

SWINE HEALTH

Title: African Swine Fever Virus: a bioinformatics resource to support comparative genomics for vaccine and virulence studies. – **NPB #15-123**

Investigator: Dr Chris Upton

Co-Investigator: Dr Juergen Richt, Kansas State University

Institution: University of Victoria, Canada

Date Submitted: Dec 19, 2016

Industry Summary

African swine fever is a very serious disease for which no vaccine is currently available. The virus (ASFV) has spread multiple times out of Africa into European countries and it is fortunate that it has not been introduced to North America where it would be disastrous for the pork industry. Vaccination is the only way to protect animals and eventually try to reduce the area in which the virus is endemic. To be able to create a vaccine that works with multiple genotypes of the virus, it is first necessary to determine the genetic sequence of all the types of virus. Fortunately, there are research projects planned by the ASFV research community to gather this information. The project described here was aimed at creating a database and bioinformatics resource to allow researchers to efficiently access and analyze ASFV genome information; specialized tools are required for viral genomes.

The project's goals were achieved by modifying a database and tools that was originally designed for other large viruses (poxviruses). One of the key achievements was to enable the resource to reside in a computer cloud environment, which will allow the resource to be moved to other laboratories if required in the future.

The database is now ready for the results of sequencing projects and it will enable researchers across the world to access the data and perform a variety of analyses.

Dr. Chris Upton; cupton@uvic.ca; virology.uvic.ca

Keywords:

ASFV, bioinformatics, software, database, virology, genome sequence, annotation.

These research results were submitted in fulfillment of checkoff-funded research projects. This report is published directly as submitted by the project's principal investigator. This report has not been peer-reviewed.

For more information contact

National Pork Board • PO Box 9114 • Des Moines, IA 50306 USA • 800-456-7675 • Fax: 515-223-2646 • pork.org

Scientific Abstract

All ASFV researchers require access to the viral genome sequences. However, the access provided by NCBI gives only FASTA or GenBank files of the data, which are static text files. If only a few genomes were sequenced, managing the data in this way would be tedious but perhaps manageable. However, there are currently 16 completed and publically available and probably in the order of 100-200 due to be sequenced in the next three years. This quantity of data certainly requires storage in a dedicated database with integrated tools to make the analysis, not only efficient, but possible at all.

The poxvirus database and bioinformatics resource, previously managed at www.virology.ca, was modified so that it could be run from a cloud computer resource at WestGrid.ca. This is the western portion of Compute Canada, a Canadian academic research compute centre that serves all research facilities in the country. The purpose of this move was:

- 1) platform stability – more secure backups.
- 2) platform portability – much easier to move the system to other platforms if required (e.g. when I retire).
- 3) expandability – more powerful resources available if required.

We also created a new model for annotation of the ASFV multi-gene families to enable better consistency in the annotation of these genes in preparation for the sequencing of large numbers of genomes.

Finally, we began an out-reach campaign to make researchers in the ASFV community aware of the project and to ask for their input on the proposed annotation scheme.

Introduction

Although African swine fever seems very remote, existing primarily in Africa, its recent transmission to Europe demonstrates that this virus is actually only a small step from invading North America. This would have disastrous consequences for the North American pork industry.

The only real way to combat this sort of viral disease is via vaccination, however, no vaccine or serious candidate vaccine is yet available. Much more information is required about the spectrum of genome variation before adequate vaccines can be developed. The goal of this project was to develop a DNA sequence database and software tools to support all researchers that are involved in research into ASFV genomics. This type of database simplifies the process of managing/retrieving large amounts of genomic information and also performing the many comparison analyses that are required to find the conserved regions that might be suitable for vaccine antigens.

Objectives

- A) Establishment of a Bioinformatics Resource for ASFV
- B) Improve the ASFV genomic/annotation data
- C) Create an outreach program for ASF virologists

Materials and Methods

The big advantage of this project was that we had previously developed a similar database and tools for the analysis of poxviruses (www.virology.ca) over many years. The database was built in MySQL and all the interfaces/tools were Java to enhance portability.

The Wordpress web site was maintained because it serves as a useful point of access for the database/tools and documentation.

Results

The www.virology.ca bioinformatics resource was modified to 1) use new ASFV genomes and 2) run in a cloud environment (www.westgrid.ca).

The new resource is available at:

<https://virology.uvic.ca/organisms/dsdna-viruses/asfarviridae/>

Access to the database and tools is still via a Wordpress website for simplicity, however, since it now can function in cloud environment the entire system is much more portable. In addition, we have included an upgrade to the security of the site requiring access to be via https. This protects passwords used by researchers and will help prevent hackers from damaging the system.

The following tools were switched to the cloud environment:

