This report should be submitted within 2 weeks after you return to Japan. Please do not change the formatting

（Abroad・~~Domestic~~）Internship report form（Student） 2022/11/30 (Year/Month/Day)

| | |
|---|---|
| Name | Patrick Reteng |
| Laboratory | Collaboration and Education |
| Year (Grade) | D4 |
| Internship institution | Center for Virus Research, Westmead Institute for Medical Research |
| Internship period | Internship period: 11/04/2022 - 11/18/2022<br>(Departure Date from Sapporo: 10/29/2022, Arrival Date in Sapporo: 11/19/2022) |
| Purpose | - To gain knowledge about the application of metagenomic NGS as a pathogen detection platform, especially in the clinical setting.<br>- To gain experience on how to leverage the information obtained from genome sequencing using next generation sequencing into a more practical knowledge, such as viral epidemiology.<br>- To experience the working condition and situation in Australia, as it is one of the desired destinations for future workplace. |

- The reason why you chose this institute

As next generation sequencing (NGS) technologies have become more available in the past decades, it has been widely used in the field of infectious disease. It has been particularly useful in the epidemiology of infectious disease. Among the wide applications, the diagnostic application is particularly of interest. Theoretically, by reading all the genetic material that exist in a sample and then match those genetic sequences to the available database, this approach can detect and identify most of the existed pathogens. This approach is known as metagenomic NGS (mNGS). However, despite mNGS seems to be promising in theory, there are many factors that affects the translation of the technology to an actual clinical service. At present, there are only handful of places with running clinical mNGS service.

Westmead Institute of Medical Research (WIMR) is a private institute located in Westmead, New South Wales (NSW). The institute is known for its clinical research, with most of the researchers are also working clinician. In addition, the institute is located in a complex, together with Westmead Hospital, one of the major hospitals in NSW (Fig. 1). The hospital also hosts one of the biggest pathology labs in the state, where they perform routine pathogen

sequencing.

There are two main reasons behind the selection of this institute. Firstly, at center of virus research WIMR, Dr. John-Sebastian Eden is developing a rapid mNGS protocol and pipeline, which has been used in surveillance of some viral disease (including Japanese encephalitis virus, Hendra virus, and respiratory syncytial virus (RSV). This protocol has a potential for diagnostic purpose because of the simplicity and short sample-to-result turnaround. The interesting part is that there is currently a project to develop an mNGS diagnostic service, in which the rapid protocol developed by Dr. Eden will likely be included. Thus, I believe I can get a valuable experience by visiting such a leading scientist at mNGS field. Secondly, there are lots of information that can be obtained from the sequencing the genome of a virus, including its evolution and transmission. The lab is maintaining a genomic epidemiological surveillance of "neglected" respiratory viruses, particularly respiratory syncytial virus (RSV) in the state of NSW, in coordination with the genomic laboratory of pathology lab of NSW. Through this approach, the value of molecular diagnostic can be leveraged. I am interested in extracting such information from the sequencing data that I have, and I think the research and work being done in this laboratory very much align with what I want to do in the future. Lastly, I believe the common value of the two reasons I have mentioned above is that through this internship, I would be able to see how research on the bench (mNGS, genome sequencing) will be communicated to the clinical side.


**Figure 1**. Location of of WIMR and the Westmead Hospital complex

- Result of the activity （about 800 words、provide photos, tables and figures that clearly show the activities during the period）

As mentioned above, there are two main interests behind the selection of the institute, and the outcome of the activity mainly centered around those.

1. Rapid detection of pathogen from respiratory samples using mNGS – the RapidPrep.

During the first week, a set of extracted RNA samples were obtained from the hospital, to be analyzed with mNGS. For the sake of actually detecting some pathogens and that both samples can be used for both purposes, only samples previously tested positive for RSV were subjected to mNGS. Seven of these extracted RNA samples and a blank control were then processed using an in-house rapid mNGS protocol developed by Dr. Eden. Due to confidentiality, the complete protocol will not be explained in detail but in short, the process includes first and second strand cDNA synthesis, host depletion, and then library preparation with a transposase-based kit. High quality mNGS library can be generated in hours (Fig. 2A). The library was then sequenced in an Illumina iSeq platform (Fig. 2B) which was completed the next day.

