Genomic Surveillance in Shiga Toxin-Producing Escherichia coli (STEC) from humans sources in Brazil

JULIANA APARECIDA CARVALHO ROSSI[1], AMANDA MARIA DE JESUS BERTANI[1], AMANDA YAEKO YAMADA[1,2], THAIS VIEIRA[1], GISELE LOZANO COSTA[1], SANDRA REGINA SCHICARIOL PINHEIRO[1], ERICA DE LIMA ORI[1], MARISA DE JESUS DE CASTRO LIMA[1], CARLOS HENRIQUE CAMARGO[1,2], MONIQUE RIBEIRO TIBA CASAS[1]

[1.] INSTITUTO ADOLFO LUTZ, CENTRAL, SÃO PAULO / SÃO PAULO, BRASIL
[2.] UNIVERSIDADE DE SÃO PAULO, FACULDADE DE MEDICINA, SÃO PAULO / SÃO PAULO, BRASIL

Shiga toxin-producing *Escherichia coli* (STEC) is one of the six diarrheagenic pathotypes of *E. coli*. STEC infections can progress to hemorrhagic colitis and Hemolytic Uremic Syndrome (HUS), potentially leading to death. Strains carrying the *eae* gene are associated with more severe cases due to their ability to induce attaching and effacing lesions and are classified as enterohemorrhagic *E. coli* (EHEC). Due to the severity of these infections, continuous epidemiological surveillance is essential to monitor the circulation of virulent serotypes and lineages, such as *E. coli* O157:H7, as well as virulence and antimicrobial resistance genes. This study aimed to perform a genetic analysis of all STEC strains available in the collection of the national reference laboratory for *E. coli* identification, to characterize virulence genes, resistance, serotypes, and phylogeny. A total of 40 human-origin *E. coli* isolates classified as STEC were included. Collected between 2010 and 2025, these isolates were initially characterized by TaqMan real-time PCR and later confirmed by whole-genome sequencing (WGS). DNA was extracted using the Wizard Genomic DNA Purification kit (Promega). WGS was performed on the Illumina NextSeq platform. Reads were trimmed and assembled using CLC Genomics Workbench v.11.0.1, and genomes were annotated using Prokka on Galaxy Europe. *In silico* serotyping and detection of resistance genes were performed using SerotypeFinder 2.0 and ResFinder 4.7.2 from the Center for Genomic Epidemiology (CGE). Virulence genes were identified using VFDB and STECFinder. MLST was performed using PubMLST (Achtman scheme). Phylogenetic analyses were based on core-genome alignment with Roary, SNP extraction with SNP-sites, a SNP distance matrix, and maximum-likelihood trees reconstructed using RAxML and visualized in Microreact. Most strains (55%) carried *stx*1, 40% carried *stx*2, and 5% carried both genes. The most frequent variants were *stx*1a (45%), *stx*2c (17.5%), and *stx*2a (12.5%). Twenty-three strains (57.5%) carried *eae* and were classified as EHEC. Additionally, 75% carried *iha*, 57.5% *ehx*A, 40% *kat*P, and 27.5% *tox*B. Eight strains (20%) carried all studied virulence genes simultaneously, strongly suggesting the presence of the pO157 plasmid. Twenty-two serotypes were identified, with O111:H8 (20%) and O157:H7 (10%) being the most frequent. Eighteen distinct Sequence Types (STs) were found, mainly ST16 (20%), associated with O111:H8, and ST11 (10%), associated with O157:H7. Five strains had undefined profiles, suggesting possible new STs. Nine strains (22.5%) were classified as MDR, mainly resistant to Sulfamethoxazole, Tetracycline, and Trimethoprim. Phylogenetic analysis showed multiple clades, indicating high genomic diversity. Several isolates clustered with international strains, mainly ST11, suggesting genetic relatedness to globally distributed lineages. These results highlight the importance of continuous genomic and epidemiological surveillance of STEC for case tracking, outbreak prevention, and

antimicrobial resistance monitoring.

Agradecimentos: