In silico dissection of Type VII Secretion System components across bacteria: New directions towards functional characterization

CHANDRANI DAS, TARINI SHANKAR GHOSH and SHARMILA S MANDE*

Bio-Sciences R&D Division, TCS Innovation Labs, Tata Research Development & Design Centre, Tata Consultancy Service Ltd., Pune 411 013, India

**Corresponding author (Email, sharmila@atc.tcs.com)*

Type VII Secretion System (T7SS) is one of the factors involved in virulence of *Mycobacteriun tuberculosis* H37Rv. Numerous research efforts have been made in the last decade towards characterizing the components of this secretion system. An extensive genome-wide analysis through compilation of isolated information is required to obtain a global view of diverse characteristics and pathogenicity-related aspects of this machinery. The present study suggests that differences in structural components (of T7SS) between Actinobacteria and Firmicutes, observed earlier in a few organisms, is indeed a global trend. A few hitherto uncharacterized T7SS-like clusters have been identified in the pathogenic bacteria *Enterococcus faecalis*, *Saccharomonospora viridis, Streptococcus equi*, *Streptococcuss gordonii* and *Streptococcus sanguinis.* Experimental verification of these clusters can shed lights on their role in bacterial pathogenesis. Similarly, verification of the identified variants of T7SS clusters consisting additional membrane components may help in unraveling new mechanism of protein translocation through T7SS. A database of various components of T7SS has been developed to facilitate easy access and interpretation of T7SS related data.

[Das C, Ghosh TS and Mande SS 2016 *In silico* dissection of Type VII Secretion System components across bacteria: New directions towards functional characterization. *J. Biosci.* **41** 133–143] **DOI** 10.1007/s12038-016-9599-8

1. Introduction

Type VII Secretion System (T7SS) was first identified in *Mycobacterium tuberculosis* H37Rv, the etiological agent of chronic human disease tuberculosis (Stanley *et al.* 2003). The genome of *M. tuberculosis* H37Rv has been reported to contain five 'gene clusters' (T7SS clusters), each encoding at least the core components of T7SS (Abdallah *et al.* 2007). The secreted proteins of T7SS have been referred to as 'WXG100' proteins since these proteins are of approximately 100 amino acid (aa) length and contain WXG100 motif (Abdallah *et al.* 2007). Several studies have suggested the involvement of the secreted proteins (like ESAT-6, CFP-10) in virulence (Stanley *et al.* 2003; Guinn *et al.* 2004). Protein secretion through T7SS has also been experimentally proven in other mycobacterial species, which include *Mycobacterium smegmatis, Mycobacterium leprae* and *Mycobacterium marinum* (Geluk *et al.* 2002; Converse and Cox 2005; Tan *et al.* 2006). Apart from mycobacteria, variants of T7SS have also been found in non-mycolata species under the phyla Actinobacteria, *Streptomyces coelicolor* (Akpe San Roman *et al.* 2010). The corresponding gene cluster was observed to contain a non-mycobacterial gene *bldB*, in addition to the homologs of seven genes of the classic mycobacterial cluster.

Apart from those in phylum Actinobacteria, systems transporting secretory proteins containing 'WXG' motif were also experimentally identified in genera Bacillus, Staphylococcus and Listeria, all belonging to the phylum Firmicutes (Sao-Jose *et al.*

Keywords. Gene cluster; HMM profile; *Mycobacterium tuberculosis* H37Rv; pathogenicity; Type VII Secretion System

Abbreviations used: HMM, Hidden Marcov Model; T7SS, Type VII Secretion System; T7SSC, Type VII Secretion System Component

Supplementary materials pertaining to this article are available on the Journal of Biosciences Website.

Published online: 12 February 2016

2004; Burts *et al.* 2005; Way and Wilson 2005; Garufi *et al.* 2008; Huppert *et al.* 2014). In addition to a few genes that are not present in the T7SS cluster of mycobacteria, interestingly the gene clusters encoding T7SS in Firmicutes were observed to contain only two gene families (encoding WXG100 proteins and ATPases of FtsK family), which are common to mycobacterial T7SS clusters. Some of the components of T7SS have also been computationally predicted in species belonging to Proteobacteria, Chloroflexi and Planctomycetes (Gey Van Pittius *et al.* 2001; Pallen 2002; Sutcliffe 2011). Such wide distribution of T7SS and its significance in virulence necessitates a comprehensive genome-wide characterization of the various classes of T7SS-like secretion systems across all bacterial lineages. Although *in silico* analyses of some of the components have been performed by a few earlier studies (Gey Van Pittius *et al.* 2001; Pallen 2002; Das *et al.* 2011; Sutcliffe 2011), an extensive analyses of T7SS components across all the completely sequenced bacterial genomes is expected to help in better understanding diverse characteristics of this machinery and its role in pathogenesis.

The present study focuses on comprehensive genome-wide analyses of the T7SS components across all bacterial clades. In this study, available data corresponding to the experimentally validated secretory, structural and regulatory components of T7SS in Actinobacteria and Firmicutes have been utilized. Genome-wide mapping and classification of probable T7SS gene clusters in various bacterial lineages have been performed utilizing these components as 'seeds'. Information on the identified components of T7SS has been organized into a database T7CD (Type VII Secretion System Component Database). The present study suggests a global trend of distinct structural components (of T7SS) in Actinobacteria and Firmicutes. In addition, we propose association of T7SS to some membrane proteins, which have not been characterized earlier as parts of T7SS. Further, the present study predicts hitherto uncharacterized T7SS-like clusters in a few pathogenic bacteria like *E. faecalis*, *S. viridis*, *S. equi*, *S. gordonii* and *S. sanguinis*.

2. Methods

The workflow used for identification and analyses of T7SS components across different bacterial phyla is illustrated in figure 1.

2.1 *Construction of 'model Type VII Secretion System clusters'*

All the experimentally identified T7SS clusters were collated and referred to as 'model T7SS clusters' (schematically represented in supplementary figure 1 of supplementary material 1). These included six clusters corresponding to *M. tuberculosis* H37Rv (as the representative of mycobacterial clusters), *S. coelicolor*, *Bacillus subtilis*, *Bacillus anthracis*, *Streptococcus aureus* and *Listeria monocytogenes* (Sao-Jose *et al.* 2004; Burts *et al.* 2005; Way and Wilson 2005; Abdallah *et al.* 2007; Garufi *et al.* 2008; Akpe San Roman *et al.* 2010).

2.2 *Identification of Type VII Secretion System components (T7SSC) across bacteria*

The homologs of T7SS components were identified through sequence comparison using BLASTp (Altschul *et al.* 1990) and functional domain search using Pfam database (Finn *et al.* 2014). The seed components used for these two approaches were all the T7SS components (total 95) from the 'model T7SS clusters'. Based on gene type/family, these components were divided into 23 groups. For simplicity, the 23 components (associated with T7SS) have been referred to as T7SSC1 (Type VII Secretion System Component-1) through T7SSC23 (supplementary table 1 of supplementary material 1).

