INNOVATING CROSS-DOMAIN SOLUTIONS TO DETECT EMERGING BIOLOGICAL THREATS

Detection Of Genomic Aberrations In Viral Genomes By Measurement Of Phylogenetic Distant Attraction

Robert Murdoch BATTELLE MEMORIAL INSTITUTE **Noah Winters** BATTELLE MEMORIAL INSTITUTE

Background Information
In modern biosurveillance and epidemiology of viral diseases, identifying the strain and origin of viruses is assessed by analysis of whole genome sequences. More rigorous analyses are conducted by building maximum likelihood (ML) or Bayesian trees against a reference collection of related viruses to determine taxonomic subgrouping, providing insights into geographic and temporal origins. Such exercises are state of the art for biosurveillance but are not sensitive for detection of viral recombination events or intentional genetic engineering.

Purpose
The outcome of phylogenetic tree-based viral analysis is the single most probable tree; each feature on this tree (i.e. branch points) has a probabilistic likelihood score assigned to it that is derived from the spectrum of the non-final tree topologies; for example, the placement of a novel virus may occur at the same location in 90% of the trees. Generation of this final tree necessarily involves inferring hundreds or thousands of non-final phylogenetic trees, each of which represents a possible phylogenetic interpretation of the multiple-sequence alignment features. Among these minority trees, which are typically not examined, may lie alternative tree placements that reflect otherwise-undetected aberrant signatures driven by recombination or engineering.

Objective and Rationale
Data streams currently outstrip the number subject matter experts required to manually curate phylogenetic trees. High throughput methods to scan and measure aberrant signals in phylogenetic data are therefore necessary to prioritize samples for expert review.

Methods and Preliminary Results
The method presented here generates a small set of metrics for detecting aberrant topologies among non-final phylogenetic trees. This is accomplished by iteratively comparing branch-length distributions from non-final tree topologies to the “true” tree. Analyses generated from sets of viral samples subjected to artificial in silico recombination will be presented.

Preliminary Conclusions
“Distant attraction” or “aberration” metrics may aid in high-throughput biosurveillance activities by enabling detection of possible natural recombination events or intentional engineering efforts which would otherwise go undetected by current practices.