Australasian Plant Pathology (2022) 51:123–133
https://doi.org/10.1007/s13313-021-00845-6

ORIGINAL RESEARCH ARTICLE

Deciphering novel potential antibacterial targets in tomato pathogen *Ralstonia solanacearum* GMI1000 through integration of in silico subtractive genomics, codon usage and protein–protein interaction analyses

Gurunathan Subramanian[1] · Umashankar Vetrivel[2] · Mohamed Imran Mohamedyousuff[3]

Received: 26 June 2020 / Accepted: 1 December 2021 / Published online: 9 January 2022

Abstract

Ralstonia solanacearum is a bacterial phytopathogen of worldwide attention as it causes wilt disease in tomato leading to huge economic loss. This poses a need for identifying host non-specific drug targets to design effective antibacterial agents. Hence, in this study novel therapeutic targets were prioritized by implementing integrative in silico analysis spanning genome to metabolic pathway levels. Initial comparison of chromosome encoded proteins of pathogen against host proteome revealed non-host homologous protein sequences in the pathogen. Further, these proteins were probed for essentiality analysis using database of essential genes as reference, which prioritized the host non-homologous proteins that are essential for survival. Among the 3380 chromosome encoded proteins of the pathogen, 115 were identified as essential and host non-homologous. Subsequent metabolic pathway analysis revealed 24 proteins to be involved in pathogen-specific pathways. On further expression profiling based on codon usage, 17 of these 24 proteins were found to be highly expressed, as per the Codon Adaptation Index (CAI) score. In addition, subcellular localization, virulence factor analysis, drugbank screening and Protein–Protein Interactions (PPIs) were also implemented in order to stringently prioritize the targets based on potentiality and druggability. On cumulative analysis of all these results, 3 unique proteins (Q8XU97, Q8Y1J8 and Q8XVG8) of this pathogen were recognized as highly potential and novel antimicrobial targets. As these proteins resulted from a multilevel stringent prioritization process and were also found to be involved in key survival pathways, these shall form as druggable targets for the design and development of safe bactericidal agents to combat this economically important pathogen.

Keywords *Ralstonia solanacearum* · Antibacterial targets · Subtractive genomics · Essential genes · Bacterial wilt disease · Protein–Protein Interactions (PPIs)

Introduction

Ralstonia solanacearum is a globally recognized phytopathogenic bacterium causing bacterial wilt disease in more than 450 plants and crop species. It is primarily soil-borne and enters susceptible hosts via their roots, proliferates in the root cortex and spreads to the vascular system leading to blockage of water and other vital nutrients movement owing to abundant bacterial colonization. Thus, the infected plant collapses suddenly and dies, eventually releasing pathogen to the soil (Elphinstone 2005; Peeters et al. 2013). It is a Gram-negative, aerobic, non-spore forming motile bacillus belonging to the class of beta-proteobacteria under the family *Pseudomonodaceae*. More than 140 different strains of *R. solanacearum* are reported (Elphinstone 2005; Tahat and

✉ Gurunathan Subramanian

[1] PG and Research Department of Biotechnology & Microbiology, National College (Autonomous), Karumandapam, Tiruchirapalli 620001, Tamil Nadu, India

[2] Department of Health Research (Govt. of India), ICMR-National Institute of Traditional Medicine, Indian Council of Medical Research, Nehru Nagar, Belagavi 590010, Karnataka, India

[3] Department of Microbiology, Research Associate, National Repository for Microalgae and Cyanobacteria - Freshwater (NRMC-F), Bharathidasan University, Tiruchirappalli 620024, Tamil Nadu, India

Sijam 2010; Tiwari et al. 2017). Based on the diversity of hosts, strains of *R. solanacearum* have been grouped into five races namely, race 1, 2, 3, 4 and 5 (Elphinstone 2005). Similarly, six biovars (1, 2, 3, 4, 5, 6) have been reported based on their biochemical properties (He et al. 1993) and four phylotypes (viz. Phylotype I, II, III and IV) have been documented based on their geographical origins (Peeters et al. 2013). *R. solanacearum* GMI1000 is a race 1, biovar 3 / Phylotype I strain isolated from tomato plant in French Guyana. It is virulent to a broad range of plants and internationally used as a model to study plant pathogenicity (Salanoubat et al. 2002).

Discovery of antimicrobial targets was traditionally dependent on gene-based information. However, employing -omics data for target discovery have gained popularity in current post-genomics era (Lin and Qian 2007). Subtractive genomics (aka called 'differential genome display') is an interesting approach in which genes (or corresponding proteins) essential for pathogen's life but not located in host cells are identified as potential antimicrobial targets (Galperin and Koonin 1999). Even though this approach has been successfully used to identify protein targets in different human bacterial pathogens (Barh et al. 2011; Vetrivel et al. 2011; Shoukat et al. 2012; Mondal et al. 2015; Hossain et al. 2016; Uddin and Jamil 2018; Sadhasivam and Vetrivel 2018; Shuvo et al. 2019), only a few researchers have attempted it on phytopathogenic bacteria (Katara et al. 2012; Keshri et al. 2014; Prabha et al. 2019). Hence, implementing this approach on available whole genome data of bacterial phytopathogens (either draft or complete) would reveal novel antibacterial targets in them (Sundin et al. 2016; Tiwari et al. 2017).

Codon usage analysis in any organism (including bacteria) provides valuable insights into its gene expression levels (Vetrivel et al. 2011). In 1987, Sharp and Li introduced a statistically derived metric called Codon Adaptation Index (CAI) to analyze codon usage bias in an organism. This widely employed unit is expressed as a value between 0 and 1. Genes with higher CAI values are likely to be highly expressed (Vetrivel et al. 2011).

Complex biological processes and structural framework of cells primarily depend on Protein–Protein Interactions (PPIs). Studying PPIs involved in biological pathways aids improved understanding of the physiological and pathological features of a cell (Petschnigg et al. 2011). Furthermore, it helps in identifying hub proteins (proteins found to partner with several other proteins in the biological network) and blocking them with specific molecules could kill the pathogen. Thus, these proteins may act as plausible antimicrobial targets in a pathogen, provided the corresponding host cells lack them (Uddin and Jamil 2018). PPIs identified hub proteins are also reported as possible therapeutic targets in bacterial pathogens of humans (Jadhav et al. 2014; Uddin and Jamil 2018) and insects (Gupta et al. 2016).

