Identification of Staphylococcus aureus secretome protein signature using logistic regression to distinguish its role in interaction with the host

Ana Carolina Barbosa Caetano[1], Sandeep Tiwari[2], Núbia Seiffert[3], Vasco a de C Azevedo[4], Thiago Luiz de Paula Castro[3],

1 Universidade Federal de Minas Gerais
2 Institute of Biological Science, Federal University of Minas Gerais, Belo Horizonte
3 Federal University of Bahia
4 Federal University of Minas Gerais

Abstract

Staphylococcus aureus is a Gram-positive pathogen and the major causing agent of mastitis in ruminants worldwide. Mastitis is often difficult to cure, leading to important losses in farm productivity. Animal isolates of S. aureus are commonly categorized into specific clonal complexes, supported by strong genetic evidences of host-dependent specialization. The extracellular proteins produced by the pathogen are known to play a role in communication with the host and comprise the arsenal used to establish infection. In this context, variant extracellular proteins are expected to be found among different groups of S. aureus. These variants are likely to be correlated with host specificity. The variety and specificity of protein structure and function depends on sequence diversity. When combined, the twenty naturally occurring amino acid residues can form 8,000 triplets and 160,000 quadruplets. A single triplet change may affect protein folding, catalytic domains, protein–ligand interactions, and protein–protein interactions. In this study, we aimed to identify amino acid triplets found in the exoproteomes of S. aureus isolated from ewe (strain O11) and bovine (RF122). Firstly, protein sequences of both strains were downloaded from National Center for Biotechnological Information (NCBI). Then, the SurfG program was used to predict the cellular localization of proteins. 91 and 96 secreted proteins were selected for strains O11 and RF122, respectively. To analyze amino acid triplets occurrence in the two strains, the logistic regression method was applied using the MATLAB software, version R2017a. As result, we found the triplets 'DQA', 'TRI', 'PVS', 'IDV' and 'DVN' in strain RR122, whereas the triplets 'MMK', 'KMK', 'MKM', 'VQA' and 'TRV' were found in O11. Furthermore, the proteins containing most of these triplets were identified. The truncated methicillin resistance-related surface protein (CAI81729.1) and the hypothetical protein tagged as EGA96785.1 were found in the RF122 and O11, respectively. The outcome of this work could facilitate in-silico functional characterization and the study of the differential interaction of the two strains with their respective hosts. We plan to include more strains from different S. aureus groups and further characterize the interaction with different hosts.
Funding: CAPES, TWAS, CNPQ, FAPEMIG