EXPLORING THE PANGENOME AND IMMUNO-PROTEOME OF URINARY TRACT INFECTION CAUSING BACTERIAL PATHOGENS TO UNRAVEL VACCINE TARGETS

Author

FATIMA KANWAL

NUST201463577MASAB92514F

Supervisor

Dr. AMJAD ALI

ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES (ASAB),

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY (NUST),

SECTOR H-12, ISLAMABAD

2016

Exploring the Pan-genome and Immuno-Proteome of Urinary Tract Infection Causing Bacterial Pathogens to Unravel Vaccine Targets

Author

FATIMA KANWAL

Reg. Number: NUST201463577MASAB92514F

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Industrial Biotechnology

Synopsis Supervisor:

DR. AMJAD ALI

Thesis Supervisor's Signature: _____

ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES (ASAB),

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY (NUST),
MAIN NUST CAMPUS,

SECTOR: H-12, ISLAMABAD

2016

Annex A to NUST Letter No. 0972/102/Exams/Thesis-Cert dated Dec 16.

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of PhD thesis written by **Ms**. **Fatima Kanwal** Registration No. **NUST201463577MASAB92514F** of **ASAB** has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations/PhD Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of PhD degree. It is further certified that necessary amendments as pointed out by GEC members and foreign/local evaluators of the scholar have also been incorporated in the said thesis.

Signature with stamp: _____

Name of Supervisor: Dr. Amjad Ali

Date: _____

Signature of HoD with stamp: _____

Date: _____

Countersign by

Signature (Dean/Principal): _____

Date: _____

SUPERVISORY HIRARCHEY

SUPERVISOR

Dr. Amjad Ali

GUIDANCE AND EXAMINATION COMMITTEE MEMBERS

Dr. Saadia Andaleeb (ASAB, NUST)

Dr. Hussnain A. Janjua (ASAB, NUST)

EXTERNAL EXAMINER

Dr. Jamil Ahmed (RCMS, NUST)

Declaration

I certify that this research work titled "*Exploring the Pan-genome and Immuno-Proteome of Urinary Tract Infection Causing Bacterial Pathogen to Unravel Vaccine Targets"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Fatima Kanwal

2014-NUST-MS-Biotech-925

Copyright Statement

Acknowledgements

All praises to **Allah Almighty**, the most **Beneficent**, and the most **Merciful**. It is only because of **Almighty Allah** that I was able to complete my research work. **Allah Almighty** has indeed granted me far more than I deserve.

قَالُوا سُبْحَانَكَ لَا عِلْمَ لَنَا إِلَّا مَا عَلَّمْتَنَا ۖ إِنَّكَ أَنتَ الْعَلِيمُ الْحَكِيمُ

They said, "Exalted are You; we have no knowledge except what You have taught us. Indeed, it is You who is the Knowing, the Wise."

The completion of my research work took vigorous efforts, help and guidance of a group of people at **National Analytical Lab**, **Atta-ur-Rahman School of Applied Biosciences (ASAB)**. I feel obligatory to acknowledge them all.

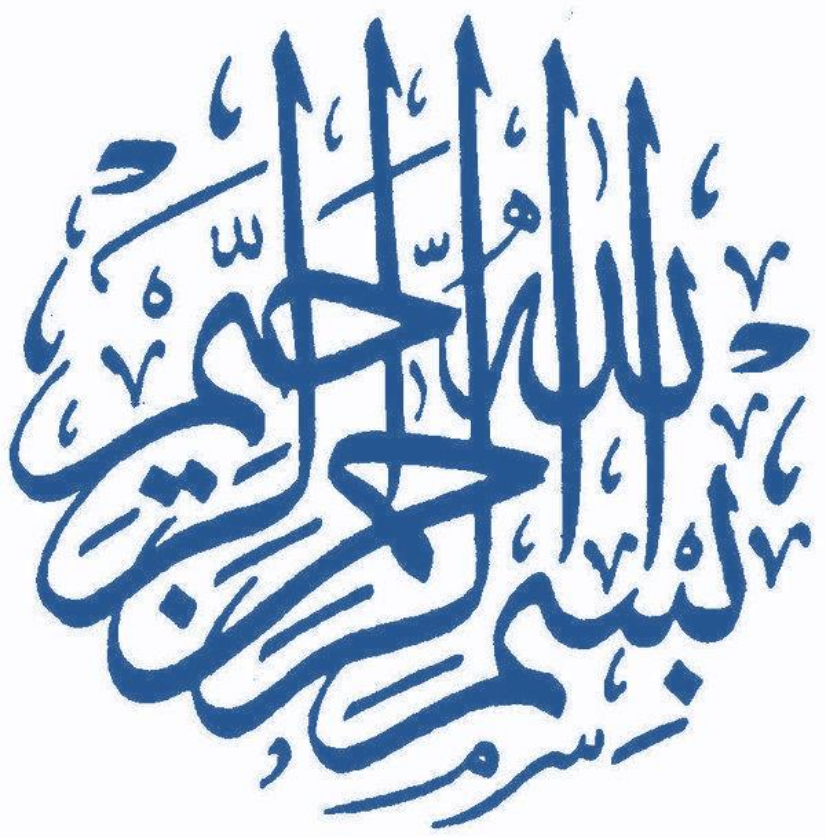
First of all, I will like to show my respect and gratitude for my supervisor, **Dr. Amjad Ali Assistant Professor, Industrial Biotechnology ASAB** for his patience and especially for his insistence and belief that I would be able to successfully complete an *in-silico* research project. I am grateful for his help throughout the thesis and especially for his course Advanced Computational Biology which has open new world of scientific imagination for me. I was lucky to work under his supervision and I gained a lot of experience regarding team work. I will always be obliged to him for imparting his wisdom and realization of importance of ethical values to me.

I am also thankful to my **Guidance Committee members**, not only for their guidance but also for **Dr. Saadia Andleeb's Assistant Professor, Industrial Biotechnology ASAB** affection and kind heartedness and, **Dr. Hussnain A. Janjua's Assistant Professor,**

Industrial Biotechnology ASAB persistent encouragement and appreciation. I am also greatful to my external supervisor **Dr. Jamil Ahmed**, **Assistant Professor, Research Center of Modeling and Simulation(RCMS),** for his supervision, support and provision of RCMS facilities. I will also like to show my gratitude to **Dr. Najam us Sahar Sadaf Zaidi, HOD, Industrial Biotechnology ASAB**, for her help both at personnel and academic level. It was her lecture on Influenza Vaccine Development and Dr. Amjad Ali's enthusiasm for *in-silico* research that inspired me both for my research and FICS project. I will also like to thank **Dr. Peter John Principal, ASAB** and all of the ASAB faculty and staff for the provision of facilities and management which has turned ASAB into an excellent research institute.

I am also grateful to my lab fellows at **National Analytical Lab** for the provision of their input and guidance especially Kanwal Naz for introducing me to the concept of Pangenomics. Afreenish Hassan, Anam Naz and Faryal Mehwish Awan also provided their valuable suggestion for my thesis.

Finally, I would also like to appreciate my friends for their motivation and encouragement, namely; Sidrah Jahangir, Saima Hanif, Sabiha Bibi, Sidra Irum, Saleh Sarwar, Danish Gul, Hammad Rai, Simra Khattak and especially my best friends Sumera Khan, Anushah Abbas, Madeeha Khalil and Madiha Amjad.


بسم الله الرحمن الرحيم

Abstract

Urinary tract infections are a growing concern for the society in terms of morbidity and health care economic burden worldwide, emphasizing the prerequisite for new vaccine interventions in the current scenario of antibiotic resistance. The current study aims at prioritizing broad-spectrum vaccine targets, present in different uropathogens, through a pan-genomics strategy coupled with a comparative genomics and immuno-proteomics analysis. The inter-specie pan-genome analysis was not fruitful in obtaining any conserved genome, therefore intra-specie pan-genome analysis was performed which revealed 13556, 9266, 10713, 291789, 4031, 4285, 4195 and 3924 genes for E. coloi, K. pneumoniae, K. oxytoca, P.aeruginosa, E. faecalis, E. faecium, S. aureus and P. mirabilis respectively. The core genomesubsets estimated to carry 2595, 4132, 3534, 4151, 2169, 1909, 1902 and 2969 for E. coloi, K. pneumoniae, K. oxytoca, P.aeruginosa, E. faecalis, E. faecium, S. aureus and P. mirabilis respectively. Moreover the pan-genome of these pathogens was also analyzed for the presence of resistance genes and 564 resistance genes were identified. The discrete core-genomes of all the pathogens under analysis were identified and considered for filtration through five different parameters such as host non homology analysis (2,1169), essentiality analysis (5,824), virulence factor analysis (3,747), <2 helices analysis (15,033) and sub-cellular localization (1488). The analyses revealed a set of promising vaccine targets present in the core-genome of individual uropathogenic species. The novel vaccine targets namely, Omp N, toluene efflux pump outer membrane precursor, zinc binding, phosphor ethyl pyramidine kinase, 50S ribosomal L13, Fe binding precursor, N-acetylneuraminic acid outer membrane channel and family RND efflux system outer membrane lipo, were found to be functionally important with at least >5 metabolic interactors, also found to be present in a wide range of other urogenital disease causing bacteria and having <110kDa molecular

weight, makes them ideal vaccine candidates. Immuno proteomic analysis further illustrated the ability of these vaccine targets to trigger both B-cell and T-cell activation. The identified vaccine targets may assist in designing and development of broad-spectrum vaccines against these uropathogens. The pipeline can also be utilized to identify vaccine trgets against other pathogens.

***Keywords*: Urinary tract infection, Pan-genome, Comparative genomics, Vaccine targets, Broad-spectrum**

Dedicated to my exceptionally brave, resilient kind and generous beloved mother

Table of Contents

List of Figures

List of Tables

List of Acronyms & Abbreviations

Artificial Neural Network Based B-Cell Epitope Prediction Server	ABCpred
Antibiotic Resistance	ABR
Absorption, Distribution, Metabolism Excretion and Toxicity	ADMET
The first web-based vaccine design program for reverse vaccinology and applications for vaccine development	Vaxign
Basic Local Alignment Search Tool is an algorithm for comparing primary biological sequence information	BLAST
Protein BLAST	BLASTp
Blood Stream Infection	BSI
Centre for Disease Dynamics, Economics & Policy	CDDEP
Sub cellular localization predictor using SVM	CELLO
A webserver for protein subcellular localization prediction with functional gene ontology annotation	CELLO2GO
Cystic Fibrosis	CF
Tool for Genome-Scale Analysis of protein functions and evolution	COG Database
Cluster of Orthologous Groups of Proteins	COGs
Carbapenem Resistance	CR
Database of Essential Genes	DEG
A text-based format for representing either nucleotide sequences or peptide sequences	FASTA
Prediction tool of trans membrane helices and topology of Proteins	HMMTOP
Lipopolysaccharide	LPS
Lower Unary Tract Infection	LUTI
Major Histocompatibility Complex	MHC
A quantitative T-Cell epitope prediction server of peptide MHC binding	MHCpred

A Microbial Database of Protein Toxins, Virulence Factors and Antibiotic Resistance Genes for Biodefense Applications	MvirDB
National Institute of Allergy and Infectious Diseases	NIAID
Outer Membrane Protein	OMp
Penicillin Binding Protein	PBP
Protein Homology/ Analogy Recognition Engine	Phyre
Protein-Protein Interactions	PPI
Predictor of Promiscuous MHC Class-I Binding Site	ProPred
MHC Class-II Binding Peptide Prediction Server	ProPred 1
Protein sub cellular localization predictor with refined localization sub categories and predictive capabilities for all prokaryotes	PSORTb
Respiratory Tract Infection	RTI
Markov Model Prediction Tool for Transmembrane Helices in Proteins	TMHMM
Universal Protein Resource	UniProt
United States Food and Drug Authority	US FDA
Urinary Tract Infection	UTI
Virulence Factor Database for Pathogenic Bacteria	VFDB
World Health Organization	WHO

CHAPTER 1
INTRODUCTION

CHAPTER 1 INTRODUCTION

1.1 Urinary Tract Infection (UTI)

Urinary Tract Infection (UTI) is the infection of urinary tract caused by micro-organisms including bacteria, fungus or viruses (Singhal, Sharma et al. 2015). When the pathogen affects the lower urinary tract, either urethra (urethritis) or bladder (cystitis), it's called lower urinary tract infection (LUTI). Infection of upper urinary tract comprising kidneys, is termed as paleonephritis (Flores-Mireles, Walker et al. 2015). Mostly patient suffer from recurrent UTI but in case of complicated UTI patient may also suffer from kidney damage or even kidney failure. Complicated paleonephritis frequently leads to bacteremia and sepsis (Phimister, Arias et al. 2015).

Empiric UTI management with antimicrobials has become complicated with antimicrobial resistance (Girard, Gaujard et al. 2015). It is also becoming a worldwide problem due to bacterial multiple mutations and their lateral gene transfer further compromising the antibacterial therapy, particularly in case of recurrent urinary tract and respiratory infections (Tang, Apisarnthanarak et al. 2014). Antibiotic susceptibility patterns for clinical pathogens vary slightly but the major antibiotic resistance patterns remains the same around the world both in case of community acquired and hospital acquired UTI's (Kumar, Singh et al. 2014).

The current prescribed antibiotics for UTI are exhausting and there is an urgent need of

developing new antibiotics (Cai, Mazzoli et al. 2012). Whereas developing new antibiotics is a very expensive process and yields less desirable results comparative to insilico vaccine or drug development, which is a much cost effective process. In insilico approach of antibiotic drug or vaccine design and development antibiotic target identification is the first step (Silverman and Holladay 2014).

1.2 Common UTI causing pathogens

UTI causing bacteria are the most common nosocomial infection pathogens around the world (Singhal, Sharma et al. 2015). Fungal UTI's are also common in women due to feminine physiology and hormonal changes (Lehman and Donlan 2015; Nitzan, Elias et al. 2015). Viruses rarely cause UTI in either men or women and mostly occur as a secondary infection in case of a compromised immune infection (Thomas and Tracy 2015).

Among bacteria the most common UTI causing organism is *Escherichia coli* (Hill, Pan et al. 2015). *Klebsiella* species are the second most common cause of UTI (Vidal, Torre-Cisneros et al. 2012). *Escherichia coli* along with Klebsiella species are responsible for more than 70% of the UTI infections all over the world (Linhares, Raposo et al. 2013). *Pseudomonas aroginosa, Enterococcus faecium, Enterococcus faecalis, Staphylococcus aureus* and *Proteus mirabilis* are also well known for hospital acquired UTI's (Guzman, Pruzzo et al. 1989; Jha and Bapat 2005; Mittal, Aggarwal et al. 2009).

1.3 Prevalence and risk factors of UTI

UTI is the second most common bacterial infection after respiratory tract infections (Sahebnasagh, Saderi et al. 2015). It is the most common hospital acquired bacterial infection (Bénet, Haesebaert et al. 2015). Infants, pediatrics, adult and geriatrics all suffer from UTI no less than once in their lifetime. Most women suffer from UTI, at least twice in some part of

their life (Dielubanza and Schaeffer 2011).

Infants, geriatrics, women in gestational phase, catheterize patients or patients with other comorbidities with compromised immune system are more prone to UTI's (Nicolle 2008; Dielubanza and Schaeffer 2011) . In gestational phase high progesterone levels in pregnant women likely to cause the flow of urine back towards the kidneys leading to asymptomatic UTI. Pregnant women with asymptomatic UTI are 40% more prone to kidney infection. Complicated paleonephritis may leads to kidney failure, premature birth or pre-eclampsia (Dielubanza and Schaeffer 2011).

