposes, identified functional sequences will be used to express antibodies for developing diagnostic reagents.

For future work, the initial basic database -created within the proof of concept budget - will be developed and expanded to generate near real-time crowd-sourcing livestock immunity maps from various epidemiologic zones using continuous sample sequence input and updates, and continuous information feedback to the farmers and veterinary services for interventions.

The initial proof of concept budget of $100,000 will be spent as follows:

Sampling, meta-data collection and farmer feedback: $10,000

Sample sequencing and data analysis: $25,000

Creation of immunity map database and machine-learning: $30,000

Short term staff and bench cost: $30,000

Miscellaneous expenses: $5,000