**Table 1.** Sequencing statistics of the mNGS run.

| Sample | Raw Reads | Dropped during QC | Human reads | Non-human reads |
|---|---|---|---|---|
| 1 | 699,307 | 19,902 | 486,829 | 192,576 |
| 2 | 663,202 | 18,134 | 530,676 | 114,392 |
| 3 | 658,632 | 17,881 | 507,803 | 132,948 |
| 4 | 672,718 | 16,529 | 553,751 | 102,438 |
| 5 | 379,935 | 9,817 | 267,333 | 102,785 |
| 6 | 696,021 | 16,882 | 571,279 | 107,860 |
| 7 | 631,111 | 15,722 | 537,428 | 77,961 |
| 8 | 51,534 | 26,106 | 619 | 24,809 |

**Table 2**. Sequencing statistics of the mNGS run, using WIMR pipeline and using

| Sample | WIMR pipeline (Contig) | Collabed pipeline | | | |
|---|---|---|---|---|---|
| | | RSVA | RSVB | Bov_CoV | others |
| 1 | Corynebacterium | | | | |
| 2 | RSVA, Corynebacterium | 509 | 1 | | |
| 3 | RSVA, Moraxella, Haemophilus | 218 | | | |
| 4 | RSVA, Moraxella, Haemophilus, Streptococcus | 1482 | | 1 | TTV (1) |
| 5 | RSVB | | 201 | | |
| 6 | RSVB | | 1099 | | |
| 7 | RSVB, Moraxella | | 5235 | | |
| 8 | - | | | | |

*RSV, respiratory synctitial virus; TTV, torque teno virus.

The sequence data were then analysed bioinformatically to classify and detect any viral RNA in the sample. The analysis started by removing the host sequence by aligning the data to a human reference genome, the unaligned read was then extracted. To provide a

fast result and to get a picture of what pathogen exist in the sample, the non-human sequences were analysed with Kraken2 – a k-mer based classification tool. While Kraken provides a fast analysis, the results tend to be less accurate. Yet, for a rapid mNGS pipeline, it could provide a bigger picture of what exists in the sample. To ensure the highest quality result, the sequences were subjected to de-novo assembly to generate a "longer" sequence (contig) which provide more accurate alignment search. The generated contigs were then subjected to database search of both nucleotide and translated amino-acid sequences.

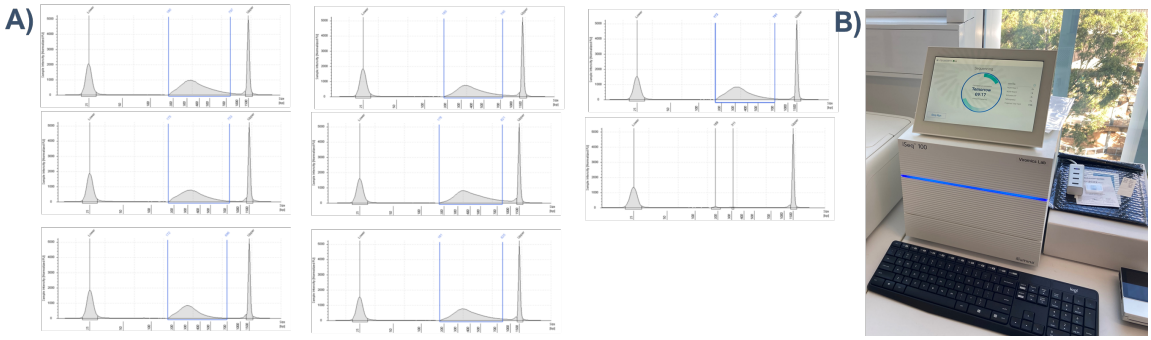


**Figure 2.** A) QC result of the libraries showed a good library quality. The QC was performed on an Agilent TapeStation platform. B) The iSeq platform in the laboratory.

