

Diagnostic metagenomics: a case study based on suspected dengue infection

LC Contevelle, MA Marín, AMB de Filippis, RMR Nogueira, MCL Mendonça, ACP Vicente

Oswaldo Cruz Institute, Oswaldo Cruz Foundation

Dengue virus infects an estimated 50–100 million people annually worldwide and hyper-endemic in Brazil, where the mortality associated to this infection has reached 12% in 2010. During epidemics in Brazil, 50% of the suspected cases remain not confirmed even by specific RT-PCR, NS1 antigen test, IgM and viral cell culture and some of them evolved to fatal cases. In this study, our aim was to analyze a set of fatal cases, in order to reveal, minimizing the bias of specific diagnosis, any infectious agent present in these cases. We applied metagenomic approaches using high-throughput sequencing data from an Illumina HiSeq 2500 run. We performed quality control of the reads with cutadapt and prinseq. Then, clean reads that mapped to the human genome (Hg38) using Bowtie 2 with default parameters were filtered. Taxonomic analysis were performed with Kraken, GOTTECHA, SURPI, Metaphlan2, Taxoner and Blastn that identified three viruses in distinct samples that were subsequently confirmed by specific PCR: Parvovirus B19, Hepatitis G virus and Torque-teno virus. Considering the Parvovirus B19, it was possible to recover its complete genome (5.6 kb) and determine that it belongs to genotype 1A, which is the predominant genotype worldwide. Two pathogenic bacteria were identified in four other samples: *N. meningitidis* serogroup C (n=2) and *Streptococcus pneumoniae* (n=2). Both bacteria have been causing outbreaks and epidemics in Brazil and infections with the former having high mortality rate. Therefore, metagenomics and bioinformatic analysis are an useful approach to reveal unpredictable epidemiological scenarios and should be applied in global emerging pathogen surveillance programs.