Pan-genome analysis of invasive *Streptococcus mutans* strains

Srinivasan Sujitha[1], Paramasamy Gunasekaran[2] and Jeyaprakash Rajendhran[1,*]

[1]Department of Genetics, School of Biological Sciences, Madurai Kamaraj University, Madurai 625 021, India
[2]VIT Bhopal University, Bhopal 466 114, India

Streptococcus mutans **is responsible for dental problems and is associated with cardiovascular co-morbidities. Only a few selected strains can adhere to and invade endothelial cells. To ascertain which strains have the capability to invade cardiovascular cells, *in silico* PCR was performed on all the 193 available strains. The genome sequences were screened for collagen-binding genes *cnm* and *cbm*. Among the 193 strains tested, only 4 showed the presence of collagen-binding gene. BPGA tool was used for pan-genome analysis of invasive strains. Results indicated an almost closed pan-genome for *S. mutans* comprising 45,654 core genes, 29,452 accessory genes and 232 unique genes. Most of the unique genes belonged to only 5 genomes amongst the 42 invasive genomes analysed. These five genomes were screened for the presence of virulence genes using the MP3 software. Protein–protein interactions between the pathogenic proteins and extracellular matrix components were analysed using HPIDB. Surface-localized proteins were predicted to interact with the human tumour suppressor gene.**

Keywords: Adhesion, invasive strains, pan-genome, *Streptococcus mutans*, virulent genes/proteins.

capacity to interact with the host tissue. The serotype *c* is the most abundant in natural ecosystems and is least efficient in adhering to human coronary artery endothelial cells[6]. *S. mutans* can gain transient entry into the bloodstream following routine cleaning procedure. The momentarily released bacteria find their way to the blood vessels, which have lipid depositions[7]. This chance event considerably increases the probability of bacteria attaching to endothelial cells of the host. In this study, a pan-genome analysis and genome mining were carried out to find the virulence genes associated with the invasive strains of *S. mutans*. A total of 193 genomes (including assemblies, contigs, chromosome and complete genomes) of *S. mutans* were available at the time of analysis. Figure 1 shows the workflow followed in the study.

Materials and methods

Genome categorization based on surface polymer composition

We used the FastPCR tool for finding the serotype and inva-