Complete genome data of *R. solanacearum* GMI1000 is available (5.8 Mb) (https://iant.toulouse.inra.fr/bacteria/annotation/cgi/ralso.cgi). The genome is organized in two independently replicating circular replicons: a 3.7 Mb chromosome and a 2.1 Mb megaplasmid. It has a coding potential for approximately 5000 proteins and contains high G+C content (average value of 67%) (Salanoubat et al. 2002). Hence, the present work aims at applying strategies coupling subtractive genomics and knowledge-based predictions (primarily CAI and PPIs) on the genome of *R. solanacearum* GMI1000 to identify novel protein targets to enable the identification and application of effective molecules against this economically important pathogen.

Materials and methods

Data collection

Whole proteome sequences of *R. solanacearum* GMI1000 (UP000001436) and *Solanum lycopersicum* (UP000004994) were downloaded from UniProtKB database (http://www.uniprot.org/taxonomy/complete-proteomes) in FASTA format. Prokaryotic essential gene sequences data were retrieved from the Database of Essential Genes (DEG) version 15.2 (http://tubic.tju.edu.cn/deg/) (Luo et al. 2014). Database comprising of known protein targets were downloaded from DrugBank Database to perform comparative analysis (Wishart et al. 2018). Supplementary Table 1 displays the genome and proteome details of the pathogen studied.

Excluding orthologs and paralogs

Entire proteome of *R. solanacearum* GMI1000 is made up of 5002 proteins. Majority of the housekeeping genes of the pathogen are located in the chromosome. However, several megaplasmid encoded proteins are also encoded by chromosomal genes of the bacterium (Salanoubat et al. 2002; Peeters et al. 2013). Thus, 3380 chromosomally encoded proteins of the pathogen were alone chosen for the study. Among the 3380 sequences selected, those with amino acid residues of 100 or lesser length were considered as orthologs and hence were excluded from the dataset (Barh et al. 2011; Vetrivel et al. 2011; Hossain et al. 2016). Similarly, duplicate proteins or paralogs present in the pathogen were removed by CD-HIT normalization (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit) (Huang et al. 2010). Here, a stringent criterion of 80% sequence identity threshold was adopted (Shoukat et al. 2012).

Selection of essential proteins among non-host homologs

The protein set resulting from the preceding step was analyzed using BLASTP tool (Altschul et al. 1990) to explore its similarity with the proteome of the host *i.e.* tomato (*Solanum lycopersicum*). A random expectation value (E-value) cutoff of 0.01 was opted and proteins with 'no hits found' alone were extracted to form the resultant dataset termed as 'Non-Host Homologs'.

All Non-Host Homologs were further screened out based on their essentiality for the basic survival of the pathogen. BLAST analysis of Non-Host Homologs vs. DEG database prokaryotic sequences eliminates pathogen's essential sequences showing homology to host proteins. A maximum threshold value of 10^{-10} (Keshri et al. 2014) and sequence identity above 30% (Barh et al. 2011) were the set parameters to obtain essential proteins in *R. solanacearum* GMI1000.

All BLASTP analyses were performed using stand-alone BLAST software (version 2.6.0+). The parameters such as e-value and sequence identity were provided using the command line option. An in-house developed PERL script was used to parse the BLASTP results for retrieving respective protein sequences in each of the above-mentioned steps.

Mining essential proteins involved in unique metabolic pathways of the pathogen

Essential proteins of *R. solanacearum* GMI1000 filtered in the previous step were uploaded to KAAS (KEGG Automatic Annotation Server) at KEGG (http://www.genome.jp/tools/kaas). The query sequences were searched in KEGG GENES database to reveal their respective functional annotation as KEGG Orthology (KO) assignments and KEGG Pathways (Moriya et al. 2007).

Biochemical pathways of *R. solanacearum* GMI1000 and tomato were recognized using the KEGG pathway database. Pathways of the pathogen were manually compared with that of the host. Metabolic pathways observed in the pathogen, but absent in host were identified as unique pathways of *R. solanacearum* GMI1000. The remaining pathways were deemed insignificant, as these were either common to both organisms or distinctive to host. Only unique pathways associated essential proteins were short listed as possible drug targets and were chosen for further investigation.

Screening of highly expressed genes/proteins based on Codon adaptation index (CAI)

Studying codon usage of an organism helps in revealing the selectivity and expression levels of its genes. Codon bias of an organism is determined by calculating its Codon Adaptation Index (CAI) (Sharp and Li 1987). CAI value indicates codon usage adaptation in highly expressed genesets. Codon Adaptation Index (CAI) of each essential gene sequence was computed using CAI tool in European Molecular Biology Software Suite (EMBOSS) GUI Server (http://bgagenomics.iicb.res.in/EMBOSS/). Inbuilt codon usage table (epsesm.cut) generated from highly expressed genesets of *Pseudomonas syringae pv tomato* strain DC3000 was used as the reference set. CAI values of genes range between 0 and 1. Higher CAI value suggests most likely presence of highly expressed genes. Hence, genes displaying $>=0.75$ CAI values were segregated based on earlier studies portraying this cutoff to be optimal (Wu et al. 2005).

Subcellular localization prediction

Bacterial proteins in Gram-negative members are known to localize at five subcellular locations viz. cytoplasm, cytoplasmic membrane (also called inner membrane), periplasmic, outer membrane and extracellular space (Yu et al. 2014). Prediction of probable target proteins' subcellular localization in the bacterium was studied using PSORTb tool (version 3.0) (Yu et al. 2010). Furthermore, CELLO2GO server (http://cello.life.nctu.edu.tw/cello2go) (Yu et al. 2014) was employed to confirm PSORTb results. Both web tools reveal results with higher accuracy. Functional annotation of the analyzed proteins was also displayed by CELLO2GO analysis. Thus, the protein datasets were sorted in accordance to their subcellular localization.

Prediction of virulence proteins

MP3 webserver was used for predicting bacterial virulence protein sequences. The server provides a fast, accurate and sensitive prediction of pathogenic proteins by integrating Support Vector Machine and Hidden Markov Model approach (Gupta et al. 2014).

Prediction of novel antibacterial targets

Each of the selected targets was compared with experimentally determined drug targets data of the DrugBank Database (Wishart et al. 2018) using the BLAST tool (E-value cut off of 10^{-4} and similarity score >70), as per the procedure demonstrated by Keshri et al. (2014). Proteins exhibiting no similarity to experimental drug targets are classified as novel antibacterial targets of *R. solanacearum* GMI1000.

