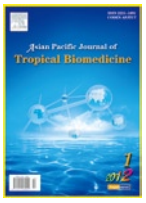
Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Biomedicine

journal homepage:www.elsevier.com/locate/apjtb



Document heading

Functional assignment of the 64 mycobacteriophages into gene families

Jerrine Joseph, Sameer Hassan, Vasanthi Rajendran,Vanaja Kumar

Tuberculosis Research Center, Chennai, India

ARTICLE INFO

Article history:
Received 23 December 2011
Received in revised form 2 January 2012
Accepted 15 February 2012
Available online 28 April 2012

Keywords:
ACLAME
Mycobacteriophage
BLAST
Genome

ABSTRACT

Objective: To functionally annotate the protein sequences for 64 mycobacteriophage genomes using the ACLAME database and to discuss how structured information on mycobacteriophage encoded proteins helps global in silico analysis. **Methods:** Each of the protein sequences of 64 mycobacteriophage genomes were searched against ACLAME database Version 0.4 using BLAST and were further classified based on function. The default setting of ACLAME 0.4 BLAST was used. **Results:** Based on ACLAME, assignment of 72 protein functional families was done for the 64 sequenced genomes, and families were compared among and across genomes. **Conclusions:** The finding of the present study provides scope for designing novel approaches to decipher this huge chunk of unknown proteins by experimental methods.

1. Introduction

The mycobacteriophages are a remarkable diverse group of viruses whose characterization has provided helpful insights into the mosaic nature of mycobacteriophage genomes and the evolutionary mechanisms that give rise to them. The observed diversity is especially intriguing given that all these phages infect a common host, although we speculate that they have different but overlapping host ranges which are reflected in their different GC% contents[1]. Nevertheless, with more than 60 complete mycobacteriophage genome sequences available it is clear that we are far from saturating the knowledge mycobacteriophage types or the knowledge of mycobacteriophage genes. Further expansion of our knowledge of mycobacteriophage genomics is thus expected to advance our understanding of their diversity and evolution[2]. But it is becoming increasingly clear that our insights into mycobacteriophage genomes are constrained by our lack of understanding of the functions of more than 1 000 new sequence families, and we hope that functional genomic approaches will help to provide answers to some of the questions regarding what these mycobacteriophage genes do[3]. In the absence of experimental data, the function of a protein is usually inferred on the basis of its sequence similarity to a protein of known function. Owing to the exhaustive data available, there is a revived interest in their sequencing. 184 phage genomes were available on the NCBI site in 2003, but they increased to 306 in 2006 and a sudden surge of 579 was observed in 2010.

The ACLAME 0.4 database aims at providing a reticulate classification of the prokaryotic 'mobilome', *i.e.*, all the proteins encoded by prokaryotic mobile genetic elements (MGEs), whether extrachromosomal or integrated in the host genome[5]. ACLAME exploits the idea that groups together different types of MGEs sharing identical functions (*e.g.* replication and site-specific recombination, transposition, conjugation, *etc.*) within a single repository, and optimizes the use of experimental evidence available on any element to support the functional annotation of another element[6].

In the present study we attempted to predict the function and group all the proteins of the 64 mycobacteriophages based on ACLAME database which currently holds information on 10 mycobacteriophage genomes only.

2. Materials and methods

Each of the protein sequences was searched against ACLAME database Version 0.4 using BLAST 7 and was further classified based on its function. The default setting of ACLAME 0.4 BLAST was used.

*Corresponding author: Dr. Vanaja Kumar, Scientist F, Department of Bacteriology, Tuberculosis Research Centre, Mayor VR Ramanathan Road, Chetput, Chennai - 600 031, India.
Tel: +91-44-2836 9659
Fax: +91-44-2836 2528
E-mail: vanaja_kumar51@yahoo.co.in

