Entropy-based approach for selecting informative regions in the L1 gene of bovine papillomavirus for phylogenetic inference and primer design

M.V.A. Batista, A.C. Freitas and V.Q. Balbino

Departamento de Genética, Centro de Ciências Biológicas,
Universidade Federal de Pernambuco, Recife, PE, Brasil

Corresponding author: M.V.A. Batista
E-mail: mvabatista@hotmail.com

ABSTRACT. Bovine papillomaviruses (BPVs) cause many benign and malignant lesions in cattle and other animals. Twelve BPV types have been identified so far, and several putative novel BPV types have been detected based on the analysis of L1 gene fragments, generated by FAP59/64 and MY11/09 primers. Phylogenetic trees are important in studies that describe novel BPV types. However, topological mistakes could be a problem in such studies. Therefore, we made use of entropy to find phylogenetic informative regions in the BPV L1 gene sequences from all 12 BPVs. Six data sets were created and phylogenetically compared to each other using neighbor-joining and maximum likelihood methods of phylogenetic tree reconstruction. We found two major regions in the L1 gene, using an entropy-based approach, which selects regions with low information complexity. More robust phylogenetic trees were obtained with these regions, when compared to the ones obtained with FAP59/64 and MY11/09 primers. More robust phylogenetic trees are important to accurately position novel BPV types, subtypes and variants. We conclude that an entropy-based approach is

a good methodology for selecting regions of the L1 gene of BPVs that could be used to design more specific and sensitive degenerate primers, for the development of improved diagnostic methods.

Key words: Bovine papillomavirus; L1 gene; Phylogenetic analysis; Entropy