The first step for identification of homologs has been schematically shown in supplementary figure 2 of supplementary material 1. The initial homolog identification of 23 T7SS components was performed using Blastp with an e-value of 10^{-4}. The homologs obtained were further filtered based on sequence identity and sequence coverage to reduce the number of false positives. An identity cut-off of 30% has previously been shown to be a fair estimator for identification of homologous protein sequences (Pearson 2013). In addition, members of some of the components of T7SS have been reported to share a moderate or low sequence similarity at protein level (Gey Van Pittius *et al.* 2001; Abdallah *et al.* 2007; Houben *et al.* 2014). For example, members of T7SSC1 from *M. tuberculosis* H37Rv were shown to have sequence similarity of around 30% - 60% to those from *Corynebacterium diphtheriae* and *S. coelicolor* (Gey Van Pittius *et al.* 2001). Again, two members of T7SSC8 from *M. smegmatis* have been shown to share an identity of ~45% (Houben *et al.* 2014). In light of the above observations, a relatively stringent sequence identity threshold of 50% was applied in the present study for the purpose of identifying close homologs of T7SS components in different bacterial species. Further, distant homologs were identified based on protein domains as described in the paragraph below. A criterion of high sequence coverage threshold of 80% was employed in combination with sequence identity threshold in order to further refine the search results.

The second step for identification of homologs of T7SS components has been schematically shown in supplementary figure 3 of supplementary material 1. The second approach for identification of T7SS components across bacteria was by mapping protein domains using 'Pfam' database (Finn *et al.* 2014). This approach ensured identification of protein sequences containing relevant functional domains, which may have escaped the first approach due to low global

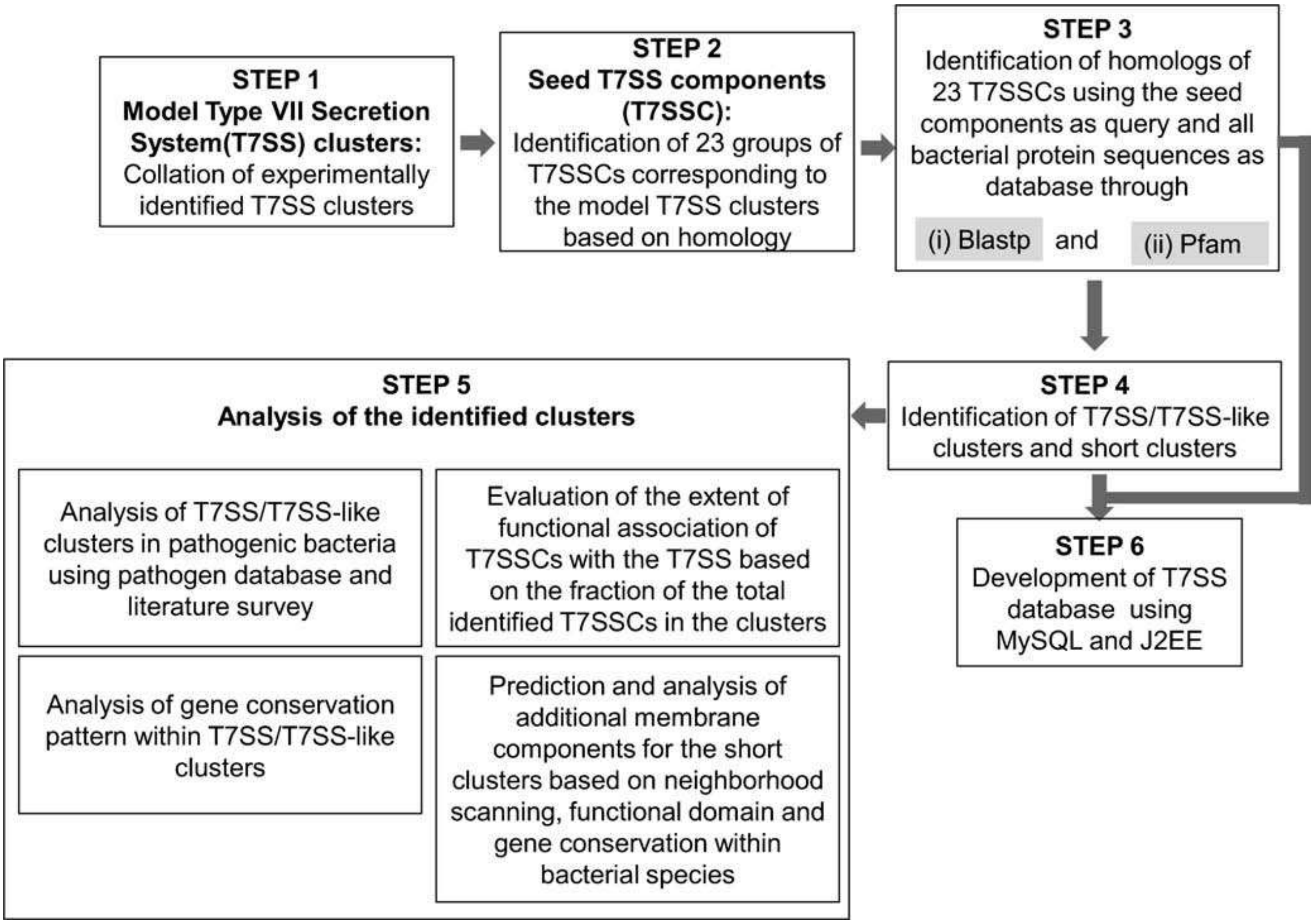
STEP 1
Model Type VII Secretion System(T7SS) clusters:
Collation of experimentally identified T7SS clusters
STEP 2
Seed T7SS components (T7SSC):
Identification of 23 groups of T7SSCs corresponding to the model T7SS clusters based on homology
STEP 3
Identification of homologs of 23 T7SSCs using the seed components as query and all bacterial protein sequences as database through
(i) Blastp and (ii) Pfam
STEP 4
Identification of T7SS/T7SS-like clusters and short clusters
STEP 5
Analysis of the identified clusters
Analysis of T7SS/T7SS-like clusters in pathogenic bacteria using pathogen database and literature survey
Evaluation of the extent of functional association of T7SSCs with the T7SS based on the fraction of the total identified T7SSCs in the clusters
Analysis of gene conservation pattern within T7SS/T7SS-like clusters
Prediction and analysis of additional membrane components for the short clusters based on neighborhood scanning, functional domain and gene conservation within bacterial species
STEP 6
Development of T7SS database using MySQL and J2EE

Figure 1. Overall methodology followed for identification and analysis of Type VII Secretion System (T7SS) clusters and short clusters.

sequence similarity with the query sequences. Pfam is a database of protein families, which are defined by groups of protein sequences having similar domains. Each protein domain is stored as a profile Hidden Markov Model ('HMM profile'). In the current study, the protein domains present in the 23 types of seed components (T7SSC1 - T7SSC23) were mapped and were utilized to search similar domains in all completely sequenced bacteria. The domains present in the seed components have been listed in supplementary table 2 of supplementary material 1. For the search strategy to be implemented, the query set was built with the protein sequences of all completely sequenced bacteria and the database was constructed with a set of domain profiles ('HMM profiles'). Details on the construction of this database have been explained below.

In the first step, the 'HMM profiles' corresponding to the domains present in the 23 types of seed components, were retrieved. However, out of 23, only 16 types of seed components have been found to have domains defined in Pfam database. When queried with other seven types of seed components (T7SSC13, T7SSC14, T7SSC18, T7SSC19, T7SSC20, T7SSC22 and T7SSC23), no significant domain hits could be retrieved. While retrieving further information on the domains identified for the 16 T7SSCs (from Pfam database), it was observed that, some of these domains may co-occur with other domains, which have not been associated to T7SS yet. For example, all the seed components corresponding to T7SSC1 contain a single domain called 'WXG100' (Pfam id: PF06013). However, according to the current version of Pfam database, the 'WXG100' domain may co-occur with several other domains (like Colicin-DNase, SLT, CHAP, Peptidase_C2, etc.) not associated to T7SS. Thus, the relevant domain architecture for T7SSC1 was considered to be only WXG100 domain. Similarly, the relevant domain architectures were identified for all the 16 types of seed components (supplementary table 2 of supplementary material 1). Subsequently, a database was created with both the relevant and co-occurring domains (not associated to T7SS) of these 16 types of seed components. The co-occurring domains were included to map false homologs containing any unwanted domain along with relevant domain(s). For the rest seven types of seed components (T7SSC13, T7SSC14, T7SSC18, T7SSC19, T7SSC20, T7SSC22 and T7SSC23) with undefined 'HMM profiles' in Pfam database, 'HMM profiles' were built using the program 'hmmbuild' provided in the standalone version of 'HMMER' package (*http://hmmer.janelia.org/*) (Eddy 1998). These 'hmm profiles' were appended to the initial database.

