

WHOLE GENOME COMPOSITION DISTANCE FOR HIV-1 GENOTYPING

Xiaomeng Wu, Randy Goebel
Department of Computing Science, University of Alberta
Edmonton, Alberta T6G 2E8, Canada

Xiu-Feng Wan
Systems Biology Laboratory, Department of Microbiology, Miami University
Oxford, OH 45056, USA.

Guohui Lin
Department of Computing Science, University of Alberta
Edmonton, Alberta T6G 2E8, Canada

Existing HIV-1 genotyping systems require a computationally expensive phase of multiple sequence alignments and the alignments must have a sufficiently high quality for accurate genotyping. This is particularly a challenge when the number of strains is large. Here we propose a whole genome composition distance (WGCD) to measure the evolutionary closeness between two HIV-1 whole genomic RNA sequences, and use that measure to develop an HIV-1 genotyping system. Such a WGCD-based genotyping system avoids multiple sequence alignments and does not require any pre-knowledge about the evolutionary rates. Experimental results showed that the system is able to correctly identify the known subtypes, sub-subtypes, and individual circulating recombinant forms.