

Mycobacteriophage Genome Database

Jerrine Joseph, Vasanthi Rajendran, Sameer Hassan, Vanaja Kumar*

Department of Bacteriology, Tuberculosis Research Centre (Indian Council of Medical Research), Mayor VR Ramanathan road, Chetput, Chennai - 600 031, India; Vanaja Kumar - Email: vanajakumar@trcchennai.in; Phone: +91-44-28369659; Fax: +91-44-2836 2528; *Corresponding author

Received November 12, 2010; Accepted January 06, 2011; Published August 02, 2011

Abstract:

Mycobacteriophage genome database (MGDB) is an exclusive repository of the 64 completely sequenced mycobacteriophages with annotated information. It is a comprehensive compilation of the various gene parameters captured from several databases pooled together to empower mycobacteriophage researchers. The MGDB (Version No.1.0) comprises of 6086 genes from 64 mycobacteriophages classified into 72 families based on ACLAME database. Manual curation was aided by information available from public databases which was enriched further by analysis. Its web interface allows browsing as well as querying the classification. The main objective is to collect and organize the complexity inherent to mycobacteriophage protein classification in a rational way. The other objective is to browse the existing and new genomes and describe their functional annotation.

Keywords: Mycobacteriophages, genome, annotation, database, MySQL, PHP

Availability: <http://mpgdb.ibioinformatics.org/mpgdb.php>

Background:

Mycobacteriophage genomes were chosen because of their abundance, their wide size distribution and their very compact genome organization. This enables them to be the ideal choice to function as genetic tool in tuberculosis research. All these features were expected to raise most of the technical challenges to be solved in building up the database. Despite their importance, little is known about the genomic diversity harbored in phages. Methods to capture complete phage genomes have been hampered by the lack of knowledge about the target genomes [1]. The sheer abundance and importance of mycobacteriophages coupled with knowledge of their genetic makeup, demands establishing comprehensive annotation methods that can be applied. This wealth of information helps to decipher the genetic framework that drives phage biology, thus providing a window into understanding how these important organisms modulate and by extension, impact human health.

Database Content:

The present version of the mycobacteriophage genome database (version 1.0) contains 6086 genes from 64 mycobacteriophages. All the sequence data used has been downloaded from the NCBI genomes section (<http://www.ncbi.nlm.nih.gov/genomes>). As a first step towards building a comprehensive mycobacteriophage functional classification of the 64 sequenced genomes, it has been subjected to several servers like PFAM [2], PROSITE [3], CATH [4], SCOP [5] and so on for annotation. The proteins from 64 mycobacteriophages were clustered into families based on ACLAME classification [6]. The output of the analysis resulted in 72 clusters covering a total of 2082 proteins representing 34.7% of all analyzed proteins.

Methodology:

Tool Design and implementation:

The system used is MySQL database to store gene and protein sequence annotated data. Programs are written in PHP enabled database search using keywords like 'Genome name' or 'mycobacteriophage genes'. The database facilitates to query and retrieve results in 'Text' or 'Table' or 'Graphical'

formats as 'Description', 'Summary', 'predicted results' and 'graphs' respectively. Detailed sequence data can be obtained from the corresponding hyperlinks. Data flow diagram for MGDB is represented in the **Figure 1**. The browsing and query web pages are shown as snapshots in **Figure 2**.

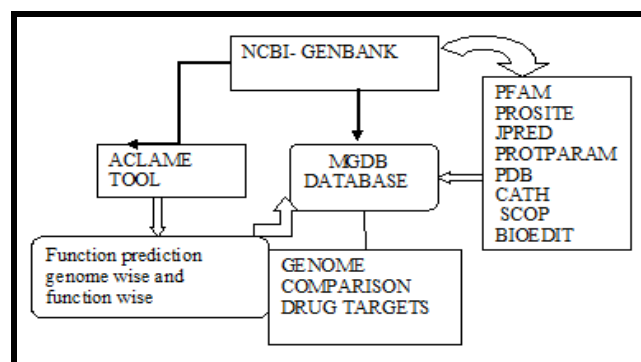


Figure 1: Data flow diagram for mycobacteriophage genome database.

Future Development:

The database will be periodically updated and enriched with more features to make it further interactive and user friendly. Tools like BLAST [7] and CLUSTAL W [8] are to be incorporated in the database, to better equip the mycobacteriophage research community. Phage genome comparison and newly identified drug targets of *M. tuberculosis* will also be updated in the future. While these resources are perused by a great number of the mycobacteriophage research community, we will undertake the initiatives to acquire curate and enhance the content of this database in service to the wider research community.