Catheterize patients more frequently suffer from LUTI because the major mechanism involve in bacterial invasive colonization is biofilm formation. Bacterial pathogens can easily form biofilm on the catheter polymeric surface. Majority of the visits by women in emergency departments is due to UTI (Tomas, Getman et al. 2015). Women with menopause suffer from more recurrent episodes of UTI. 10% people suffer from UTI in their childhood.

In U.S. alone it is the reason for seven million hospital visits, one million emergency department visits and one thousand hospitalizations (Salvatore, Salvatore et al. 2011). The cost of these infections is substantially equal in rapports of lost time at work and expenses of medical care. In the United States, the direct budget of UTI management is estimated at 1.6 billion USD yearly (Shahzada, Shehzada et al. 2014).

1.4 Sign, symptoms and pathogenesis of UTI

Initially UTI remains asymptomatic and sign and symptoms vary according to the site of infection. The most common sign and symptoms of UTI are dysuria, (Demetriou, Emans et al. 1982) increased frequency of urination (Foxman 2002), vaginal discharge (Greenfield, Friedland et al. 1974), flank pain (Seixas-Mikelus, Jenkins et al. 2009), vomiting (Kim, Wie

et al. 2014), hematuria (Kim, Wie et al. 2014), pyrexia (Wan, Skoog et al. 2012) and pyuria (Shigemura, Tanaka et al. 2014).UTI is initiated by uropathogen residing in the gut and contaminate periurethral area (Figure 1).

The uropathogen may migrate to bladder and colonize it by pili and adhesion mediated invasion. Neutrophils are released in human (host) inflammation response. The uropathogen resistant to neutrophil response would undergo multiplication and biofilm formation. The residing uropathogen causes epithelium damage and release of pathogenic toxins and proteases. The uropathogen can further migrate to kidney and can cause colonization of the kidney, releasing the toxins in kidney causing tissue damage leading towards bacteremia (Flores-Mireles, Walker et al. 2015).

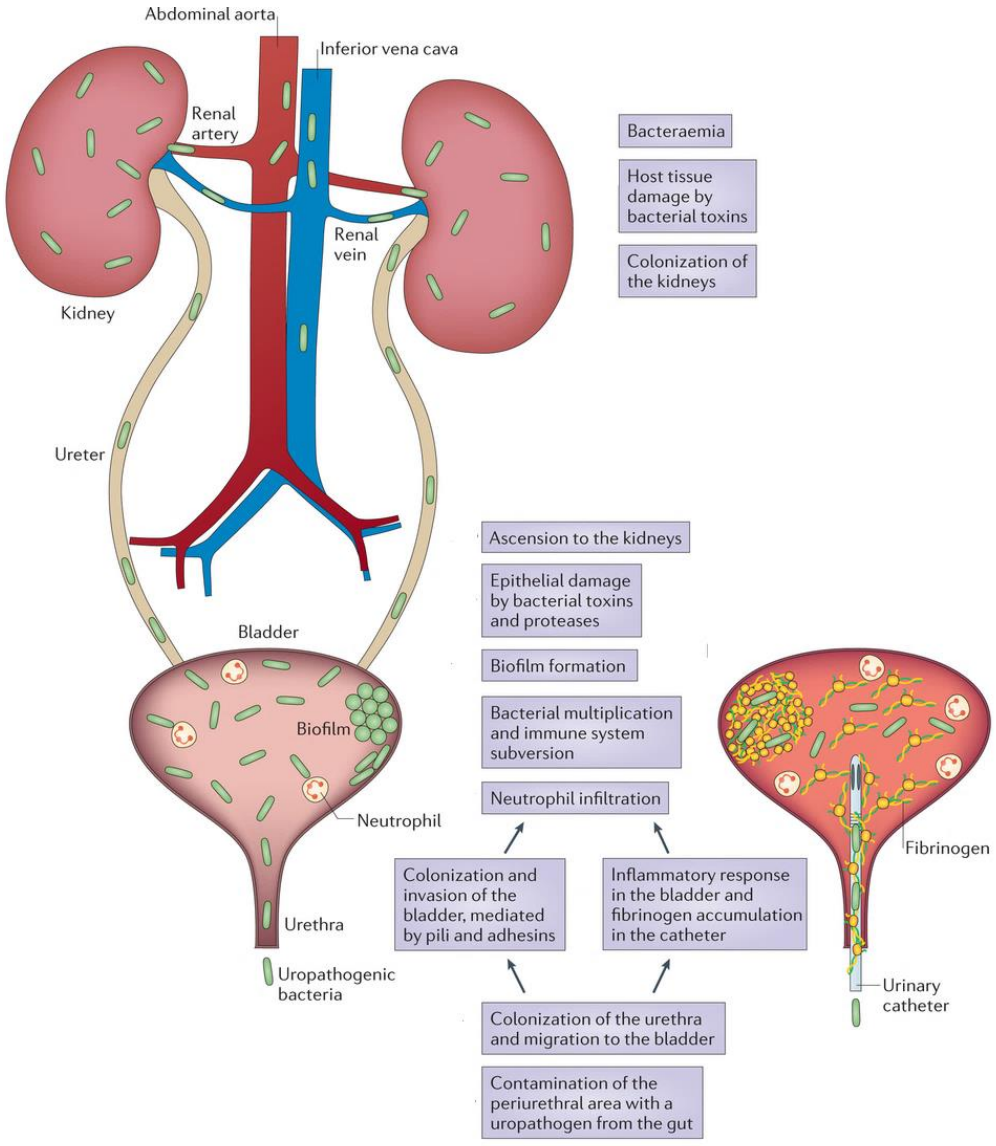


Figure 1: Urinary tract infections (UTIs) pathology. The figure illustrates the different stages of uncomplicated UTI leading to complicated UTI and catheter induced UTI (Adapted from Nature Reviews Microbiology).

1.5 UTI treatment and interventions

Main stay of treatment is antibiotics (Sakran, Smolkin et al. 2015). Commonly used antibiotics for UTI treatment are cefmetazole (CMZ), ceftriaxone(CRO), gentamicin (GEN), amoxicillin-clavulanic acid (AMC), trimethoprim-sulfamethoxazole (SXT) and ampicillin (AMP) (Chen, Chang et al. 2014). The rate of effective antibiotic treatment is declining due

to increase in the antibiotic resistance (Graham 2015).

Other than the usual contemporary management panache, antibiotic stewardship program in pathological labs, research labs and pharmaceutical industry are also being utilized to prevent increased ABR rates (Laxminarayan, et al;2015). In different health care systems controlled antibiotic prescriptions and patient compliance are being enforced to prevent recurrent infections. Phage based treatment for bacterial infections are also available and proved to be efficacious.

Despite the fact they had successfully completed its clinical trials, it's still not given mass level authorization because of the chances of ABR blooms by plasmid transfer (Chanishvili 2012). Natural remedies like cranberry juice and supplements are also prescribed but they do not successfully treat the infection nor their mechanism of efficacy is established (Raz, Chazan et al. 2004; Barbosa-Cesnik, Brown et al. 2011).

1.6 Delinquencies in effective UTI treatment

UTI is a complex infection and requires aggressive treatment, but antibiotic resistance makes it more complicated and require rapid progress in development of new drugs (Shepherd and Pottinger 2013). No vaccine based treatment has yet been successfully developed for UTI (Brotman, Ravel et al. 2014). Recurrent infections suggest a need for both vaccine and drug based treatment (Hancock, Nijnik et al. 2012).

In case of drugs we are facing a huge antibiotic resistance challenge and in case of vaccine we need more efficient candidates (Keshavjee and Farmer 2012). Due to resistance, drugs which were very effective earlier have become ineffective (Laxminarayan, Duse et al. 2013). The antibiotic resistance, susceptibility and epidemiological studies around the world

are indicating that the current effective antibiotics would also exhaust in the near future (Laxminarayan, Duse et al. 2013).

1.7 Antibiotic resistance and its health care cost

The phenomena of ABR have led to the need of development of new antibiotics (Blair, Bavro et al. 2015). According to the World Health Organization (WHO), the ABR is the resistance is the decreased susceptibility of antibiotic which was once used for the treatment of an infection caused by that organism (Llor and Bjerrum 2014). ABR is basically the characteristic of the bacterium not the host (Do Thi, Chuc et al. 2014). Antibiotic resistance can be natural (Figure 2) or acquired (Figure 3) (Finley, Collignon et al. 2013).

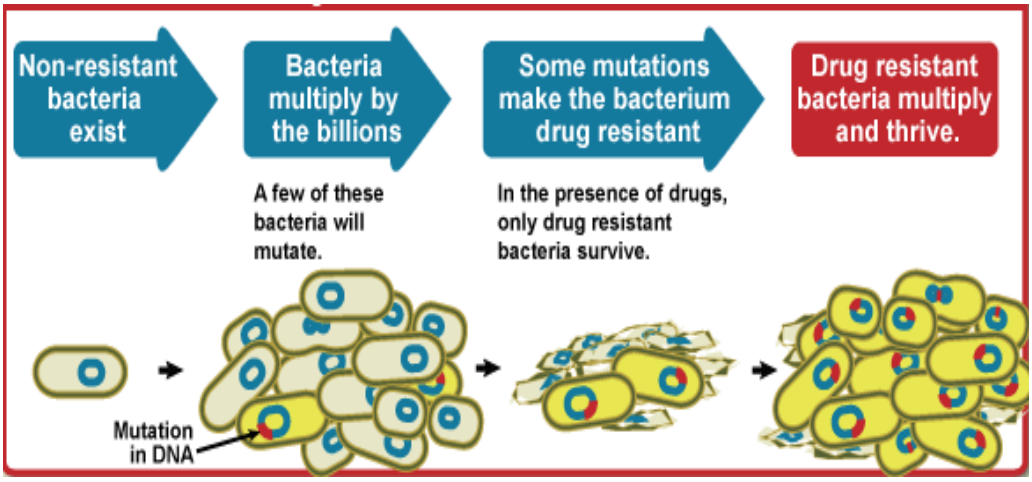


Figure 2: Genetic mutation cause drug resistance. The figure illustrates the development of natural antibiotic resistance phenomenon in bacteria (Courtesy of National Institute of Allergy and Infectious Diseases (NIAID)).

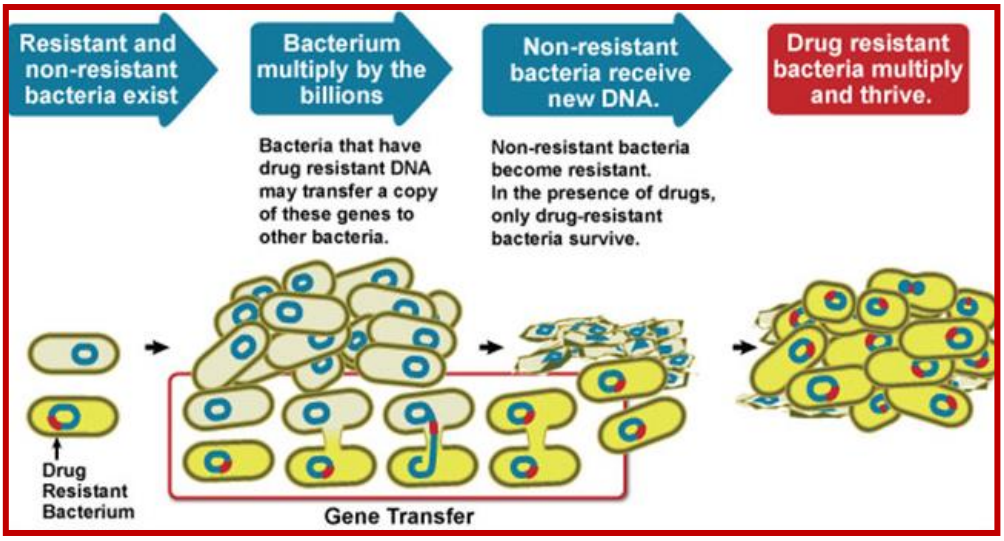


Figure 3: Gene transfer facilitating spread of drug resistance. The figure illustrates the development of antibiotic resistance phenomenon via gene transfer (Courtesy of National Institute of Allergy and Infectious Diseases (NIAID)).

It is estimated that in 2050 it would cost $100 trillion (£64 trillion) to the global economy. In 2050, the death toll due to untreatable bacterial infection would surpass mortality rate due to cancer. With these scenarios it is presumed that if there would be no improvement in the susceptibility patterns of antibiotics then any bacterial epidemic condition would be worse than Ebola epidemic (Phimister, Arias et al. 2015). The antibiotic resistance, susceptibility and epidemiological studies around the world are indicating that the current effective antibiotics would also exhaust in the near future (Laxminarayan, Duse et al. 2013).

1.8 Current trend of antibiotic drug development

Research and development of novel antibiotic drug development is facing a decline (Högberg, Heddini et al. 2010). More than half of the pharmaceutical companies have left research in regards of new antibiotic development. The pipeline is becoming sparse;

worldwide the progress is not satisfactory. Market for antibiotic drugs is growing 5% where as there is progression of more than 16% for antiviral drugs and vaccines.

In the past, utilization of high throughput technology have resulted a very insufficient drug candidates with high costs for development and research. These factors discourage the pharmaceutical industry for major investment in antibiotic drug development. Pharmaceutical companies even go for a research program on a non-profit basis they have being cooperate organizations they have to progress in a businesslike manner (Schäberle and Hack 2014).

1.9 Vaccines, a solution to pharmaceutical retaliation

In 2010, FDA approved two antibacterial vaccines (meningococcal conjugate vaccine) and Prevnar (pneumococcal conjugate vaccine) (Mullard 2011). Contrary to antibiotic drug development vaccine development is a different scenario. According to an estimate, a pharmaceutical company with a successful vaccine is earning US $3.6 billion annually (Jaffe, Lerner et al. 2001). Whereas the cost for development project to marketing is US$ 500mn – 1bn, over the period 12-15 years.

In contrast to antibiotic dwindling market, sales for vaccines are persistently increasing and have reached from $6bn in 2000 to $33bn in 2014 (Organization 2000). Governments in developed countries are also promoting vaccine development by completely bearing the cost of a successful vaccine development program and also offering other initiatives like patent extensions and additional prize money (Jaffe, Lerner et al. 2001).

1.10 Academics, research and pharmaceutical industry

To overcome the deadlock there is a requirement of co-operation between research, academia and industry joint efforts are necessary. Antibacterial target identification is the first step in the modern development of drugs or vaccines. Less than 20% of the antibacterial

targets have been discovered up till know. Target identification stage of antibiotic development is more suitable for academic research. It requires limited resources and can be facilitated by biotechnology industry. The toxicological studies and animal trials require major funding's that can be supported by pharmaceutical industry. The total cost for any antimicrobial development is 500-800 million € and requires more than 4-6 years after its discovery. The research in novel antibiotic vaccine or drug development cannot proceed without collaborating the academic research, biotech and pharmaceutical industry (Schäberle and Hack 2014).

OBJECTIVES

- Intra-specie pan-genomic features analysis of all the strains for a common uropathogen (*Escrishia coli, K.pneumoniae, K.oxytocca, P.aroginosa, E.faecalis, E.faceium, S.aureus* and *P.mirabilus*).
- Inter-specie pan-genomic features analysis of all the strains for a common uropathogen (*Escrishia coli, K.pneumoniae, K.oxytocca, P.aroginosa, E.faecalis, E.faceium, S.aureus* and *P.mirabilus).*
- *In-silico* identification of vaccine-candidate proteins utilizing pan-genomics and reverse vaccinology approach and, Epitope mapping of prioritized vaccine candidates.

CHAPTER 2

LITRARTURE REVIEW

CHAPTER 2 LITRATURE REVIEW

2.1. Antibacterial vaccines

Vaccines are still the most prevailing agents for infectious disease prevention. Their use precludes the necessity for antibiotic by averting infection in the first place, and they have been used in the clinical setup from long ago. A seven-mere pneumococcal conjugate vaccine was developed for UTI treatment though it was efficacious but then there was the problem of new serotypes that were not covered by that vaccine. Vaccines should be developed against the pathogens giving major clinical problems due to antibiotic resistance (McKenna 2012; Laxminarayan, Duse et al. 2013).

2.2. Reverse vaccinology

Reverse Vaccinology is the advancement in conventional vaccinology which involves fastidious methods of pathogen cultivation, harvesting, identification of both most immunogenic and protective antigen suitable as vaccine target among thousands of proteins and then finally its screening in animals and clinical trials (Figure 4). Reverse vaccinology came into focus when Rino Rappuoli developed a successful vaccine against *B meningococcus*. It involves the screening of an entire pathogenic genome utilizing bioinformatics approaches to identify the genes. The provision of completely sequenced bacterial and host genomes has made the genomic based approaches far less expensive.

These approaches are used to prioritize the vaccine target usually focusing on extracellular localization, signal peptides and B-cell epitopes.

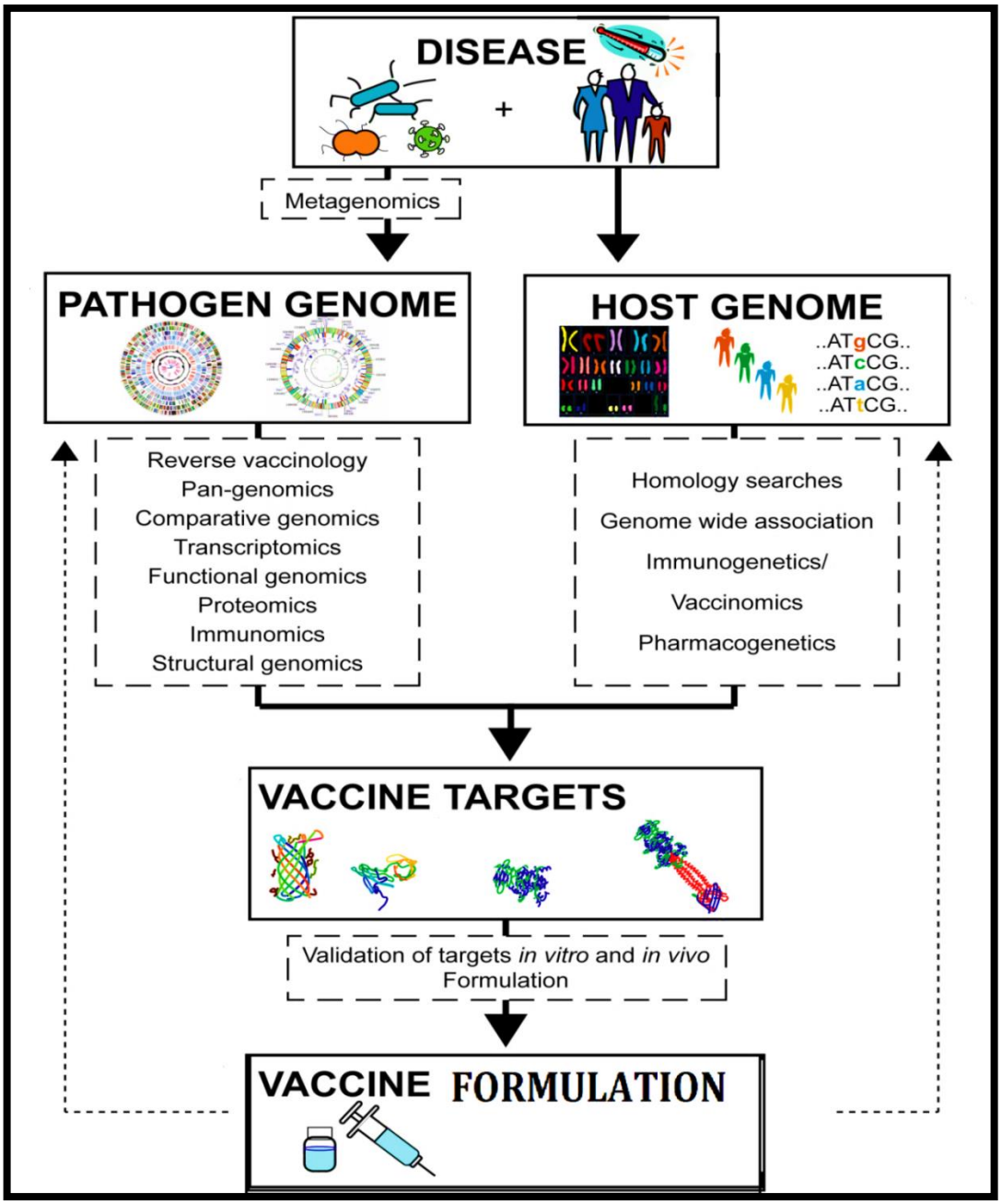


Figure 4: Genomics-based approaches used in the control of a disease to the development of a vaccine. The genomic based disease control strategy involves the identification of the causative agent. Utilization of transcriptomeic, immuno-proteomic or structural-genomic analysis for vaccine target identification. Screening of human genome is also carried out to avoid human homologs as vaccine targets. Vaccine clinical response and the safety of the vaccine is analyzed. Then the vaccine formulation is forwarded for licensing (Courtesy of Seib KL, Dougan G, Rappuoli R (2009)).

The proteins with desirable attributes are selected as vaccine candidates and further screened in animal and human clinical trials. It provides solution to many problems arrived during vaccine development in conventional vaccinology as it provides access to antigens which either expressed in less concentrations or do not express at all invitro. (Pizza, Scarlato et al. 2000; Capecchi, Serruto et al. 2004; Rinaudo, Telford et al. 2009; Delany, Rappuoli et al. 2013; Kanampalliwar, Soni et al. 2013). Major advantage is its utilization for vaccine development against non- culturable microbes. It also has a drawback that it can only be applied to identify non- proteinaceous antigens like lipopolysaccharides and glycolipids.

2.3. Pan-genome analysis

Pan-genome encodes all the possible lifestyles carried out by the organism under observation. Advances in technologies of next generation sequencing have enabled genome-analysis at a huge scale. But such an analysis requires a frame work specifically for predicting and modeling genomic diversity and pan-genome analysis provides such a frame work. It involves the estimation of genomic diversity, prediction of additional whole genome sequences and may even necessary for characterizing such diversity. (Medini, Donati et al. 2005; Reinhardt, Baltrus et al. 2009; Reno, Held et al. 2009; Donati, Hiller et al. 2010).

Pan-genome consists of a core-genome, dispensable genome and strain specific genes as shown in Figure 5. Core-genome is the sub-set of genes shared by all the strains of the clad in the data set at hand, responsible for carrying out all the necessary activities for the organism survival. The dispensable genome consists of a sub-set of genes shared by a group of strains in the data set under analysis including accessory and unique subsets. Such a sub-set is contributing either to bio-chemical pathways or biological functions which may not be

necessary for the survival of the organism but convene selective advantages such as niche adaptation, antibiotic resistance or new host colonization.

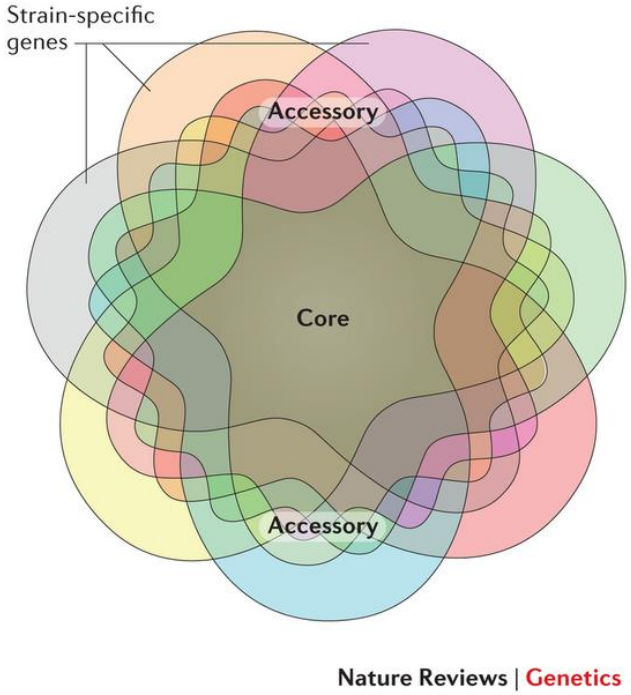


Figure 5: Pan-genome and core genome diagrammatic depiction. The pan-genome of an assembly denotes to the totality of all the genes that are existent in members of the assemblage. Pan-genomes consist of the core genome, which comprises the genes conserved in all members of a group of interest, and the accessory genome comprising the genes that are present in only one or a few members of the group. (Courtesy of Soucy, S. M., J. Huang, et al. 2015)

But pan-genome analysis has a limitation of drawing a resilient conclusion from a big complex data. Still it is regarded as the best framework design in case of closely related species genome analysis (Medini, Donati et al. 2005; Tettelin, Masignani et al. 2005; Hiller, Janto et al. 2007; Lapierre and Gogarten 2009; Rodriguez-Valera and Ussery 2012; Donati and Rappuoli 2013).

2.4. Vaccine against Escherichia coli

E. coli alone is responsible for more than 90% of extraintestinal infections. UTI complications even lead to bacteremia and sepsis. Decrease in efficacy of the current

therapeutic regimes due to ABR against *E. coli* (Figure 6) and recurrent infections have focused the research in uropathology towards vaccines (Kunin 1987; Litwin, Saigal et al. 2005; Pitts, Niska et al. 2008; Poolman and Wacker 2015). Most commonly prescribed antibiotics against *E. coli* in UTI are 3rd generation cephalosporins. *E. coli* isolates from different parts of the world indicate the presence of ABR against cephalosporins around the world (Figure7).

Another problem with cephalosporins are their adverse effects including suppression of endogenous gut flora leading to reduction in vitamin K synthesis and occasionally causing clinically significant bleeding, pseudomembranous colitis, and inhibition of acetaldehyde dehydrogenase, causing acute ethanol intolerance. Immediate and delayed hypersensitivity reactions, various hematological disorders, nephrotoxicity are also occasionally reported (Platt 1982).

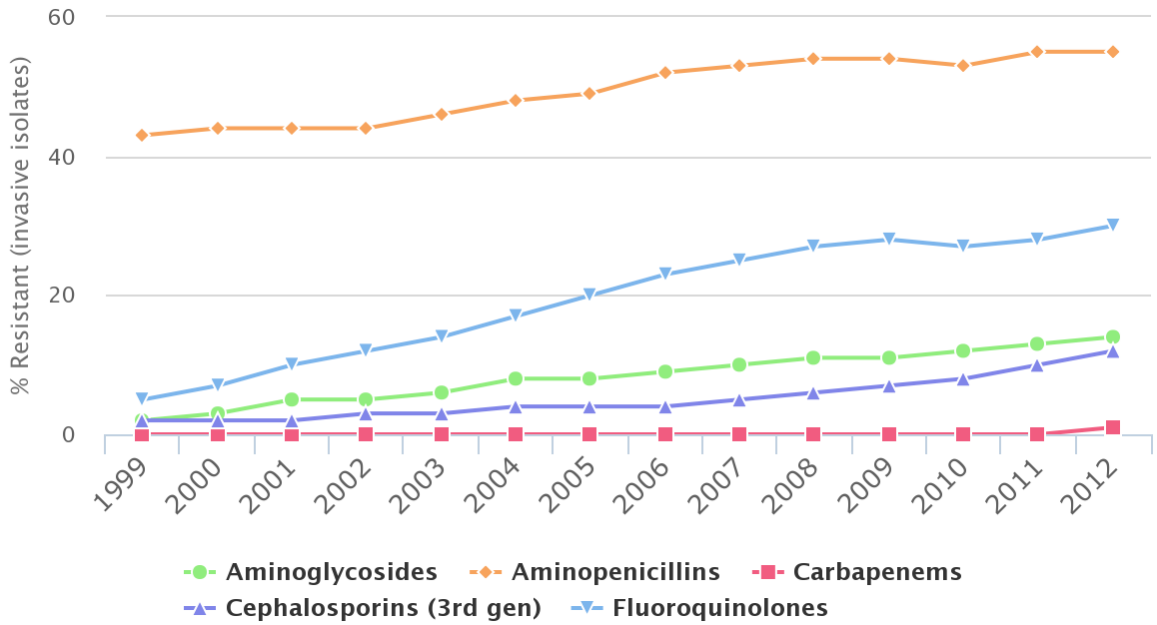


Figure 6: ABR causing decrease efficacy of antibiotics against E. coli. The figure illustrates the decrease in efficiency of antibiotic regime used against E. coli due to persistent increase in ABR rates. (Courtesy of: Center for Disease Dynamics, Economics & Policy CDDEP).

In case of vaccine development most important is the identification a suitable VF as a vaccine candidate and different approaches are applied in this regard. One classical approach is the genomic comparison of archetypical pathogenic *E. coli* with a commensal *E. coli* such as K-12 strain MG1655. The sequences present in the pathogenic but not in the commensal strain are further prioritized for the selection of a suitable virulence factor. Another methodology was to sequence the whole pathogenicity islands (PAIs) and then prioritize the resulting sequences for a vaccine candidate but it is a very limited approach.

An alternate approach is subtractive hybridization in which both pathogenic and non-pathogenic genomes are hybridized. This allows the unattached regions of pathogenic genome to remain available in hybridized genome for polymerase chain reaction (PCR) which can be further detected and analyzed for protein prioritization for a suitable vaccine candidate (Laupland, Bagshaw et al. 2005; Bagshaw and Laupland 2006; De Las Rivas and Fontanillo 2010; Salvatore, Salvatore et al. 2011; Sievert, Ricks et al. 2013; Szklarczyk, Franceschini et al. 2014) .