Once the database of 'HMM profiles' was created, protein sequences of completely sequenced bacteria were queried against the database using the program 'hmmscan' provided in the standalone version of 'HMMER' package (Eddy 1998) with an e-value of e^{-5}. The result obtained from 'hmmscan' was filtered to recognize only the sequences with relevant domain architectures.

2.3 *Identification of Type VII Secretion System clusters*

The genes encoding the components of bacterial secretion systems have been reported to be located as gene clusters (Kostakioti *et al.* 2005), with no exception for Type VII Secretion System (T7SS) (Abdallah *et al.* 2007). Thus, after the identification of homologs of T7SSCs across bacterial genomes, it was necessary to evaluate their genomic distances in order to demarcate functionally associated components forming probable T7SS-like clusters. For this purpose, the significant cluster length (distance between the start sites of the two genes located at the boundaries of T7SS-like cluster) was decided based on the lengths of the model T7SS clusters. The lengths of these model T7SS clusters have been listed in supplementary table 3 of supplementary material 1. The methodology adopted for identification of T7SS/T7SS-like clusters has been schematically represented in supplementary figure 2 of supplementary material 1.

The lengths as well as the number of components of the model clusters were observed to vary noticeably, suggesting their functional diversity. While, the largest cluster belonged to *M. tuberculosis* H37Rv with a length of 21,537 bp, the smallest one belonged to *S. aureus* with a length of 9,454 bp. Thus, the cut-off score for cluster length was defined to be 25,000 bp. In other words, if the predicted T7SS components in an organism were observed to be located in proximity to each other forming a genomic patch of less than (or equal to) 25,000 bp length, it was considered as a gene cluster. Further, another constraint was applied to the identified gene clusters in order to designate them as 'potential T7SS-like clusters'. This constraint was based on the types of the different components present in the T7SS-like clusters (rather than their lengths). The experimentally identified T7SS/T7SS-like clusters (model T7SS cluster) with minimum number of components belonged to *B. anthracis* (Garufi *et al.* 2008). It was observed to contain only two T7SS components (T7SSC1 and T7SSC5). Surprisingly, no membrane spanning component was reported to be associated to this particular cluster, although, membrane components are indispensable in protein translocation. In the current study, any predicted gene cluster was considered as 'potential T7SS cluster' if, it contained at least three T7SS components corresponding to WXG100 family of secreted proteins (T7SSC1), FtsK family of motor ATPases (T7SSC5) and one membrane protein (any one of T7SSC7, T7SSC11, T7SSC12 and T7SSC13). Apart from predicting potential T7SS/T7SS-like clusters, the genomic islands lacking any component known to be involved in the formation of membrane pore (of the Type VII secretion machinery), but containing homologs of at least T7SSC1 (secreted proteins belonging to WXG100 family) and T7SSC5 (ATPase of FtsK family), were mapped. Such clusters have been referred to as 'short cluster' throughout the manuscript. As one such cluster in *B. anthracis* has been reported to be functional (Garufi *et al.* 2008), it is possible that, one or more hitherto uncharacterized membrane component(s) participate(s) in the Type VII-like secretion pathway in this organism. Thus, demarcation of short clusters may provide cues for identification of novel membrane components associated to Type VII-like secretion pathway.

2.4 *Analysis of potential T7SS clusters and short clusters*

2.4.1 *Analysis of T7SS/T7SS-like clusters in pathogenic bacteria:* In order to check existence of any novel 'T7SS–pathogenicity' association, the pathogen information was downloaded from *http://ftp.cbi.pku.edu.cn/pub/database/Genome/genomeprj_archive/* and the pathogenic organisms containing T7SS/T7SS-like clusters were identified.

2.4.2 *Insights into functional association of Type VII Secretion System components with T7SS:* Once the potential T7SS/T7SS-like clusters were identified across different bacterial phyla, it was interesting to study, for each of the 19 T7SSCs, the fraction of total homologs that were actually observed to occur specifically within the predicted clusters. Four remaining components, namely, T7SSC20, T7SSC21, T7SSC22 and T7SSC23 were not considered for the current analysis. Although, the corresponding seed components were reported to be involved in the secretion of the proteins using T7SS (supplementary table 1 of supplementary material 1), they were observed to be located outside the respective T7SS cluster (Fortune *et al.* 2005; MacGurn *et al.* 2005; Raghavan *et al.* 2008; Millington *et al.* 2011; Chen *et al.* 2012). In addition, this analysis was performed with only those phyla identified to have 'potential T7SS/T7SS-like clusters'. The premise of this analysis is as follows. If a majority of homologs identified for any particular T7SSC are present as a part of 'potential T7SS/T7SS-like clusters', it indicates that the function of the T7SSC is probably specific to T7SS. On the other hand, for any T7SSC, if a large fraction of the identified homologs are not observed to lie within the 'potential T7SS/T7SS-like clusters', it is likely that they participate in other pathways (besides Type VII secretion pathway). Thus, for each T7SS components, the fraction of homologs identified within the 'potential T7SS/T7SS-like clusters' were evaluated in order to better understand the functional association of T7SSCs with T7SS.

2.4.3 *Gene conservation pattern within Type VII Secretion System clusters:* The main focus of this analysis was to interpret the relation between bacterial species and the gene conservation pattern of the identified potential T7SS/T7SS-like clusters. Thus, a single representative cluster from each bacterial species was considered to avoid complexity. In the bacterial species containing multiple T7SS/T7SS-like clusters with varied length and number of components, the functional T7SS clusters identified experimentally mostly correspond to the larger ones with maximum types of T7SS components. For example, among the five T7SS clusters (ESX1- ESX5 regions) in *M. tuberculosis* H37Rv, only three comparatively large clusters (ESX1, ESX3 and ESX5) with more diverse components than the other two clusters (ESX2 and ESX4), have been reported as functional (Abdallah *et al.* 2007). Thus, based on such observation, for each species, the largest cluster with maximum types of T7SS components was selected as the representative cluster. After the representative clusters were selected for each bacterial species, they were grouped based on the component conservation pattern. Each group, consisting of one or more representative T7SS/T7SS-like clusters having identical component conservation pattern, was represented by a Boolean vector of size 23 (corresponding to 23 T7SS components), where '0' was assigned for absence of a particular T7SS component and '1' for its presence. A tree was constructed based on the edit distance (difference in presence status of the T7SS components) of the boolean vectors representing the groups. The 'PHYLIP' package (Felsenstein 1989) was used to compute distances among the boolean vectors and to generate a rooted tree.