Protein–protein interactions (PPIs) studies

PPIs of novel potential targets were studied using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) v11.0 database (Szklarczyk et al. 2019). It contains 2.45 billion proteins from 5,090 (4,445 bacteria)

organisms and possesses direct (physical) and indirect (functional) interactions of proteins. PPIs possessing high confidence score i.e. ≥ 0.7 were only chosen so as to avoid any false-positive results. On the basis of node degree (K), the hub proteins were identified (Gupta et al. 2016). Figure 1 outlines the methodology adopted in this research work.

Results

Identification of essential proteins

Among 5002 proteins of *R. solanacearum* GMI1000 proteome, only 3380 chromosomally encoded proteins were utilized in this study. A total of 310 and 80 proteins were identified as orthologs and paralogs, respectively. The remaining 2990 non-orthologs cum non-paralogs present in the pathogen were BLAST aligned against 34,648 proteins of the tomato host proteome to reveal 2760 Non-Host Homologous proteins. Thus, a total of 620 proteins *i.e.* 18.34% of the pathogen's chromosome proteins were excluded from the data set. A total of 115 essential proteins were mined among the Non-Host Homologs during BLAST search against DEG. Supplementary Table 2 lists the number of proteins identified at different stages of analyses performed in this study.

Mapping essential proteins as metabolites of unique pathways

During KAAS analysis, a total of 105 sequences among the 115 essential proteins were identified to involve in 46 different biochemical pathways of the pathogen (Supplementary Table 3). The results clearly indicate that majority of the proteins analysed have roles in basic cellular mechanism pathways such as citric acid cycle, amino acid and nucleotide biosynthesis, peptidoglycan biosynthesis, lipopolysaccharide synthesis, etc. Also, involvement of proteins in cell signalling pathways such as quorum sensing and biofilm formation is also revealed. A few proteins were identified to have roles in resistance conferring pathways such as CAMP resistance, vancomycin and *beta*-lactum resistance. Maximum of 14 proteins have been found to be involved in two-component system pathways followed by 7 and 6 proteins in biosynthesis of secondary metabolites and ABC transporters pathways respectively. KEGG Orthologs could not be determined for ten essential proteins and hence, were excluded from further studies.

Involvement of essential proteins as metabolites of *R. solanacearum* GMI1000 unique metabolic pathways was also studied. Among 124 biochemical pathways of the pathogen, 40 (32.25%) were found to be unique and 24 essential proteins were participants in them. Table 1 provides details of the 24 selected essential proteins and their corresponding metabolic pathway(s) in the pathogen. Two-component system, bacterial secretion system, phosphotransferase system, peptidoglycan and lipopolysaccharide biosynthesis, quorum sensing and resistance mechanism pathways are the identified unique pathways of the pathogen. Among the 24 essential proteins, 5 were found to be involved in only one unique metabolic pathway and the rest involved in more than one unique pathway. Since the remaining 84 pathways were

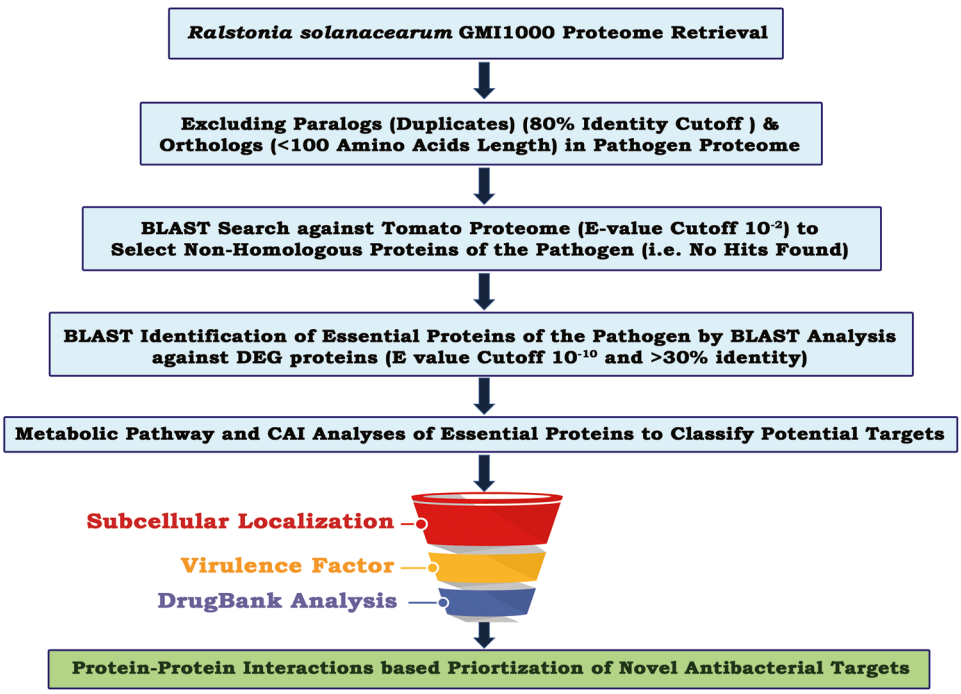

Fig. 1 Steps involved in Potential Antibacterial Targets Identification in *Ralstonia solanacearum* GMI1000

Table 1 Metabolic Pathway and CAI Analyses Results of the 24 selected essential proteins of *R. solanacearum* GMI1000

Protein ID	KO	KO Description	Pathway(s) Involved	CAI
Q8XW15	K01626	3-deoxy-7-phosphoheptulonate synthase	Biosynthesis of secondary metabolites, Biosynthesis of aminoacids, Quorum sensing, Two component system, biosynthesis of antibiotics, Phenylalanine, tyrosine and tryptophan biosynthesis	0.895
Q8XZ75	K02040	Phosphate transport system substrate-binding protein	ABC transporters, Two Component System	0.869
Q8Y1T3	K03406	Methyl-accepting chemotaxis protein	Two component system, Bacterial chemotaxis	0.843
Q8Y3B8	K05515	Penicillin-binding protein 2	Peptidoglycan biosynthesis, Beta-lactum resistance	0.843
Q8XU12	K01546	Potassium-transporting ATPase potassium-binding subunit	Two Component System	0.836
Q8XZE1	K05875	Methyl-accepting chemotaxis protein II, aspartate sensor receptor	Two component system, Bacterial Chemotaxis	0.834
Q8XZJ3	K00990	Uridylyltransferase	Two Component System	0.825
Q8Y253	K10002	Glutamate/aspartate transport system permease protein	Two Component System, ABC Transporters	0.821
Q8XUB6	K11103	Aerobic C4-dicarboxylate transport protein	Two Component System	0.819
Q8XUS0	K02454	General secretion pathway protein E	Biofilm Formation, Bacterial Secretion System	0.816
Q8XWC4	K00075	UDP-N-acetyl muramate dehydrogenase	Peptidoglycan Synthesis, Amino sugar and nucleotide sugar metabolism	0.813
Q8XU97	K01752	L-serine dehydratase	Biosynthesis of Secondary Metabolites, Cysteine and Methionine Metabolism, Biosynthesis of aminoacids, Glycine, serine and threonine metabolism, Carbon Metabolism, Biosynthesis of Antibiotics	0.809
Q8Y089	K03406	Methyl-accepting chemotaxis protein	Two component system, Bacterial chemotaxis	0.805
Q8XVG8	K02770	Fructose PTS system EIIBC or EIIC component	PTS, Fructose and mannose metabolism	0.801
Q8XZW9	K00404	Cytochrome c oxidase cbb3-type subunit I	Two Component System, Oxidative Phosphorylation	0.800
Q8Y1J8	K02841	Heptosyltransferase I	LPS Biosynthesis	0.800
Q8XTY3	K03406	Methyl-accepting chemotaxis protein	Two component system, Bacterial chemotaxis	0.796
Q8Y1E7	K01626	3-deoxy-7-phosphoheptulonate synthase	Biosynthesis of secondary metabolites, Biosynthesis of aminoacids, Quorum sensing, Two component system, biosynthesis of antibiotics, Phenylalanine, tyrosine and tryptophan biosynthesis	0.699
Q8Y3G9	K03585	Membrane fusion protein, multidrug efflux system	Beta-lactam resistance, CAMP resistance	0.636
Q8XZS1	K07264	4-amino-4-deoxy-L-arabinose transferase	LPS biosynthesis, CAMP resistance	0.628
Q8Y3F3	K00641	Homoserine O-acetyltransferase/O-succinyltransferase	Biosynthesis of Antibiotics, Sulfur Metabolism, Biosynthesis of secondary metabolites, Cysteine and Methionine Metabolism	0.603
Q8Y1X2	K02843	Heptosyltransferase II	LPS Biosynthesis	0.584
Q8Y2C8	K03092	RNA polymerase sigma-54 factor	Two Component System, Biofilm formation	0.573
Q8Y3H1	K18139	Outer membrane protein, multidrug efflux system	Beta-lactum resistance, Quorum Sensing	0.569

common to host and pathogen, they were excluded from further investigation.

Detection of highly expressed genes/proteins based on CAI value

The 24 essential proteins identified represent plausible therapeutic targets of *R. solanacearum* GMI1000. However, further filtration is needed so as to reduce the time, labour and resources involved in the later stages of drug development process. Hence, CAI was chosen as a filtering criterion. Each of the 24 essential gene's CAI value was calculated. Significant variations were visible in the overall CAI value range of the pathogen (0.569–0.895) as shown in Table 1. Thus, among 24 sequences, 17 were shortlisted as highly expressed and can be considered as significant protein targets against the *R. solanacearum* GMI1000.

Prioritization of protein targets

All 17 proteins are interesting targets for antimicrobials development against *R. solanacearum* strain GMI1000. However, further prioritization of these targets on the basis of subcellular localization, virulence prediction, drugbank database search and PPI studies will provide a handful

of highly significant targets for improved antimicrobial development. Hence, prioritization of protein targets was performed.

Subcellular localization and virulence prediction

Consensus results of the localization predictions made by PSORTb and CELLO2GO tools of the 17 highly expressed protein target sequences is reported in Table 2. Results reveal that most of the query proteins were present in the inner membrane i.e. 59%. Besides, 35% of proteins were predicted to localize in cytoplasm. Only 6% of proteins were predicted to reside in the periplasm of the pathogen. However, no proteins displayed outer membrane or extracellular localization (Fig. 2).

A total of 7 proteins (1 cytoplasmic and 6 membranous) were identified as virulent and the remaining 10 proteins appeared as non-virulent during the MP3 server prediction procedure (Table 2).

Novel targets identification and mining PPIs

Among 17 targets, 7 were found as novel as evidenced by 'no hits found' during the BLAST analysis against the 2933 experimental drug targets of DrugBank Database. The results further reveal 3 cytoplasm localizing proteins (Q8XZJ3, Q8XU97 and Q8Y1J8) and 4 membrane localizing proteins (Q8XU12, Q8Y253, Q8XUB6, Q8XVG8) as highly potential novel protein targets present in *R. solanacearum* GMI1000 bacterium (Table 2).

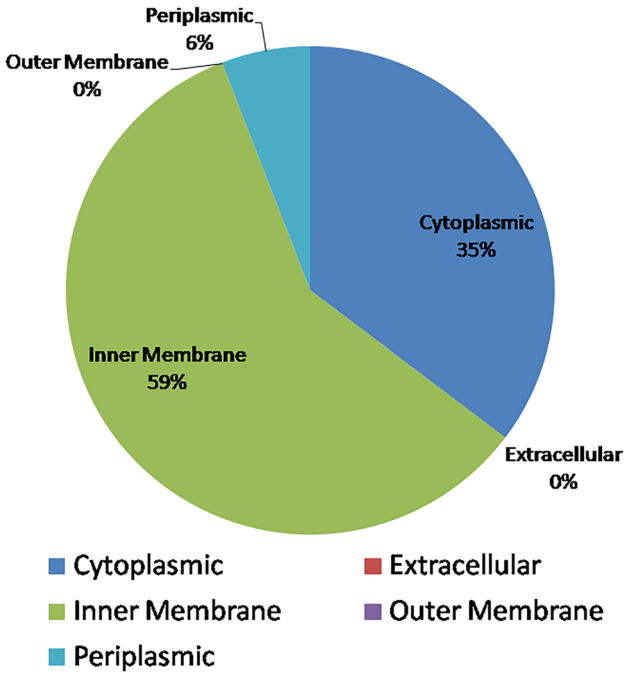

Fig. 2 Subcellular Localization of 17 Plausible Antibacterial Targets in *Ralstonia solanacearum* GMI1000

Table 2 Results of Subcellular Localization, Virulence Factor and DrugBank Database Analyses of *R. solanacearum* GMI1000 17 significant essential proteins