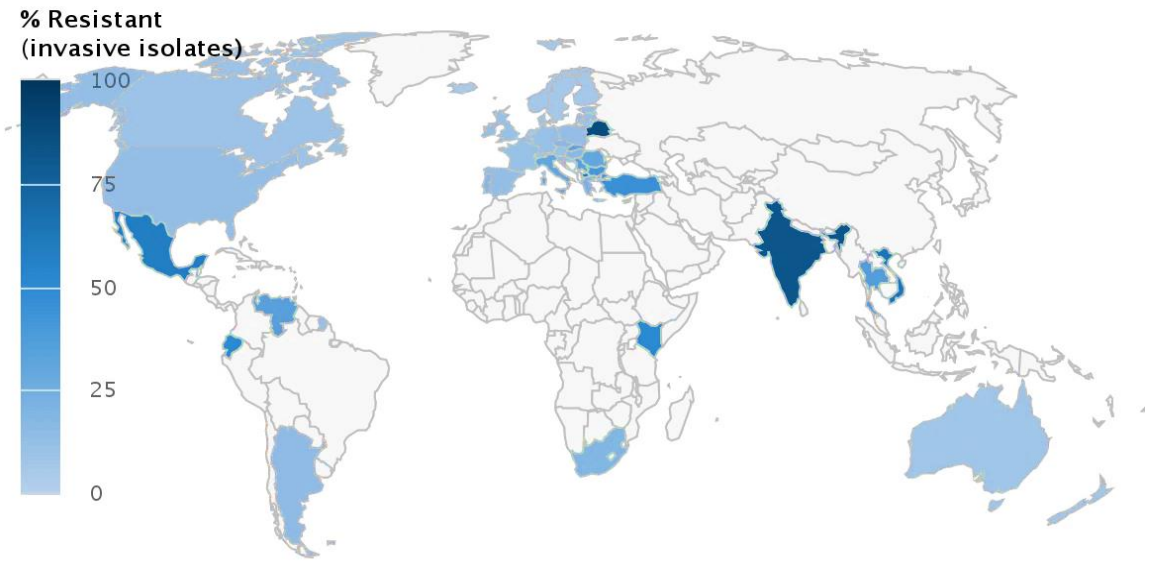


Figure 7: *E. coli* resistance to 3rd generation cephalosporins. The figure illustrates the world wide cephalosporin resistance developed by *E. coli. (Courtesy of: Center for Disease Dynamics, Economics & Policy CDDEP).*

But the advancement in research has rendered that an ideal VF cannot be selected on basis of presence or absence in pathogenic and commensal strains. It also depends upon their variable expression and qualitative features. Initially Type 1 fimbria was considered a promising vaccine candidate but it became controversial due to its presence in the commensal strains. Later on further research has revealed that significant difference exist in the commensal, pathogenic and among the clinical isolates with respect to the Fim H adhesion molecules. Currently different vaccines against ExPEC are under progress in different levels of clinical trials (Johnson 1991; Lim, Gunther et al. 1998; Sokurenko, Chesnokova et al. 1998; Gunther, Lockatell et al. 2001).

2.5. Advances in vaccine development for Klebsiella pneumonia and Klebsiella oxytoca

Extended spectrum beta lactamases (ESBL) carrying Gram-negative pathogens confer complicated infections which have become cause of concern from decades around the world (Figure 8).

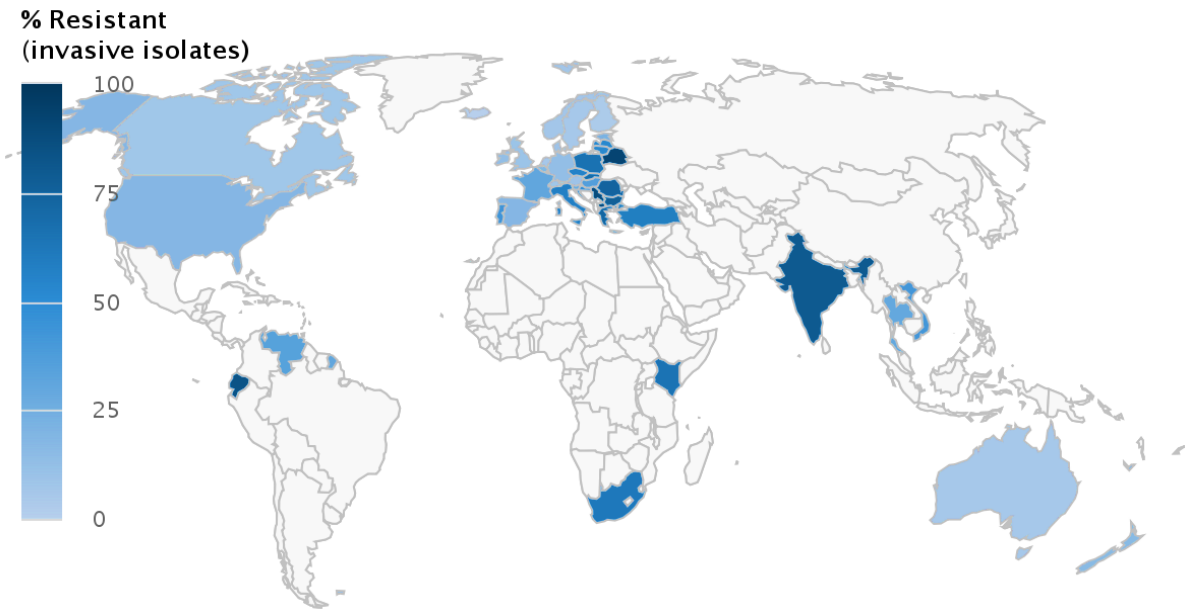


Figure 8: ***K. pneumoniae*** **resistance to 3rd generation cephalosporins.** The figure illustrates the world wide cephalosporin resistance developed by *K. pneumoniae* (Courtesy of: Center for Disease Dynamics, Economics & Policy CDDEP).

Major reason behind this problem is the rapid spread of antibiotic resistance genes due to the spread of antibiotic resistant plasmids among Enterobacteriaceae. Carbapenem resistance (CR) has also become very common in hospital acquired infections and resulted lack of efficacy in currently prescribed antibiotics (Figure 9). The commensal microbes are already resistant to major therapeutic regimes. *K. pneumoniae* resides in the intestinal flora and is part of the Enterobacteriaceae.

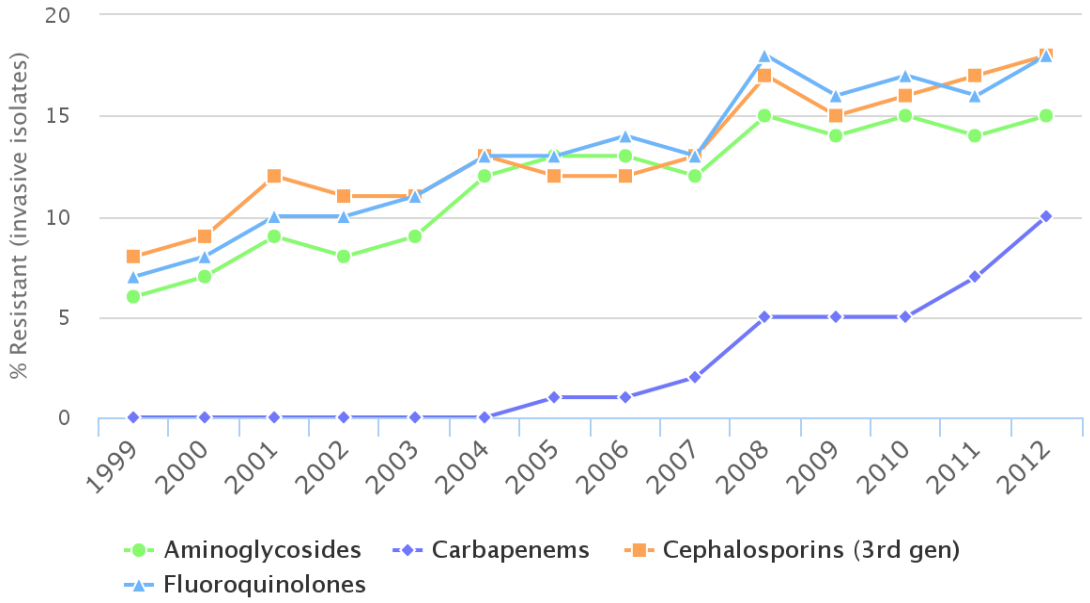


Figure 9: ABR causing decrease effiecncy of antibiotics *K. pneumoniae.* The figure illustratea the decrease in efficency of antibiotic regeime used against *K. pneumoniae* due to persistant increase in ABR rates (Courtesy of: Center for Disease Dynamics, Economics & Policy CDDEP).

K. pneumoniae is responsible for UTI, blood stream infections (BSI) and RTI in patients with compromised patients (Sader, Pfaller et al. 1994; Chong, Ito et al. 2011; Mathers, Cox et al. 2011; Gona, Barbera et al. 2014). Conditions like diabetes mellitus, cancer hepatic disorders, biliary tract disorders and excessive alcohol consumption disturb the host immune response and promote *K. pneumoniae* infection. Lack of early diagnosis and effective treatment may lead to higher chances of morbidity and mortality in bacterial pneumoniae. This has created a demand for an effective vaccine against *K. pneumoniae* (Tsay, Siu et al. 2002; Tsai, Huang et al. 2010).

A series of channel proteins called outer membrane proteins (Omps) are present within the outer membrane of *K. pneumoniae* to facilitate the uptake of components that are necessary for their growth and survival. Examples of such Omp molecules include OmpW, OmpX, OmpC, OmpX and FepA. These are able to induce an efficient humoral and cellular

immune response. Generally the vaccines for *K. pneumoniae* comprise native components like fimbriae, capsular polysaccharides and also lipopolysaccharides (Jeannin, Renno et al. 2000; Hong, Patel et al. 2006; Clements, Jenney et al. 2008; Baghal, Gargari et al. 2010).

K. pneumonia is known to invade both the innate and adaptive immune response of the host and the vaccines have proved to be unstable for such a persistent response. Epitope based vaccine is considered an efficient strategy to combat *K. pneumoniae* hazard. It is well known that an Epitope base vaccine of T-cell and B-cell epitopes integrated in a multi epitope vaccine can be recognized by the immune system. This would result in a cellular response induced by T-cells and a humoral response induced by B-cells.

The advancement in computational biology and bioinformatics along with the availability of completely sequenced genomes allow us to scrutinize these genomes for the presence of immune-dominant B-cell and T-cell epitopes for designing an efficient vaccine (Sbai, Mehta et al. 2001; van der Burg, Bijker et al. 2006; Bijker, Melief et al. 2007). No significant work has yet appeared in regard to vaccine development against *K. oxytoca*. An E. coli derivative vaccine is under animal trials but it is designed to prevent mastitis in bovine (Gonzalez, Cullor et al. 1989). *K. oxytoca* induced UTI is still treatable but its susceptibility profile is decreasing without a suitable treatment it can also become a problem like *E.coli* (Jean, Coombs et al. 2016).

2.6. Deficiency of efficacious Enterococcus faecalis and Enterococcus faecium vaccine

Enterococcus has recently emerged as multiple antibiotic resistant pathogens in hospital acquired infections. *E. faecalis* and *E. faecium* are the third and fourth most commonly isolated nonsocial pathogen worldwide. 100% vencomycin resistant isolates of Enterococcus

species have been isolated in different parts of the world. These species have immense ability to survive under unfavorable environmental conditions (Murray 2000; Treitman, Yarnold et al. 2005; Werner, Coque et al. 2008; Arias and Murray 2009; Zarb, Coignard et al. 2012).

Due to the high viability character they have emerged as an antibiotic resistant pathogen worldwide. This dramatic increase in antibiotic resistance signifies the need alternate treatment approaches. In case of vaccine development different surface antigens have been identified in *E. faecalis* and *E. faecium* but only three are considered as promising candidates Sag, Ace and ABC transporter sand were tested in animal trials. A peptidoglycan associated penicillin binding protein (PPB) and few other proteins Ddcp, LysM, PpiC are also under consideration but their mechanism are still unknown (Burnie, Carter et al. 2002; Treitman, Yarnold et al. 2005; Hyyryläinen, Marciniak et al. 2010; Kropec, Sava et al. 2011; Zhang, Paganelli et al. 2012).

2.7. Pseudomonas aroginosa antibiotic resistance and vaccine development

Pseudomonas aroginosa is one of the pathogens known to cause UTI (Jarvis and Martone 1992). In case of nosocomial catheter associated infections *P.aroginosa* is the third most common pathogen. *P.aroginosa* has the ability to form biofilms on the surface of indwelling catheters in cauterized patients. Their growth results in the formation of micro biomes which then fuse to form biofilms.

These biofilms are resistant to antibiotics and host defense mechanism. This phenomenon has resulted in worldwide *P. aroginosa* antibiotic resistance (Figure 10). Biofilms are the main contributors of *P.aroginosa* pathogenicity and often lead to persistent and recurrent infections. Other UTI *P.aroginosa* is also the major cause of pneumonia,

bacteremia, and burn-site infection and is also the most prevent pathogen for causing cystic fibrosis (CF)(Donlan 2001; Drenkard 2003; Wagner and Iglewski 2008).

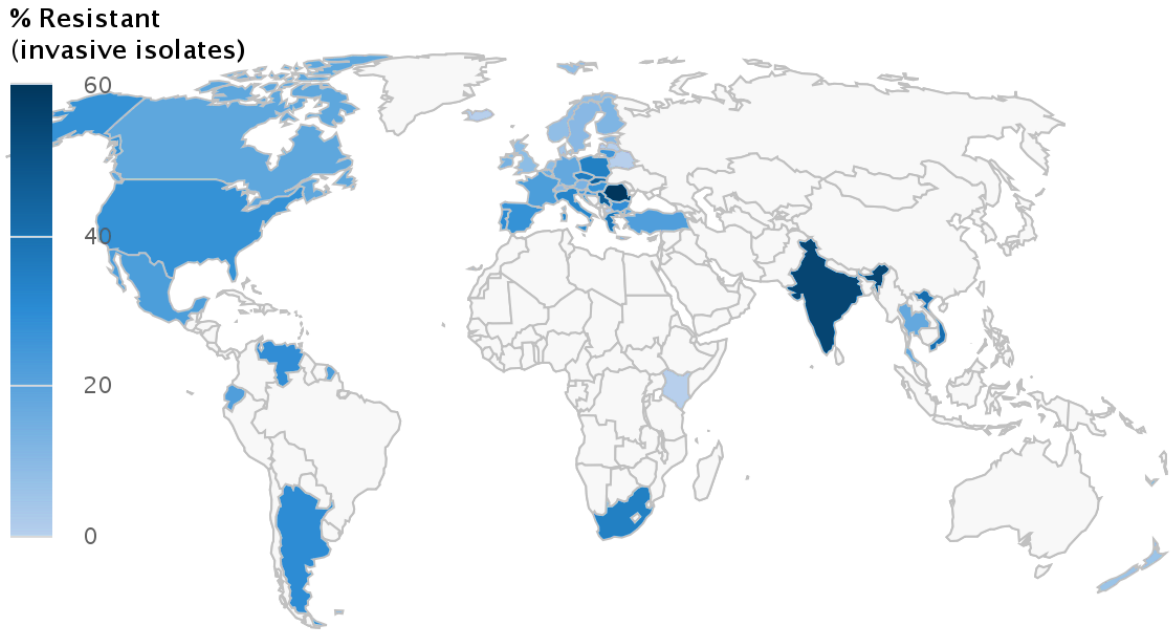


Figure 10: *P. aroginosa* resistance to fluoroquinolones. The figure illustrates the world wide cephalosporin resistance developed by *P.aroginosa*. *(Courtesy of: Center for Disease Dynamics, Economics & Policy CDDEP.)*

Even new antibiotics are being compromised by the emergence and rapid spread of antibiotic resistant strains of *P.aroginosa* (Figure 11) and other bacterial pathogens in immune compromised patients. The increased incidences of emergence of new antibiotic resistant strains and diminishing pipeline of antibiotics in development have triggered the research towards vaccine development.

Efforts are being made from decades to develop a vaccine against *P.aroginosa* but there is no substantial development and no extended spectrum vaccine against *P.aroginosa* is yet available in the market. The efforts for vaccine development were mainly focused toward

active immunotherapy rather than passive immunotherapy (Howett and Kreider 2002; Sorichter, Baumann et al. 2009; Sharma, Krause et al. 2011; Prevention, Control et al. 2013).

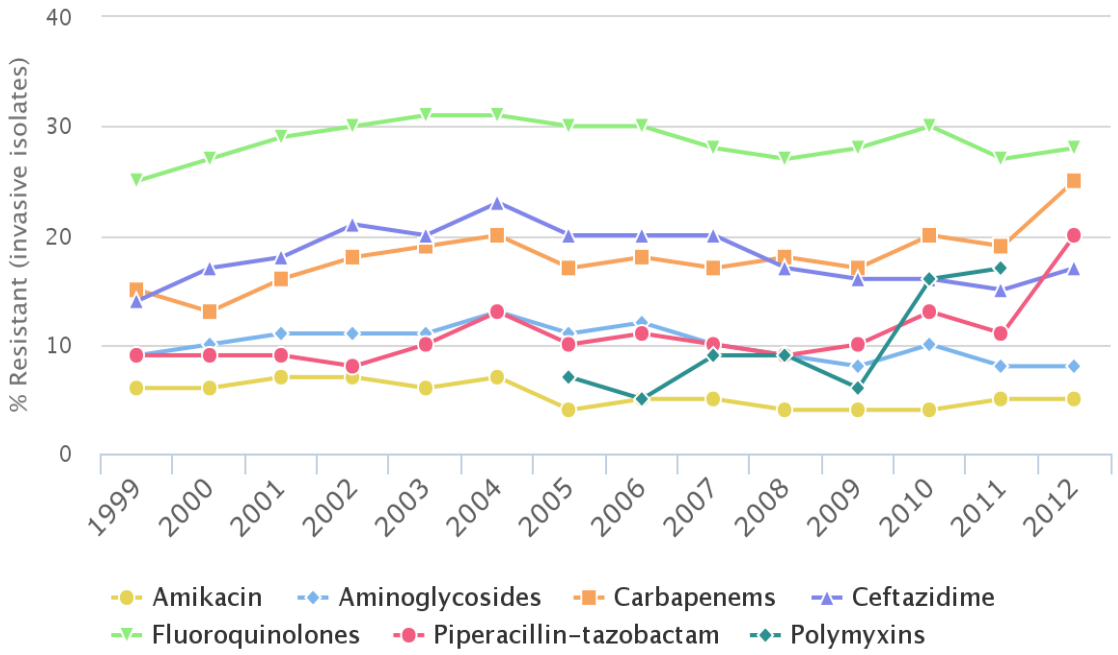


Figure 11: ABR causing decrease effiecncy of antibiotics *P.aroginosa* The figure illustratea the decrease in efficency of antibiotic regeime used against *P. aroginosa* due to persistant increase in ABR rates. (Courtesy of: Center for Disease Dynamics, Economics & Policy CDDEP.)

The vaccine antigens identified so far include lipopolysaccharide (LPS) O antigens, outer membrane proteins, flagella, PCrV and antigens of type III secretory system. LPS are one of the outer membrane proteins they were first to be investigated for P.aroginosa vaccine and were also lead towards passive immunity. But the problem with their use was that they were not conserved in different serotypes. Flagellin was comparatively more conserved (Rumbaugh, Sawa et al. 2003; Sorichter, Baumann et al. 2009; Sharma, Krause et al. 2011).

The ideal antigen required for *P.aroginosa* should have two main characteristics of conservancy in all serotypes and ability to generate both cellular and humoral response of the host immune system to combat strong pathogenic mechanism of *P.aroginosa*. The

advancement in computational biology and completely sequenced genomes allow us to scrutinize those genomes for identifying immune dominant B and T cell epitope for designing an efficient vaccine (García-Quintanilla, Pulido et al. 2016).

2.8. Vaccine against Proteus mirabilis

Proteus mirabilis is the common cause of complicated UTI especially in patients with comorbidities and in case of catheterized for longer period of time. It was isolated in 44% of the complicated UTI cases in catheterized patients. Its pathogenesis mechanism involves the high level production of urease enzymes. Urease converts urea into ammonia and increase the urine pH which if remains untreated leads to urolithiasis (Li and Mobley 2002).

P.mirabilus is also responsible for other nosocomial infections like sepsis and pneumonia. Though 10-20% of the *P.mirabilus* isolates were found to be antibiotic resistance because they hide within the stones crystallized by urease unaffected by antibiotics and often become the cause of recurrent infections. Vaccination is evidenced to be a suitable option. Different vaccines carrying outer membrane protein antigen and heat killed pathogen were tested but gave 70% protection in case of antigens the antigen genes were conserved but there was possibility of variation in expression in different strains (Li, Lockatell et al. 2004; Scavone, Miyoshi et al. 2007).

2.9. Broad spectrum vaccine against uropathogenic bacteria

Increased ABR, side effects to certain pharmaceuticals, recurrent UTI infections and normal flora alterations demand potent vaccines for crossing barriers towards efficacious treatment. Vaccine targeting Fim H of type 1fimbriae was earlier controversial due to the presence of the target in the commensal *E.coli* latter variable expression and regulation of

target profile was enough justification to lead the vaccine towards animal trials and it has generated promising results in animal models. But still there is a requirement for a broad-spectrum anti-uropathogenic vaccine (Mobley and Alteri 2015).

Uro-Vaxon is a preventive vaccine for UTI by PM Parma, a Swedish pharmaceutical company. The vaccine is currently available in Europe and Canada. The vaccine provides modest protection and it encompasses extracts of 8 bacterial uropathogens in a single tablet dosage form. Even for the most protection the subject need to complete a course of 90 days at a frequency of one tablet per day. Lack of patient compliance towards therapeutic regimes of long duration is already a problem in health care sector.

Its lack of ability to provide immunogenic coverage against all the genome strains is another difficulty in providing broad spectrum immune response against the pathogens. Another vaccine Salco Urova, comprising heat killed mixture 10 uropathogens (6 are *E. coli* strains) in a suppository dosage form. In clinical trials the vaccine has reported to significantly decrease the reoccurrence of UTI. Decrease in the reoccurrence rate is not satisfactory for an efficacious vaccine (Langermann, Möllby et al. 2000; Brumbaugh and Mobley 2012).

A vaccine product of Greek and Bulgarian collaboration (BB-NCIPD) Limited, consisting of attenuated representative strains of *E. coli*, *P. mirabilis, K. pneumoniae* and *P.aroginosa*, is in animal trial phase. The organization has applied for the approval of freeze dried formulation so that the vaccine can be given as an oral solid dosage form. Another vaccine carrying a combination of four individual antigens Ire, Hme, Iuta and FyuA with cholera toxin CT is in process. Research is required to determine its appropriate adjuvant,

route of administration and dosage form. Also the mechanism of protection still need to be identified which would be necessary for dose adjustment and other pharmacokinetic parameters before moving forward to clinical trials (Langermann, Möllby et al. 2000; Bauer, Rahlfs et al. 2002; Brumbaugh and Mobley 2012; Mobley and Alteri 2015).

The problem raised by uropathogens is huge and the horizontal gene transfer of pathogenic and antibiotic resistance along with natural process of evolution is further complicating the current situation of anti-UTI vaccine therapy. A problem adding a heath care burden of $billions cannot be left on the hope of few vaccine with specific combinations, the fate of which are not yet known. The current scenario requires exploration in vaccine development sector, in order to fill the gap between ideal potent and broad-spectrum vaccines.

2.10. Essential characteristics of ideal antibacterial targets

Following are considered essential characteristics for a vaccine target.

- Essential non-host homolog.
- Non-human analogs considered to minimize possible off target side effects and absorption, distribution, metabolism, excretion and toxicity (ADMET).
- Preferable if involved in more than one pathway.
- Preferable if involved in pathogenesis.
- Membrane exposed enzymes or transporters are considered target for both drug and vaccine development.
- Targets common to most pathogen strains and related species are more preferable (Barh, Tiwari et al. 2011).
- Target should have fully elucidated structure and function along with physical

  properties which make it amiable to biochemical assays (Koehn and Carter 2005).

CHAPTER 3

METHODOLOGY

CHAPTER 3 METHODOLOGY

The present finding and categorization strategy encompasses four phases of analysis. (Figure 12) Phase: I involve the A) Mining of core proteome data set, through pan-core genome analysis of completely sequenced strains available for each common uropathogen (Table 1) at the time of analysis along with analyzing the important genomic features and confirming the presence of core genes in COG and KEGG Database B) Identification of presence of ABR genes in the core-genome. Both inter-specie and discrete pan-core analysis were carried out separately for ensured and comprehensive study. Phase: II passes the extracted core proteomes of the bacterial pathogens through subtractive channel of analysis in order to prioritize the proteins as an efficient vaccine candidate.

Table 1: The eight common UTI causing pathogens analyzed in the current study. The organisms analyzed in this study, the number of genomes analyzed and their reference genome.

Sr. No.	Organism	Pathogen staining type	Analyzed genomes	Reference genome
1	*E. coli*	Gram negative	159	*E.coli IAI39*
2	*K.pneumoniae*	Gram negative	52	*K. pneumoniae HS11286*
3	*K. oxytoca*	Gram negative	9	*K. oxytoca ASM102219v1*
4	*P. aroginosa*	Gram negative	59	*P. aeruginosa PAO1*
5	*E. faecalis*	Gram positive	7	*E. faecalis V583*
6	*E. faecium*	Gram positive	12	*E faecium DO*
7	*S. aureus*	Gram positive	109	*S. aureus NCTC 8325*
8	*P. mirabilis*	Gram negative	3	*P. mirabilis HI4320*

Prioritized protein mining phase comprises five sequence level analysis for filtering proteins that are A) non-homologous to humans, B) are important factors responsible for the virulence of the pathogen, C) essential for the survival of the pathogens, D) having trans membrane helices <2 for easy colonization and expression and lastly E) identification of protein localization in the bacterial cell. The prioritized vaccine candidate proteins resulting from phase I and phase II form the final list of protein targets. The final list of efficacious novel vaccine targets is quantitatively categorized in phase III.

Phase: III includes A) Molecular weight estimation, B) Identification of the interactome, C) Determination of broad spectrum targets by homology search against a list of urogenital pathogens and D) Functionality analysis F) Patent protein homology analysis. Phase: IV involves A) Epitope Analysis and B) Structural homology analysis of the targeted proteins Figure 12).

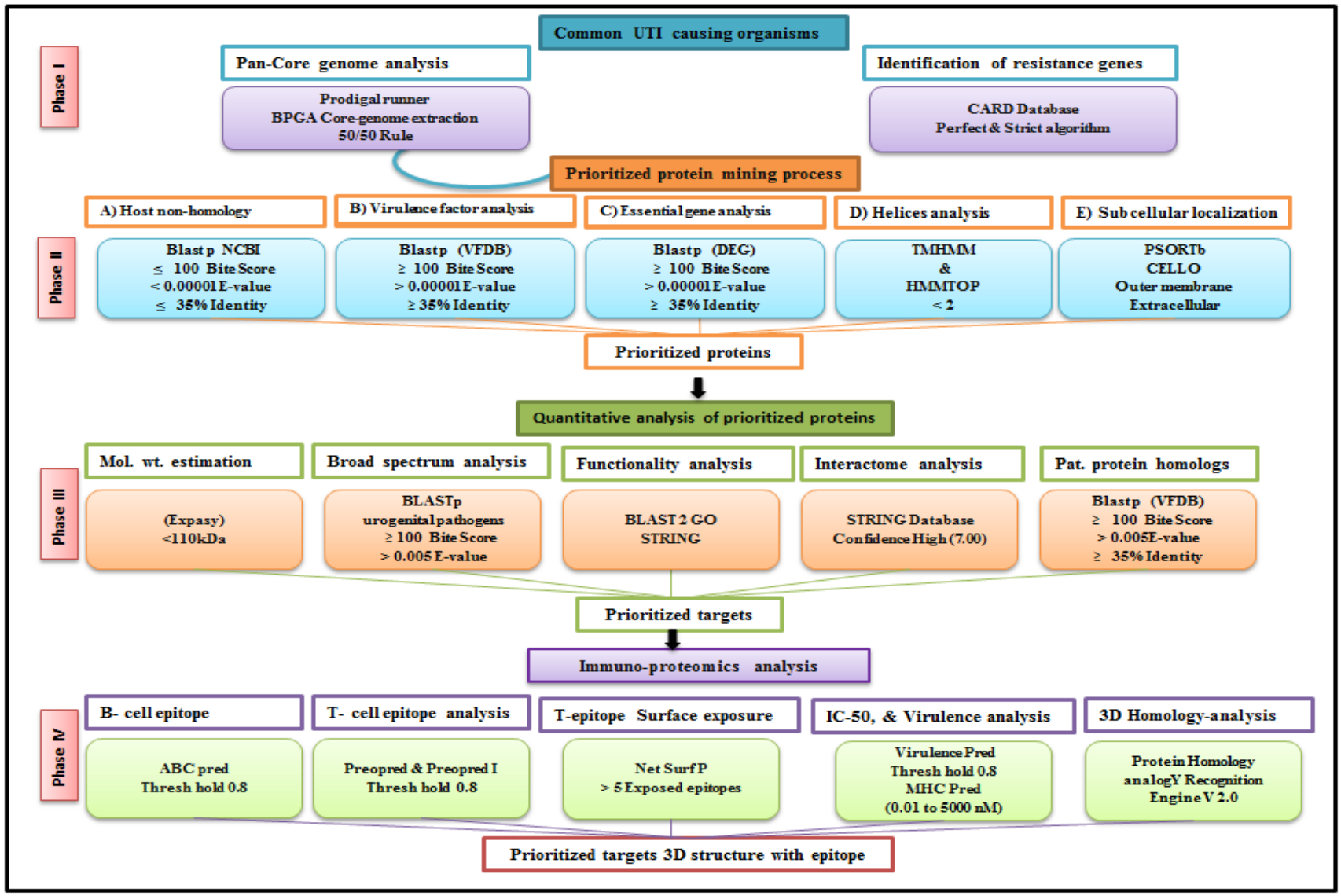


Figure 12: Flow chart for vaccine target prediction. The complete workflow illustrates each step and the selection condition followed in the four phases of analysis for the identification and characterization of promising vaccine targets.

3.1 Phase I: Mining of Core-proteome of antibiotic resistant UTI causing organisms

A. Genome selection and pan-core analysis

The whole genomes of 8 UTI causing pathogens under study were discretely retrieved from NCBI Genbank (http://www.ncbi.nlm.nih.gov/genome) (NCBI Genome, RRID: SCR_002474). Most common UTI causing and prospective antibiotic resistant organisms were selected for the analysis. Pan-core genome analysis was performed to extract core-genome of UTI causing organisms (Figure 13). Core-genome is the set of persistently conserved genes in all the strains of an organism. Solitarily the drafts of completely sequenced genomes were analyzed as a calibration for the pan-core exploration.

Both the chromosomal and plasmid DNA sequences were recovered and concatenated (Figure 13b) in a single FASTA file for each genome (Pop and Kosack 2004). The program Prodigal (Prokaryotic dynamic programming gene finding algorithm) was used for gene finding (http://compbio.ornl.gov/prodigal/Prodigal) (RRID: SCR_011936). Prodigal was employed because of its robust gene structure prediction, refine translation initiation site identification and decreased false positives (Hyatt, Chen et al. 2010).

Prodigal was accessed utilizing CMG workbench interface available at (www.cbs.dtu.dk/staff/dave/CMGtools/). It provided us with functional genome/proteome which was further analyzed. Prodigal runner generated fast (.fsa) files which were utilized as input for BPGA. USEARCH clustering tool was selected for being the fastest clustering tool for pan-genome categorization into conserved genome and dispensable genome subsets (Chaudhari, Gupta et al. 2016).

Both inter-specie and intra-specie pan-genome analysis was performed by BPGA- an ultra-fast pan-genome analysis pipeline. The conserved (both inter-specie and intra-specie core)genome of the UTI pathogens was assessed established on BLAST similarities between the genomes succeeding 50/50 rule. Simultaneously BPGA filtered the pan-genome based on the presence of their corresponding reactions in KEGG (Kyoto Encyclopedia of Genes and Genomes) available at http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/kegg/ and the database of Clusters of Orthologous Groups of proteins (COGs) available at http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/cog/.

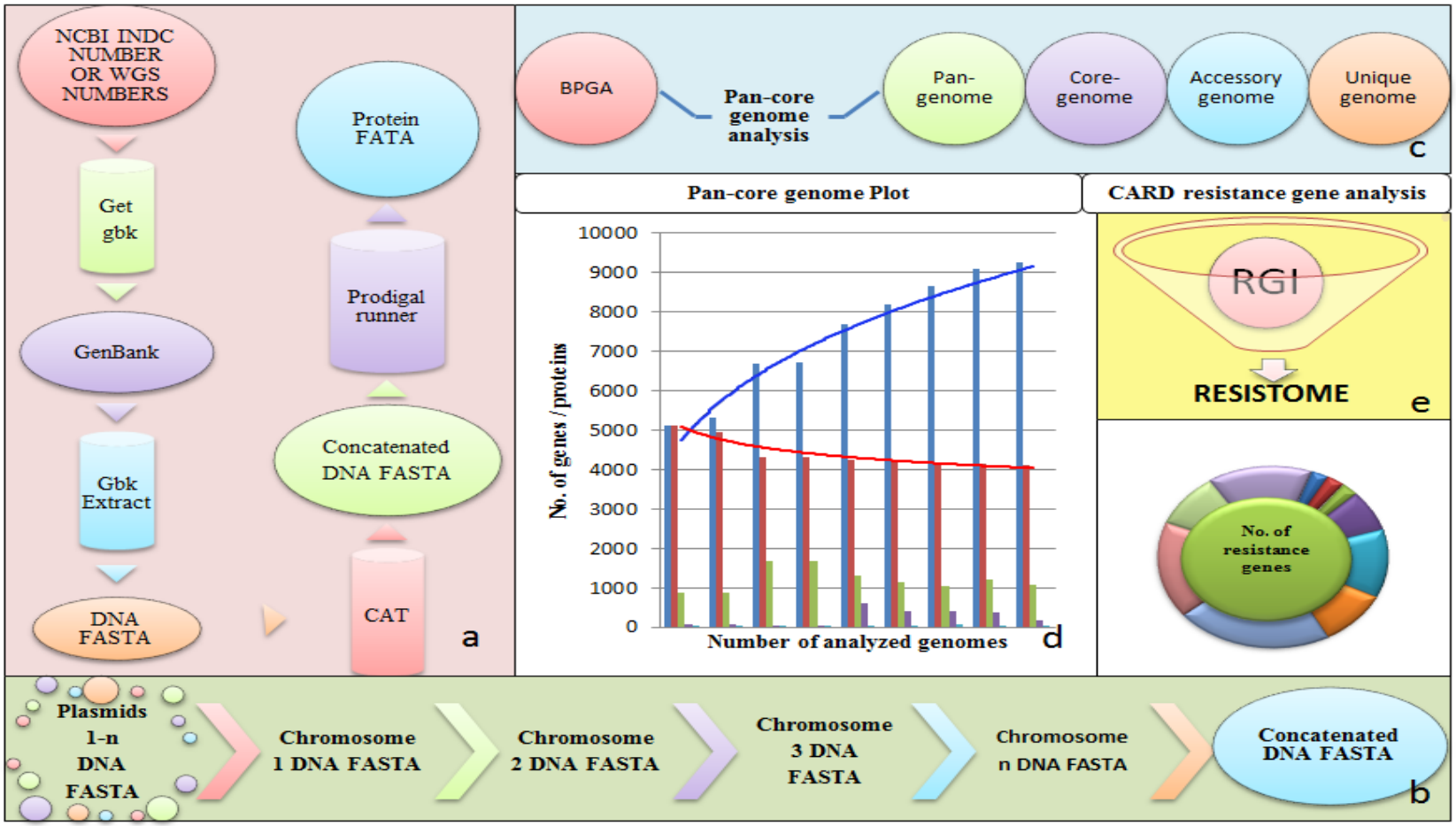


Figure 13: Schematic flow chart for Phase: I. The figure gives the visual illustration of the steps involved in (a) Obtaining functional genome from NCBI gbk files. (b) Concatenation of FASTA sequences of a genome (c) Pan-core genome analysis (d) Pan-core plot (e) Resistance genes analysis.

B. Antibiotic resistance genes analysis

Antibiotic resistance gene analysis was performed to identify the resistome in the core-genome. A gene or gene product either by drug class or resistance mechanism produces ABR is referred as resistance genes and the subset of proteins involved in such mechanism as resistome (D'Costa, McGrann et al. 2006). By identifying the resistant genes in the core genome the necessity of vaccine based solution of ABR hazard is demonstrated.

The identification of presence of resistant genes was done by utilizing the Comprehensive Antibiotic Resistance Database (CARD; http://arpcard.mcmaster.ca) carrying a repository of more than 1600 known antibiotic resistance genes (McArthur, Waglechner et al. 2013). It includes a common framework for the sharing of antimicrobial resistance data in the form of the novel Antibiotic Resistance Ontology (ARO) (McArthur, Waglechner et al. 2013). RGI was utilized to analyze the core, accessory and unique sub-sets of pan genome under Perfect and Strict paradigms.

3.2 Phase II: Priotoization of core-proteome for the redemption of suitable vaccine candidate

Five steps of protein mining were carried out to filter suitable vaccine candidates, as illustrated in Figure 15.

A. Human non-homology analysis

The core proteome of the uropathogens were subjected to a protein BLAST (BLASTp) search (Altschul, Gish et al. 1990) against the human proteome non-redundant database via Basic Local Alignment Search Tool available at http://blast.ncbi.nlm.nih.gov. Sequences which showed percentage identity <35 % human non-redundant proteome at an E-value of

<0.00001 with ≤100 bit score were marked as proteins having no significant homology with human proteins were selected for further analysis. This step is to minimize undesirable auto-immune reactions and cross-relativity to prevent the prioritize protein from binding at the active sites of a host homologs protein (Hassan, Naz et al. 2016).

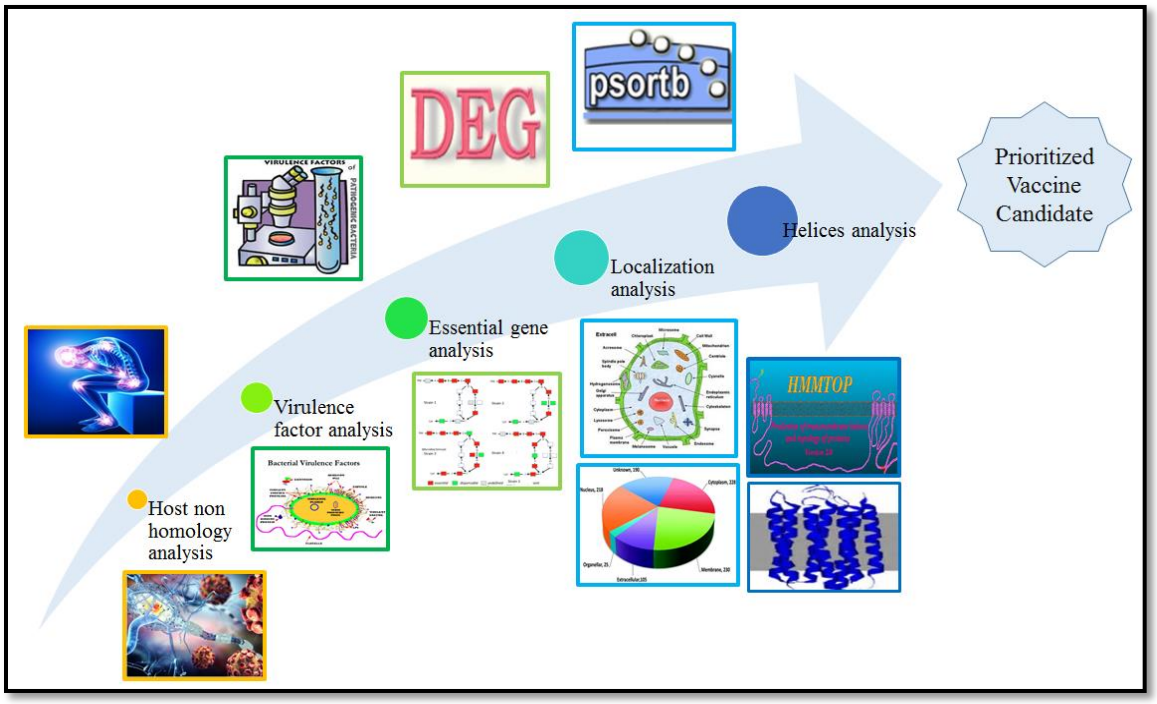


Figure 14: Schematic flow chart of Phase: II. The figure provides visual illustration of five steps carried out for protein priotoization.

B. Virulence factor analysis

Virulent factors (VFs) are characterized as potential vaccine targets for vaccine and drug development (Baron and Coombes 2007; He 2012). Virulence is the basic factor of pathogenesis responsible for some severe human diseases. Consequently virulence factor analysis is given high priority in insilco vaccine candidate priotoization (Naz, Awan et al. 2015). The core genomes of uropathogens were analyzed for the presence of virulent factors by performing a BLASTp search (Altschul, Gish et al. 1990) against Virulence Factor

Database (VFDB) available at (http://www.mgc.ac.cn/VFs/) and MvirDB (http://mvirdb.llnl.gov/). Sequences which showed e-value < 0.00001, with bit Score 100 and >35% percentage identity with annotated virulence factors in VFDB were marked as proteins as significantly virulent proteins.

C. Essential gene scrutiny

Essential genes are genes that are indispensable to support cellular life and constitute a minimal gene set required for a living cell to survive. The inactivation of a targeted protein should be lethal for the pathogen there for a prospective vaccine target and must be indispensible for the existence and growth of the pathogen signifying essential genes as potential vaccine targets (Jadhav, Shanmugham et al. 2014). The core genomes of uropathogens were also screened by a BLASTp search (Altschul, Gish et al. 1990) against Database of Essential Genes (DEG) http://tubic.tju.edu.cn/deg/, a repository of genes indispensable for the existence of an organism. DEG 6.1, contains 10,618 essential genes from prokaryotic and eukaryotic organisms (Zhang and Lin 2009). Core proteins which showed e-value <1.0 e-5, with bit Score >100 and >35% percentage identity were discernible as essential proteins based on the assumption that are essential in one organism are likely to be essential in another.

D. Protein helices analysis

Vaccine candidate protein should be able to easily colonize and express in the host cell so that it can play is pivotal role in host defense interactions and interplay (Naz, Awan et al. 2015). Therefore each core-genome sub-set was also screened through TMHMM data base (Fernando, Selvarani et al. 2004) and HMMTOP (Lu, Szafron et al. 2004) via default standards of their parameters. Proteins devouring 2 or more transmembrane helices were not

considered to be convenient targets due to the presence of more than one trans-membrane helical region, generally be unsuccessful to clone and express in the course of validation examination (Naz, Awan et al. 2015).

E. Sub-cellular localization

UTI causing organisms are diverse and belong to gram positive and gram negative genera (Table 1). Thus the proteins can possibly be located in five subcellular localizations including cytoplasm, plasma membrane, periplasm, outer membrane, or extracellular. The implication of the localization analysis is to illustrate the protein as drug or vaccine target. Cytoplasmic proteins can perform as probable drug targets, whereas surface membrane proteins can be considered as vaccine targets (Barh, Tiwari et al. 2011).

Subcellular location evidence of some proteins is established in protein databases like UniProt (Apweiler, Bairoch et al. 2004). In case of lack of experimental verification, subcellular localization prediction tools such as PSORTb (http://www.psort.org/psortb) and CELLO (http://cello.life.nctu.edu.tw) can be utilized. In the present study, localities of the short-listed proteins were acknowledged using PSORTb 3.0.2 and CELLO (Yu, Lin et al. 2004; Gardy, Laird et al. 2005).

Endorsement of prioritized proteins mining parameters

Scrutinizing all the five parameters of prioritized protein mining including host non-homology analysis, essentiality analysis, virulence factor analysis, sub-cellular localization and trans-membrane helices analysis through the above mentioned databases and predefined parameters will filter out proteins were extracted out as prioritized vaccine candidates from the core genome of the analyzed uropathogens. The filtered proteins will be analyzed for their quantitative and functional analysis, Figure 16.

3.3 Phase III: Quantitative and functional annotation of novel prioritized proteins

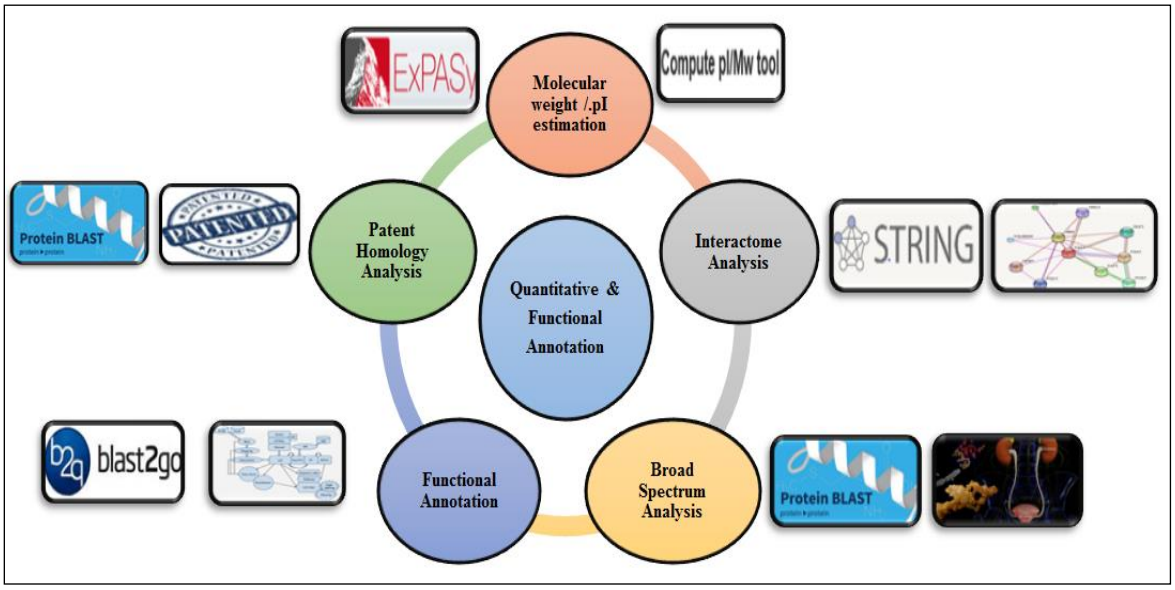


Figure 15: Schematic illustration of Phase: III. The figure gives the visual representation of the analysis carried out for quantitative and functional annotation of novel prioritized proteins.

A. Molecular weight estimation

It is essential for vaccine candidates to be of small molecular size because large peptides cannot pass through glomerulus filtrate and are excreted. This would prevent the circulation of the prioritized proteins in the blood circulation for generating an effective immune response. Small sized proteins less than 110kD are required to limit the amount of acrylamide a carcinogenic neurotoxin required during purification process.

Molecular weight estimation was carried out through Compute pI/MW a theoretical pI and Mw computation tool PI/M(RRID:SCR_012880), which permits the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for user entered sequences (Gasteiger, Hoogland et al. 2005).

B. Interactome analysis

Protein interactome analysis was performed to analyze the functional importance of short listed vaccine targets in the core-genome subsets. Protein-protein interaction network was assembled for each of the short-listed protein using Search Tool for the Retrieval of Interacting Genes (STRING, RRID:SCR_005223), accessible at (http://string-db.org) (Szklarczyk, Franceschini et al. 2011) which includes more than 1100 completely sequenced organisms.

Protein connotation information of STRING is resulting from experimental data and computational techniques (co-expression, gene fusion, co-occurrence, and neighborhood), numerous protein interactions, and curated pathway databases (Jensen, Kuhn et al. 2009). High confidence interactors (not more than 50) with score more than or equal to 0.700 alone were included in the protein network (Raman, Yeturu et al. 2008). To evade false positives and false negatives, all interactors with low as well as medium confidence score were excluded from the network.

C. Broad spectrum activity analysis

Sequence alignment comparison of these vaccine candidates with additional medically significant pathogens expedites the evolution of the proposed targets as ideal broad spectrum targets. Prioritized proteins were analyzed using BLASTp search (Altschul, Gish et al. 1990) against a wide-range of pathogenic bacteria with an expected threshold e-value of < 1.0 e-5 and > 100 Bite score for the identification of broad spectrum targets

. A list of urogenital isolates reported in the human micro biome project was considered in this analysis. The pathogenic character of the urogenital isolates was established from the literature. A total of 101 disease-causing bacteria from different genus were used in the broad

spectrum analysis. From the homology analysis against each of the pathogen it is theorized that close homologs present in more number of pathogens are more likely to be a 'promising broad spectrum target' (Raman, Yeturu et al. 2008).

D. Functionality analysis

Proteomic analyses often generate a vast amount of data. Simple identification and quantification of proteins from a cell proteome or subproteome is not adequate for the full understanding of intricate mechanisms taking place in the living systems. Therefore, the functional annotation of protein datasets via bioinformatics tools is essential for interpreting the results of high-throughput proteomics (Carnielli, Winck et al. 2015). The function of the selected prioritized proteins was annotated (Vetrivel, Subramanian et al. 2011) using INTERPROSCAN (Mulder and Apweiler 2007), a tool that assimilates numerous protein signature recognition approaches and databases.

E. Patent protein homology analysis

The NR (non-redundant) patent (pat) sequence databases (http://www.ebi.ac.uk/patentdata/nr/) are an important resource for patent-related searches, especially for determining potential commercial use of biological sequences (Seeber 2007). The NR pat sequence databases are publicly available collection of NR patent sequences, at mutually the sequence and patent-family stages, carrying >30 million nucleotide and protein sequences dig out from patent documents in the public domain (Li, Kondratowicz et al. 2013). The BLASTp analysis (Altschul, Gish et al. 1990) at predetermined cut-off e-value < 1.0 e-5, with bit Score > 100 and > 35% was carried out for determining potential of the prioritized proteins for commercial use of biological sequences and their patentability.

3.4 Phase IV: Epitope analysis and homology modeling of the vaccine targets

A. Epitope mapping of the prioritized targeted proteins

A successful peptide based vaccine must also include immune dominant T- and B-cell epitopes that is, epitopes that are readily accessed and recognized by the immune system and that have influence on the specificity of the induced antibody. The prioritized potential vaccine candidates were subjected to epitope mapping to reveal that only a small number of regions in a protein are immunogenic and capable of provoking humoral and cellular immune responses (Figure 16).

ABC-pred server (www.imtech.res.in/raghava/abcpred/) was utilized with the aim to predict B cell epitopes in an antigen sequence via artificial neural network (Saha and Raghava 2006). This server is developed on the basis of recurrent neural network (machine based technique) using fixed length patterns. The cutoff value for B-cell epitope prediction was set at > 0.8. The selected B-cell epitopes were further analyzed for T-cell epitopes for binding. Propred is a graphical web tool for predicting MHC class II binding regions in antigenic protein sequences.

Propred server (http://www.imtech.res.in/raghava/propred/) is a useful tool in locating the promiscuous binding regions that can bind to several HLA-DR alleles (Hassan, Naz et al. 2016). ProPred1 (http://www.imtech.res.in/raghava/propred1) is an on-line web tool for the prediction of peptide binding to MHC class-I alleles, a matrix-based method that allows the prediction of MHC binding sites in an antigenic sequence for 47 MHC class-I alleles, available at. T-cell $_{epitope}$ surface exposure and surface accessibility was estimated by using NetSurfP (http://www.cbs.dtu.dk/services/NetSurfP).

VaxiJen (http://www.jenner.ac.uk/VaxiJen) is the first ever server for alignment-independent prediction of protective antigens and classify antigens solely based on the physicochemical properties of proteins without recourse to sequence alignment (Doytchinova and Flower 2007). The antigenicity of the prioritized epitopes was determined through VaxiJen v2.0 at a cut off value of 0.4. The half maximal inhibitory concentration (IC50) is a measure of the effectiveness of a substance in inhibiting a specific biological or biochemical function.

According to the FDA, IC50 represents the molar concentration of a drug that is required for 50% inhibition in vitro. MHC pred (http://www.ddg-pharmfac.net/mhcpred/MHCPred/) was used to calculate the IC50 score for DRB1*0101 allele (Guan, Doytchinova et al. 2003). DRB1*0101was selected as it was worldwide the most commonly found allele and therefore give a strong immune reaction after robust antigen recognition (Hassan, Naz et al. 2016). Virulence of potential epitopes was determined by using VirulentPred (http://bioinfo.icgeb.res.in/virulent/), a SVM based method to predict bacterial virulent proteins sequences, which can be used to screen virulent proteins in proteomes (Garg and Gupta 2008).

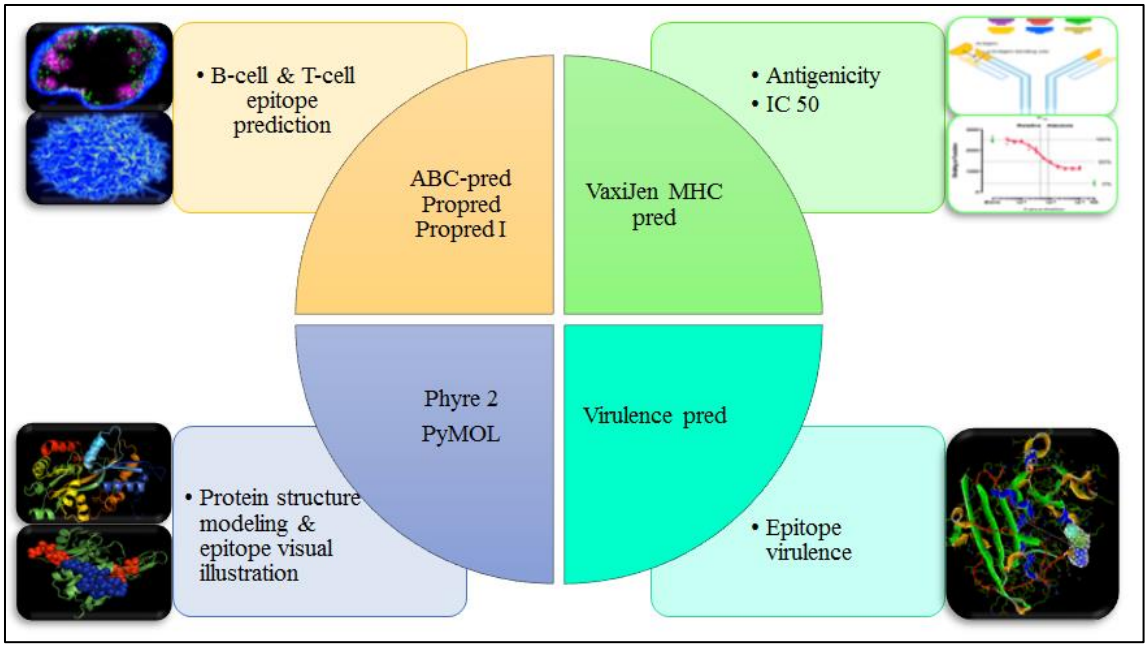


Figure 16: Schematic illustration of Phase: IV. The figure provides the visual representation of epitope analysis and protein structure modeling.

B. Protein structure modeling

The 3D structure of the protein was obtained by utilizing Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) to determine the pattern of the epitope sequence and its pharmacologically active sites of prioritized epitope with other pharmacologically active targets. Phyre2 customs advanced remote homology discovery approaches to build 3D models, ligand binding sites and analyze the effect of amino acid variants for a user's protein sequence. As these prioritized proteins are conserved in all the genome strains of each targeted uropathogen it was preferred to obtain their structure via homology modeling comparative to accessing a 3D structure from PDB usually belonging to a single strain. The B-cell and T-cell epitopes were also depicted on the 3D structure via PyMOL molecular visualization system rendering and animating 3D molecular structures (DeLano 2002).

CHAPTER 4

RESULTS

s

CHAPTER 4 RESULTS

The current study embodies an innovative hierarchal insilco maneuver to discover and categorize potential vaccine candidates in most common UTI causing pathogens (mentioned in Table 1). These uropathogens are also most imperative multi-drug resistant, community and hospital acquired pathogens. Considering the core genome of epidemiologically important UTI causing bacterial pathogens, as input datasets of hierarchy of study ensued the identification of a list of efficient vaccine targets. The present finding and categorization strategy encompasses four phases of analysis.

4.1. Phase: I Pan-core genome and resistance gene analysis

A. Excavation of core proteome dataset for each bacterial pathogen (whole genome)

Pan genome analysis signified the full complement of genes in the global genome repository of each analyzed pathogen, illustrated in Table 1, Chapter 3. Pan, core, accessory and unique genes were steadfast in each bacterial genome. Pan-core genome plot created by analyzing the total set of all genomes (410) revealed an increasing pan- genome with no conserved genome; this is due to the huge diversity of the analyzed strains and apparent from the pan-genome tree as illustrated in Figure 18.

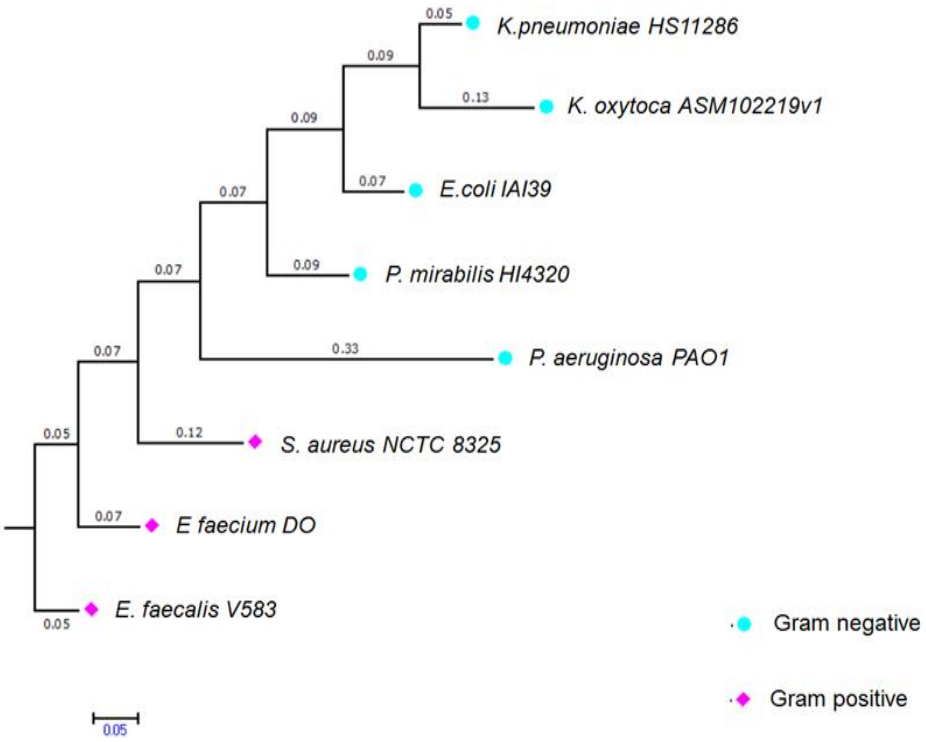


Figure 17: Phylogenetic variation of the analyzed genomes. Phylogenetic tree of reference strains pan-genome of the analyzed organisms illustrating huge genomic diversity.

As expected, in case of representative (reference strains) for all the above species explore for pan-core genome estimation similar behavior of increasing the pan-genome was observed and no core genes (zero) were obtained as depicted in the Figure 19. Low or no conserved genome was evident as the analyzed species belonged to different taxonomic genera. In case of discrete pan-genome analysis on species level a persistent increase in the pan-genome repository and an equally persistent decrease in the core genome are observed with the addition of a new genome in each individual pan-core genome analysis. Unlike to the whole set (410 inter-species pan-genome analyses) the significant core genes (core genomes) were observed for, each analyzed specie.

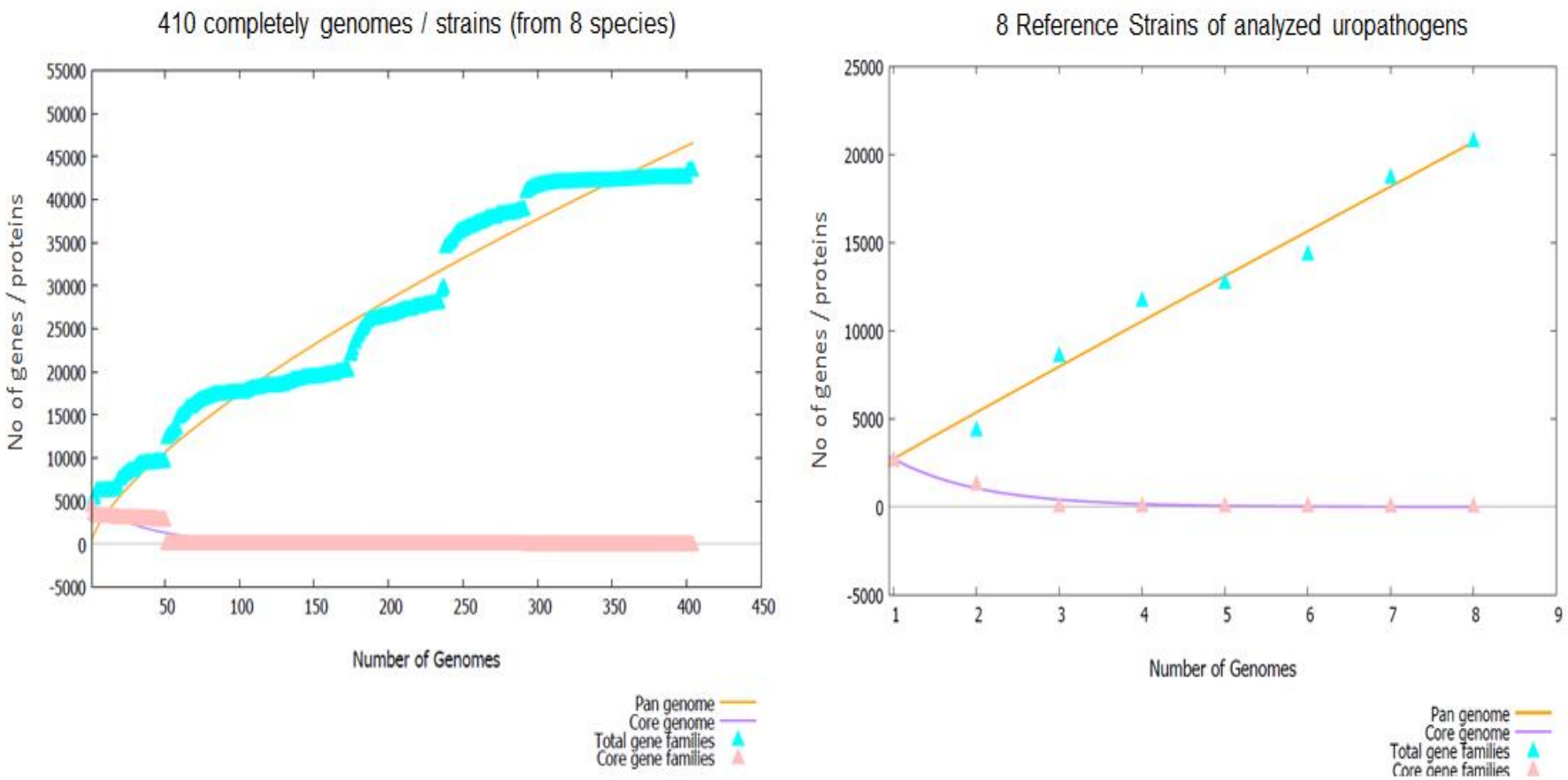


Figure 18: Inter-species pan-core genome plots. The pan-genome plot of 410 complete genomes of the analyzed organisms and pan-genome plot of reference strains of analyzed organisms. Both plots illustrate an increasing pan-genome lacking a persistent conserved genome.

a. Pan-genome estimation of *E .coli* genomes

The complete pan-genome analysis gives the idea about the essential and non-essential genomic characters of the bacterial specie. The information can be utilized to apprehend pathogenic variations. In the current study pan-genome analysis was applied on 6,25,387 coding sequences. Pan-genome is the consortium of all the genes present in the genome families. 13,556 genes were obtained which represented the pan-genome repertoire, shown in blue bars in the following Figure 20.