2.4.4 *Identification of additional membrane components of T7SS:* One of the most crucial T7SS components is the membrane pore which facilitates protein translocation through cytoplasmic membrane. The membrane pore in the experimentally identified T7SS clusters belonging to Actinobacteria is well defined and contains an N-terminal YukD domain followed by 10/11 transmembrane domains. On the other hand, the corresponding component(s) in Firmicutes is not well characterized and it has been suggested that, more than one transmembrane protein may contribute to the formation of the membrane pore (Sao-Jose *et al.* 2004; Burts *et al.* 2005; Way and Wilson 2005). In the current study, many genomic regions (referred to as 'short clusters') were identified which contained homologs of T7SSC1 and T7SSC5, but lacked any membrane spanning components in their vicinity. An example of such a region is the one found in *B. anthracis,* for which secretion of the WXG100 proteins has been experimentally reported (Garufi *et al.* 2008). These observations indicate that, the WXG100 proteins, secreted using the machinery encoded by such regions, probably use alternative membrane components, which have not been associated to T7SS till date. Thus, in order to identify the presence of any auxiliary membrane components, the neighbourhood of the 'short clusters' were explored in the six phyla (Actinobacteria, Firmicutes, Chloroflexi, Proteobacteria, Planctomycetes and Cyanobacteria) predicted (in the current study) to contain such short clusters. The neighbourhood of each of these short clusters (ie. a genomic distance of 25 kbp as described earlier) was scanned in order to identify proteins containing putative membrane components. For this purpose, the annotations of the neighboring proteins were first obtained from the 'protein table file' (.ptt) for each genome obtained from NCBI (*http://www.ncbi.nlm.nih.gov/Ftp/*). Subsequently proteins with annotation keywords 'hypothetical', 'membrane' and 'ABC transporter' were retrieved, as any of these three annotations indicate the presence of membrane spanning regions in the corresponding proteins. The identified set of neighbouring proteins, annotated as 'Hypothetical', in each region were then subjected to TMHMM (*http://www.cbs.dtu.dk/services/TMHMM/*) search in order to predict the presence of transmembrane regions. Similarly, the sets of neighbouring membrane and ABC transporter proteins were also queried against TMHMM to evaluate the number and location of the transmembrane regions in these protein sequences. Finally, if one or more of such membrane protein(s) identified in the neighbourhood of short clusters, were also seen to be conserved across closely related bacterial strains or have been suggested to be involved in transport related functions, they were marked as probable membrane components associated to T7SS. The results of such analysis are expected to help in short listing probable hitherto uncharacterized membrane components associated with such secretion systems.

2.5 *Development of T7SS database*

A database of the identified T7SS components, called T7CD (Type VII Secretion System Component Database) has been developed using MySQL. The database contains only those organisms which possess at least one predicted T7SS/T7SS-like or short clusters. A web interface developed using J2EE has been embedded with facilities like 'Browse', 'Search', in order to facilitate easy retrieval and interpretation of the data stored in the database.

3. Results and discussion

3.1 *Probable Type VII Secretion System components (T7SSC) across bacterial phyla*

Homologs of at least 23 components of T7SS (T7SSC1 through T7SSC23) were found to be present in 24 phyla (figure 2). However, homologs of the signature component T7SSC1 (secretory protein of WXG100 family), were observed to be restricted to only seven (out of the 24) phyla, namely, Actinobacteria, Firmicutes, Chloroflexi, Proteobacteria, Planctomycetes, Cyanobacteria and Spirochaetes. This indicates a higher likelihood of the presence of Type VII-like Secretion Systems in bacterial species belonging to these seven phyla.

Apart from the secretory proteins of WXG100 family, homologs of ATPase component (T7SSC5) and the protease component (T7SSC8) were observed to be present in 23 (out of 24) phyla (figure 2), indicating probable involvement of these components in house-keeping pathways (other than virulence related pathways like secretion systems). This is in line with what has been reported in earlier literature (Siezen and Leunissen 1997; Massey *et al.* 2006).

The phyla distribution of all other T7SS components (membrane components, secretory proteins not belonging to WXG100 family, regulators, ATPase and the components of unknown function) were observed to be confined within a few number of phyla (at most three). This indicates that, these components probably have organism-specific functions. Furthermore, while three membrane components (T7SSC4, T7SSC7 and T7SSC9) were found only in Actinobacteria, two other (T7SSC11 and T7SSC12) were specifically detected in Firmicutes. This suggests that, species belonging to Actinobacteria and Firmicutes might consist of structurally (and functionally) distinct membrane translocons for transporting similar secretory proteins through T7SS. Although, such speculation has been reported in earlier studies based on the membrane components found in a few organisms under Actinobacteria and Firmicutes (Abdallah AM *et al.* 2007), the current study suggests that, use of distinct membrane translocon for transport of similar secretory proteins in these two phyla is probably a global trend.

3.2 *Probable Type VII Secretion System clusters (T7SS) in bacteria*

A total of 428 potential T7SS/T7SS-like clusters were identified in 254 organisms belonging to two bacterial phyla,

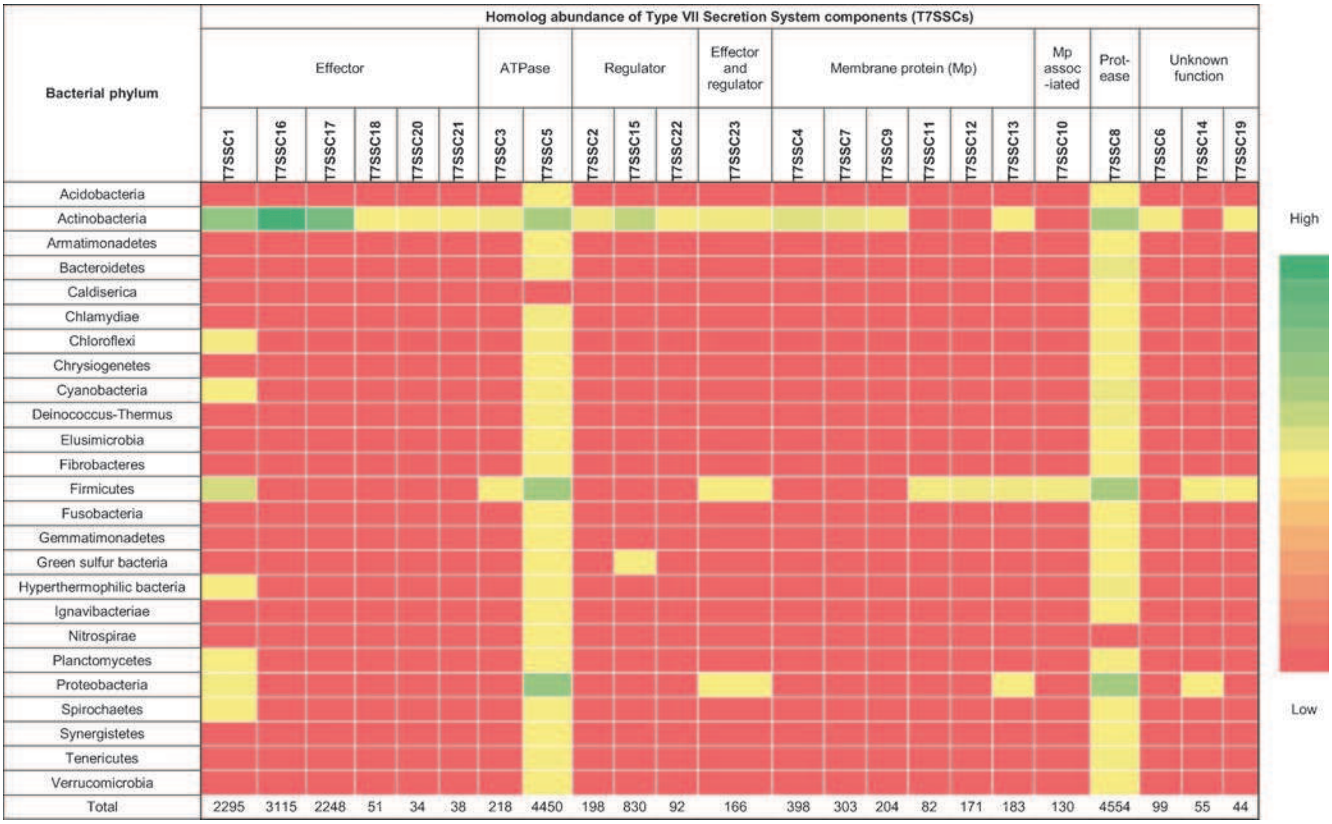

Figure 2. Heatmap depicting the abundance of the 23 Type VII Secretion System components (T7SSCs) across different bacterial phyla.

namely Actinobacteria and Firmicutes (supplementary table 1 of supplementary material 2). Many organisms belonging to the phyla Actinobacteria were observed to possess multiple clusters. On the other hand, each of the organisms (identified to have T7SS-like clusters) belonging to Firmicutes were observed to possess single copy of T7SS cluster within its genome. Both these results corroborate with the previously reported observations of T7SS clusters in mycobacterial strains (Abdallah *et al.* 2007), as well as in species belonging to Firmicutes (Sao-Jose *et al.* 2004; Burts *et al.* 2005; Way and Wilson 2005; Huppert *et al.* 2014).

The observation that, besides Actinobacteria and Firmicutes, other five phyla (Chloroflexi, Proteobacteria, Planctomycetes, Cyanobacteria and Spirochaetes) encode WXG100 family proteins not belonging to any of the identified T7SS/T7SS-like clusters, led to analysis of short clusters (as mentioned in Methods section) in these phyla. Interestingly, these phyla (except Spirochaetes) were observed to harbor short clusters (supplementary table 2 of supplementary material 2). Given that, such short cluster has been reported to be functional in *Bacillus anthracis* (Garufi *et al.* 2008), it is likely that, the corresponding secretion machinery uses transmembrane domain containing ATPase component T7SSC5 as membrane pore. Alternatively, the machinery probably utilizes a membrane component that has not been characterized yet. Thus, the identified short clusters were further investigated to identify probable membrane components associated to them and are discussed in section 3.3.4.

3.3 *Analysis of T7SS clusters and short clusters*

3.3.1 *Pathogenic status of T7SS clusters:* Analysis of the identified T7SS/T7SS-like clusters indicated presence of such clusters in 109 and 40 pathogenic bacteria belonging to Actinobacteria and Firmicutes, respectively (supplementary table 4 of supplementary material 1). Interestingly, presence of such T7SS-like Secretion Systems in some of the pathogenic organisms, including *Enterococcus faecalis*, *Saccharomonospora viridis, Streptococcus equi*, *Streptococcuss gordonii* and *Streptococcus sanguinis,* has not been reported earlier. A schematic representation of the predicted T7SS-like clusters in these four species is further shown in figure 3. In addition, it is interesting to observe that a significant number of organisms (105) identified to contain T7SS/T7SS-like clusters are non-pathogenic. This probably indicates involvement of the secreted proteins in growth related functions, similar to what has been reported for *S. coelicolor* and *L. monocytogenes* (Way and Wilson 2005; Akpe San Roman *et al.* 2010). Thus, the results of the current study open up avenues for the experimental verification of the functional capabilities of the identified probable T7SS-like clusters in these pathogenic as well as non-pathogenic species.

3.3.2 *Insights into functional association of Type VII Secretion System components with T7SS:* Amongst the identified homologs of the secretory protein (T7SSC1) of WXG100 family, while around 56% were observed to be present within the predicted

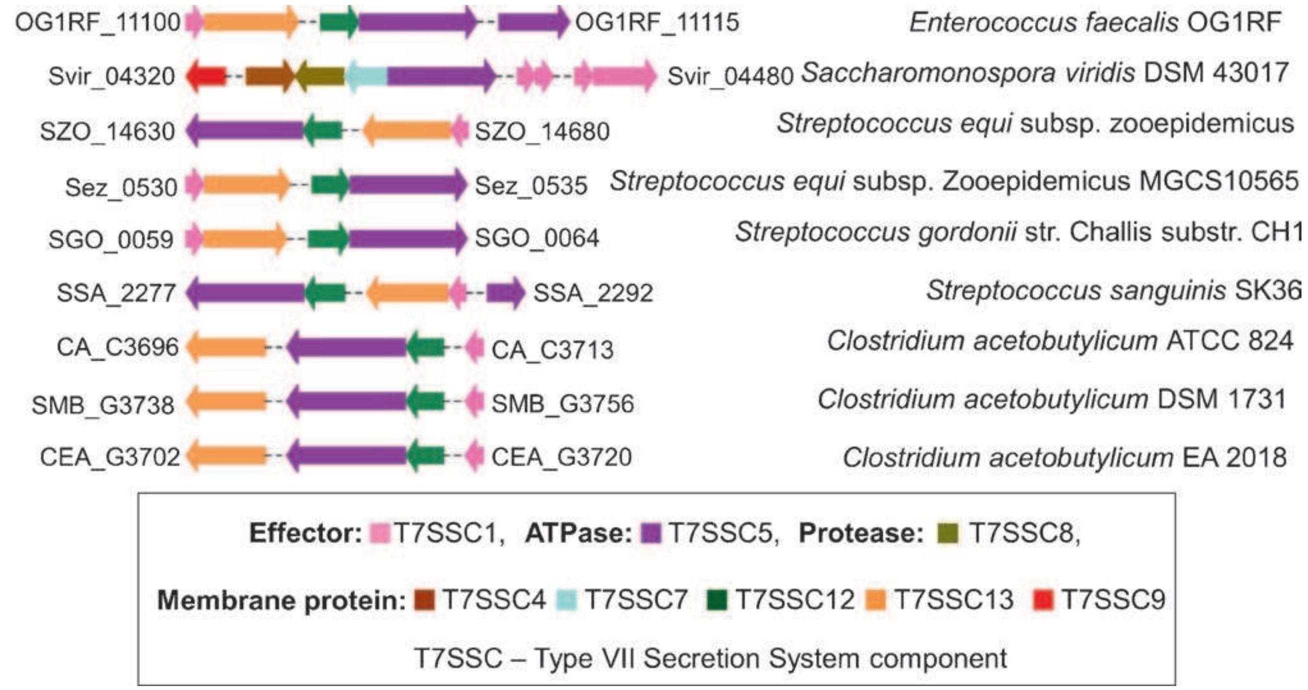
OG1RF_11100
OG1RF_11115
Enterococcus faecalis OG1RF
Svir_04320
Svir_04480
Saccharomonospora viridis DSM 43017
SZO_14630
SZO_14680
Streptococcus equi subsp. zooepidemicus
Sez_0530
Sez_0535
Streptococcus equi subsp. Zooepidemicus MGCS10565
SGO_0059
SGO_0064
Streptococcus gordonii str. Challis substr. CH1
SSA_2277
SSA_2292
Streptococcus sanguinis SK36
CA_C3696
CA_C3713
Clostridium acetobutylicum ATCC 824
SMB_G3738
SMB_G3756
Clostridium acetobutylicum DSM 1731
CEA_G3702
CEA_G3720
Clostridium acetobutylicum EA 2018
Effector: T7SSC1, ATPase: T7SSC5, Protease: T7SSC8,
Membrane protein: T7SSC4 T7SSC7 T7SSC12 T7SSC13 T7SSC9
T7SSC – Type VII Secretion System component

Figure 3. Genomic organization of the T7SS/T7SS-like clusters in certain pathogenic bacteria predicted in the present study.

clusters (16% being in short cluster), remaining 44% were detected as isolated components (supplementary figure 4 of supplementary material 1). Similarly, a large number of the homologs of the secretory components (T7SSC16 and T7SSC17) belonging to PE and PPE family were found to be located outside the predicted T7SS/T7SS-like clusters as isolated components. Interestingly, a number of isolated components of PE and PPE family have been reported earlier to be functionally associated to one of the T7SS regions (ESX-5) in *M. tuberculosis* (Sayes *et al.* 2012). The above results indicate the need for further experimental investigations for understanding the translocation mechanisms of the above mentioned secretory components.

