Available online at www.elixirpublishers.com (Elixir International Journal)

Bio Technology

Elixir Bio Tech. 69 (2014) 23105-23110

ISSN: 2229-712X

Comparative Genome Analysis of *Streptococcus pyogenes* and *Streptococcus equi*

Rohit Lall[1], Nitendra[1] and Satendra Singh[2]

[1]Department of Molecular and Cellular Engineering, JSBB, SHUATS, Allahabad-211007,

[2]Department of Computational Biology & Bioinformatics, JSBB, SHUATS, Allahabad-211007.

ARTICLE INFO

Article history:

Received: 4 February 2014,

Received in revised form:

29 March 2014;

Accepted: 10 April 2014;

Keywords

UTR,

CNS,

Childhood meningitis.

ABSTRACT

The comparative genome analysis of *Streptococcus pyogenes* and *Streptococcus equi* sps equi *4047* depict exon, Untranslated Regions (UTR), conserved noncoding sequences (CNS), contigs, mobile genetic elements, Insertion sequence (IS) elements, transposons, plasmids and shows high degree of heterogeneity evidenced by the large number of single nucleotide polymorphisms (SNPs). Thus UTR, CNS, and IS elements can be used as drug target.

Introduction

The *Streptococcus pyogenes 370* [4] is a gram positive [2, 3, 12], nonmotile, facultative, mesophilic obligate parasitic bacterium with cocci in shape. The genome is circular with a size of 1,852,442 base pairs and an average G+C content of 38.5%. The average G+C content of the protein-coding sequences is 39.1%. The bacteriophage genome contains SLP, PAG encoding virulence factors. The transcription starts in both directions from *oriC* [8]. The presence of alternate transcription signals allows the streptococcus to respond to environmental changes [1, 7]. The salivaricin A (bacteriocin) was present in 90% of *S. pyogenes* strains [13]. The major virulence regulon was controlled by Mga and located upstream to immunogenic secreted protein gene, *isp* [11]. The *Streptococcus pyogenes MIGAS* contains 1,752 protein encoding genes (PEG), 40 putative virulence associated genes, prophage associated genes (PAG), superantigens like proteins (SLP) [9] causing toxic shock syndrome, cellulitis, and rheumatic fever [15-18]. The six potential virulence factors are responsible for horizontal gene transfer (HGT) and generate new strains with increased pathogenic potential. *S.equi*.subsp.equi. *4047* cause childhood meningitis [5] due to rupture of abscesses in retropharyngeal lymph nodes. The functional loss, pathogenic specialization, and genetic exchange results in evolution [10] of *S.equi*.subsp.equi. *4047*.

Comparative genome analysis added value of complete genomes in more sequences, large scale pattern detection in genomes, function/orthology prediction by bi-directional best hit approaches, presence/absence/variation of pathways and prediction of new pathway. Analysis and Comparison of genomes at various levels of DNA (e.g. GC content) (we actually do not need sequencing for that), dinucleotide frequencies, coding densities of leading/lagging strands, GC skew etc.), protein coding potential (e.g. coding density), presence/absence/size of Protein families, presence/absence of genes/comparing at the level of orthologs and Gene Order evolution. Evolution of gene content is identified by Quantitative approaches. Count the number of genes that two genomes share (orthology) and relate to phylogenetic distance. The rate of gene content reconstructs genome evolution. Qualitative approaches: interpret the differences between two genomes in terms of the functions of the encoded proteins to explain the differences between the phenotypes in terms of the genomes' gene content. Genome phylogeny is based on gene content. Count the number of shared orthologs between genomes using bi-directional best, significant, hit approach (include fusion/fission). Create a similarity matrix by dividing number of shared orthologs by the genome size of the smallest genome. Create a distance based phylogeny from the similarity matrix [14].

Materials and methods

Comparative analysis of DNA sequences was done by selecting *Streptococcus pyogenes* as reference organism and *Streptococcus equi* as compared organism in microbial vista. The microbial vista displays conserved region identity alignment, synteny viewer, and dot plot. The conserved region describes the exon, UTR, conserved noncoding sequences, contigs, and protein coding region alignment between *Streptococcus pyogenes* and *Streptococcus equi*.

The synteny viewer represents the alignment density at particular location in chromosomes. The dot plot describes the chromosome boundaries/ scaffold, the blue color line describes the forward strand and red color line describes the opposite strand. The grey color predict the *Streptococcus pyogenes* chromosome and colored as *Streptococcus equi* chromosome.

The VISTA tools allows to align DNA sequences, quickly visualize conservation levels between them, identify highly conserved regions, and analyze sequences of interest through one of the following approaches: Precomputed whole-genome sequences of reference and compared organisms were browsed and submitted to Genome VISTA or mVISTA to align them with each other (a variety of alignment programs with several distinct capabilities are made available). The web page http://genome.lbl.gov/vista/ serves as a portal for access to all VISTA tools.

Results and discussion

The visual comparative analysis of *Streptococcus pyogenes* and *Streptococcus equi* genome assemblies were done at

Tele:

E-mail addresses: nitendra_20011@rediffmail.com