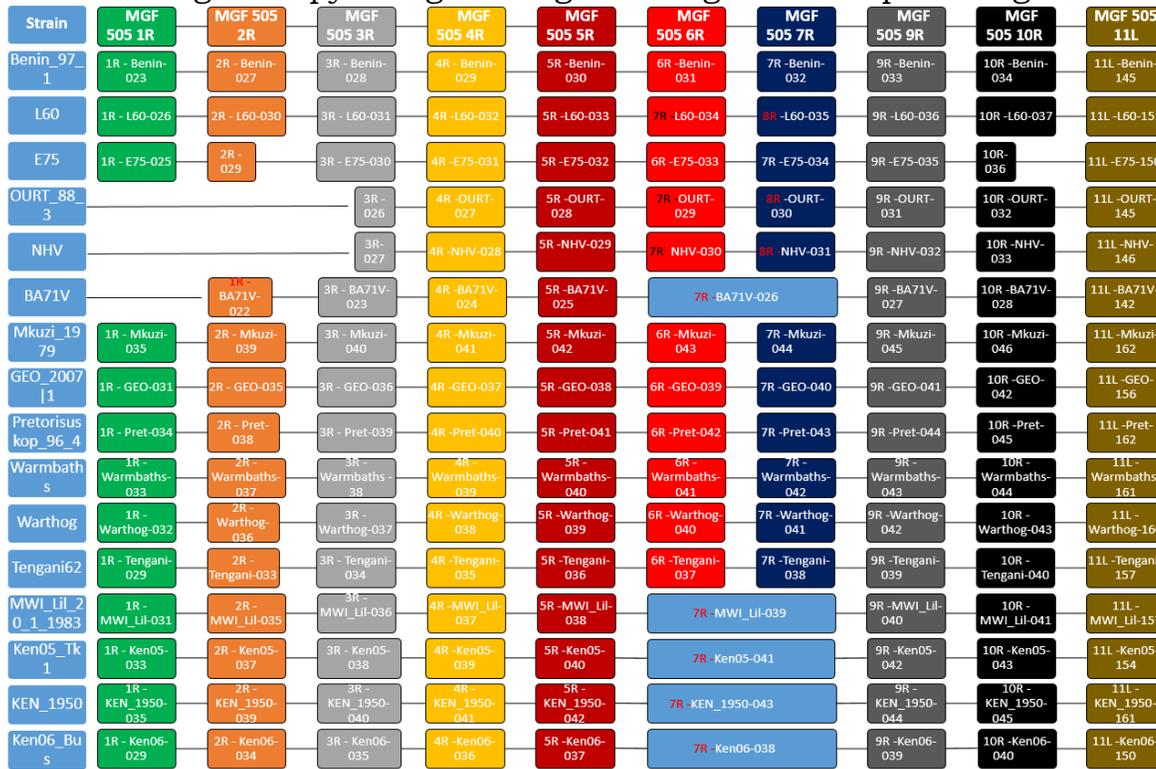
- Viral Orthologous Clusters (VOCs): An easy-to-use Java GUI used to access the ASFV genome database. Starting point for gene analysis pipelines.
- Viral Genome Organizer (VGO): An easy-to-use Java genome browser for viruses that graphically displays and performs searches within a genome and talks to the ASFV database for display of annotations and genome comparisons.
- Base-By-Base (BBB): A tool used to align/visualize/edit and search genes-proteins-genomes, and visualize differences between the sequences. The program can also be used to view RNA-Seq data and analyze for recombination.
- JDotter: a program for generating dotplots. This can be used for whole genomes, sub-genomes or protein sequences.
- Genome Annotation Transfer Utility (GATU): Annotates genomes based on a closely related reference genome. The proteins of the reference genome are BLASTed against the genome to be annotated in order to find the corresponding genes in the genome to be annotated. GATU also suggests novel genes to the human annotator, who has last word on the annotation process.
- Sequence Searcher (SSeq): Sequence Searcher is an easy-to-use Java tool for searching protein and DNA sequences for user specified sequence motifs.

In addition, CODEHOP, a tool for finding primers to be used for detection of distantly related genes was re-coded in Java and added to Base-By-Base. This tool was developed several years ago but the code broke due to changes in Javascript libraries, it could be useful for determining methods of screening the most divergent ASFVs.

iCODEHOP: a new interactive program for designing Consensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences R. Boyce, P. Chilana, and T. M. Rose; 2009 NAR 37 W222.

All the publically available ASFV genomes were reannotated and a plan for describing the multi-gene families (MGF) developed. This will allow better comparison of ASFV genomes

and be especially useful when trying to correlate virulence and genome characteristics. The following figure illustrates one (#505) of the five sets of MGFs and highlights the variation in gene copy and gene length among the 16 sequenced genomes.



During the course of this project and the testing of the software, we made the observation that there appears to be examples of natural recombination present in the currently known genomes. The regions involved are relatively small (500-1000 nt) and therefore not detected by standard software. The small size of these swapped regions may also indicate that this process is not unusual and that it might be an important driver of the evolution of ASFV genomes. We plan to investigate this further.

The following list of ASFV researchers offered to give feedback about our annotation options, however, in practice the group was not greatly used since several researchers provided most of the comments. We did advertise the new resource to as many ASFV researchers as we could find so that they could test out the new versions of the tools.

Linda Dixon	Linda.dixon@pirbright.ac.uk
Dave Chapman	dave.chapman@pirbright.ac.uk
Armanda Bastos	ADBastos@zoology.up.ac.za
Richard Bishop	bishop5030@gmail.com
Raquel Portugal	raquel.portugal@fli.bund.de
Douglas Gladue	Douglas.Gladue@ARS.USDA.GOV
Yolanda Revilla	yrevilla@cbm.uam.es

Denis Kolbasov	kolbasovdenis@gmail.com
Dan Rock	dlrock@illinois.edu
Claudia Torresi	c.torresi@izsum.it
Dave Brake	david.brake@st.dhs.gov
Fredrik Granberg	Fredrik.Granberg@slu.se
Jenna Achenbach	J.E.Achenbach@iaea.org
Charles Lamien	C.Lamien@iaea.org
Sandra Blome	Sandra.blome@fli.bund.de

Discussion

The new bioinformatics resource is now available worldwide, to all researchers working with ASFV. This should simplify their analysis of the ASFV genomes when working on aspects of virulence determination and also vaccine design.

We are now part of a new initiative (starting Jan 2017) that has proposed to sequence approximately 100 new ASFV genomes (over the next 3 years) to get a better understanding of the antigenic variation among these viruses. This work will be performed by researchers in Kenya, Uganda, South Africa and Canada, my group, that will continue to provide bioinformatics support and provide analysis of the ASFV genomes. It is being funded by the Defense Threat Reduction Agency (DTRA, USA).