Another interesting observation from the current analysis pertains to the membrane components of T7SS. Unlike, the above mentioned secretory proteins of T7SS, more than 70% of homologs for all the membrane components (T7SSC4, T7SSC7, T7SSC9, T7SSC11-T7SSC13) and the membrane-associated component (T7SSC10), were found to be located inside the predicted T7SS/T7SS-like clusters (supplementary figure 4 of supplementary material 1). This suggests that, these components are likely to perform T7SS-specific functions and thus, can be utilized as additional markers to identify Type VII-like secretion pathways in newly sequenced bacterial species.

3.3.3 *Gene conservation pattern within Type VII Secretion System clusters:* The distance tree generated with 30 groups of identified T7SS/T7SS-like clusters (described in Methods section), was observed to contain two major branches (figure 4). While one branch contained the T7SS/T7SS-like clusters from Actinobacteria, the other branch consisted of two sub-branches containing Actinobacteria and Firmicutes. This suggests that, while Actinobacteria and Firmicutes have distinct gene conservation patterns within T7SS/T7SS-like clusters, the conservation pattern within Firmicutes is much stronger as compared to that in Actinobacteria.

Based on the component conservation patterns across the two major branches (figure 5), a few interesting findings were obtained. While the homologs of the regulator T7SSC2 were found to be present specifically in the branch composed of only Actinobacteria, homologs of the other

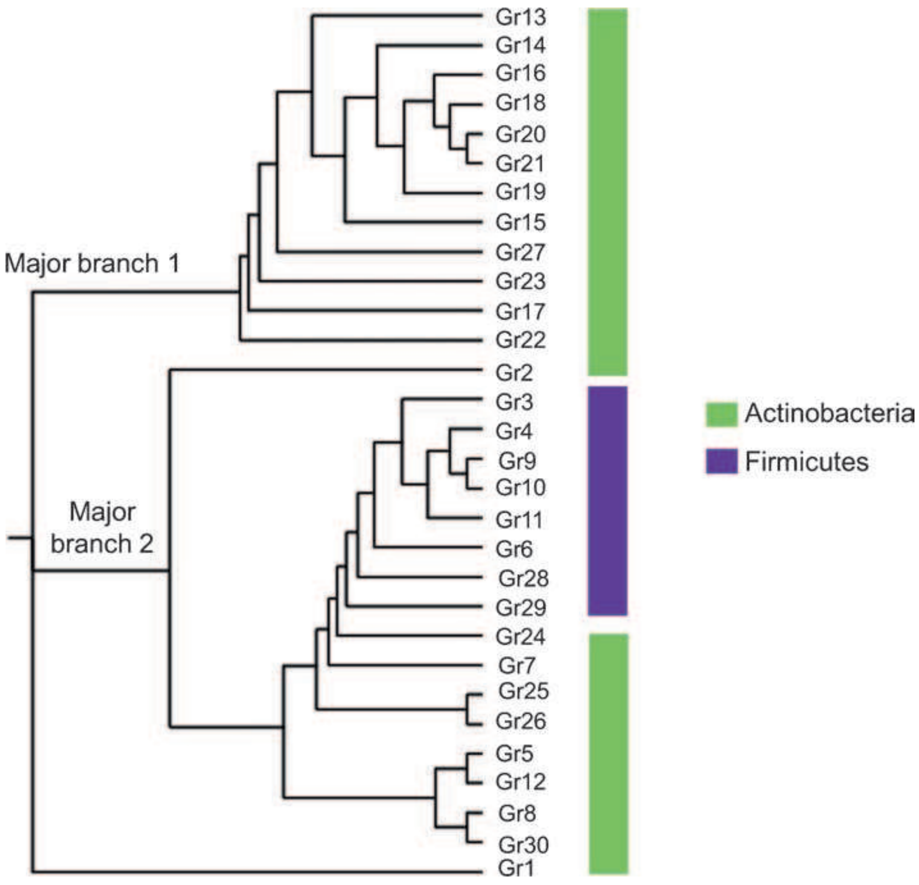

Figure 4. Tree generated based on the T7SS component conservation patterns of 101 representative T7SS/T7SS-like clusters. Each node of the tree corresponds to a single group of T7SS/T7SS-like clusters with a particular component conservation pattern.

regulator T7SSC15 were observed in a few actinobacterial groups belonging to the second branch. The results of the current analysis (figure 5A) indicate a probable regulatory function of the homologs of T7SSC2 in Nocardia and Segniliparus, similar to *M. tuberculosis.* In addition, insights from the present analyses (figure 5B) along with previous reports (Akpe San Roman *et al.* 2010) suggest probable regulatory role of T7SSC15 in a few genera (Salinispora, Catenulispora and Micromonospora), similar to that in Streptomyces. In summary, the current study suggests that, even species within Actinobacterial clade, may use alternate (and varied) regulatory mechanisms for the expression of secretory proteins.

3.3.4 *Additional membrane components in Type VII-like Secretion System:* Analysis of the genomic neighbourhood of 'short clusters' identified a few membrane components, which were not known to be associated with T7SS, in six phyla (Actinobacteria, Firmicutes, Proteobacteria, Chloroflexi, Planctomycetes and Cyanobacteria). Some of these membrane components were found to contain functional domains (supplementary table 5 and supplementary figure 5 of supplementary material 1) which are reported to be involved in membrane transport events (Thanassi and Hultgren 2000; Py *et al.* 2001; Kobiler *et al.* 2002; Sanchez-Pulido *et al.* 2002; Tseng *et al.* 2009; Vastermark *et al.* 2011). Thus, the results of the present study suggest likely involvement of more than one type of