3. Results

The sequencing of more than 60 mycobacteriophage genomes has triggered the interest to map the functions of their genes. The findings throw insight into their diversity and evolution. Here we discuss how structured information on mycobacteriophage encoded proteins helps global in silico analysis. Based on ACLAME, assignment of 72 protein functional families was done for the 64 sequenced genomes and families were compared among and across genomes. Based on the current analysis, 72 unique functional families were obtained from 64 mycobacteriophages as shown in Table 1 depicting each individual copy number of genes for each function per genome and the number of genomes having the same function. The highest copy number was found in phage tail fibre assembly gene with 141 copy numbers, followed by phage DNA replication/helicase gene with 99 copy numbers and phage release by lysis gene with 94 copy numbers. These three genes were found to be present in all 64 phage genomes studied. Among these genomes, few of them have more than one copy of the gene invariably. The lysis gene is present in 45 genomes only. The 72 assigned functions consist of 1 143 proteins out of 6 000 genes of 64 phage genomes. This leaves a large chunk of protein families without being assigned functions. The remaining unknown families need to be characterized which pose great challenge to phage research. We further grouped the 72 phage families into 8 broad categories such as DNA metabolism, phage tail proteins, phage fibres and fibres assembly, holins and endolysins, site-specific recombinases, phage head /capsid proteins, transcriptional regulation and post -transcriptional regulation(Figure 1). This gives a holistic view of the functional classification of proteins involved in phage life cycle. The diversity and mosaicism was indeed baffling as well as posing more questions about their trend in course of evolution. Functional annotation based on ACLAME protein families allows for a more robust annotation of phage proteins and will greatly help to answer the intriguing queries of mycobacteriophage genomes.

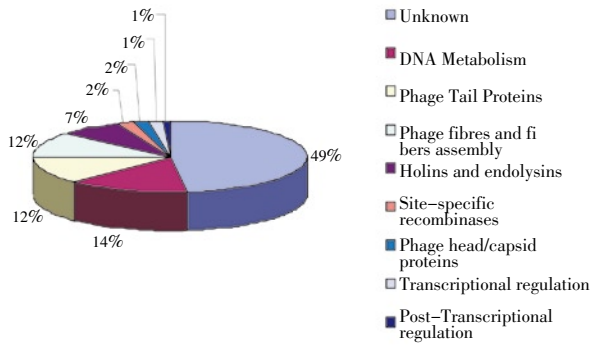


Figure1. Distribution of the main functional categories in mycobacteriophages.

Table 1
The consolidated list of 72 functional families for the 64 genomes.

List of mycobacteriophage genome functions	No. of copy numbers present across 64 genomes	No. of genomes having same function across 64 genomes
Phage tail fiber assembly	141	64
Phage DNA replication/helicase	99	62
Phage release by lysis	94	45
Phage tail tape measure protein	63	46
Phage terminase large subunit	45	44
Phage DNA maturation	47	45
Phage head tail joining	25	25
Phage minor tail protein	37	37
Phage endolysin activity	43	41
Phage portal protein	27	27
Phage head/capsid major protein	28	24
Phage prohead/capsid assembly	29	28
Deoxyribonuclease I activity	34	25
Crossover junction endodeoxyribonuclease activity	14	14
Phage major tail protein	19	19
Holin activity	16	13
Tyrosine-based site-specific recombinase activity	15	15
Transcription factor activity	22	15
Phage scaffolding protein	23	22
Prohead protease activity	10	8
DNA replication proofreading	10	10
Exonuclease activity	10	10
DNA strand annealing activity	10	10
DNA recombination	12	12
DNA-directed DNA polymerase activity	16	16
Transcription repressor activity	11	11
Thymidylate synthase activity	14	12
5′-3′ exonuclease activity	13	10
Homologous DNA recombination	10	10
Holliday junction resolvase complex	11	11
DNA-methyltransferase activity	21	21
Polypeptide N-acetylgalactosaminy ltransferase activity	11	11
Site-specific DNA inversion	7	7
Prophage DNA integration	20	20
Serine-based site-specific recombinase activity	3	3
DNA primase activity	7	7
Thioredoxin	7	7
Integral to membrane	6	6
Thiol-disulfide exchange intermediate activity	4	4
Phage entry exclusion	5	5
Transcriptional repressor activity	2	2
Negative transcriptional regulation of phage gene expression	7	7
3′-5′-exodeoxyribonuclease activity	3	3
DNA-methyltransferase activity	2	2
Oxidoreductase activity	2	2
IS3 family	7	7
Single-stranded DNA binding	3	3
Serine-type peptidase activity	6	6
DDE transposase activity	6	6

