A random priming amplification method for whole genome sequencing of SARS-CoV-2 and H1N1 influenza A virus

Main author: Klaudia Chrzastek (The Pirbright Institute)

Co-authors: Klaudia Chrzastek, Chandana Tennakoon, Dagmara Bialy, Graham Freimanis, John Flannery, Holly Shelton

INTRODUCTION

Non-targeted whole genome sequencing is a powerful tool to comprehensively identify constituents of microbial communities in a sample. There is no need to direct the analysis to any identification before sequencing which can decrease the introduction of bias and false negatives results. It also allows the assessment of genetic aberrations in the genome (e.g. single nucleotide variants, deletions, insertions and copy number variants), including in non-coding protein regions. Those changes might occur following vaccination or anti-viral treatment for example, which could lead to therapy-related failures.

METHODOLOGY

The performance of four different random priming amplification methods to recover RNA viral genetic material of SARS-CoV-2 coupled with whole genome sequencing was compared in this study. In method 1 (H-P), the reverse transcriptase (RT) step was performed with random hexamers whereas in methods 2-4, RT incorporated an octamer primer with a known tag. In methods 1 and 2 (K-P), sequencing was carried out on material derived from the RT-PCR step whereas in methods 3 (SISPA) and 4 (S-P), an additional amplification was incorporated before sequencing.

Metagenomics detection. Three independent methods were used to detect the presence of the viruses in the samples using an in-house generated pipeline that will be discussed. The first assembly method can identify organisms if they are present in the sequencing data in a sufficiently high concentration to be assembled. The second method can detect viruses at a lower concentration. The final method would be sensitive if the references were close to the isolates in the samples.

RESULTS

The SISPA method was the most effective and efficient method that we tested for non-targeted/random priming whole genome sequencing of COVID. The SISPA method

described in this study allowed for whole genome assembly of SARS-CoV-2 and influenza A(H1N1)pdm09 in mixed samples. We determined the limit of detection and characterisation of SARS-CoV-2 virus, which was 103 pfu/ml (Ct, 22.4) for whole genome assembly and 101 pfu/ml (Ct, 30) for metagenomics detection.

DISCUSSION

The performance of four different random priming amplification methods to recover RNA viral genetic material was compared in this study. The SISPA technique allowed for whole genome assembly of SARS-CoV-2 and influenza A(H1N1)pdm09 in single mixed virus samples. The whole genome sequences recovered after applying the SISPA (or S-P) method presented in this study are free of primer bias and allowed for a polymorphism analysis. This method is predominantly useful for obtaining genome sequences from RNA viruses or investigating complex clinical samples (such as mixed infections in a single reaction) as no prior sequence information is required. The method might be useful to monitor SARS-CoV-2 virus changes such as mutation or deletions in virus genome, to perform simple and fast metagenomics detection and to assess the general picture of different microbes within the sample that might be useful to identify the other co-factors that correspond to COVID-19 infection. Notably, the method presented here does not rely on primer specificity unlike conventional qRT-PCR and therefore any changes in viral genomes (mutations or deletion) do not impact the detection of pathogens.