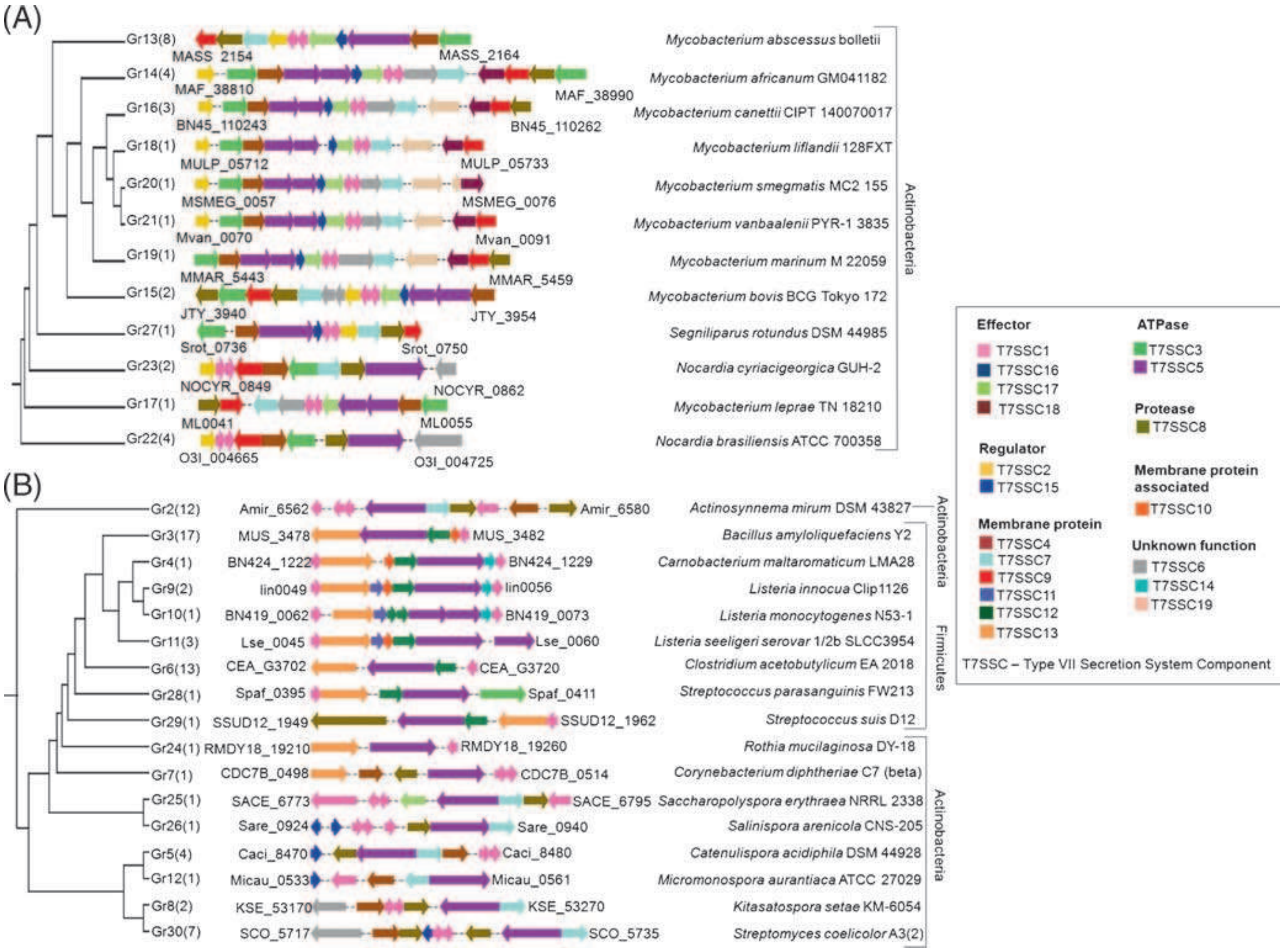
(A)
Gr13(8)
MASS_2154
MASS_2164
Mycobacterium abscessus bolletii
Gr14(4)
MAF_38810
MAF_38990
Mycobacterium africanum GM041182
Gr16(3)
BN45_110243
BN45_110262
Mycobacterium canettii CIPT 140070017
Gr18(1)
MULP_05712
MULP_05733
Mycobacterium liflandii 128FXT
Gr20(1)
MSMEG_0057
MSMEG_0076
Mycobacterium smegmatis MC2 155
Gr21(1)
Mvan_0070
Mvan_0091
Mycobacterium vanbaalenii PYR-1 3835
Gr19(1)
MMAR_5443
MMAR_5459
Mycobacterium marinum M 22059
Gr15(2)
JTY_3940
JTY_3954
Mycobacterium bovis BCG Tokyo 172
Gr27(1)
Srot_0736
Srot_0750
Segniliparus rotundus DSM 44985
Gr23(2)
NOCYR_0849
NOCYR_0862
Nocardia cyriacigeorgica GUH-2
Gr17(1)
ML0041
ML0055
Mycobacterium leprae TN 18210
Gr22(4)
O3I_004665
O3I_004725
Nocardia brasiliensis ATCC 700358
Actinobacteria
(B)
Gr2(12) Amir_6562 Amir_6580 Actinosynnema mirum DSM 43827
Gr3(17) MUS_3478 MUS_3482 Bacillus amyloliquefaciens Y2
Gr4(1) BN424_1222 BN424_1229 Carnobacterium maltaromaticum LMA28
Gr9(2) lin0049 lin0056 Listeria innocua Clip1126
Gr10(1) BN419_0062 BN419_0073 Listeria monocytogenes N53-1
Gr11(3) Lse_0045 Lse_0060 Listeria seeligeri serovar 1/2b SLCC3954
Gr6(13) CEA_G3702 CEA_G3720 Clostridium acetobutylicum EA 2018
Gr28(1) Spaf_0395 Spaf_0411 Streptococcus parasanguinis FW213
Gr29(1) SSUD12_1949 SSUD12_1962 Streptococcus suis D12
Gr24(1) RMDY18_19210 RMDY18_19260 Rothia mucilaginosa DY-18
Gr7(1) CDC7B_0498 CDC7B_0514 Corynebacterium diphtheriae C7 (beta)
Gr25(1) SACE_6773 SACE_6795 Saccharopolyspora erythraea NRRL 2338
Gr26(1) Sare_0924 Sare_0940 Salinispora arenicola CNS-205
Gr5(4) Caci_8470 Caci_8480 Catenulispora acidiphila DSM 44928
Gr12(1) Micau_0533 Micau_0561 Micromonospora aurantiaca ATCC 27029
Gr8(2) KSE_53170 KSE_53270 Kitasatospora setae KM-6054
Gr30(7) SCO_5717 SCO_5735 Streptomyces coelicolor A3(2)
Actinobacteria
Firmicutes
Actinobacteria
Effector
T7SSC1
T7SSC16
T7SSC17
T7SSC18
ATPase
T7SSC3
T7SSC5
Protease
T7SSC8
Regulator
T7SSC2
T7SSC15
Membrane protein associated
T7SSC10
Membrane protein
T7SSC4
T7SSC7
T7SSC9
T7SSC11
T7SSC12
T7SSC13
Unknown function
T7SSC6
T7SSC14
T7SSC19
T7SSC – Type VII Secretion System Component

Figure 5. Elaborated representation of the tree generated based on T7SS component conservation patterns. (**A**) Elaborated representation of first major branch. (**B**) Elaborated representation of second major branch. Each node of the trees corresponds to a single group of T7SS/T7SS-like clusters with a particular component conservation pattern. The number of members and genetic organization of the largest cluster have been shown for each group. In the figures 'T7SSC' has been used as an abbreviation of 'Type VII Secretion System component'. In these figures, for each leaf node representing one group of T7SS clusters, a schematic diagram of representative T7SS/T7SS-like cluster (with the cluster having highest length for the particular group being selected as the representative), the corresponding organism containing this cluster and number of members in the group have been shown.

cytoplasmic membrane pore in Type VII-like secretion pathway. However, further experimental verification will be required to explain functional relatedness of the identified membrane components with T7SS.

3.4 *Database of Type VII Secretion System components*

Information on the T7SS components and T7SS/T7SS-like clusters identified in the current study and the ones identified experimentally by previous studies (supplementary figure 1 of supplementary material 1) have been organized into a database called T7CD (Type VII Secretion System Component Database). The web interface of the database can be accessed at *http://metagenomics.atc.tcs.com/T7SS/*. The web interface has been designed to facilitate easy retrieval and interpretation of the stored data.