Table 1, continued

List of mycobacteriophage genome functions	No. of copy numbers present across 64 genomes	No. of genomes having same function across 64 genomes
ATP-dependent polydeoxyribonucleotide 5′-hydroxyl-kinase activity	8	8
Ribonucleoside-triphosphate reductase activity	3	3
ATP-dependent DNA helicase activity	8	8
Phage tail assembly	8	8
Phage tail chaperon activity	8	8
Peroxidase activity	3	3
Phage terminase small subunit	3	3
DNA binding activity	7	7
6-pyruvoyltetrahydropterin synthase activity	2	2
GTP cyclohydrolase activity	2	2
Excisionase activity	2	2
ATPase activity	1	1
Peptidase activity	1	1
Toxin activity	1	1
Post-segregational killing	1	1
NDNA-directed DNA polymerase activity	1	1
Hydrolase activity	1	1
Gueuosine biosynthetic process	1	1
Nicotinate phosphoribosyltransferase activity	1	1
Transpositional DNA recombination	1	1
IS607 family	1	1
Negative transcriptional regulation of phage gene expression	1	1
Transcriptional-regulation of phage gene expression	1	1

The accession number for the 64 phage genomes are as follows:

NC_008194, NC_010763, NC_012788, NC_010762, NC_004689, NC_009878, NC_011054, NC_011291, NC_011286, NC_002656, NC_004687, NC_004682, NC_011271, NC_008207, NC_011284, NC_008203, NC_004680, NC_004683, NC_004686, NC_004681, NC_008195, NC_004685, NC_001900, NC_011022, NC_011288, NC_009993, NC_011290, NC_008202, NC_011020, NC_011019, NC_011292, NC_011056, NC_001335, NC_008196, NC_011021, NC_011273, NC_011044, NC_004688, NC_008197, NC_008198, NC_005259, NC_008200, NC_008205, NC_011287, NC_011057, NC_012027, NC_008199, NC_011055, NC_011039, NC_011023, NC_008204, NC_011289, NC_011272, NC_004684, NC_011269, NC_011267, NC_011270, NC_003387, NC_011285, NC_009820, NC_009877, NC_008206, NC_013650, NC_012027.

4. Discussion

Functional assignment exclusively for the 6 000 mycobacteriophage proteins revealed that the proteins with known functions are miniscule compared to the unknown. The phage DNA metabolism constituted the major chunk of the assigned portion which means it is needless to say that they are critical in the phage lifecycle. Their role can be extrapolated for the survival tactics of the phages over generations. The phage tail fibre and tail assembly genes also give a hint as they could be the harboring factors like adhesives or spikes or receptors that are critical for adhesion to the host cell wall. Hence their pan presence per se across the genomes speaks volumes of their role which needs to be elucidated in detail. The finding of the present study provides scope for designing novel approaches to decipher this huge chunk of unknown proteins by experimental methods. This will further pave the way to better understanding of mycobacteriophage biology.

In conclusion, functional assignment exclusively for the 64 mycobacteriophage proteins revealed that the proteins with known function are miniscule compared with the unknown. The finding of the present study provides scope for designing novel approaches to decipher this huge chunk of unknown proteins by experimental methods, which paves the way for better understanding of mycobacteriophage biology.

Conflict of interest statement

We declare that we have no conflict of interest.

References

[1] Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Weber RJ, et al. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J Mol Biol* 2010; **397**(1): 119-143.

[2] Hatfull GF, Cresawn SG, Hendrix RW. Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. *Res Microbiol* 2008;**159**(5): 332-339.

[3] Hatfull GF. Mycobacteriophages: genes and genomes. *Ann Rev Microbiol* 2010; **64**: 331-356.

[4] Leplae R, Hebrant A, Wodak SJ, Toussaint A. ACLAME: a classification of mobile genetic elements. *Nucleic Acids Res* 2004; **32**: D45-D49.

[5] Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* 2008; **24**(6): 863-865.

[6] Toussaint A, Lima-Mendez G, Leplae R. PhiGO, a phage ontology associated with the ACLAME database. *Res Microbiol* 2007; **158**(7): 567-571.

[7] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**(17): 3389-3402.