ABSTRACT

HIV is a relatively new pandemic, responsible for causing AIDS and killing millions of people around the world in a variety of different strains and subtypes. However, phylogenetic trees documenting HIV's evolutionary history have long been focused on strains related to the Western world. They also often have poor bootstrap and/or posterior probability support, strongly indicating inaccurate trees. The goal of this project was to produce state-of-the-art trees with a focus on Western Africa, considered the origin of the virus, with the expectation that both patterns found in other trees and brand new ones would be uncovered; this would prove to be difficult after almost 50 years of research in the field. Sequences from the Los Alamos HIV database were aligned using MAFFT, and converted to a NEXUS file by SeaView. Running the NEXUS file through PAUP*, a parsimony-constraint phylogenetic tree was obtained. The NEXUS file was then converted into XML format by BEAUTi and annotated with informative priors. The XML file was then run through BEAST where a phylogram was produced. Both of these programs, especially BEAST, are heavily computational. In BEAST, the Monte Carlo Markov Chain was set to 20,000,000, taking almost 1 day and 12 hours to complete. The two trees had near perfect bootstrap and posterior probability values, which indicate the certainty with which the clades represent accurate groupings. This, combined with their regional focus, makes the trees revolutionary findings themselves. The trees demonstrate general patterns found in earlier HIV research, as well as new evidence for the theory that HIV-1 Type O came from gorilla SIV and that HIV-2 Types A and B might have come from a single transmission. Both of these discoveries completely change our view of HIV's evolutionary history. This research can be used to better understand the origin of the virus as well as design new treatments targeting specific similarities or continuities between strains and subtypes.