

TITLE: METAGENOMIC DATA COMPILATION REVEALS NEW EVOLUTIONARY ASPECTS OF HOLIN-INDEPENDENT LYSINS

AUTHORS: GONTIJO, M.T.P.¹; VIDIGAL, P.M.P.²; BROCCHI, M.¹

INSTITUTION: ¹ UNIVERSIDADE ESTADUAL DE CAMPINAS, CAMPINAS, SP (CIDADE UNIVERSITÁRIA ZEFERINO VAZ – BARÃO GERALDO, CEP 13083-970, CAMPINAS – SP, BRAZIL) ² UNIVERSIDADE FEDERAL DE VIÇOSA, VIÇOSA, MG (AV. PETER HENRY ROLFS, S/N - CAMPUS UNIVERSITÁRIO, CEP 36570-900, VIÇOSA – MG, BRAZIL)

ABSTRACT:

Bacteriophage-bacteria co-evolution drives phage genetic diversity. However, the evolution of the bacteriophage lysis cassette remains unclear. Aiming to unravel evolutionary aspects of the primitive holin-independent lysis system, we investigated 2,628 putative lysin genes previously identified in 183,298 genomic sequences gathered from a metagenomic compilation study. The proteins were screened using the HHpred webserver. We characterized signal peptides using SignalP and PrediSi. In addition, we screened transmembrane domains using SOSUI, TMHMM, Phobius, and Topcons. Putative lysins containing transmembrane domains comprising between 40% and 60% of Gly and Ala and up to 2 basic residues (Arg, Lys, and His) were annotated as presenting a putative SAR domain. The protein sequences were aligned using MAFFT and then the alignment was trimmed using Gblocks. We then used ProtTest to obtain the amino acid substitution model and MrBayes to construct the lysin cladogram using the Markov Chain Monte Carlo method with 5,000,000 generations and a sampling frequency of 1,000. The final tree was visualized using FigTree. The dataset comprised 379 (14.42%) holin-independent lysins: 320 proteins (12.18%) containing canonical signal peptides and 59 (2.24%) containing signal-arrest-release (SAR) sequences. Our analysis also revealed one putative multimodular holin-independent lysin containing a holin-associated domain. In addition, the data also revealed one putative lysin gene containing a spanin domain. Phylogenetic analysis showed that these multimodular proteins are found within monophyletic clusters of single-domain holin-independent lysins. These results suggest that the acquisition of genes encoding transmembrane proteins (holins and spanins) began with multimodular enzymes containing both enzymatic catalytic domains and transmembrane regions responsible for *Sec*-independent peptide export. Throughout evolution, these genes were separated as independent proteins. These data also highlight the importance of metagenomic data in the study of lysin evolutionary markers.

Keywords: bacteriophage-mediated lysis, bacteriophage-derived lytic enzyme, enzymatic, antibiotic substitute, lysozyme.

Development Agency: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).