4. Conclusions

In the present study, *in silico* approaches have been used to identify Type VII Secretion System components and probable T7SS clusters in various bacterial groups. Based on the conservation of the T7SS components, the present study identified differences in modes of protein translocation through T7SS in organisms belonging to two phyla namely, Actinobacteria and Firmicutes. Further experimental validations of the probable additional membrane components identified in some of the organisms may help in improving our current knowledge of protein export mechanism through Type VII-like secretion pathway. In addition, the probable presence of T7SS-like clusters identified in a few pathogenic bacteria is likely to help in understanding the pathogenicity related aspects mediated through Type VII-like secretion pathway in these pathogens. The comprehensive database of T7SS components and T7SS/T7SS-like clusters would facilitate easy access of the data from a single platform, which in turn is expected to aid in advancements in related fields of research.

Acknowledgments

We thank Hemang Gandhi for development of the web interface. CD is also a PhD student at the Department of Chemical Engineering, Indian Institute of Technology (IIT), Bombay, and would like to acknowledge the department for its support.

References

Abdallah AM, Gey van Pittius NC, Champion PAD, Cox J, Luirink J, Vandenbroucke-Grauls CMJE, Appelmelk BJ and Bitter W 2007 Type VII secretion–mycobacteria show the way. *Nat. Rev. Microbiol.* **5** 883–891

Akpe San Roman S, Facey PD, Fernandez-Martinez L, Rodriguez C, Vallin C, Del Sol R and Dyson P 2010 A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in *Streptomyces coelicolor. Microbiology (Reading, Engl).* **156** 1719–1729

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ 1990 Basic local alignment search tool. *J. Mol. Biol.* **215** 403–410

Burts ML, Williams WA, DeBord K and Missiakas DM 2005 EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc. Natl. Acad. Sci. USA* **102** 1169–1174

Chen JM, Boy-Röttger S, Dhar N, Sweeney N, Buxton RS, Pojer F, Rosenkrands I and Cole ST 2012 EspD is critical for the virulence-mediating ESX-1 secretion system in *Mycobacterium tuberculosis. J. Bacteriol.* **194** 884–893

Converse SE and Cox JS 2005 A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis. J. Bacteriol.* **187** 1238–1245

Das C, Ghosh TS and Mande SS 2011 Computational analysis of the ESX-1 region of *Mycobacterium tuberculosis*: insights into the mechanism of type VII secretion system. *PLoS ONE.* **6**, e27980

Eddy SR 1998 Profile hidden Markov models. *Bioinformatics* **14** 755–763

Felsenstein J 1989 PHYLIP - phylogeny inference package (Version 3.2). *Cladistics* **5** 164–166

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, *et al.* 2014 Pfam: the protein families database. *Nucleic Acids Res.* **42** D222–D230

Fortune SM, Jaeger A, Sarracino DA, Chase MR, Sassetti CM, Sherman DR, Bloom BR and Rubin EJ 2005 Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc. Natl. Acad. Sci. USA* **102** 10676–10681

Garufi G, Butler E and Missiakas D 2008 ESAT-6-like protein secretion in *Bacillus anthracis. J. Bacteriol.* **190** 7004–7011

Geluk A, van Meijgaarden KE, Franken KLMC, Subronto YW, Wieles B, Arend SM, Sampaio EP, de Boer T, *et al.* 2002 Identification and characterization of the ESAT-6 homologue of *Mycobacterium leprae* and T-cell cross-reactivity with *Mycobacterium tuberculosis*. *Infect. Immun.* **70** 2544–2548

Gey Van Pittius NC, Gamieldien J, Hide W, Brown GD, Siezen RJ and Beyers AD 2001 The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol.* **2** RESEARCH0044

Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S and Sherman DR 2004 Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis. Mol. Microbiol.* **51** 359–370

Houben ENG, Korotkov KV and Bitter W 2014 Take five —Type VII secretion systems of Mycobacteria. *Biochim. Biophys. Acta (BBA) Mol. Cell Res.* **1843** 1707–1716

Huppert LA, Ramsdell TL, Chase MR, Sarracino DA, Fortune SM and Burton BM 2014 The ESX system in *Bacillus subtilis* mediates protein secretion. *PLoS ONE* **9** 5

Kobiler O, Koby S, Teff D, Court D and Oppenheim AB 2002 The phage lambda CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis. *Proc. Natl. Acad. Sci. USA* **99** 14964–14969

Kostakioti M, Newman CL, Thanassi DG and Stathopoulos C 2005 Mechanisms of protein export across the bacterial outer membrane. *J. Bacteriol.* **187** 4306–4314

MacGurn JA *et al*. 2005 A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis. Mol. Microbiol.* **57** 1653–1663

Massey TH, Mercogliano CP, Yates J, Sherratt DJ and Löwe J 2006 Double-stranded DNA translocation, structure and mechanism of hexameric FtsK. *Mol. Cell* **23** 457–469

Millington KA, Fortune SM, Low J, Garces A, Hingley-Wilson SM, Wickremasinghe M, Kon OM and Lalvani A 2011 Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for *Mycobacterium tuberculosis* infection. *Proc. Natl. Acad. Sci. USA* **108** 5730–5735

Pallen MJ 2002 The ESAT-6/WXG100 superfamily – and a new Gram-positive secretion syste? *Trends Microbiol.* **10** 209–212

Pearson WR 2013 An introduction to sequence similarity (“Homology”) searching. *Curr. Protoc. Bioinformatics* **3** 3.1

Py B, Loiseau L and Barras F 2001 An inner membrane platform in the type II secretion machinery of Gram-negative bacteria. *EMBO Rep.* **2** 244–248

Raghavan S, Manzanillo P, Chan K, Dovey C and Cox JS 2008 Secreted transcription factor controls *Mycobacterium tuberculosis virulence. Nature* **454** 717–721

Sanchez-Pulido L, Martın-Belmonte F, Valencia A and Alonso MA 2002 MARVEL, a conserved domain involved in membrane apposition events. *Trends Biochem. Sci.* **27** 599–601

Sao-Jose C, Baptista C and Santos MA 2004 *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. *J. Bacteriol.* **186** 8337–8346

Sayes F, Sun L, Di Luca M, Simeone R, Degaiffier N, Fiette L, Esin S, Brosch R, *et al*. 2012 Strong immunogenicity and cross-reactivity of *Mycobacterium tuberculosis* ESX-5 type VII secretion: encoded PE-PPE proteins predicts vaccine potential. *Cell Host Microbe* **11** 352–363

Siezen RJ and Leunissen JA 1997 Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6** 501–523

Stanley SA, Raghavan S, Hwang WW and Cox JS 2003 Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc. Natl. Acad. Sci. USA* **100** 13001–13006

Sutcliffe IC 2011 New insights into the distribution of WXG100 protein secretion systems. *Antonie Van Leeuwenhoek* **99** 127–131

Tan T, Lee WL, Alexander DC, Grinstein S and Liu J 2006 The ESAT-6/CFP-10 secretion system of *Mycobacterium marinum* modulates phagosome maturation. *Cell. Microbiol.* **8** 1417–1429

Thanassi DG and Hultgren SJ 2000 Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr. Opin. Cell Biol.* **12** 420–430

Tseng T-T, Tyler BM and Setubal JC 2009 Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol.* **9** S2

Vastermark Å, Almen MS, Simmen MW, Fredriksson R and Schiöth HB 2011 Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster EamA (DUF6) before the radiation of Viridiplantae. *BMC Evol. Biol.* **11** 123

Way SS and Wilson CB 2005 The *Mycobacterium tuberculosis* ESAT-6 homologue in *Listeria monocytogenes* is dispensable for growth in vitro and in vivo. *Infect. Immun.* **73** 6151–6153

MS received 05 November 2015; accepted 22 January 2016

Corresponding editor: Seyed E Hasnain