

A Structural bioinformatics approach for Functional characterization of *Treponema pallidum* subspecies hypothetical proteins

Arun Kumar Jaiswal, sandeep tiwari, Vasco A de C Azevedo, Siomar de Castro Soares

Universidade Federal do Triângulo Mineiro

Abstract

In the 21st century, there is yet an unsatisfactorily high worldwide rate of Sexually transmitted infections (STIs), around the globe, in excess of a million STIs are obtained each day. As per World Health Organization (WHO), in 2016 there were an approximated 376 million new cases of the four treatable STIs-Chlamydia, Gonorrhea, Syphilis and Trichomoniasis and among those 6 million cases has been accounted for just of syphilis and more than 11 million new infections of syphilis occur each year. The gram negative medically important spirochete bacterium called *Treponema pallidum* (highly virulent bacterium) subspecies *pallidum* is causative agents of Syphilis. There are three progressively known species from same Genus are causes human treponemal infections, for example, *Treponema pallidum* subspecies *pertenue* that causes yaws, *Treponema pallidum* subspecies *carateum* causes pinta and *Treponema pallidum* subspecies *endemicum* causes bejel or endemic (Nonvenereal-transmission of sickness without Sexually contact) syphilis. The molecular mechanisms mainly pathogenesis of *Treponema pallidum* are not well known. This is because *Treponema pallidum* can not be cultured in laboratory, naturally fragile behavior and phylogenetically no conventional virulence factor homologous with other pathogens makes it difficult to work with. Actually almost 30% of its proteins coding genes are not functionally known and characterized as hypothetical protein. In this work we applied a structural bioinformatics technique using Phyre2 web-based homology modelling (tertiary structure homology modelling) to better understand and annotate the hypothetical proteins based on proteome wide scale of *Treponema pallidum* subspecies level. In this work we used genome of each subspecies (complete genomes of *Treponema pallidum* subspecies *pallidum*, *Treponema pallidum* subspecies *endemicum* and *Treponema pallidum* subspecies *pertenue*). We found 297 (30%) of protein coding genes were hypothetical in *Treponema pallidum* subspecies *pallidum* strain Nichols and after comparing these hypothetical protein coding genes with *Treponema pallidum* subspecies *endemicum* and *Treponema pallidum* subspecies *pertenue*, 253 (26%) of protein coding genes were common. The 297 out of 969 *Treponema pallidum* subspecies *pallidum* strain Nichols protein modeled with Phyre2-based tertiary structure modeling with high-confidence score which were assigned as hypothetical proteins with no functions in published proteomes. Hypothetical proteins modeled in this work with high-confidence were predicted that showed remarkable structural similarity with proteins that experimentally confirmed to be required for virulence in other pathogens. Significantly, our tertiary structure modeling approach was also able to predict structural models based on functionally annotated templates for over of all hypothetical *T. pallidum* proteins, which will help to better understand the structure-function relationships and fundamental molecular mechanisms of *T. pallidum* pathogenesis at subspecies level.

Funding: