



A Web Resource for Designing Subunit Vaccine Against Major Pathogenic Species of Bacteria

Gandharva Nagpal^{1,2†}, Salman Sadullah Usmani^{1,3†} and Gajendra P. S. Raghava^{1,3*}

¹ Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India, ² Centre for Bioinformatics, Computational and Systems Biology, Pathfinder Research and Training Foundation, Greater Noida, India, ³ Center for Computational Biology, Indraprastha Institute of Information Technology, Okhla, India

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Giampiero Pietrocola,
University of Pavia, Italy
Ghita Ghislat,
INSERM U1104 Centre
d'immunologie de Marseille-Luminy,
France

*Correspondence:

Gajendra P. S. Raghava
raghava@iiitd.ac.in

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 29 June 2018

Accepted: 13 September 2018

Published: 02 October 2018

Citation:

Nagpal G, Usmani SS and
Raghava GPS (2018) A Web Resource
for Designing Subunit Vaccine Against
Major Pathogenic Species of Bacteria.
Front. Immunol. 9:2280.
doi: 10.3389/fimmu.2018.02280

Evolution has led to the expansion of survival strategies in pathogens including bacteria and emergence of drug resistant strains proved to be a major global threat. Vaccination is a promising strategy to protect human population. Reverse vaccinology is a more robust vaccine development approach especially with the availability of large-scale sequencing data and rapidly dropping cost of the techniques for acquiring such data from various organisms. The present study implements an immunoinformatic approach for screening the possible antigenic proteins among various pathogenic bacteria to systemically arrive at epitope-based vaccine candidates against 14 pathogenic bacteria. Thousand four hundred and fifty nine virulence factors and Five hundred and forty six products of essential genes were appraised as target proteins to predict potential epitopes with potential to stimulate different arms of the immune system. To address the self-tolerance, self-epitopes were identified by mapping on 1000 human proteome and were removed. Our analysis revealed that 21 proteins from 5 bacterial species were found as virulent as well as essential to their survival, proved to be most suitable vaccine target against these species. In addition to the prediction of MHC-II binders, B cell and T cell epitopes as well as adjuvants individually from proteins of all 14 bacterial species, a stringent criteria lead us to identify 252 unique epitopes, which are predicted to be T-cell epitopes, B-cell epitopes, MHC II binders and Vaccine Adjuvants. In order to provide service to scientific community, we developed a web server VacTarBac for designing of vaccines against above species of bacteria. This platform integrates a number of tools that includes visualization tools to present antigenicity/epitopes density on an antigenic sequence. These tools will help users to identify most promiscuous vaccine candidates in a pathogenic antigen. This server VacTarBac is available from URL (<http://webs.iiitd.edu.in/raghava/vactarbac/>).

Keywords: reverse vaccinology, vaccine designing, immunotherapeutic, epitopes, antigen, virulence factor, essential genes

INTRODUCTION

Evolution of existing bacterial pathogens and emergence of new pathogenic strains are continuously causing problems to mankind. To address the pathogenic challenges to the human health, researchers developed various vaccines and antibiotics during the twentieth century. A worldwide usage of such therapeutic strategies led to the expansion of structural as well as

behavioral properties in various bacterial species. Diverse defense mechanisms were acquired by the known bacterial species that resulted in antibiotic tolerance or drug resistance, currently a major concern during administration of known antibiotics in pathogenic conditions. Lord Jim O'Neil and his team estimated that antimicrobial resistance could cause 10 million deaths a year by 2050 (1).

Vaccination has proved to be a promising strategem to protect human population from dreadful diseases like smallpox, polio, etc. Yet, vaccines are currently unavailable for many infectious diseases such as melioidosis caused by *Burkholderia pseudomallei* (2). Even among the existing vaccines, BCG which is heavily used vaccine against tuberculosis, is less effective in the immunocompromised and adults in tropical and subtropical region (3) and its efficacy is rapidly decreasing (4). Anthrax, DPT, Hib, Meningococcal vaccine, Pneumococcus heptavalent conjugate polysaccharide vaccine (PCV7), Pneumococcus 13-valent conjugate polysaccharide vaccine (PCV13) etc are generally considered effective at controlling disease symptoms. However, a major concern of recent vaccination programs tends to be the preference of vaccine escape mutants, i.e., the repertoire of immune targets in the pathogen is different from that used in the vaccine formulation (5). Antigenic variation and continuous bacterial evolution has necessitated designing of better vaccine candidates.

Virulence factors (VFs) have been explored to design vaccines against several pathogens such as Porcine enteric coronaviruses (CoVs) (6), *Vibrio cholerae*, enterotoxigenic *Escherichia coli* (7) etc. Bacterial virulence factors (VFs) are captivating, as they play a major role in establishment of infection to cause pathological conditions while surviving in a hostile environment (8). Virulence factors are thus defined as the gene products enabling the pathogen to enter, replicate, and persist in a host even in a small inoculum. These include cell surface proteins that facilitate bacterial attachment, cell surface carbohydrates which protect a bacterium, bacterial toxins, and hydrolytic enzymes that subsidize pathogenicity of bacterium (8).

In addition to virulence factors, genes essential for the living of the bacterial cell, have proved to be attractive drug targets in various pathogens such as the fungus *Aspergillus fumigatus* (9, 10), protozoan *Leishmania* species, (11) as well as bacteria like *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* (11). Essential genes are indispensable for cellular metabolism.

The last two decades witnessed an embracement of a more revolutionary vaccine development approach, namely the information driven “reverse vaccinology” which involves identification of vaccine candidates by genome or proteome sequence analysis as the primary step. High cost and practical limitations of feasibility and time have engendered a predilection toward computational techniques for rational reduction in the number of vaccine candidates to be experimentally tested. The epitope prediction methods have made possible searching of sequence sets as large as the whole proteomes of organisms with various restriction filters to arrive at novel vaccine candidates.

In order to facilitate the scientific community, researchers have developed numerous pipelines, databases, and tools for

identifying vaccine candidates against specific types of diseases (12, 13). Recently, researchers have proposed potential vaccine candidates against diverse pathogens like *Ebola* (14), *Zika* (15), *M. tuberculosis* (16), *Shigella* (17), *Hepatitis B Virus* (18), *Helicobacter pylori* (19), *Vibrio cholera* (20), *Dengue* (21), etc. using *in silico* predictions. These predicted candidates help reduce the burden in experimental labs for the researchers, thus saving time and cost of vaccine development. Most of the above-mentioned studies only focus on a single pathogen and lacked identification of adjuvant candidates.

In this study, we have systemically arrived at epitope based vaccine candidate against major pathogenic bacterial species. **Table 1** provides a summary of the major pathogenic manifestations of the bacterial species considered in the present study. Both, the virulence factors and the essential genes, have been considered to predict epitopes. It is a well-known fact that, diverse bacterial species and strains have evolved multiple mechanisms for evading the host immune responses, not only by producing likes of host inhibition pathway receptor ligands (“molecular mimicry”) but also by producing virulence factors resembling the intermediates of the host inhibitory pathways. Hence, peptides found in proteomes of the 1000 human genomes were removed while implementing the pipeline. To assist the scientific community all the potential vaccine candidates have been compiled and presented in the form of a database (<http://webs.iitd.edu.in/raghava/vactarbac/>).

MATERIALS AND METHODS

Raw Data Source

Proteins were taken as raw data to predict the potential vaccine candidates. We considered both virulence factors as well as essential genes of pathogenic bacteria.

Virulence Factors

The Virulence Factor Database (VFDB) is a resource for housing the Virulence Factors (VFs) that enable a microorganism to establish itself on or within a host thus bestowing the pathogen with a potential to cause disease (22). VFDB provides two hierarchical non-redundant datasets, i.e., sets A and B. Set A is a core dataset, covering genes associated with experimentally verified virulence factors. Therefore, we have selected protein sequences of core dataset A for further analysis. The link http://www.mgc.ac.cn/VFs/Down/VFDB_setA_pro.fas.gz provides protein sequences of the VFDB database in FASTA format when uncompressed using the command “gunzip” on a unix command line or using the tool “Winzip” on a Windows platform.

Essential Genes/ Proteins

Database of Essential Genes (DEG) maintains currently known essential genomic elements, such as protein-coding genes and non-coding RNAs, among bacteria, archaea, and eukaryotes (22). These genes are mandatory for survival of the corresponding bacterium. DEG contains 18286 proteins encoded from genes experimentally found to be essential for the survival of the 32 bacteria. For our study, we restricted the analysis of bacterial

TABLE 1 | Bacterial species considered in the study with the pathological conditions caused by them.

S. No.	Bacterial Species	Human Disease
1	<i>Bacillus subtilis</i>	food poisoning
2	<i>Burkholderia pseudomallei</i>	Melioidosis
3	<i>Campylobacter jejuni</i>	Campylobacteriosis
4	<i>Escherichia coli</i>	cholecystitis, bacteremia, cholangitis, urinary tract infection (UTI), and traveler's diarrhea, and other clinical infections such as neonatal meningitis and pneumonia
5	<i>Haemophilus influenzae</i>	pneumonia, bacteremia, meningitis, epiglottitis, septic arthritis, cellulitis, otitis media, and purulent pericarditis
6	<i>Helicobacter pylori</i>	ulcers in the stomach and small intestine
7	<i>Mycobacterium tuberculosis</i>	Tuberculosis
8	<i>Pseudomonas aeruginosa</i>	urinary tract infections, respiratory system infections, dermatitis, soft tissue infections, bacteremia, bone and joint infections, gastrointestinal infection
9	<i>Salmonella enteric</i>	Food poisoning
10	<i>Staphylococcus aureus</i>	pneumonia, meningitis, osteomyelitis, endocarditis, toxic shock syndrome, bacteremia, and sepsis
11	<i>Streptococcus agalactiae</i>	neonatal infection, septicemia, pneumonia, meningitis, chorioamnionitis
12	<i>Streptococcus pneumoniae</i>	pneumonia (infection of the lungs), ear infections, sinus infections, meningitis (infection of the covering around the brain and spinal cord), and bacteremia (blood stream infection)
13	<i>Streptococcus pyogenes</i>	streptococcal pharyngitis, rheumatic fever, rheumatic heart disease, and scarlet fever
14	<i>Vibrio cholera</i>	Cholera

essential genes belonging to 14 species that have also been included in VFDB, for the convenience of comparison and combining with VFDB analysis.

Target Selection for Vaccine Identification

The most important aspect of subunit vaccine designing is vaccine target selection. In case of data downloaded from VFDB (set A), it contains 2595 proteins (11) from 46 bacterial species. We have matched these bacterial species into DEG database that resulted into 14 common bacterial species. Therefore, our dataset contains only those bacterial species, which have at least one virulence factor and one protein encoded by essential genes from VFDB and DEG respectively. DEG contains 9664 proteins corresponding to these pathogenic bacteria. These are the essential proteins of pathogenic bacteria. Some specific bacterial components, because of their function proved to be better possible candidate for designing subunit vaccine (20, 23–25) therefore; we have selected some specific significant proteins.

Membrane Protein

These are the protein emerged from the membrane of various pathogenic bacteria. There are evidences that membrane proteins

mediate pathogen entry and colonization and are also the primary target of adaptive immune response (24). A total of 420 membranous proteins were selected from DEG for further processing.

Envelope Protein

These proteins are derived from the envelope of various pathogenic bacteria and are essential for their survival. These are important for attachment to host cell surface and also responsible for facilitating an immune response in the host cell (25). A total of 36 proteins were screened for further processing.

Secretion Protein

Fifty-two secretory proteins corresponding to bacterial species of our interest were deposited in DEG and play important role in their survival. Secretory system of pathogenic bacteria has been heavily exploited for screening of new possible vaccine (26, 27). These proteins play many roles in promoting bacterial virulence, from enhancing attachment to eukaryotic cells, to scavenging resources in an environmental niche, to directly intoxicating target cells and disrupting their functions (28).

Repair Protein

Bacterial damage repair mechanisms have broader roles encompassing responses to virulence as well as stress (29). Therefore we have also considered 38 proteins associated with the repair mechanism of various pathogenic bacteria.

Epitope Prediction Pipeline Generation of Nona-Peptides

In order to design new vaccine candidate, first we have generated nona-peptides. Nona-peptides were 9-mer sequences (9 residues continuous stretch of peptide) originated from the essential or virulent proteins selected for the study. In order to avoid redundancy, we have removed all duplicated nona-peptides.

Removal of Self-Epitopes

Self-tolerance must also be considered in vaccine designing, as body's immune system rarely acts against self-epitopes. Therefore, all the nona-peptides, which are also present in human body need to be removed. To achieve this, we mapped nona-peptides on 1000 human proteome, and removed all the nona-peptides which are 100% identical.

Epitope Prediction

Our subsequent objective was to select peptides, which could activate human immune system, and generate memory cells. Therefore, we have used a pipeline for predicting different kinds of immune epitopes among these nona-peptides. This pipeline predicts; (i) B-cell epitopes using LBtope (30) (ii) MHC class II binders using ProPred (31) (iii) T-cell epitopes using CTLpred (32) and (iv) vaccine adjuvant using VaxinPAD (1).

IEDB Mapping

In addition to these tools, the experimentally reported epitopes present in the Immune Epitope Database (IEDB) were also mapped on the vaccine target proteins selected for this study. **Figure 1** shows the complete workflow of this study.

Database Web Interface

All the predicted epitopes have been stored and presented on a web-portal named as VacTarBac. It is built on Apache HTTP server (version 2.2.17), which is installed on machine with Ubuntu as operating system. The responsive front-end, which is suitable for mobile, tablet, and desktop, was developed using HTML5, CSS3, PHP5, and JavaScript. MySQL (a relational database management system, version 5.0.51 b) was used at the back-end to manage the data.

RESULTS

Screening of Bacterial Species

The VFDB virulence factor proteins download link is available at http://www.mgc.ac.cn/VFs/Down/VFDB_setA_pro.fas.gz, providing all the sequences in FASTA format with the headers providing the information for the protein, for example, the VFDB id, source organism, etc. From these headers, the source organism names were extracted. The DEG is available for download at <http://www.essentialgene.org> in the form of FASTA sequences too. Only those sequences from DEG were extracted that sourced from organism species (indicated by the headers) present in the VFDB headers too. Finally, the protein sequences from the VFDB and DEG belonged to 14 bacterial species.

Screening of Target Proteins

These 14 bacterial species contain 1459 virulence factor as stated in VFDB. All these 1459 virulent factors were selected

for the further study. Previous studies have advocated that bacterial proteins at specific cellular locations (20, 23–25, 33) or performing specific functions essential for cell (34) are potential candidates for designing subunit vaccine. Therefore, on the basis of function and localization within the cell, 546 Proteins from the 14 bacterial species, were categorized as membrane, envelope, repair, and secretory proteins. **Table 2** shows distribution of screened target proteins among selected 14 bacteria.

The most important aspect of this rational determination of vaccine targets was to extract proteins common in both VFDB and DEG dataset. Our analysis shows that 21 IDs from VFDB dataset were identical to 35 IDs of DEG dataset. (**Supplementary Table 1**) These are 21 proteins from 5 bacterial species. We consider these as the most suitable vaccine targets, being both, virulence factors as well as gene products essential for survival.

Epitope Predictions

After generating the nona-peptides, redundant nona-peptides as well as self-antigens were removed by mapping on 1000 human proteome. For rest of the peptides, we predicted the immunogenicity using our prediction pipeline. LBtope, ProPred, CTLpred, and VaxinPAD helped us identify B-cell epitopes, MHC class II binders, T-cell epitopes, and adjuvants respectively. These tools integrated and implemented as a pipeline helped identify the immunogenic regions of the individual proteins in the form of nona-mer epitopes. On the other hand, the visualization tools integrated in the platform “VacTarBac” helped envisage the

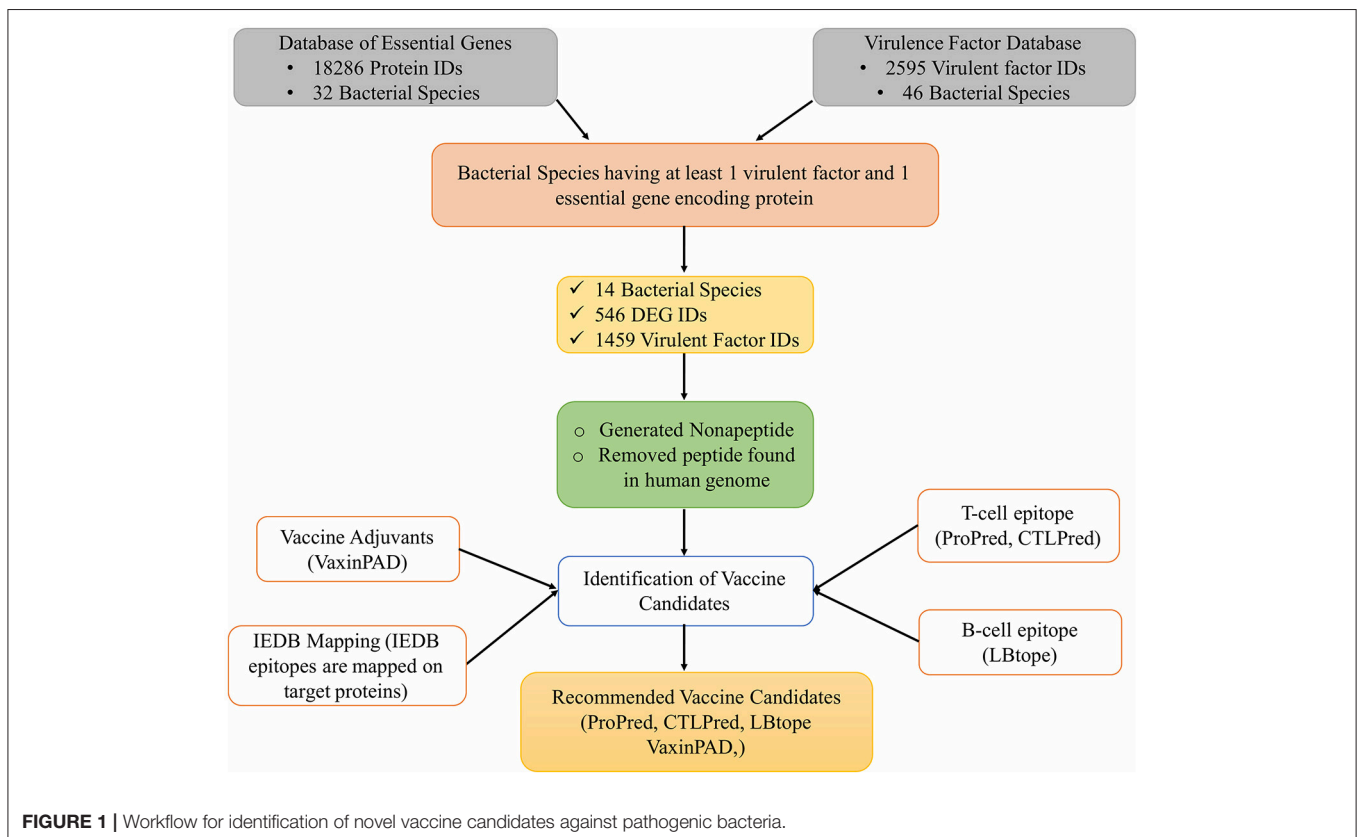


FIGURE 1 | Workflow for identification of novel vaccine candidates against pathogenic bacteria.

stretches of lengths longer than nona-mer having high density of predicted epitopes. The user may anticipate that these regions could prove to be antigens effective in the form of vaccines. Thus, for experimental investigation of the sequences that may prove to be effective vaccines, sequences longer than 9-mers containing numerous predicted epitopes could be taken up as a better strategy for the development of vaccines.

The epitopes predicted individually by each tool of our pipeline are too many to be tested experimentally. (Table 3) Therefore, we have applied an intuitive method to arrive at a reasonable number. (Figure 2 and Supplementary Table 2) Epitopes were separately identified for T-Cell and B-Cell categories. In case of T-Cell epitopes only those were retained that were predicted T-Cell epitopes, MHC-II Binders as well as adjuvants. For B-Cell epitopes, peptides predicted to be B-Cell epitopes, MHC-II Binders and adjuvants were finalized for recommendation as therapeutics.

Epitopes Assessed to be the Best Vaccine Candidates

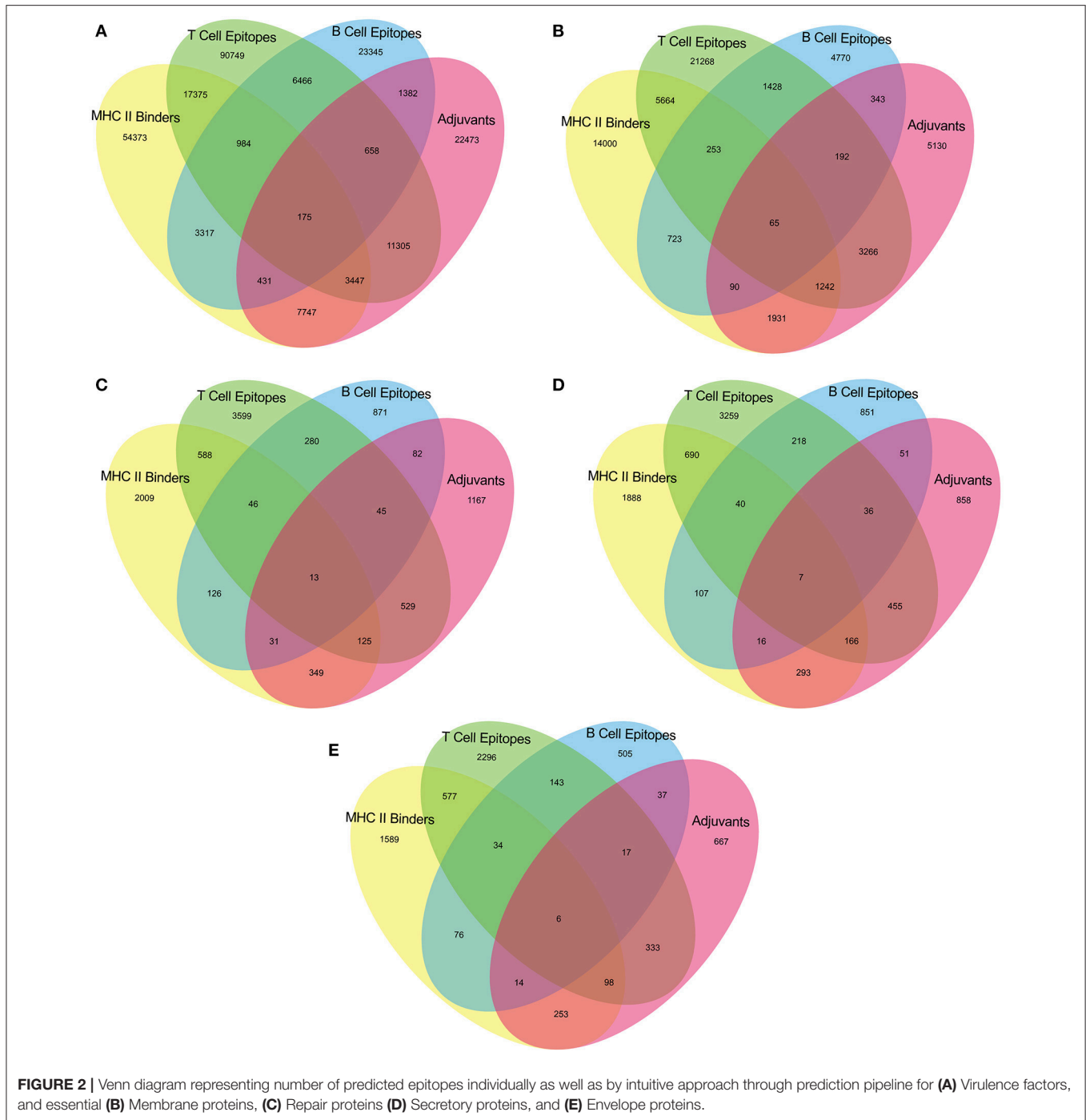
The aim of this study is to enhance the vaccine designing against various pathogenic bacteria by predicting potential MHC binder, B cell and T cell epitopes as well as vaccine adjuvants using a prediction pipeline. Yet, the idea of using a pipeline instead of the prediction tools individually was intended to pose stepwise filters on the numerous peptides that may be predicted as epitopes for different arms of the immune system. These filters reduce the number of epitopes to be experimentally tested and lead to a final set of peptide sequences with more possibility of activating immune system as these are recommended by all the individual tools in the prediction pipeline. Epitopes positively predicted by all the prediction tools, provided coverage of 13 out of 14 bacterial species. An organism-wise as well as vaccine target protein category-wise representation as shown in Table 4 provides more comprehensive analysis of the epitopes assessed

TABLE 2 | Bacterial species-wise distribution of Target Proteins.

Bacteria	Virulent Factors	Protein encoded by Essential Genes				Virulent Factors which are Essential Gene products
		Membrane Protein	Repair Protein	Secretory Protein	Envelope Protein	
<i>Bacillus subtilis</i>	1	2	0	6	0	0
<i>Burkholderia pseudomallei</i>	148	27	0	4	0	0
<i>Campylobacter jejuni</i>	128	7	0	0	0	1
<i>Escherichia coli</i>	293	37	1	2	0	0
<i>Haemophilus influenzae</i>	85	10	4	0	0	0
<i>Helicobacter pylori</i>	114	8	0	0	12	1
<i>Mycobacterium tuberculosis</i>	65	241	3	23	0	14
<i>Pseudomonas aeruginosa</i>	242	18	10	8	0	2
<i>Salmonella enteric</i>	123	35	0	3	0	3
<i>Staphylococcus aureus</i>	80	11	13	6	12	0
<i>Streptococcus agalactiae</i>	45	1	0	0	0	0
<i>Streptococcus pneumoniae</i>	43	14	7	0	12	0
<i>Streptococcus pyogenes</i>	43	2	0	0	0	0
<i>Vibrio cholera</i>	49	7	0	0	0	0
Total	1459	420	38	52	36	21

TABLE 3 | Distribution of proteins, nona-peptides and predicted epitopes among various categories.

Source	VFDB	Database of Essential Genes (DEG)			
	Virulence factor	Membrane Protein	Repair Proteins	Secretory Protein	Envelope Protein
Number of Proteins	1459	420	38	52	36
Number of Nona-peptides	1058876	482420	37823	53517	30497
Number of unique Nona-peptides	244347	60436	9866	8935	6660
Number of exclusive Nona-peptides absent in 1000 human proteome	228589	60365	9860	8935	6645
T-Cell Epitopes	131159	33378	5225	5141	3504
B-Cell Epitopes	36758	7864	1494	1326	832
MHC-II Binders	87849	23968	3287	3207	2647
Adjuvants	47618	12259	2341	1882	1425



to be the “best vaccine candidate” using all the prediction tools.

Best Possible Antigenic Proteins

As stated earlier, the study started with 1459 virulence factor and 546 essential proteins, we have predicted epitopes from all of the generated nona-peptides. Some proteins, have high number of predicted MHC binders, T cell eptopes, B cell epitopes as well as vaccine adjuvants. For example, the peptide syntase (pyoverdine)

protein, a virulence factor of *Pseudomaonas aeruginosa* results into 8197 predicted epitopes. Such proteins or regions within the proteins could be wisely used as potential antigens to stimulate the immune system (Supplementary Table 3). Figure 3 shows the portion of an essential protein from *Bacillus subtilis*, which has a high number of predicted epitopes. This indicates that instead of experimenatly validating a single nona-peptide, taking a longer portion of the protein, having overlapping predicted epitopes, could prove to be a better approach.

TABLE 4 | Species-wise distribution of the best epitopes by filtration through all the *in silico* tools used in the study.

Bacteria	Best Epitopes (T-cell Epitope+B-Cell Epitope+MHC II Binder+Adjuvant)				
	Essential Gene Proteins				Virulent Factors
	Membrane	Repair	Secretory	Envelope	
<i>Bacillus subtilis</i>	0	0	0	0	0
<i>Burkholderia pseudomallei</i>	8	0	0	0	25
<i>Campylobacter jejuni</i>	0	0	0	0	17
<i>Escherichia coli</i>	7	0	1	0	42
<i>Haemophilus influenzae</i>	3	3	0	0	6
<i>Helicobacter pylori</i>	0	0	0	2	4
<i>Mycobacterium tuberculosis</i>	61	1	11	0	11
<i>Pseudomonas aeruginosa</i>	4	4	0	0	38
<i>Salmonella enterica</i>	2	0	0	0	15
<i>Staphylococcus aureus</i>	2	4	0	2	5
<i>Streptococcus agalactiae</i>	1	0	0	0	4
<i>Streptococcus pneumoniae</i>	3	1	0	2	3
<i>Streptococcus pyogenes</i>	1	0	0	0	3
<i>Vibrio cholerae</i>	0	0	0	0	6
Total number of Epitopes	92	13	12	6	179
Number of Unique Epitopes	65	13	7	6	175

Considering all the epitopes together in this table, these belonged to 13 bacterial species out of the total 14 species considered for the study.

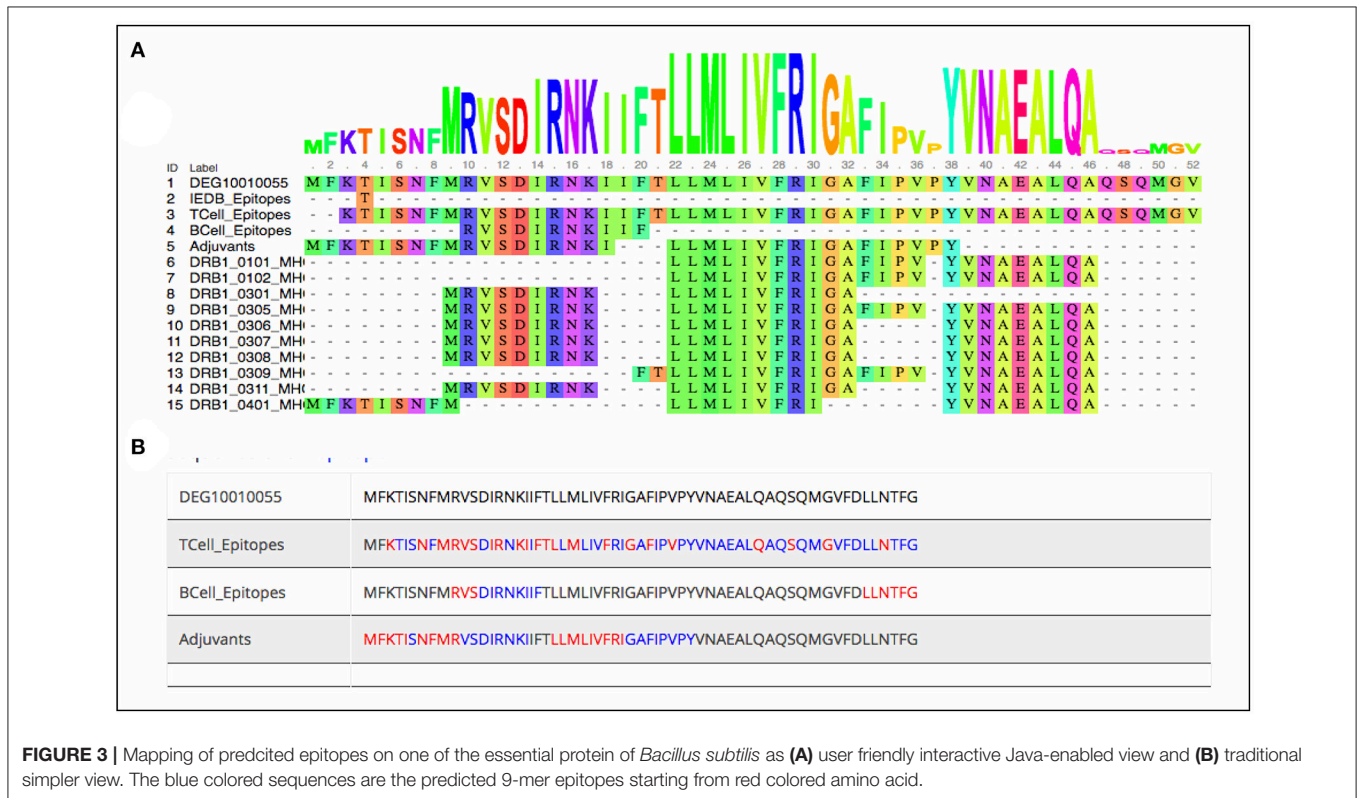


FIGURE 3 | Mapping of predicted epitopes on one of the essential protein of *Bacillus subtilis* as (A) user friendly interactive Java-enabled view and (B) traditional simpler view. The blue colored sequences are the predicted 9-mer epitopes starting from red colored amino acid.

Webserver Implementation

To assist the scientific community in expediting the peptide-based vaccine designing against pathogenic bacteria, all the

predicted potential epitopes were compiled in the form of a database and were provided as a web-based service (<http://webs.iitd.edu.in/raghava/vactarbac/>). A user can browse all the

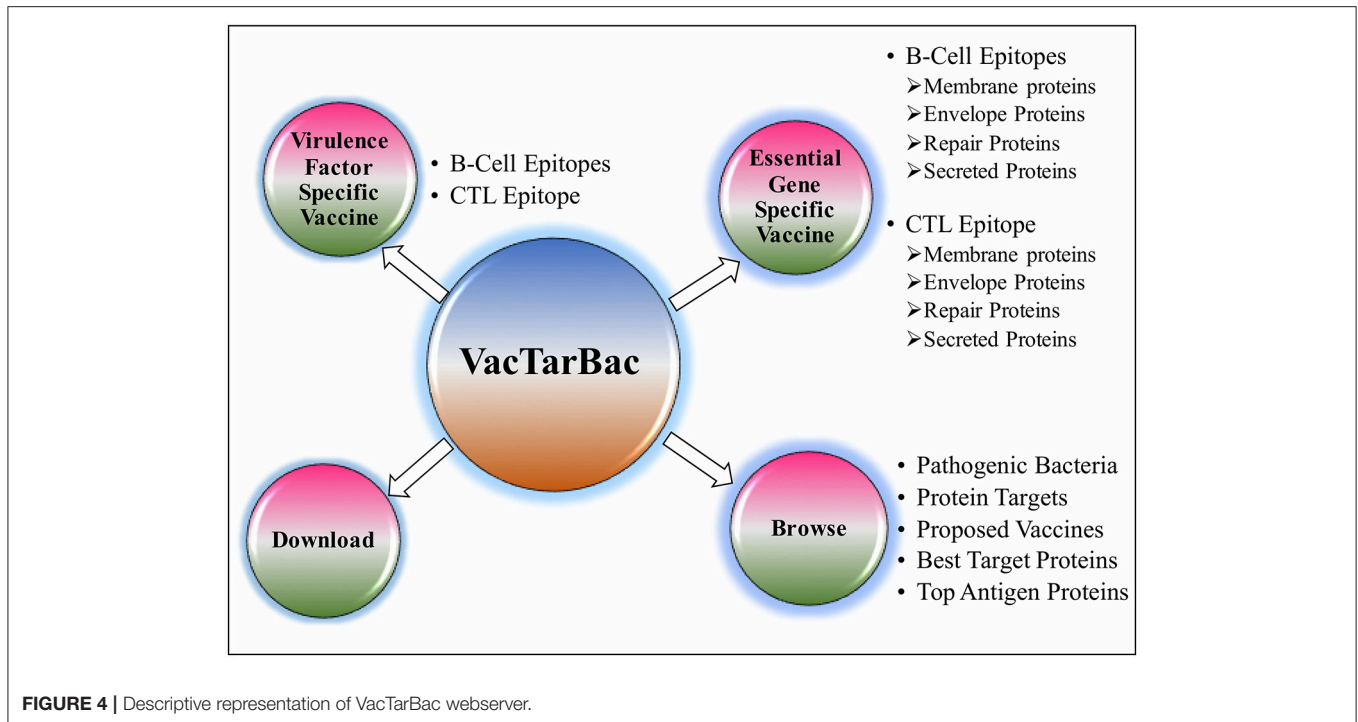


FIGURE 4 | Descriptive representation of VacTarBac webservice.

recommended potential B cell and T cell epitopes designed by targeting virulence factors as well as proteins encoded by essential genes of all 14 bacterial species. Browsing by the pathogenic bacteria as well as targeted proteins considered while designing of potential epitopes is also implemented. A user can also browse the top 5 antigens from the enlisted bacteria. In addition, a list of proposed vaccine candidates is also provided in the browse option (Figure 4). For the ease of user, results have been displayed in tabular and graphical form as well as an interactive visualization.

DISCUSSION AND CONCLUSION

The twentieth century witnessed a remarkable success in development of antibiotics against various pathogenic bacteria but the emergence of drug resistance and toxicity caused a shift in the antimicrobial strategies toward peptide-based therapies. Several anti-bacterial peptides have been studied to check the pestiferous effects of pathogenic bacteria (35, 36). Rapid bacteriocidal activity and low propensity for resistance development are some of the major assets of these peptides, but high cost, limited stability and unknown toxicology, and pharmacokinetics are the major disadvantages (37). Contrary to antibiotics or peptide-based drugs, instigating host immune responses against the pathogen offers multifaceted management of the invasion and has been successfully achieved in the past by the use of vaccines like DPT, BCG etc against few deadly infectious diseases like tetanus, diphtheria, tuberculosis etc. Consequently, active search for vaccines is underway currently for many infectious diseases caused by pathogenic bacteria, for example, melioidosis caused by *Burkholderia pseudomallei* (2).

Even in case of tuberculosis, for which multiple lines of drugs and BCG vaccines are available; reduced efficacy of BCG and antibiotic resistance forced investigation of better and more potent vaccines. A recent computational effort has gone up to the strain level comparison and considered proteins as vaccine targets that were shared among tuberculoid, non-tuberculoid, and vaccine strains (16). Continuous efforts are being made by researchers to identify novel vaccine candidates against several pathogens such as *Ebola* virus (14), *Zika* virus (15), *vaccinia* virus (38), *Neisseria meningitidis* (39), *Corynebacterium pseudotuberculosis* (40) *Edwardsiella tarda*, and *Flavobacterium columnare*; fish pathogens (41, 42).

In past, several *in silico*, protein-based vaccine candidates have been identified, using peptide conservation score and predicted peptide-MHC binding (43) as well as molecular docking and MHC-peptide complex stabilization assay (18). In this study, we employed immunoinformatic tools to identify antigen and epitope-based vaccine candidates, having the potential to evoke one of the many arms of immune system. The present study includes proteins; virulence factors as well as proteins encoded by genes essential for survival from multiple pathogenic bacteria. The pipeline created for the identification of vaccine candidates, consists of widely used, accurate and recommended tools (12, 13). Our pipeline also includes a tool to identify peptide-based adjuvant candidates, a major advantage over other studies. The numbers of epitopes predicted by the individual tools within the pipeline that are not present in the proteome derived from the 1000 human genomes, were in thousands. For arriving at economical number of epitopes that could be tested experimentally, an intuitive rational criterion was to extract epitopes predicted positive by more than one tool in the pipeline. This would effectively mean that the selected epitope could have

a predictive capability of evoking different arms of the immune system.

Upon executing the rational criteria, 1459 virulent factors from VFDB yielded 3622 nona-peptides having ability to bind with MHC-II, activate T-cell response and could act as self-adjuvants. Beside this, peptides are predicted to bind with MHC-II, activate B-cell response as well as self-adjuvants. Similarly, 420 membrane proteins from the DEG provided 1307 predicted T-cell epitopes and 155 predicted B-cell epitopes that are also MHC II binders and adjuvants. The 36 DEG envelope proteins when subjected to the epitope prediction pipeline, yield 104 predicted T-cell epitopes and 20 predicted B-cell epitopes with probability that these will bind MHC Class II molecules and would also be adjuvants. In cases of 52 secretory proteins and 38 repair proteins taken from the DEG, the predicted T-cell epitopes, and B-cell epitopes were respectively 173 and 23 for secretory proteins and 138 and 44 for the repair proteins, all of them being also the positively predicted MHC II binders as well as adjuvants.

