

An automated method for the identification of Dengue, Zika, Yellow Fever and Chikungunya virus species and genotypes

Luiz Carlos Júnior Alcântara¹, Nuno R. Faria², Marta Giovanetti¹, Vagner Fonseca¹, Maria Inés Restovic¹, Murilo Freire¹ & Tulio de Oliveira³

¹*Oswaldo Cruz Foundation (FIOCRUZ), Salvador, Bahia, Brazil.* ²*Oxford University, UK.* ³*Africa Centre, University of KwaZulu-Natal, Durban, South Africa.*

In recent years, an increasing number of outbreaks of Dengue virus (DENV), Zika virus (ZIKV) and Chikungunya virus (CHIKV) have been reported in Asia and the Americas, while the Yellow Fever virus (YFV) continues endemic in Africa. The geographical distribution of ZIKV has expanded significantly reported now in at least 41 countries. Since these arboviruses share many clinical symptoms, such as febrile illness with rash, myalgia, or arthralgia, and current serological tests lack the power to discriminate between ZIKV and other flaviviruses, such as DENV and YFV, genetic testing during acute infection has become the standard method to identify the cause of infection. To facilitate diagnosis and the development of prevention and treatment strategies that efficiently target the diversity of these viruses, we developed a rapid high-throughput-genotyping system. The method involves the alignment of a query sequence with a carefully selected set of predefined reference strains, followed by phylogenetic analysis of multiple overlapping segments of the alignment using a sliding window. Each segment of the query sequence is assigned the genotype and sub-genotype of the reference strain with the highest bootstrap (>70%) and bootscanning (>90%) scores. The new Arbovirus-Genotyping Tools provide accurate classification of these arboviruses and are currently being assessed for their diagnostic utility. In conclusion, our new computational method allows the high-throughput classification of DENV, ZIKV, YFV and CHIKV species and genotypes in seconds. Species can be classified using short reads from any NGS platform, such as metagenomics Illumina's RNA-seq, and genotypes can be classified most confidently when using envelope gene or complete genome sequences. The framework's is freely available online from a dedicated server (<http://www.bioafrica.net/software.php>).

Area: Software Development and Databases