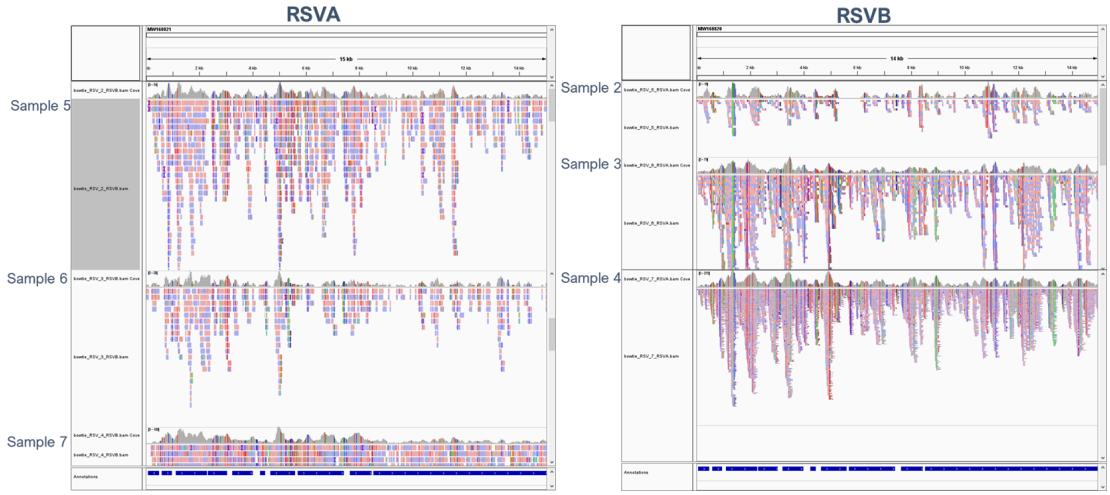


**Figure 3.** mNGS results show good coverage across the RSVA or RSVB genome achieved by RapidPrep.

Turnaround-time wise, this platform provides a rapid sample-to-result turnaround which will greatly benefit for diagnostic purpose. Within the next day, sequences of RSV were able to be detected from the data. One of the challenges in mNGS is in ensuring adequate depth for the pathogen sequence to be detected given the large amount of noise. The total read output from iSeq platform might not provide enough depth for mNGS of seven samples. Nevertheless, RSV sequences were detected from six samples (Table 1), and

when the reads were aligned to the reference genome, they provide enough depth for further analysis (Fig. 3). As shown in the kraken output, the detection tends to be less accurate, however the contig-based analysis further confirms the findings. Interestingly, in one sample no viral reads were able to be detected and in one sample, a co-infection with Human Coronavirus OC63 (HCoV-O63) was detected. However, due to limited sequencing depth, a contig was not able to be assembled for the HCoV virus. Because I also performed similar research, I tried to run the data through the pipeline that I established and found similar result. The HCoV-O43 was also detected, as Bovine Coronavirus, a closely related virus. Thus, a skilled operator that interprets the results is still required to call for the diagnosis. Nevertheless, altogether these results further showcasing how mNGS can improve the diagnostic result.

2. Epidemiological study of RSVA and RSVB in New South Wales.

   In 2021, an off-season RSV outbreak was reported from several Australian states. Interestingly, this outbreak happened to be resulted from a clonal expansion of a certain RSVA lineage in NSW (North South Wales) once the covid-19 lockdown is eased. However, when the border is open, several lineages has been reintroduced back, including some RSVB lineage. To monitor the circulation of RSV in NSW, a protocol to generate a whole genome sequence of RSV were developed by WIMR. A four different primer sets were generated to provide a whole coverage of approximately 15 kB long RSVA and RSVB genome (Fig. 4). Equal amounts of amplicons from 24 samples were then pooled into the library then sequence with iSeq.

