

TITLE: Comparative genomics of ST5 and ST105 lineages of methicillin-resistant *Staphylococcus aureus*

AUTHORS: Botelho, A. M. N.¹, Viana, A. S.¹, Mansoldo, F. R. P.¹, Vermelho, A. B.¹, Narechania, A.², Moustafa, A.³, Planet, P. J.^{2,3}, Ferreira-Carvalho, B. T.¹, Figueiredo, A. M. S.^{1*}

INSTITUTION: 1. Instituto de Microbiologia Paulo de Góes, Universidade Federal do Rio de Janeiro (Av. Carlos Chagas Filho, 373, Bloco I. Rio de Janeiro, RJ, Brasil, CEP: 21941-902). 2. Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA. 3- University of Pennsylvania, Children's Hospital of Philadelphia, Philadelphia, PA, USA.

ABSTRACT

Methicillin-resistant *Staphylococcus aureus* (MRSA) are major pathogens in hospital and community settings. An important characteristic of MRSA is the clonal structure of its population. MRSA clones are usually defined by MLST and SCCmec typing. The Brazilian epidemic clone [(ST239 (CC8)-SCCmecIII)] used to be the predominant MRSA in hospitals located in different Brazilian States. Nevertheless, recent studies have reported the entrance of CC5 MRSA, specially ST5 and ST105, in this country. Despite the reduced number of isolates tested thus far, CC5 seems to have supplanted ST239, in the Southeast region of Brazil. The aim of this study was to use comparative genomics to map important differences in the genomes of ST5 and ST105 MRSA, mainly in genes associated with virulence and gene regulation. One hundred and twelve strains belonging to CC5, isolated from patients admitted in hospitals from Rio de Janeiro State, were selected for whole genome sequencing, and the sequences were analyzed using bioinformatic platforms and tools. RAxML was used to generate a SNP-based phylogenetic tree for the CC5 genomes sequenced and those obtained from GenBank (n= 260). The maximum likelihood tree grouped the genomes in four clades. ST5-SCCmecIV were found exclusively in clade 1 and ST105-SCCmecII in clade 4. Genomes associated to ST5-SCCmecII could be found within all clades, suggesting more than one entrance of this SCCmec, as previously reported. The tree topology seems to indicate that ST105-SCCmecII from Brazil originated from a ST5-SCCmecII ancestor possibly from North-America. Local BLASTn was run against a database built with all 260 CC5 genomes, using the 2593 ORFs of N315 genome as query. In the analysis of polymorphisms of each ORFs that are specific for the groups under analysis (ST5 or ST105) a volcano plot was constructed, allowing mapping of ORF mutations that could have driven the divergence of the ST105 that emerged in Brazil. Concluding, this study shows that the ST105 from Brazil diverged from CC5 from USA in about 70 ORFs including those associated with virulence regulation and metabolic processes. Besides that, mutations were detected in several ORFs associated to autolysis and proteolysis, which are mechanisms recognized to interfere with the host immune response.

Keywords: Comparative genomics, Hospital infection, MRSA, Phylogenetic analysis, Whole-genome sequencing.

Development Agencies: FAPERJ, CAPES, CNPq, Fundação Bill & Melinda Gates.