As of 2024.3

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) 2024/07/02 (Year/Month/Day)

Name	Pawarut NARONGPUN
Laboratory	Division of Bioresources, International Institute for Zoonosis Control
Year (Grade)	DC3
Internship institution	Research Group for Genomic Epidemiology, Technical University of Denmark (DTU)
Internship period	Internship period: 05/06/2024 - 06/28/2024 (Departure Date from Sapporo: 05/04/2024, Arrival Date in Sapporo: 07/30/2024)

- **Purpose and the reason why you chose this institute**

The primary goal of this internship is to gain proficiency in analyzing sequencing data under the guidance of experienced professionals. I am eager to develop my bioinformatics skills and learn how to troubleshoot computational problems bioinformatically because, in today's scientific research, researcher has a tendency to integrate sequencing technologies and data analysis into their works. Moreover, I believe that this internship will provide valuable benefits for my future career.

The DTU Research Group for Genomic Epidemiology specializes in preventing the spread of infectious diseases in humans, with a strong focus on antimicrobial resistance, which aligns closely with my research interests. Moreover, this group is recognized as a leader in Europe for its extensive research on *Staphylococcus aureus* isolated from both patients and animals, evidenced by numerous published studies. Therefore, I am eager to engage in discussions with experts in *S. aureus* research to further enrich my understanding in this area."

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

During my internship at the Research Group for Genomic Epidemiology, DTU, I conducted all my research using my personal computer. The focus was on computational work, and there were no wet lab experiments involved. Proficiency in Python and other computational skills is crucial for anyone interested in this lab, as learning these skills later on could be challenging and time-consuming. Before traveling to Denmark, *S. aureus* DNA was extracted

using a commercial extraction and purification kit in Japan. The extracted DNA underwent thorough qualification using multiple steps and protocols. Subsequently, it was sequenced using two platforms with different technologies: Illumina and Oxford Nanopore sequencing.

1. Quality assessment and preprocessing of sequencing reads

At DTU, students use Linux operating system, which is available on Computerome, to analyze their sequencing data. After uploading my data, I employed the FoodQC pipeline to prepare my short-read sequencing data for further analysis. This pipeline consists of three key steps: trimming, genome assembly, and quality control - both before and after the trimming and assembly process. However, long-read data preprocessing was completed in Japan. Consequently, I focused solely on genome assembly followed by quality control.

2. Genome polishing

To improve the draft genome assembled from long Nanopore reads, I employed a technique called genome polishing. This process leverages high-quality contigs generated from Illumina short-read sequencing to correct small-scale errors present in the Nanopore assembly. The resulting complete genome is then suitable for downstream analyses such as gene annotation, synteny analysis, and phylogenetic construction.

3. Phylogenetic tree analysis

After completing the initial two steps, my sequencing data was prepared for downstream analysis. Next, I employed two phylogenetic methods, namely maximum-likelihood and Bayesian inference, to construct phylogenetic trees based on core genome alignment. Initially, I created multiple sequence alignments using 1,197 draft genomes of *S. aureus* sourced from an international database, alongside 18 complete and draft genomes from Thailand. The resulting maximum-likelihood tree was then visualized using the Interactive Tree of Life (iTOL). Subsequently, 168 out of 1,125 genomes were selectively chosen for Bayesian phylogenetic tree analysis. The results from this step could reveal the geographical relationships of *S. aureus* strains originating in Thailand with those from other countries.

4. Structural comparison of staphylococcal cassette chromosome *mec* (SCC*mec*)

In my previous experiment, I identified an uncommon SCC*mec* type. In my current study, I aim to elucidate the origin of this unknown SCC*mec*

element. This research question prompted me to conduct structural comparisons of SCC*mec*. To address this question, I employed two bioinformatic tools: Flankophile and Clinker. However, following discussions with my host professor and mentor, they advised me to conduct further analysis, as the initial results were not sufficient to answer my research question. I am planning to continue these analyses upon my return to Japan.