   Amplicons were not generated from one sample, and this is also the same sample in which zero RSV reads were detected from metagenomic. This sample is originated from a rural part of Australia, and possibly has been degraded during sample transfer or that simply the virus amount is too low to be detected. Some amplicons drop out also were observed from several samples, resulting in only partially assembled genome. Complete RSV genome were successfully assembled from 21 samples, while partial genomes obtained from two samples. Both RSVA and RSVB are known to co-circulate, but the outbreaks in 2021 was dominated by RSVA. It is interesting that the current dominant type has shifted to RSVB (Table 3). This result was informed to a public health officer at the genomic lab of pathology NSW, and a phylogeny tree was constructed to be used by the public health officer (Fig. 5).

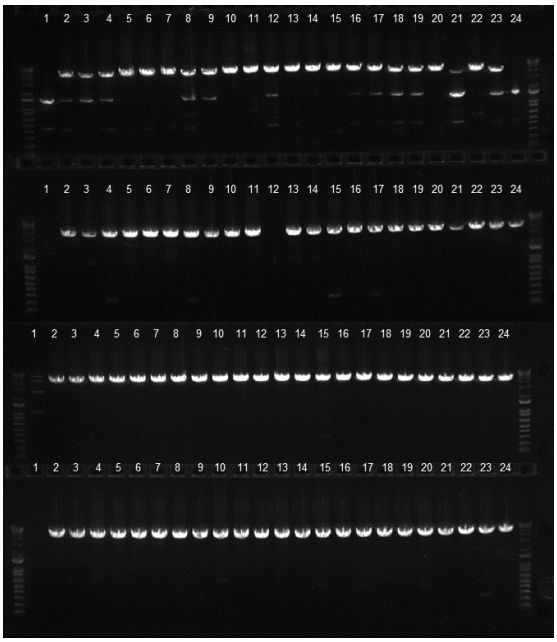


**Figure 4.** Visualization of the amplification result for RSV-WGS. The whole genome of RSV is divided into four segments, in which a primer set is used to amplify the 4 to 5 kB fragment. Several fragments were not amplified, likely due to mismatch at the primer site.

**Table 3.** Type of RSV obtained from WGS.

| Sample | RSV type | Collection date | Sample | RSV type | Collection date |
|---|---|---|---|---|---|
| 1 | - | 27/07/2022 | 13 | B | 13/07/2022 |
| 2 | B | 23/08/2022 | 14 | B | 12/07/2022 |
| 3 | B | 23/08/2022 | 15 | B | 12/07/2022 |
| 4 | B | 16/08/2022 | 16 | B | 12/07/2022 |
| 5 | A | 22/10/2022 | 17 | B | 12/07/2022 |
| 6 | A | 21/10/2022 | 18 | B | 11/07/2022 |
| 7 | A | 25/10/2022 | 19 | B | 14/07/2022 |
| 8 | B | 17/07/2022 | 20 | B | 12/07/2022 |
| 9 | B | 13/07/2022 | 21 | B | 06/07/2022 |
| 10 | B | 15/07/2022 | 22 | B | 09/07/2022 |
| 11 | A | 17/07/2022 | 23 | B | 04/07/2022 |
| 12 | B | 12/07/2022 | 24 | B | 07/07/2022 |

Additionally, to further visualize the phylodynamic of RSVA and RSVB in NSW, the sequences were subjected to NextStrain analysis. This analysis was able to predict the when the viruses in a branch of the tree last shared common ancestor, which is useful in order to predict the transmission of the virus. To get a global picture of the current RSV situation, addition, 576 RSVB sequences and 874 RSVA sequences from 2010 were extracted from GenBank and then analysed together using the NextStrain pipeline (Fig. 6 and Fig. 7). At the time, sequences from GisAid database were not able to be accessed due to server issue. Nevertheless, some interesting findings can be seen. Firstly, from the phylogenetic tree of

