

# A reliable Epstein-Barr Virus classification based on phylogenomic and population analyses

Zanella L.

Riquelme I.

Buchegger K.

Abanto M.

Ili C.

Brebi P.

The Epstein-Barr virus (EBV) infects more than 90% of the human population, playing a key role in the origin and progression of malignant and non-malignant diseases. Many attempts have been made to classify EBV according to clinical or epidemiological information; however, these classifications show frequent incongruences. For instance, they use a small subset of genes for sorting strains but fail to consider the enormous genomic variability and abundant recombinant regions present in the EBV genome. These could lead to diversity overestimation, alter the tree topology and misinterpret viral types when classified, therefore, a reliable EBV phylogenetic classification is needed to minimize recombination signals. Recombination events occur 2.5-times more often than mutation events, suggesting that recombination has a much stronger impact than mutation in EBV genomic diversity, detected within common ancestral node positions. The Hierarchical Bayesian Analysis of Population Structure (hierBAPS) resulted in the differentiation of 12 EBV populations showed seven monophyletic and five paraphyletic. The populations identified were related to geographic location, of which three populations (EBV-p1/Asia/GC, EBV-p2/Asia II/Tumors and EBV-p4/China/NPC) were related to tumor development. Therefore, we proposed a new consistent and non-simplistic EBV classification, beneficial in minimizing the recombination signal in the phylogeny reconstruction, investigating geography relationship and even infer associations to human diseases. These EBV classifications could also be useful in developing diagnostic applications or defining which strains need epidemiological surveillance. © 2019, The

Author(s).