During my time at DTU, I have become proficient in several software and tools essential for my research:

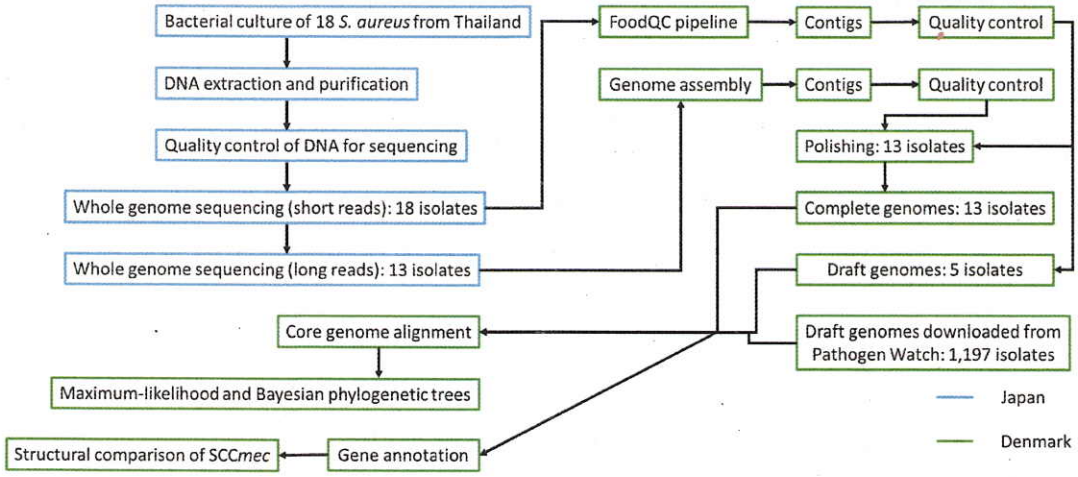
1. FoodQC pipeline (QUAST, SPAdes, BBDuk): quality control and *de novo* assembly for short reads
2. Flye: *de novo* assembly for long reads
3. Raven: *de novo* assembly for long reads
4. Polypolish: genome polishing
5. CSI Phylogeny: phylogenetic tree construction
6. Flankophile: to visualize flanking region synteny and gene synteny-based analysis
7. Clinker: gene synteny-based analysis
8. Prokka: gene annotation
9. NotePad ++: gene annotation, text editor, metadata
10. Git Bash: Flankophile
11. BEAST: Bayesian phylogenetic tree analysis
12. BEAUti: Bayesian phylogenetic tree analysis
13. Tracer: Bayesian phylogenetic tree analysis
14. TreeAnnotator: to construct a maximum clade credibilty tree
15. iTOL: tree visualization
16. FigTree: tree visualization
17. ResFinder: identification of acquired antibiotic resistance genes
18. SCC*mec*Finder: identification of SCC*mec* elements in sequenced *S. aureus* isolates
19. KmerFinder: prediction of bacterial species using a fast K-mer algorithm
20. MyDbFinder: typing using BLAST based on a user defined database

In addition to focusing on my own research, I had the opportunity to attend lectures and workshops hosted by Dr. Pimlapas.

The course, titled "Whole Genome Sequencing Analysis for Microbial

Diagnostic, Identification, and Cluster Analysis," lasted four days and included hands-on workshops on bacterial genomic analysis and an introduction to metagenomics. Dr. Pimlapas began with foundational concepts in sequencing analysis, sample collection for bacterial sequencing, and an overview of whole genome sequencing technologies. During the course, students practiced using various tools for bacterial typing with sequencing data, such as multi-locus sequencing typing (MLST) and plasmid MLST. Additionally, Dr. Pimlapas lectured on the detection of antimicrobial resistance genes using ResFinder and phylogenetic tree analysis. She also invited Dr. Miena Ivanova, a postdoc, to deliver a lecture on antibiotics for bacteria. Eventually, she closed the class with lectures on metagenomics and machine lerning for food safety.