RSVA the three sequences that were obtained are genetically distinct from the outbreak strain in 2021, suggesting that the current sequences resulted from introduction. However, given the long branch of the sequences, one might speculate that this clade might has been circulating under the radar in Australia, or maybe at the other (possibly) European countries prior to introduction to Australia. Secondly, the 2021 outbreak strain appeared to be discontinued, which might be resulted of the low number of samples being sequenced. It is also possible that this linage has gone extinct. Thirdly, majority of RSVB sequences are clustered in the same clade and showing a diversification event. From this phylogenetic analysis, it is also shown that interstate spillover is less common for RSVB, but rather introduction from outside of Australia is more common. This observation needs to be taken with a grain of salt, because the number of sequences is limited and because the particular branch showing a long distance, hinting to sampling-bias. Yet, the timing of the possible introduction aligned to the time when lockdown was initiated, and people resided in foreign country were coming back. It also seemed that this lineage does not stay in circulation, as a result of pressure from the strict lockdown that was imposed.


**Figure 5.** The phylogenetic tree of the obtained RSVB sequences that was reported to the hospital This tree including the sequences from an indigenous community (highlighted in pink).

Besides the two main activities above, I got to visit the genomic lab of pathogen and see how the work there was done by pathologists. It interesting how Australia did they genomic surveillance in hospital, something that I was not able to participate in Japan nor Indonesia. I

also joined a "plate" round, with the genomic team, where they gather and give update on what pathogen is being sequence and if there are some anomalies in their sequence (including possible antimicrobial resistance and concerning mutation), or how they are related to other cases that have been sequenced based on the phylogeny. For example, during that meeting, the coronavirus team gave an updated about a novel strain of SARS-CoV-2 (BR.2.1) which is highly likely to originate from NSW.

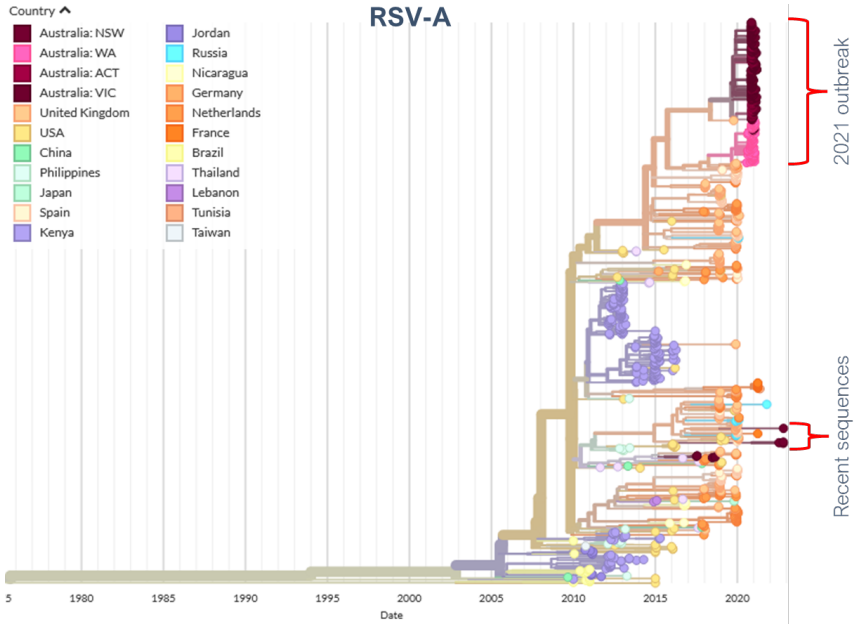


**Figure 6.** The dynamic of RSVA evolution and transmission as resulted from the NextStrain pipeline. The analysis replicated the finding of outbreak in 2021 and find that the current sequences are genetically distinct from the 2021 outbreak line.

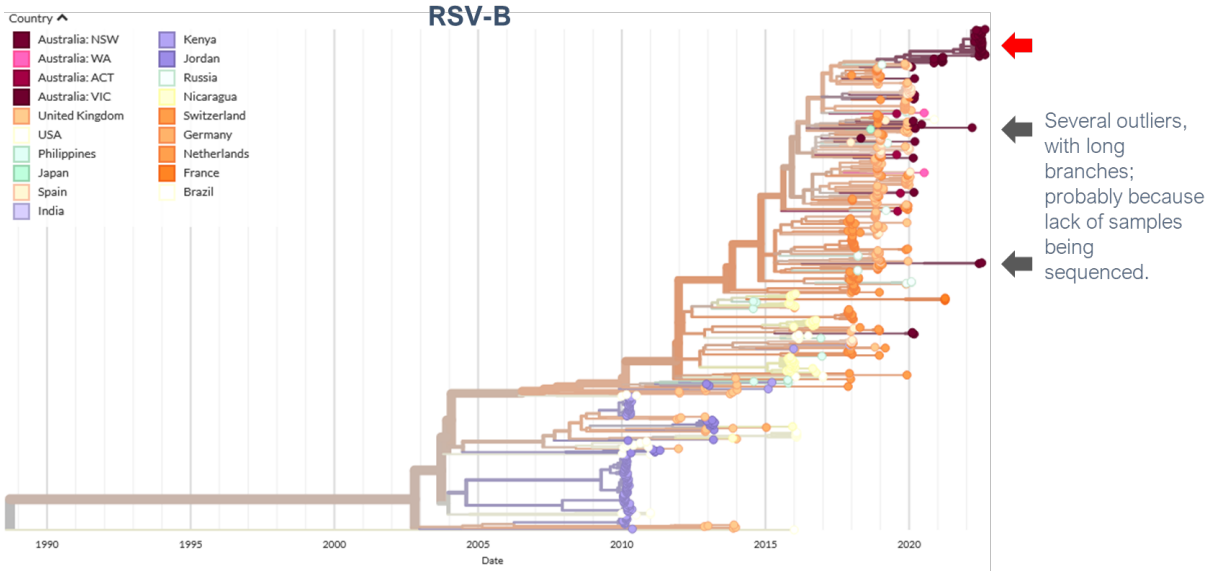


**Figure 7.** The dynamic of RSVB evolution and transmission as resulted from the NextStrain pipeline. The recent sequences are clustered together in one clade. Several outliers with long branch can also be

observed.

- What do you think the positive impact of the activity will have on your further career path?

I believe the most impactful outcome is the connection that I made with the Dr. Eden and his team, as a leading clinical metagenomic scientist. Through this internship I was also able to observe the working conditions of post-doctoral fellow in Australia or even working situation in Australia. I would very much like to work in Australia, shall a chance arise. Though I was not offered a job, the host did suggest me to get in contact with some principal investigators which might be beneficial for me. Additionally, I was able to observe how genomic epidemiologies is being integrated to clinical practice, and also get some insight on the challenges of applying mNGS in routine clinical practice.

- Advice for your junior fellows

Several advices to those who have not completed internship:

- "The world is your oyster" – you can go **anywhere** due to this opportunity, which does not present in other graduate school program, and you should make the most out of this opportunity. Aim big, do not limit yourself, go for the place that you always dream of, and I am sure you will get a lot from the short visit, including a motivation to pursue further career.
- Do not hesitate to contact potential supervisor; after all, Lady Luck favors the one who tries. Once you send an email, do not forget to send a follow up email because they are busy people, and your email might be drowned in hundreds of emails that they receive. Do not give up until you sent your follow up email.
- Prepare early, especially regarding Visa for those who need it.
- Prepare some savings, in case of expensive budget that you have to pay first or unexpected spending (i.e. health check for Visa, cost to visit the embassy).
- If possible, secure another source of funding so you can conduct your internship longer.

| | |
|---|---|
| Approval of supervisor | Institution・Official title・Name<br>International Institute of Zoonosis Control, Division of Collaboration and Education – Associate Professor – Junya Yamagishi |

※1 Send the electronic file to the WISE Program Office

※2 Attach a copy certificate of the content of internship activity that is prepared by the counterpart at the internship institution (any form with a signature of the counterpart).

※3 The Steering Committee for the WISE Program will first confirm the content of this report and report will be forwarded to the Educational Affairs Committee for credits evaluation.