In conclusion, the ultimate aim of the present study was the identification of epitopes capable of activating multiple wings of the human immune system arrived at by filtering the epitopes through stringent criteria. Combining the results of the VFDB and DEG datasets, this study was able to identify 252 unique epitopes predicted to be T-cell epitopes, B-cell epitopes, MHC II binders and Vaccine Adjuvants not present in the human proteome (belonging to the 1000 genomes) extracted from proteins (essential gene products and/or virulence factors) of 13 bacterial species. All the recommended vaccine candidates have been stored in a repository; VacTarBac (<http://webs.iitd.edu.in/raghava/vactarbac/>) freely available on the world-wide web.

Peptides are the promising candidate as immunotherapeutic but are less immunogenic when used alone as vaccine. They need potent immunostimulatory adjuvants to effectively activate the innate and adaptive arms of immune system. In this study, we have implemented VaxinPAD in our prediction pipeline, which aids in predicting based vaccine adjuvant, thus strengthening the study. Beside this, peptide formulation is a critical task and a formulation scientist must overcome the chemical instability of peptides. The conformational fluxionality and propensity to self-associate makes peptide more difficult to formulate. Various computational tools predict the aggregation propensities of polypeptide chains such as Zyggregator (44), PASTA 2.0 (45), etc. These can be used prior to formulating the peptide-based vaccine. THPdb is a database of FDA approved protein and peptide therapeutics, and provide detailed formulation strategies of peptide-based drug available in the market (46). Although little literature exist about formulating peptide drug product but strategies used in successful peptide drugs will be helpful in designing newer formulation techniques and constituent for peptide based vaccines.

REFERENCES

1. Neill JO'. *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations The Review on Antimicrobial Resistance Chaired* (2014).

We have predicted epitopes based on highly cited, published and accurate immune epitope prediction tools. Yet, these prediction algorithms have their own limitations. Thus, the antigen or epitope should be experimentally validated before suggesting it for medical purpose. Apart from this, the platform, VacTarBac, would require continuous revamp owing to the rapidly growing sequencing data particularly that of the novel and emerging pathogenic strains. A future work for this platform may include comparison of the pathogenic proteomes at the strain level. This may lead to identification of recently acquired proteins/peptides of the pathogen that may have rendered existing therapeutic strategies against the emergent strain ineffective. Moreover, the platform, VacTarBac, still lacks information of vaccine formulations that may be added in future to help in the actual development of the vaccines. Despite such limitations, we anticipate that the current study and the current data in the repository VacTarBac, will be helpful for researchers and will boost and hasten the vaccine designing against pathogenic bacteria considered in the study.

AVAILABILITY AND UPDATE OF THE RESOURCE

VacTarBac is freely available at <http://webs.iitd.edu.in/raghava/vactarbac/>. We will update the platform at every 6 months depending upon the availability of virulence factor and essential genes information in other resources.

AUTHOR CONTRIBUTIONS

SU and GN downloaded and processed the data. GN prepared the pipeline and predicted epitopes. SU and GN analyzed results and prepared tables and figures. SU, GN, and GR wrote the manuscript. SU developed web interface. GR conceived the idea and coordinated the project.

FUNDING

Authors are thankful to funding agencies J.C. Bose National Fellowship (DST), Department of Biotechnology (DBT) and Council of Scientific and Industrial research (CSIR) for fellowships and financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02280/full#supplementary-material>

2. Peacock SJ, Limmathurotsakul D, Lubell Y, Koh GCKW, White LJ, Day NPJ, et al. Melioidosis vaccines: a systematic review and appraisal of the potential to exploit biodefense vaccines for public health purposes. *PLoS Negl Trop Dis.* (2012) 6:e1488. doi: 10.1371/journal.pntd.0001488

3. Andersen P, Doherty TM. The success and failure of BCG - implications for a novel tuberculosis vaccine. *Nat Rev Microbiol.* (2005) 3:656–62. doi: 10.1038/nrmicro1211
4. Kernodle DS. Decrease in the effectiveness of Bacille Calmette-Guérin vaccine against pulmonary tuberculosis: a consequence of increased immune suppression by microbial antioxidants, not overattenuation. *Clin Infect Dis.* (2010) 51:177–84. doi: 10.1086/653533
5. Barnett TC, Lim JY, Soderholm AT, Rivera-Hernandez T, West NP, Walker MJ. Host-pathogen interaction during bacterial vaccination. *Curr Opin Immunol.* (2015) 36:1–7. doi: 10.1016/j.coi.2015.04.002
6. Zuñiga S, Pascual-Iglesias A, Sanchez CM, Sola I, Enjuanes L. Virulence factors in porcine coronaviruses and vaccine design. *Virus Res.* (2016) 226:142–51. doi: 10.1016/j.virusres.2016.07.003
7. Sánchez J, Holmgren J. Virulence factors, pathogenesis and vaccine protection in cholera and ETEC diarrhea. *Curr Opin Immunol.* (2005) 17:388–98. doi: 10.1016/j.coi.2005.06.007
8. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res.* (2012) 40:D641–5. doi: 10.1093/nar/gkr989
9. Lu Y, Deng J, Rhodes JC, Lu H, Lu LJ. Predicting essential genes for identifying potential drug targets in *Aspergillus fumigatus*. *Comput Biol Chem.* (2014) 50:29–40. doi: 10.1016/j.compbiolchem.2014.01.011
10. Hu W, Sillaots S, Lemieux S, Davison J, Kauffman S, Breton A, et al. Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. *PLoS Pathog.* (2007) 3:e24. doi: 10.1371/journal.ppat.0030024
11. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* (2014) 42:D574–80. doi: 10.1093/nar/gkt1131
12. Usmani SS, Kumar R, Bhalla S, Kumar V, Raghava GPS. *In silico* tools and databases for designing peptide-based vaccine and drugs. *Adv Protein Chem Struct Biol.* (2018) 112:221–63. doi: 10.1016/bs.apcsb.2018.01.006
13. Dhanda SK, Usmani SS, Agrawal P, Nagpal G, Gautam A, Raghava GPS. Novel *in silico* tools for designing peptide-based subunit vaccines and immunotherapeutics. *Brief Bioinformatics* (2017) 18:467–78. doi: 10.1093/bib/bbw025
14. Dhanda SK, Chaudhary K, Gupta S, Brahmachari SK, Raghava GPS. A web-based resource for designing therapeutics against Ebola Virus. *Sci Rep.* (2016) 6:24782. doi: 10.1038/srep24782
15. Gupta AK, Kaur K, Rajput A, Dhanda SK, Sehgal M, Khan MS, et al. ZikaVR: an integrated zika virus resource for genomics, proteomics, phylogenetic and therapeutic analysis. *Sci Rep.* (2016) 6:32713. doi: 10.1038/srep32713
16. Dhanda SK, Vir P, Singla D, Gupta S, Kumar S, Raghava GPS. A web-based platform for designing vaccines against existing and emerging strains of *Mycobacterium tuberculosis*. *PLoS ONE* (2016) 11:e0153771. doi: 10.1371/journal.pone.0153771
17. Danielson RE, Cooper RC. The effect of water activity on *Legionella* spp. growth and survival. *J Appl Bacteriol.* (1986) 61:157–61.
18. Zheng J, Lin X, Wang X, Zheng L, Lan S, Jin S, et al. *In silico* analysis of epitope-based vaccine candidates against hepatitis B virus polymerase protein. *Viruses* (2017) 9:112. doi: 10.3390/v9050112
19. Meza B, Ascencio F, Sierra-Beltrán AP, Torres J, Angulo C. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against *Helicobacter pylori*: an *in silico* approach. *Infect Genet Evol.* (2017) 49:309–17. doi: 10.1016/j.meegid.2017.02.007
20. Rauta PR, Ashe S, Nayak D, Nayak B. *In silico* identification of outer membrane protein (Omp) and subunit vaccine design against pathogenic *Vibrio cholerae*. *Comput Biol Chem.* (2016) 65:61–8. doi: 10.1016/j.compbiolchem.2016.10.004
21. Fahimi H, Sadeghzadeh M, Mohammadipour M. *In silico* analysis of an envelope domain III-based multivalent fusion protein as a potential dengue vaccine candidate. *Clin Exp Vaccine Res.* (2016) 5:41–9. doi: 10.7774/cevr.2016.5.1.41
22. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* (2016) 44:D694–7. doi: 10.1093/nar/gkv1239
23. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatbadi SS, Ghasemi Y. Immunoinformatics analysis and *in silico* designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol.* (2017) 48:83–94. doi: 10.1016/j.meegid.2016.12.010
24. Yasmin T, Akter S, Debnath M, Ebihara A, Nakagawa T, Nabi AHMN. *In silico* proposition to predict cluster of B- and T-cell epitopes for the usefulness of vaccine design from invasive, virulent and membrane associated proteins of *C. jejuni*. *In silico Pharmacol.* (2016) 4:5. doi: 10.1186/s40203-016-0020-y
25. Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R. From ZikV genome to vaccine: *in silico* approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. *Immunology* (2016) 149:386–99. doi: 10.1111/imm.12656
26. Li M, Song S, Yang D, Li C, Li G. Identification of secreted proteins as novel antigenic vaccine candidates of *Haemophilus parasuis* serovar 5. *Vaccine* (2015) 33:1695–701. doi: 10.1016/j.vaccine.2015.02.023
27. Blander SJ, Horwitz MA. Vaccination with the major secretory protein of *Legionella* induces humoral and cell-mediated immune responses and protective immunity across different serogroups of *Legionella pneumophila* and different species of *Legionella*. *J Immunol.* (1991) 147:285–91.
28. Green ER, Mecsas J. Bacterial secretion systems: an overview. *Microbiol Spectr.* (2016) 4:1236–7. doi: 10.1128/microbiolspec.VMBF-0012-2015
29. Žgur-Bertok D. DNA damage repair and bacterial pathogens. *PLoS Pathog.* (2013) 9:e1003711. doi: 10.1371/journal.ppat.1003711
30. Singh H, Ansari HR, Raghava GPS. Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS ONE* (2013) 8:e62216. doi: 10.1371/journal.pone.0062216
31. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* (2001) 17:1236–7. doi: 10.1093/bioinformatics/17.12.1236
32. Bhasin VM, Raghava GPS. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* (2004) 22:3195–204. doi: 10.1016/j.vaccine.2004.02.005
33. Rana A, Thakur S, Bhardwaj N, Kumar D, Akhter Y. Excavating the surface-associated and secretory proteome of *Mycobacterium leprae* for identifying vaccines and diagnostic markers relevant immunodominant epitopes. *Pathog Dis.* (2016) 74:ftw110. doi: 10.1093/femspd/ftw110
34. Damte D, Suh JW, Lee SJ, Yohannes SB, Hossain MA, Park SC. Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of *Mycoplasma hyopneumoniae*. *Genomics* (2013) 102:47–56. doi: 10.1016/j.ygeno.2013.04.011
35. Usmani SS, Kumar R, Kumar V, Singh S, Raghava GPS. AntiTbPdb: a knowledgebase of anti-tubercular peptides. *Database* (2018) 2018. doi: 10.1093/database/bay025
36. Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* (2016) 44:D1087–93. doi: 10.1093/nar/gkv1278
37. Gordon YJ, Romanowski EG, McDermott AM. A review of antimicrobial peptides and their therapeutic potential as anti-infective drugs. *Curr Eye Res.* (2005) 30:505–15. doi: 10.1080/02713680590968637
38. Sánchez-Sampedro L, Gómez CE, Mejías-Pérez E, Pérez-Jiménez E, Oliveros JC, Esteban M. Attenuated and replication-competent vaccinia virus strains M65 and M101 with distinct biology and immunogenicity as potential vaccine candidates against pathogens. *J Virol.* (2013) 87:6955–74. doi: 10.1128/JVI.03013-12
39. Christodoulides M, Heckels J. Novel approaches to *Neisseria meningitidis* vaccine design. *Pathog Dis* (2017) 75:ftx033. doi: 10.1093/femspd/ftx033
40. Barh D, Jain N, Tiwari S, Parida BP, D'Afonseca V, Li L, et al. A novel comparative genomics analysis for common drug and vaccine targets in *Corynebacterium pseudotuberculosis* and other CMN group of human pathogens. *Chem Biol Drug Des* (2011) 78:73–84. doi: 10.1111/j.1747-0285.2011.01118.x
41. Liu X, Sun J, Wu H. Glycolysis-related proteins are broad spectrum vaccine candidates against aquacultural pathogens. *Vaccine* (2017) 35:3813–6. doi: 10.1016/j.vaccine.2017.05.066
42. Mahendran R, Jeyabaskar S, Sitharaman G, Michael RD, Paul AV. Computer-aided vaccine designing approach against fish pathogens *Edwardsiella tarda* and *Flavobacterium columnare* using bioinformatics softwares. *Drug Des Devel Ther.* (2016) 10:1703–14. doi: 10.2147/DDDT.S95691
43. Goodswen SJ, Kennedy PJ, Ellis JT. Enhancing *in silico* protein-based vaccine discovery for eukaryotic pathogens using predicted peptide-MHC

- binding and peptide conservation scores. *PLoS ONE* (2014) 9:e115745. doi: 10.1371/journal.pone.0115745
44. Tartaglia GG, Vendruscolo M. The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev.* (2008) 37:1395–401. doi: 10.1039/b706784b
 45. Walsh I, Seno F, Tosatto SC, Trovato A. PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* (2014) 42:W301–7. doi: 10.1093/nar/gku399
 46. Usmani SS, Bedi G, Samuel JS, Singh S, Kalra S, Kumar P, et al. THPdb: Database of FDA-approved peptide and protein therapeutics. *PLoS ONE* (2017) 12:e0181748. doi: 10.1371/journal.pone.0181748

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Nagpal, Usmani and Raghava. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.