Protein ID	Subcellular Localization	Virulent (V) / Non-virulent (NV)	DrugBank Target Id / Details
8XZJ3	Cytoplasm	NV	No hits
Q8XU97	Cytoplasm	NV	No hits
Q8Y1J8	Cytoplasm	NV	No hits
Q8XU12	Inner Membrane	V	No hits
Q8Y253	Inner Membrane	V	No hits
Q8XUB6	Inner Membrane	NV	No hits
Q8XVG8	Inner Membrane	NV	No hits
Q8Y1T3	Inner Membrane	V	P02941 / Methyl-accepting chemotaxis protein II
Q8XZE1	Inner Membrane	NV	P02941 / Methyl-accepting chemotaxis protein II
Q8Y089	Inner Membrane	V	P02941 / Methyl-accepting chemotaxis protein II
Q8XTY3	Inner Membrane	NV	P02941 / Methyl-accepting chemotaxis protein II
Q8XWC4	Cytoplasm	NV	P08373 / UDP-N-acetylenolpyruvoylglucosaminereductase
Q8XW15	Cytoplasm	NV	P0AB91 / Phospho-2-dehydro-3-deoxyheptonate aldolase
Q8XZ75	Periplasm	V	P0AG82 / Phosphate-binding protein PstS
Q8XUS0	Cytoplasm	V	P37093 / Type II secretion system protein E, Q7BK04 / Cag alpha
Q8XZW9	Inner Membrane	NV	Q5SJ79 / Cytochrome c oxidase subunit 1
8Y3B8	Inner Membrane	V	Q70KI2 / Penicillin-binding protein 2, P07944 / Beta-lactam-inducible penicillin-binding protein, P0A3M6 / Penicillin-binding protein 2B, P59676 / Penicillin-binding protein 2X

All proteins investigated using STRING database (Szklarczyk et al. 2019) were found to have PPIs with a maximum value of 0.939 as the average local clustering coefficient. Among the novel candidates analyzed, those with an average node degree (K) > = 5 were shortlisted as hub proteins (Vallabhajosyula et al. 2009). Thus, three proteins namely Q8XU97, Q8Y1J8 and Q8XVG8 were identified as hub proteins (Table 3, Fig. 3).

Maximum average node degree (K) value of 7.27 was observed for Q8XU97 protein. It is a probable l-serine dehydratase enzyme (E.C. 4.3.1.7) (aka serine deaminase) molecule encoded by sdaA2 gene. This 458 aminoacids containing cytoplasmic protein exhibits both lyase and ion binding capacity. Q8Y1J8 (probable lipopolysaccharide heptosyltransferase) is the second best hub protein revealed in this study. This cytoplasm localizing molecule is encoded by rfaC1 gene present in *R. solanacearum* GMI1000 and is involved in the biosynthesis of LPS core domain. Of the three hub proteins identified, Q8XVG8 is a transmembrane protein of the tomato pathogen under study. The three hub proteins reported can be considered as highly significant novel protein targets of *R. solanacearum* GMI1000 strain.

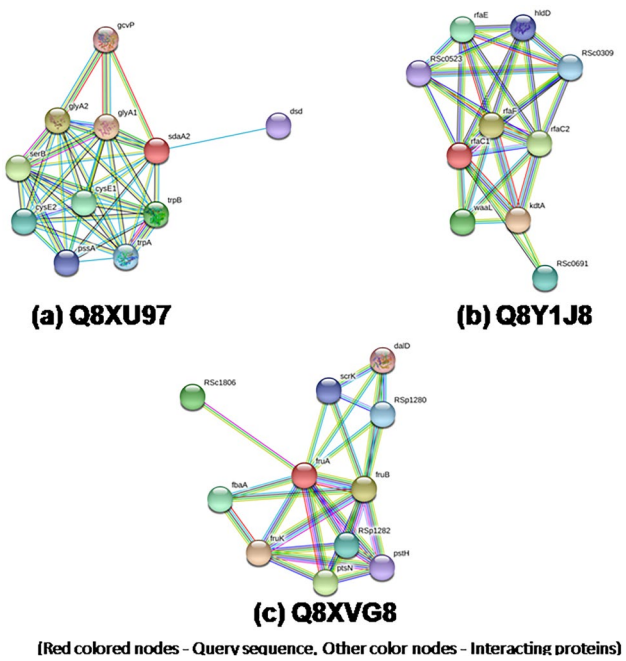

Fig. 3 Network Topology of Top Three Novel Potential Antibacterial Target Proteins (Hub Proteins) identified in *Ralstonia solanacearum* GMI1000

Discussion

Bacterial diseases account for a 20% (approx.) overall loss in plant productivity (Tiwari et al. 2017). Present approaches in dealing virus-, fungi- and nematode-borne plant diseases are far superior than the methods of yester years. However, same pace was not observed in the improvements regarding strategies to control bacterial plant diseases (Sundin et al. 2016). *R. solanacearum* strains possess peculiar ability to infect a wide range of plant species, exhibit unusual survival capacity in extreme environments and are genetically diverse (Tahat and Sijam 2010; Peeters et al. 2013). Furthermore, lack of suitable bactericide against *R. solanacearum* strains have challenged their disease management strategies (Tahat and Sijam 2010). Hence, developing appropriate therapeutics to control bacterial wilt disease caused by *R. solanacearum* GMI1000 depends on determining antibacterial targets in the pathogen. In the present study, in silico genomics subtraction integrated with other knowledge-based approaches such as codon usage and PPI analyses was carried out on *R.*

Table 3 Results of Protein–Protein Interactions (PPI) Analyses of the 7 highly plausible novel antibacterial targets of *R. solanacearum* GMI1000

Protein ID / Name	No. of Nodes	No. of Edges	Average Node Degree	Average Local Clustering Coefficient
Q8XZJ3 / GLND_RALSO Bifunctional Uridylyltransferase/Uridylyl-removing Enzyme	7	9	2.57	0.886
Q8XU97 / Probable L-Serine Dehydratase 2 (L-serine deaminase) Protein	11	40	7.27	0.939
Q8Y1J8 / Probable Lipopolysaccharide Heptosyltransferase Protein	10	30	6	0.871
Q8XU12 / Potassium-transporting ATPase potassium-bindingSubunit	7	13	3.71	0.924
Q8Y253 / Probable Glutamate/Aspartate TransmembraneABC Transporter protein	11	25	4.55	0.83
Q8XUB6 / C4-Dicarboxylate Transport Protein 2	5	6	2.4	0.867
Q8XVG8 / Probable PTS system, Fructose-specific IIbc component (EIIbc-fru) (Fructose-permeaseIIbc component) Transmembrane Protein	11	28	5.09	0.882

solanacearum GMI1000 genome to predict novel antibacterial targets. This combined approach attempted is different from earlier methods as it has multilevel stringent prioritization processes (Katara et al. 2012; Keshri et al. 2014; Prabha et al. 2019) and can be considered as a promising strategy in finding novel targets.