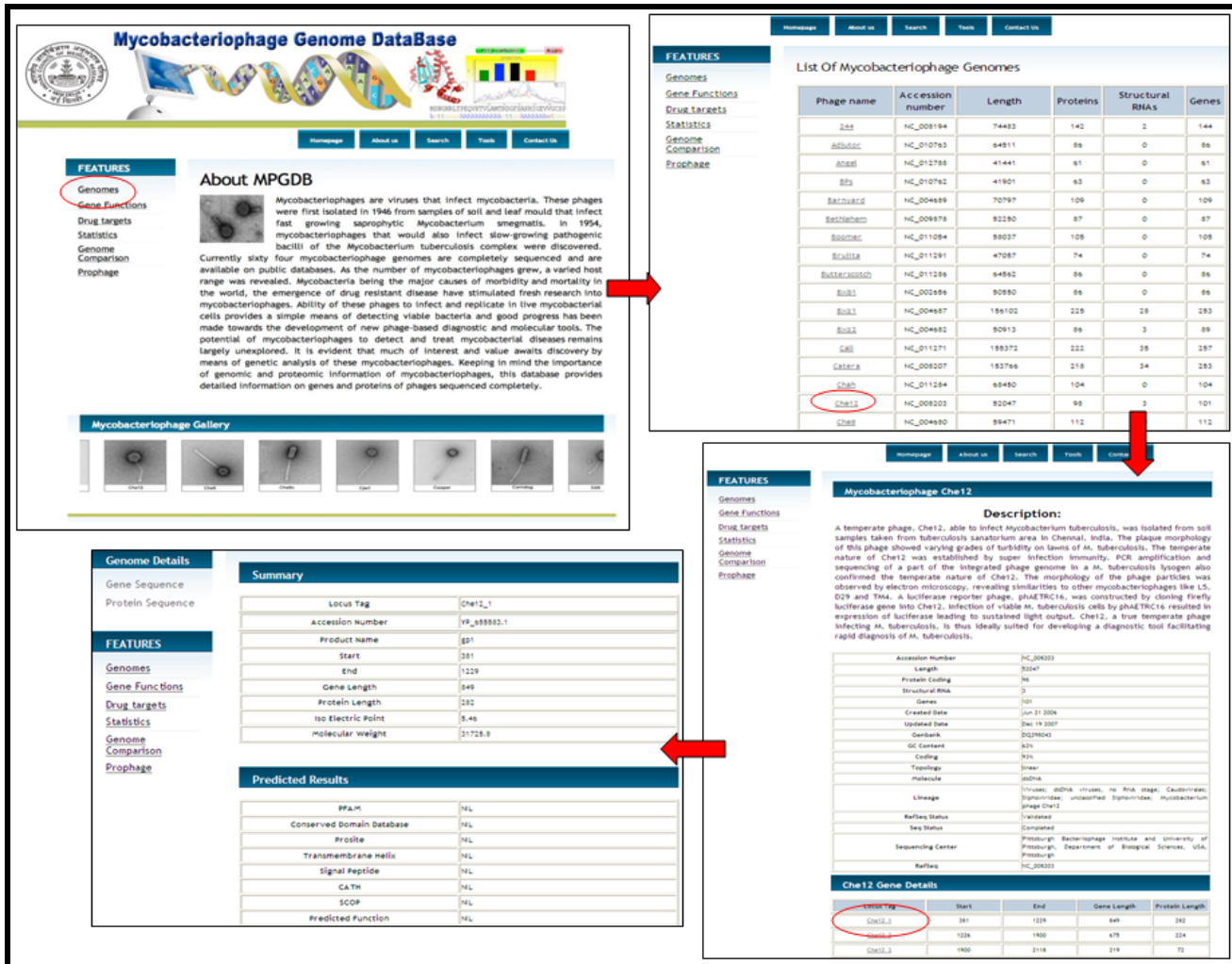


Figure 2: Mycobacteriophage genome database query and browser sample web pages.

Acknowledgement:

The authors wish to acknowledge the Department of Biotechnology (DBT) for their financial support for this work.

References:

[1] Henn MR *et al. PLoS One.* 2010 **5**: e9083 [PMID: 20140207]
 [2] Sonnhammer EL *et al. Nucleic Acids Res.* 1998 **26**: 320 [PMID: 9399864]

[3] de Castro E *et al. Nucleic Acids Res.* 2006 **34**: W362 [PMID: 16845026]
 [4] Orengo CA *et al. Structure* 1997 **5**: 1093 [PMID: 9309224]
 [5] Murzin AG *et al. J Mol Biol.* 1995 **247**: 536 [PMID: 7723011]
 [6] Leplae R *et al. Nucleic Acids Res.* 2004 **32**: D45 [PMID: 14681355]
 [7] Altschul SF *et al. Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
 [8] Higgins DG *et al. Comput Appl Biosci.* 1992 **8**: 189 [PMID: 1591615]

Edited by W Cuff

Citation: Joseph *et al. Bioinformatics* 6(10): 393-394 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.