2017 CFID Undergraduate Summer Research Award Summary Report

Whole Genome Sequencing of Mumps Virus from an Adult Outbreak in British Columbia

- **Student**: Thenuga Sasitharan, Department of Pathology and Laboratory Medicine, University of British Columbia
- **Supervisors**: Drs. Agatha Jassem and Richard Harrigan (co-supervisors)

Background

Mumps is a highly contagious infectious disease that occurred commonly in school-aged children prior to the introduction of the measles, mumps and rubella (MMR) vaccine in the 1970s. Mumps causes parotitis (inflammation of the salivary glands), but symptoms can also include fever, headache, and sore muscles. Mumps is diagnosed with serology or by detection of mumps virus (MuV) RNA in buccal swabs and urine using reverse transcription (RT) PCR, and viruses can be genotyped by small hydrophobic (SH) gene sequencing. Despite vaccination programs, outbreaks of mumps continue to occur and in fact have become more frequent. According to the US Centers for Disease Control and Prevention, in 2016 there were more confirmed cases of mumps than in any other year of the preceding decade.

In the spring of 2016, a cluster of mumps cases was identified following a mass gathering in British Columbia, Canada. The outbreak then spread to two other regions within BC. From April 1 to October 31, 140 laboratory confirmed mumps cases were identified with a median age of 27 years, including 33 (24%) with documented doses of MMR vaccine and 72 (51%) assumed immune. The outbreak strain was identified as genotype G, but with a unique five nucleotide signature not present in the SH gene of its nearest relative, the MuVi/Sheffield.GBR/1.05 strain that is endemic in North America. The outbreak strain was identified in 85 (61%) of the confirmed cases by RT-PCR.

We hypothesize that the recent British Columbia mumps outbreak was attributed to a combination of an under-immunized population and a novel strain introduction.

Objective, Aims, and Experimental Approach

A comprehensive molecular description of the outbreak strain would provide more insight into the recent British Columbia mumps outbreak. We propose to perform whole genome sequencing (WGS) of the outbreak stain from patient samples positive for MuV RNA to generate high-resolution data that can i) reveal molecular changes in the outbreak strain that may explain the increased incidence and ii) identify epidemiological links between cases not apparent from contact tracing.

Aim 1: Characterize the mumps virus from a BC outbreak

The BC mumps outbreak strain differs from the endemic genotype G strain within the SH gene, but the full extent of divergence is unknown. Although the SH gene is considered the most variable in the MuV genome among all genotypes, for genotype G strains the greatest diversity is within the noncoding region between the matrix and fusion (F) genes[1]. Also, the SH gene does not appear to be involved in viral replication or virulence, while the F and haemagglutininneuraminidase (HN) proteins play significant roles in viral pathogenesis; mutations in both affect fusion activity and neurovirulence[1]. The (HN) protein is also the major target of humoral immunity; mutations in the HN gene reduce cross neutralization between strains[1]. By providing data on all MuV genes, WGS of the outbreak strain can potentially identify markers associated with increased virulence and antigenic drift. Moreover, collection of WGS data on a large number of isolates provides an opportunity to describe additional MuV gene

[1] Jin, L., Örvell, C., Myers, R., Rota, P. A., Nakayama, T., Forcic, D., . . . Brown, K. E. (2014). Genomic diversity of mumps virus and global distribution of the 12 genotypes. *Reviews in Medical Virology,25*(2), 85-101. doi:10.1002/rmv.1819. Review.

functions, including those related to transmissibility, as well as the mutation rate of MuV and the extent of within-host genetic diversity as was previously done for a measles outbreak[2].

Methods: WGS will be initially performed from isolates and then also from primary specimens (RNA viruses mutate in cell culture passage). Briefly, after RNA extraction and reverse transcription multiple amplicons covering the entire genome will be generated by nested PCR. Amplicons will be purified, fragmented, and tagged using a DNA Library preparation kit for sequencing by the Illumina MiSeq. Genomes will be aligned and assembled against a reference database of MuV genome sequences deposited on GenBank (n=108) using BWA-MEM. Consensus sequences will be generated from MPILEUP files using SAMTOOLS. The BEAST package will be used for phylogenetic analysis. Variants will also be assessed for their biological impact by examining both amino acids changes and alterations to know regulatory regions.