During the subtractive stage of analysis, orthologs, paralogs and host-specific homologs were identified and excluded resulting in 2760 non-host homologs. Orthologs are less likely to be essential (Barh et al. 2011; Vetrivel et al. 2011; Hossain et al. 2016, 2017), paralogs being redundant as antimicrobial targets (Shoukat et al. 2012) and host-specific homologs may initiate adverse reactions, if chosen as targets (Barh et al. 2011). Earlier, Katara et al. (2012) found only 158 sequences as non-homologs in *Pseudomonas syringae* pv. *phaseolicola.* Whereas, 406 non-homologs were located in *Xanthomonas oryzae* pv.*oryzae* PX099A (Keshri et al. 2014). Recently, Prabha et al. (2019) identified 152 proteins as non-homologs during their study with *Xanthomonas oryzae pv. oryzae* KACC10331. A few hundred to thousand proteins are reported as non-host homologs in different Gram-negative bacteria (Barh et al. 2011; Vetrivel et al. 2011; Shoukat et al. 2012; Mondal et al. 2015; Hossain et al. 2016, 2017; Folador et al. 2016; Sadhasivam and Vetrivel 2018; Shuvo et al. 2019). In the final step of subtractive analysis, 115 essential proteins are revealed, similar to results seen in *Pseudomonas syringae pv. phaseolicola* (Katara et al. 2012) and *Pseudomonas aeruginosa* (Sakharkar et al. 2004). Less than 40 proteins were reported as essential in two rice pathogenic strains of *Xanthomonas* genus (Keshri et al. 2014; Prabha et al. 2019). Variation in the number of essential proteins are documented in different Gram-negative bacteria (Barh et al. 2011; Vetrivel et al. 2011; Shoukat et al. 2012; Mondal et al. 2015; Hossain et al. 2016, 2017; Folador et al. 2016; Sadhasivam and Vetrivel 2018; Shuvo et al. 2019). Variations in the number of non-homologs and/or essential protein sequences identified in earlier studies could be attributed to the variety of experimental parameters and/or computational protocols employed by the respective investigators.

In the present study, 105 essential proteins involved in 46 biochemical pathways are revealed during KAAS analysis. Involvement of essential proteins in varied pathways such as primary metabolism, cell signalling and resistance conferring are either directly or indirectly linked to virulence factor, pathogenesis, nutrient mobilization/uptake and motility of the organism (Rodriguez and Smith 2006; Maranhão et al. 2009).

Several investigators have earlier reported essential proteins as therapeutic targets during their in silico genomics analysis (Barh et al. 2011). However, additional filters reveal better targets (Barh et al. 2011; Vetrivel et al. 2011; Katara et al. 2012; Keshri et al. 2014; Hosen et al. 2014; Sivashanmugam et al. 2015; Mondal et al. 2015; Hossain et al. 2016, 2017; Uddin and Jamil 2018; Sadhasivam and Vetrivel 2018; Prabha et al. 2019). Thus, similar to earlier approaches, additional factors such as pathogen-specific biochemical pathways, higher expressivity, subcellular localization, virulence and druggablility of essential proteins were also determined to narrow potential targets in *R. solanacearum* GMI1000.

In this study, 24 unique metabolic pathways involved essential proteins can be regarded as plausible protein targets. These findings are consistent with those reported in *Pseudomonas syringae* pv *phaseolicola* (Katara et al. 2012). Less than 50 unique pathway essential proteins are recorded by several researchers in various Gram-negative bacterial species (Barh et al. 2011; Vetrivel et al. 2011; Mondal et al. 2015; Hossain et al. 2016, 2017; Uddin and Jamil 2018; Shuvo et al. 2019). As in several previous studies, essential proteins involved in common pathways were identified (Barh et al. 2011; Hossain et al. 2016, 2017; Shuvo et al. 2019) and excluded because they may initiate cross reactivity in host cells.

Subtractive genomics coupled with codon usage analysis have revealed significant therapeutic targets in bacterial pathogens of humans (Vetrivel et al. 2011; Sadhasivam and Vetrivel 2018). Thus, CAI values were calculated and the significant variations in it (0.569—0.895) suggests probable selective pressure among essential genes of the bacterium. In this study, 17 essential proteins were shortlisted as significant based on their CAI value. Twenty two highly expressed genes and their products were reported as plausible drug targets in the bean pathogen *Pseudomonas syringae pv phaseolicola* by Katara et al. (2012) following codon biasing studies. To the best of our knowledge, effective utilization of codon usage coupled subtractive genomics to analyze other phytopathogenic bacteria is not available in the literature.

Further prioritization of the 17 highly expressed essential protein targets in *R. solanacearum* GMI1000 was done based on subcellular localization, virulence, DrugBank search and PPIs analysis. In 2012, Katara and co-workers have reported intracellular localization of few proteins in *Pseudomonas syringae*. However, there is no mention regarding either the proteins or their exact intracellular positions in the bean pathogen. No other literature pertaining to computational prediction of bacterial phytopathogen proteins' subcellular localization is available till date. Nevertheless, many researchers have successfully utilized in silico intracellular prediction approaches among gram-negative human bacterial pathogens (Barh et al. 2011; Mondal et al. 2015; Hossain et al. 2016, 2017; Uddin and Jamil 2018; Shuvo et al. 2019). Many of them preferred more than one tool for their prediction so that results are more reliable. Also, several investigators have predicted target proteins at least in one intracellular location in Gram-negative bacteria studied by them. Keshri et al. (2014) reported three virulent proteins in rice pathogen *Xanthomonas oryzae* pv

oryzae PXO99A. A recent investigation on another rice infecting strain *Xanthomonas oryzae* pv *oryzae* KACC10331, revealed four proteins as virulent (Prabha et al. 2019).

Our results recognize 7 proteins of *R. solanacearum* GMI1000 as novel based on DEG analysis. Earlier, two proteins and three proteins were identified as novel targets in the bacterial pathogens affecting bean (Katara et al. 2012) and rice (Keshri et al. 2014), respectively. Similarly, a maximum of 20 novel targets in each Gram-negative human bacterial pathogen were reported in earlier studies (Barh et al. 2011; Vetrivel et al. 2011; Mondal et al. 2015; Uddin and Jamil 2018; Sadhasivam and Vetrivel 2018; Shuvo et al. 2019).