Workflow summary of my research



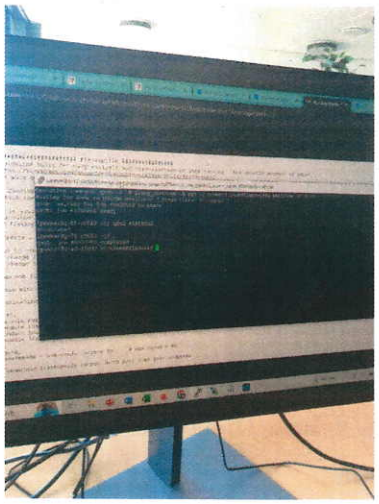


Figure 1 illustrates my work in the lab. My experiments primarily consisted of computer-based tasks, without any wet lab work. I spent my entire time at my computer, writing command lines and utilizing various software tools.

As of 2024.3

Detailed schedule

Date	Activity	Mentor
May 6th to 10th	Arrive in Denmark Meet Prof. Frank and Dr. Pimlapas Planning Quality control and preprocessing of sequencing reads	Prof. Frank Dr. Pimlapas
May 13th to 17th	Hybrid genome assembly Quality assessment Polishing	Dr. Pimlapas
May 20th to 24th	Gene annotation Synteny-based analyses	Mr. Narong
May 27th to 31st	Genome downloading from Pathogen Watch Databases Quality control and preprocessing of sequencing reads Phylogenetic tree reconstruction Metadata preparation	Dr. Pimlapas
June 3rd to 7th	Phylogenetic tree reconstruction Lectures on whole genome sequencing analysis	Dr. Pimlapas
June 10th to 14th	Phylogenetic tree reconstruction Synteny-based analyses	Dr. Pimlapas Mr. Narong
June 17th to 27th	Phylogenetic tree reconstruction Synteny-based analyses Presentation	Prof. Frank Dr. Pimlapas Mr. Narong
June 28th	Bound for Sapporo	

Figure 2. Photos with Dr. Pimlapas Leekitcharoenphon and Mr. Narong Nuanmuang. Dr. Pimlapas is a senior researcher and bioinformatician working in this lab. Her research projects are involved in antimicrobial resistance in *Salmonella* spp. and metagenomics. Mr. Narong also supported and took care of me during my stay at DTU.

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

I have gained practical experience in using essential bioinformatics tools and some techniques for bacterial genomics that were previously unfamiliar to me, including phylogenetic tree analysis using Bayesian inference, gene synteny-based analysis and flanking sequence-based clustering. Moving forward in my career, I am confident in applying these skills to my future research, regardless of the infectious disease under investigation. Also, I believe that these skills will enhance my competitiveness in future research endeavors. Additionally, my time at DTU provided valuable networking opportunities with experts and peers in the field, reinforcing the importance of staying updated and continuously improving. This experience is crucial for advancing my qualifications for potential career opportunities in academia, industry, or research institutions specializing in microbiology, genomics, or bioinformatics. Overall, participating in these activities has significantly contributed to my professional development, equipping me with essential skills, knowledge, and connections necessary for a successful career in microbiology and genomic research.

- **Report how your activity could link to One Health Approach (If applicable)**

If you also conducted OH onsite training (Ally Module4), please describe some of the examples of One Health approach you implemented in your activity. Or explain the possibility(ies) how you could link this activity to One Health approach for your future.

Understanding bacterial genomics and antimicrobial resistance gives me insights into the genetic diversity and evolution of resistance bacteria affecting both human and animal health. This understanding is crucial for addressing zoonotic diseases and shared health risks between humans and animals. Moreover, whole genome sequencing technologies allow us to access more information for monitoring and surveillance of pathogens in different hosts and scenarios. This is very significant for the early detection of disease outbreaks, tracking transmission routes, and identifying emerging or re-emerging infectious diseases that can be harmful to humans, animals, and ecosystems.

- **Advice for your junior fellows**

1. If you wish to have an internship in Europe, you should save up money for this trip due to higher living costs compared to Japan.
2. You should always have a plan B. Although I had started work permit 4 months before the date of departure, I was granted a work permit very late.

As of 2024.3

Approval of supervisor	Institution・Official title・Name Division of Bioresources, International Institute for Zoonosis Control Professor NAKAJIMA Chie 中島千絵

※1 A certification form from the host should also be submitted.
※2 The Career Path Committee will first confirm the content of this report and report will be forwarded to the Educational Affairs Committee for credits evaluation.