学位論文内容の要旨
Abstract of the dissertation

博士の専攻分野の名称：博士(感染症学)　　氏名：PATRICK RETENG
Name

学位論文題名
Development of Highly Efficient Methods for Comprehensive Pathogen Detection Using Next Generation Sequencing
(次世代シーケンサーを用いた高効率網羅的病原体検出法の開発)

Febrile illness is a common manifestation of infectious disease. The vast aetiologies, including infectious and non-infectious aetiologies, can complicate the diagnosis. For infectious aetiologies, next generation sequencing (NGS) is a powerful tool that can provide broad range pathogen detection using approaches such as shotgun metagenomic sequencing or targeted sequencing. However, its application is limited by complicated library preparation, turn-around time, cost, among others. This study explored several methods to make NGS-based pathogen detection more efficient.

Although targeted sequencing is not widely applicable for viruses, conserved sequence can be found at a family or a genus level. This conserved region provides a semi-comprehensive target for detection of viruses. Genus Flavivirus was chosen as a model for this approach because of the clinical significance and because this genus has a conserved region in the region that encodes Non-structural 5 protein. Scalability of the sample was increased by attaching unique nucleotide sequences at the 5'-end of the primers. The system was able to amplify and identify a broad range of flavivirus while having a low demultiplexing error rate (0.02%). The system successfully detected several serotypes of dengue virus (DENV) from 114 samples obtained in Vietnam. The observed positive and negative agreement in comparison to a commercial nucleic acid test were 66.7% and 95.4%, respectively. In another sample set, diverse flaviviruses were able to be detected and thus, supporting the broad-range aspect of the system.

An amplification step is often required because of the limited amount of genetic material in the sample and because NGS requires a certain input amount of genetic material. Comprehensive amplification that relies on PCR has several limitations, including requirement of a precise temperature control and risk of introducing bias. This study introduces circle whole transcriptome amplification (cWTA) as an alternative unbiased amplification that targets the RNAome. In cWTA, the synthesized cDNA was circularized using an enzyme that has a high affinity towards ssDNA then amplified using phi29 polymerase, isothermally. This study showed that 1) circularization of cDNA template improved the amplification, 2) the resulting amplicons were suitable for NGS analysis, and 3) viral sequences (DENV2 and chikungunya virus (CHIKV)) were able to be detected in the clinical samples.

Library construction is a major bottleneck for NGS analysis because it is costly and complicated. To reduce the number of libraries that need to be prepared, a combination of cWTA, Illumina sequencer, and a group testing algorithm (or metagenomic NGS enhanced by a group testing algorithm (mEGA)) were employed to detect viral pathogens in 44 serum samples obtained in Vietnam, which including one sample positive for DENV1 as a positive control. Using group testing algorithm, the number of sequencing libraries that needed to be constructed were reduced to 11 libraries, while retaining the sample information. The sample origin of viral reads detected in the pools were able to be traced back and then validated using pathogen-specific PCR. From these 43 samples, DENV2, hepatitis B virus (HBV), and human parvovirus B19, were able to be detected without prior knowledge.