In this study, PPIs network of the seven novel targets were analyzed using STRING database (Szklarczyk et al. 2019), as PPIs play a key role in predicting the druggability of the protein targets. Through topological analysis, PPIs also recognize hub proteins (possessing a large number of interacting partners), which are likely to be served by essential proteins in an organism (Vallabhajosyula et al. 2009; Hosen et al. 2014; Raman et al. 2014; Uddin and Jamil 2018). Thus, in the lines of results reported earlier (Hosen et al. 2014; Uddin and Jamil 2018), three proteins namely Q8XU97, Q8Y1J8 and Q8XVG8 are listed as hub proteins on the basis of average node degree (K) > = 5 (Vallabhajosyula et al. 2009). These three hub proteins are suggested as the best plausible novel targets in *R. solanacearum* GMI1000, since obstructing or removing hub proteins tend to adversely affect pathogen's survival (centrality-lethality rule) (He and Zhang 2006; Zoraghi and Reiner 2013).

Serine deaminase (Q8XU97) catalyzes deamination of serine into pyruvic acid in gluconeogenesis process (Salanoubat et al. 2002; Moriya et al. 2007; Yu et al. 2014). L-Serine deaminase is shown essential for survival of several bacterial genera. Inhibitors against serine deaminase leads to serine molecule build up inside bacterial cells and eventually, their high concentration will be detrimental to the bacterium (Thoden et al. 2014). According to Barh et al. (2011), nearly 70% targets commonly reported in bacterial pathogens are associated to their unique pathways that includes LPS biosynthetic and PTS pathways. Our study results reveal Q8Y1J8 as the second best hub protein of the tomato pathogen. It is involved in converting LPS into heptosylated LPS in the inner core region of LPS (Salanoubat et al. 2002; Moriya et al. 2007; Yu et al. 2014). Absence of heptose in LPS region renders the bacterium weak and increases its susceptibility to antibacterial agents (Gronow and Brade 2001). Q8XVG8 is a 573 aminoacids containing protein product of FruA gene (Salanoubat et al. 2002; Moriya et al. 2007; Yu et al. 2014). It is one among several molecules associated with unique pathways of pathogen identified to be promising targets (Barh et al., 2011). Thus, these three proteins form the best plausible targets in *R. solanacearum* GMI1000.

An integrated computational approach encompassing primarily of subtractive genomics, codon usage and PPI analyses for the identification of antibacterial protein targets in *R. solanacearum* GMI1000 was attempted in this research study. Overall results propose three proteins (Q8XU97, Q8Y1J8 and Q8XVG8) as the best plausible targets of *R. solanacearum* GMI1000. Thus, therapeutics targeting either of these proteins can be exploited in order to manage bacterial wilt disease of tomato caused by *R. solanacearum* strain GMI1000. Nevertheless, validating these interesting targets by experimental studies is also recommended. The results of this study are likely to gain significance among the active agricultural genomics researchers of today's era. Applying the stringent multilevel in silico methodology discussed above on several other economically important phytopathogenic bacterial members may reveal their significant protein targets as well, thereby enabling identification of suitable antimicrobials too.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s13313-021-00845-6.

Acknowledgements This research work is supported by a financial grant from the University Grants Commission (UGC), India in the form of UGC Minor Research Project (MRP-6785/16). The authors thank the anonymous reviewers for their valuable inputs.

Funding University Grants Commission, MRP-6785/16 (GURUNATHAN SUBRAMANIAN)

References

Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Barh D, Tiwari S, Jain N et al (2011) In silico subtractive genomics for target identification in human bacterial pathogens. Drug Dev Res 72:162–177. https://doi.org/10.1002/ddr.20413

Elphinstone JG (2005) The current bacterial wilt situation: a global overview. In: Allen C, Piror P, Hayward AC (eds) Bacterial wilt disease and the Ralstonia solanacearum complex. APS press, Minnesota USA, pp 9–28

Folador EL, de Carvalho PVSD, Silva WM et al (2016) In silico identification of essential proteins in Corynebacterium pseudotuberculosis based on protein-protein interaction networks. BMC Syst Biol 10:103. https://doi.org/10.1186/s12918-016-0346-4

Galperin MY, Koonin EV (1999) Searching for drug targets in microbial genomes. Curr Opin Biotechnol 10:571–578. https://doi.org/10.1016/S0958-1669(99)00035-X

Gronow S, Brade H (2001) Lipopolysaccharide biosynthesis: which steps do bacteria need to survive? J Endotoxin Res 7:3–23

Gupta A, Kapil R, Dhakan DB, Sharma VK (2014) MP3: A Software Tool for the Prediction of Pathogenic Proteins in Genomic and Metagenomic Data. PLoS ONE 9:e93907. https://doi.org/10.1371/journal.pone.0093907

Gupta V, Haider S, Sood U et al (2016) Comparative genomic analysis of novel Acinetobacter symbionts: A combined systems biology and genomics approach. Sci Rep 6:29043. https://doi.org/10.1038/srep29043

He LY, Sequeira L, Kelman A (1993) Characteristics of strains of Pseudomonas solanacearum from China. Plant Dis 67:1357–1361

He X, Zhang J (2006) Why Do Hubs Tend to Be Essential in Protein Networks? PLoS Genet 2:e88. https://doi.org/10.1371/journal.pgen.0020088

Hosen MdI, Tanmoy AM, Mahbuba D-A et al (2014) Application of a subtractive genomics approach for in silico identification and characterization of novel drug targets in Mycobacterium tuberculosis F11. Interdiscip Sci Comput Life Sci 6:48–56. https://doi.org/10.1007/s12539-014-0188-y

Hossain MU, Khan MdA, Hashem A et al (2016) Finding Potential Therapeutic Targets against Shigella flexneri through Proteome Exploration. Front Microbiol 7.https://doi.org/10.3389/fmicb.2016.01817

Hossain T, Kamruzzaman M, Choudhury TZ et al (2017) Application of the Subtractive Genomics and Molecular Docking Analysis for the Identification of Novel Putative Drug Targets against *Salmonella enterica* subsp. *enterica serovar* Poona. Biomed Res Int 2017:1–9. https://doi.org/10.1155/2017/3783714

Huang Y, Niu B, Gao Y et al (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26:680–682. https://doi.org/10.1093/bioinformatics/btq003

Jadhav A, Shanmugham B, Rajendiran A, Pan A (2014) Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis. Infect Genet Evol 27:300–308. https://doi.org/10.1016/j.meegid.2014.08.007

Katara P, Grover A, Sharma V V (2012) *In Silico* Prediction of Drug Targets in Phytopathogenic *Pseudomonas syringae* pv. phaseolicola: Charting a Course for Agrigenomics Translation Research. OMICS A J Integr Biol 16:700–706. https://doi.org/10.1089/omi.2011.0141

Keshri V, Singh DP, Prabha R et al (2014) Genome subtraction for the identification of potential antimicrobial targets in Xanthomonas oryzae pv. oryzae PXO99A pathogenic to rice. 3 Biotech 4:91–95. https://doi.org/10.1007/s13205-013-0131-7

Lin J, Qian J (2007) Systems biology approach to integrative comparative genomics. Expert Rev Proteomics 4:107–119. https://doi.org/10.1586/14789450.4.1.107

Luo H, Lin Y, Gao F et al (2014) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements: Table 1. Nucl Acids Res 42:D574–D580. https://doi.org/10.1093/nar/gkt1131

Maranhão FCA, Paião FG, Fachin AL, Martinez-Rossi NM (2009) Membrane transporter proteins are involved in Trichophyton rubrum pathogenesis. J Med Microbiol 58:163–168. https://doi.org/10.1099/jmm.0.002907-0

Mondal SI, Ferdous S, Akter A et al (2015) Identification of potential drug targets by subtractive genome analysis of Escherichia coli O157:H7: an in silico approach. AABC 49.https://doi.org/10.2147/AABC.S88522

Moriya Y, Itoh M, Okuda S et al (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185. https://doi.org/10.1093/nar/gkm321

Peeters N, Guidot A, Vailleau F, Valls M (2013) Ralstonia solanacearum, a widespread bacterial plant pathogen in the post-genomic era: Ralstonia solanacearum and bacterial wilt disease. Mol Plant Pathol 14:651–662. https://doi.org/10.1111/mpp.12038

Petschnigg J, Snider J, Stagljar I (2011) Interactive proteomics research technologies: recent applications and advances. Curr Opin Biotechnol 22:50–58. https://doi.org/10.1016/j.copbio.2010.09.001

Prabha R, Singh DP, Ahmad K et al (2019) Subtractive genomics approach for identification of putative antimicrobial targets in *Xanthomonas oryzae* pv. *oryzae* KACC10331. Arch Phytopathol Plant Protect 52:863–872. https://doi.org/10.1080/03235408.2018.1562674

Raman K, Damaraju N, Joshi GK (2014) The organisational structure of protein networks: revisiting the centrality–lethality hypothesis. Syst Synth Biol 8:73–81. https://doi.org/10.1007/s11693-013-9123-5

Rodriguez GM, Smith I (2006) Identification of an ABC Transporter Required for Iron Acquisition and Virulence in Mycobacterium tuberculosis. J Bacteriol 188:424–430. https://doi.org/10.1128/JB.188.2.424-430.2006

Sadhasivam A, Vetrivel U (2018) Genome-wide codon usage profiling of ocular infective *Chlamydia trachomatis* serovars and drug target identification. J Biomol Struct Dyn 36:1979–2003. https://doi.org/10.1080/07391102.2017.1343685

Sakharkar KR, Sakharkar MK, Chow VTK (2004) A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biol (gedrukt) 4:355–360

Salanoubat M, Genin S, Artiguenave F et al (2002) Genome sequence of the plant pathogen Ralstonia solanacearum. Nature 415:497–502. https://doi.org/10.1038/415497a

Sharp PM, Li WH (1987) The Codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281–1295. https://doi.org/10.1093/nar/15.3.1281

Shoukat K, Rasheed N, Sajid M (2012) Subtractive genome analysis for in silico identification and characterization of novel drug targets in C. trachomatis STRAIN D/UW-3/Cx. Int J Curr Res 4:17–21

Shuvo MSR, Shakil SK, Ahmed F (2019) Potential Drug Target Identification of Legionella pneumophila by Subtractive Genome Analysis: An In Silico Approach. Bangla J Microbiol 35:102–107. https://doi.org/10.3329/bjm.v35i2.42638

Sivashanmugam M, Nagarajan H, Vetrivel U et al (2015) In silico analysis and prioritization of drug targets in Fusarium solani. Med Hypotheses 84:81–84. https://doi.org/10.1016/j.mehy.2014.12.015

Sundin GW, Wang N, Charkowski AO et al (2016) Perspectives on the Transition From Bacterial Phytopathogen Genomics Studies to Applications Enhancing Disease Management: From Promise to Practice. Phytopathology 106:1071–1082. https://doi.org/10.1094/PHYTO-03-16-0117-FI

Szklarczyk D, Gable AL, Lyon D et al (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613. https://doi.org/10.1093/nar/gky1131

Tahat MM, Sijam K (2010) *Ralstonia solanacearum*: The Bacterial Wilt Causal Agent. Asian J of Plant Sciences 9:385–393. https://doi.org/10.3923/ajps.2010.385.393

Thoden JB, Holden HM, Grant GA (2014) Structure of l-Serine Dehydratase from Legionella pneumophila: Novel Use of the C-Terminal Cysteine as an Intrinsic Competitive Inhibitor. Biochemistry 53:7615–7624. https://doi.org/10.1021/bi501253w

Tiwari S, Awasthi M, Pandey VP, Dwivedi UN (2017) Genomics Based Approaches towards Management of Plant Diseases with Emphasis on in silico Methods as a Prudent Approach. J Agri Sci Food Technol 3:39–51

Uddin R, Jamil F (2018) Prioritization of potential drug targets against P. aeruginosa by core proteomic analysis using computational subtractive genomics and Protein-Protein interaction network. Comput Biol Chem 74:115–122. https://doi.org/10.1016/j.compbiolchem.2018.02.017

Vallabhajosyula RR, Chakravarti D, Lutfeali S et al (2009) Identifying Hubs in Protein Interaction Networks. PLoS ONE 4:e5344. https://doi.org/10.1371/journal.pone.0005344

Vetrivel U, Subramanian G, Dorairaj S (2011) A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. HUGO J 5:25–34. https://doi.org/10.1007/s11568-011-9152-7

Wishart DS, Feunang YD, Guo AC et al (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46:D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Wu G, Culley DE, Zhang W (2005) Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. Microbiology 151:2175–2187. https://doi.org/10.1099/mic.0.27833-0

Yu C-S, Cheng C-W, Su W-C et al (2014) CELLO2GO: A Web Server for Protein subCELlular LOcalization Prediction with Functional Gene Ontology Annotation. PLoS ONE 9:e99368. https://doi.org/10.1371/journal.pone.0099368

Yu NY, Wagner JR, Laird MR et al (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. https://doi.org/10.1093/bioinformatics/btq249

Zoraghi R, Reiner NE (2013) Protein interaction networks as starting points to identify novel antimicrobial drug targets. Curr Opin Microbiol 16:566–572. https://doi.org/10.1016/j.mib.2013.07.010