Results: To date, 13 isolates and 91 primary specimens have been extracted, amplified, and sequenced. These represent all the 2016 outbreak samples, excluding ones with PCR Ct values >33 where sequencing was shown not to yield good quality data, as well as 6 samples from other infections for comparison (including genotypes A, K, and H). Another 5 samples representing additional samples for comparison have been located and will be extracted, amplified, and sequenced between June 28 and July 7. These are samples recently identified with 1 nucleotide difference from the endemic strain, as well as 4 outbreak strains representing a 2008 outbreak in a faith based unvaccinated community, and 4 samples from the Manitoba Cadham provincial laboratory representing a recent Manitoba outbreak. Initial sequencing results of 8 isolates and matching primary specimens revealed that sequencing from specimens can yield good quality data and is preferred to isolate sequencing as isolates appeared to accumulate mutations in areas encoding the RNA polymerase during passage in culture. Previous genotyping data revealed that the 2016 BC mumps outbreak strain differed from the endemic strain in North America by five nucleotides in the SH gene. The SH gene of the MuV is located in the ~6000 base pair range of the MuV genome. Preliminary bioinformatics analysis revealed that the 2016 BC mumps outbreak strain differs remarkably from the endemic MuVi/Sheffield.GBR/1.05 strain in various regions across the genome: differences are not limited to the SH gene and numerous single nucleotide variations were identified when comparing the two throughout the entire genome. Genomic variability was also observed among samples positive for the outbreak MuV strain, but to a lesser extent, and mostly in the 4,000-6,000 base pair range. Future biofinformatic analysis will focus on assessing the potential impacts of the observed variations, particularly if they occur in F and HN genes. Bioinformatic analysis will be completed once all samples listed are sequenced (after July 7).

Aim 2: Identify epidemiological links within BC and between North American outbreaks

The BC mumps outbreak was observed in three geographic regions, but links between regional clusters were not always identified using traditional case investigation methods that identify key persons, places, and behaviors contributing to an outbreak. Since more than half (61%) of the lab confirmed outbreak cases could not be epidemiologically linked to another known mumps case, the transmission dynamics of the outbreak were not well defined. Also, while it is known that the genotype G endemic to North America originated from Europe[1], the origin of the recent BC outbreak strain is unknown. By describing the outbreak strain relative to historical strains as described in Aim 1, WGS data coupled with patient, spatial, and temporal information could enable high-resolution description of the direction and timeline of transmission as well as potential identification of superspreaders as was previously done for a

[2] Gardy, J. L., Naus, M., Amlani, A., Chung, W., Kim, H., Tan, M., . . . Tang, P. (2015). Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following the 2010 Olympic Winter Games Reveals Viral Transmission Routes. *Journal of Infectious Diseases,212*(10), 1574-1578. doi:10.1093/infdis/jiv271

tuberculosis outbreak[3]. Evolutionary analyses can also potentially describe the origins and spread of the outbreak strain in a global context.

Methods: Briefly, consensus MuV sequences generated in Aim 1 will be aligned to generate a phylogenetic tree using RAxML. This phylogenetic tree will be decorated with epidemiological information to enable the visualization and dynamic exploration of the data similar to nextflu.org. We will also reconstruct transmission pathways using TransPhylo.

Results: Briefly, preliminary phylogenetic analysis has described a large distance between the 2016 BC mumps outbreak strain and the endemic MuVi/Sheffield.GBR/1.05 strain, as mentioned above in Aim 1. Additional bioinformatic analyses pertaining to overall strain phylogeny as well as description of the transmission dynamics within the 2016 BC mumps outbreak will be completed once all samples listed in Aim 1 are sequenced (after July 7).

Significance and Knowledge Translation

The findings of this proposal are expected to contribute to the understanding of MuV, guide routine clinical laboratory practice, and inform strategies for future outbreak analyses. Mumps genotyping protocols could provide more value to surveillance programs if they regularly included sequencing of additional genes, as the WHO[4] has previously proposed for when a new lineage is suspected. Real-time WGS data could assist epidemiologists in identifying links between cases that could influence public health intervention efforts. Overall, more comprehensive and accurate surveillance is necessary to monitor trends in the incidence of infection, identify at risk populations for infection, and evaluate patient outcomes. Findings from this proposal would be submitted for presentation at the 2018 CACMID-AMMI Canada Annual Meeting and publication in a peer-reviewed journal such as The Journal of the Association of Medical Microbiology and Infectious Disease Canada. This project would be performed in collaboration with colleagues from the BCCDC, as well as investigators from the National Microbiology Laboratory and Manitoba Cadham provincial laboratory.

[3] Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. (2011). *New England Journal of Medicine,364*(22), 2174-2174. doi:10.1056/nejmx110050

[4] WHO. (2013). Mumps virus nomenclature update: 2012. *Weekly Epidemiological Record 2013, 87*(22), 217-224.