Core-genome is consortium of genes which are common among all the strains. Identification of core-genes is integral for the determination of vaccine candidates. It would ensure the presence of the selected vaccine target in all the completely sequenced strains of *Escherichia coli.* 2,595 genes were obtained which represented the core genome collection, shown in red bars in Figure 20. The core-genome was 54 % of the average genome.

Accessory gene group comprise the genes other than the core genes and unique genes. The total number of all the accessory genes comprising the accessory gene collection was calculated to be 2,36,558, among all 159 genomes of *E. coli* shown in green bars in Figure 20. Total 12,929 unique genes were also identified which were only present in specific strains, show in purple bars in Figure 20. The Pan-genome was further scrutinized through COG and KEGG analysis in order to access the authentication of the analysis.

In order to acquire adaptation to the new or evolving external environment, bacteria have to show adaptation for survival. Adaptation is attained either through mutation or more commonly through acquiring new genes through horizontal gene transfer. *E.coli* show high

adaptability in variable in vivo and invitro environments. Inclusion of large number of foreign genetic elements from other bacteria through horizontal gene transfer is highly expected for *E. coli*.

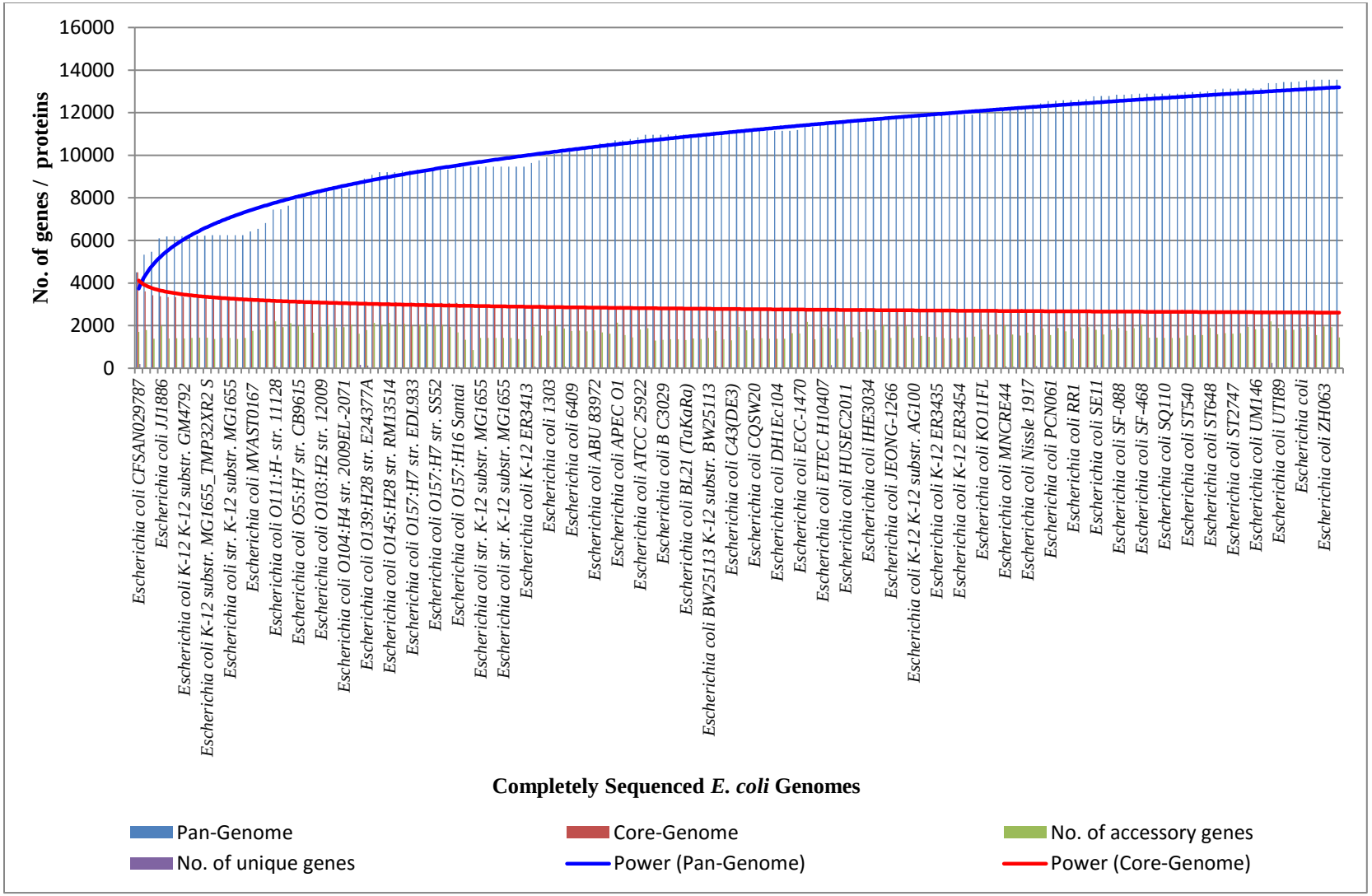


Figure 19: Pan-genome Plot of *E.coli* completely sequenced genomes. The X-axis in the graphs represents *E. coli* completely sequenced genomes and Y-axis represents the number of genes articulated in each genome. The blue power line exemplifies the pan genome persistant increase and stability indicating an open pan-genome. The red power line exemplifies the core genome conservation and stability indicative of possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the numer of genes in pan-genome, core-genome, accessory-genome, unique-genome and exclusively abscent genes for eacf strain.

b. Pan-genome of *K. pneumoniae* genomes

Pan-core genome analysis over whole *K. pneumoniae* genome was carried out, to get accurate intuition about gene diffusion in *K. pneumoniae*. This would benefit us by identifying the essential and non-essential elements in *K. pneumoniae* genome. In the current study the pan-core genome analysis was carried out on 52 completely sequenced *K. pneumoniae* genomes available at the time of analysis. The analysis was down streamed by using BPGA (Bacterial Pan Genome Analysis Tool) for efficient multi genome study. The results of the pan-core genome analysis are illustrated in a bar graph plot in the Figure 21.

BPGA carried the analysis based on famous 50/50 rule of the bio-informatics world. Local similarity between the sequences was determined on the principal of 50% similarity over 50% length of the sequence.

Pan-genome is the gene set carrying all protein coding genes of the bacteria (Sharma, Narwani et al. 2016). The *K. pneumoniae* pan-genome was found to comprise 10,713 protein coding genes. *K. pneumoniae* pan-genome is represented in blue bars in the figure 21. The blue power line is representing the increase in the pan-genome as well as highlighting the stability of the pan-genome and accuracy of the analysis as well as persistent capacity of *K. pneumoniae* to evolve.

Core genome present the group of gene conserved in the whole subjugated genome (Sharma, Narwani et al. 2016). The core genome was found to carry 3534 conserved genes and is represented by red bars in the figure. The red power line is representing the decrease in the core-genome. The highlighted increase in the pan genome and the decrease in the core genome is the authentication of an obstinate open pan-genome for *K. pneumoniae*. Core genome was found to be 63.27% of the average pan-genome, indicating a vastly established

genome. Accessory genome or dispensable genome carries the genes which are not common among all the organism of the genome (Sharma, Narwani et al. 2016). *K. pneumoniae* accessory genes are represented as green bars in the figure. Highest no. of accessory genes, 1601genes was found in *Klebsiella pneumonia CAV1193*.

Unique genome carries the genes which are present only in a single organism of the genome also known as the singleton genes (Sharma, Narwani et al. 2016). The unique genes of *K. pneumoniae* genome are represented in purple bars in Figure 21. The highest no. of unique genes, 268 genes was identified in *Klebsiella pneumoniae 342*. Exclusively absent genes are the genes which are only present in a single organism of the genome but absent in all the genomes. Exclusively absent genes of *K. pneumoniae* are represented with light blue bars in figure. The highest no of exclusively absent genes, 69 genes was found in *Klebsiella pneumoniae 30660*.

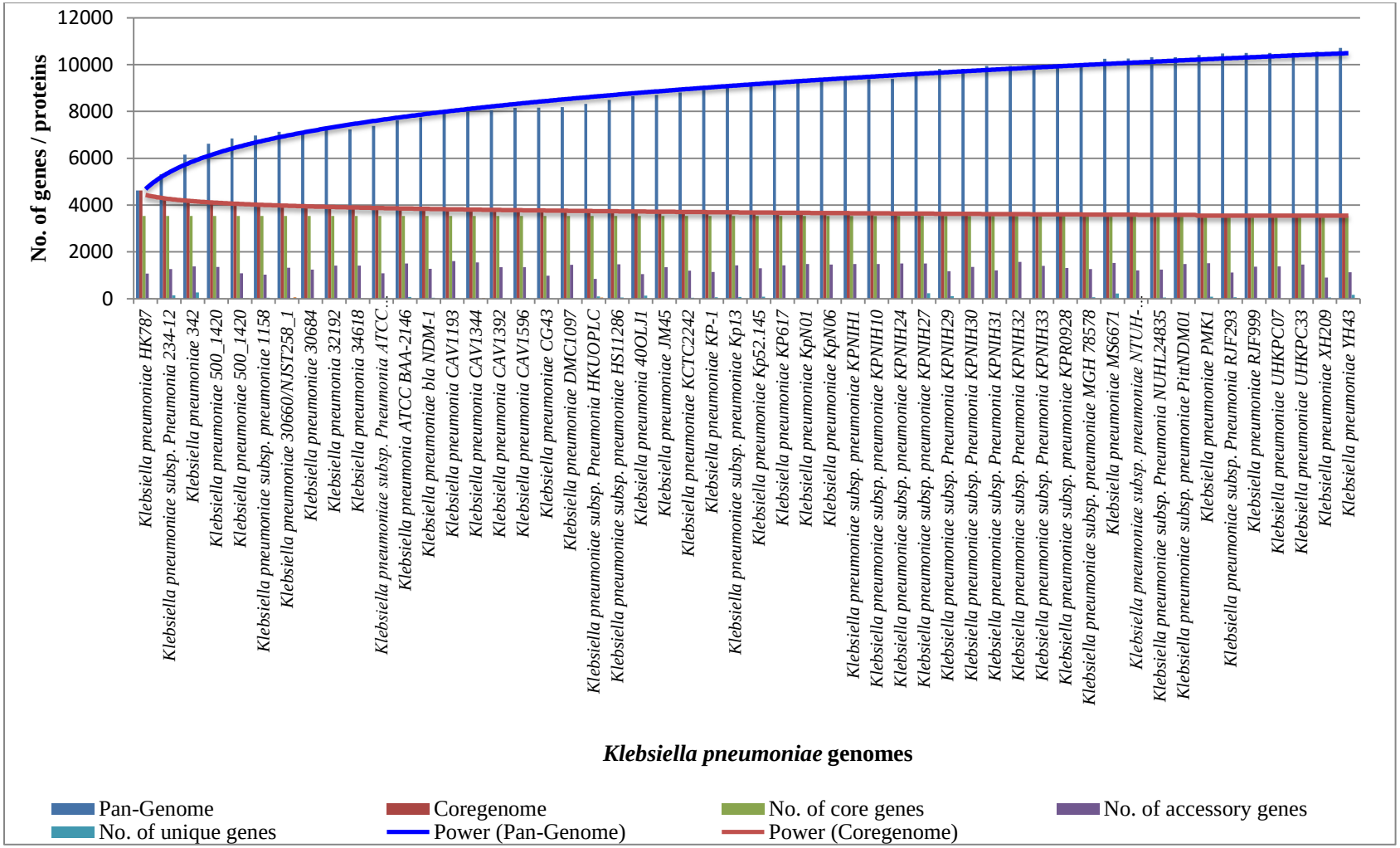


Figure 20: Pan-genome Plot of *K.pneumoniae* completely sequenced genomes. The X-axis in the graphs represents *K. pneumoniae* completely sequenced genomes and Y-axsia represents the number of genes expressed in eacg genome. The blue power line represents the persistant pan genome persistant increase and stability indicating an open pan-genome. The red power line represents the core genome conservation and stability sugesting possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the numer of genes in pan-genome, core-genome, accessory-genome, unique-genome and exclusively abscent genes for eacf strain.

c. Pan-genome estimstion of *K.oxytoca* genomes.

Interspecies pan-genome analysis was pragmatic on all the completely sequenced *Klebsiella oxytoca* genomes to govern the essential and non-essential genomic characteristics of the pathogen to capture pathogenic variability. Pan-genome exploration is a crucial step in computational based genomic approach towards vaccine development or drug target identification as it provides the understanding of functional adaptation of bacterial species (Muzzi, Masignani et al. 2007). Pan-genome analysis was completed using BPGA http://www.iicb.res.in/bpga/index.htm. BPGA categorized the pan-genome on the basis of BLAST similarities among the genome following 50/50 rule.

It's conjoint for bacterial species to inflate their genomes under selective environmental and antibiotic pressure through horizontal gene transfer in order to endure the new hostile conditions (Martínez, Baquero et al. 2007). Total 9,265 conserved genes were unraveled. Pan-genome for *K. oxytoca* comprising 66,717 genes was attained represented in blue bar in Figure 22. The *K. oxytoca* core genome consortium was analyzed to cart 39,607 genes represented in red bar in Figure 22. The totaled core genome is 72% of the average pan genome, which is enlightening the presence of remarkably conserved genomic features. As speculated earlier all the genomes share huge chunk of their genomic content.

A constant upsurge in the pan-genome content and constant diminution for the core-genome content was observed with the tallying of new genome as represented by stable power lines in the pan-core genome plot as shown in the Figure 22. This constant upsurge of the genes on the addition of genomes confirms a rich and highly diverse open genome for *Klebsiella oxytoca*. The accessory genes for *K. oxytoca* were calculated to be 10,941 genes

and the highest number of accessory genes, 1673 genes was found in *Klebsiella oxytoca CAV1335* as represented in green bar in Figure 22. The unique gene repository for *K. oxytoca* was estimated to carry 2187 genes with highest number of unique genes 615 genes were observed in *Klebsiella oxytoca CAV1374* as represented in purple bar in Figure 22.

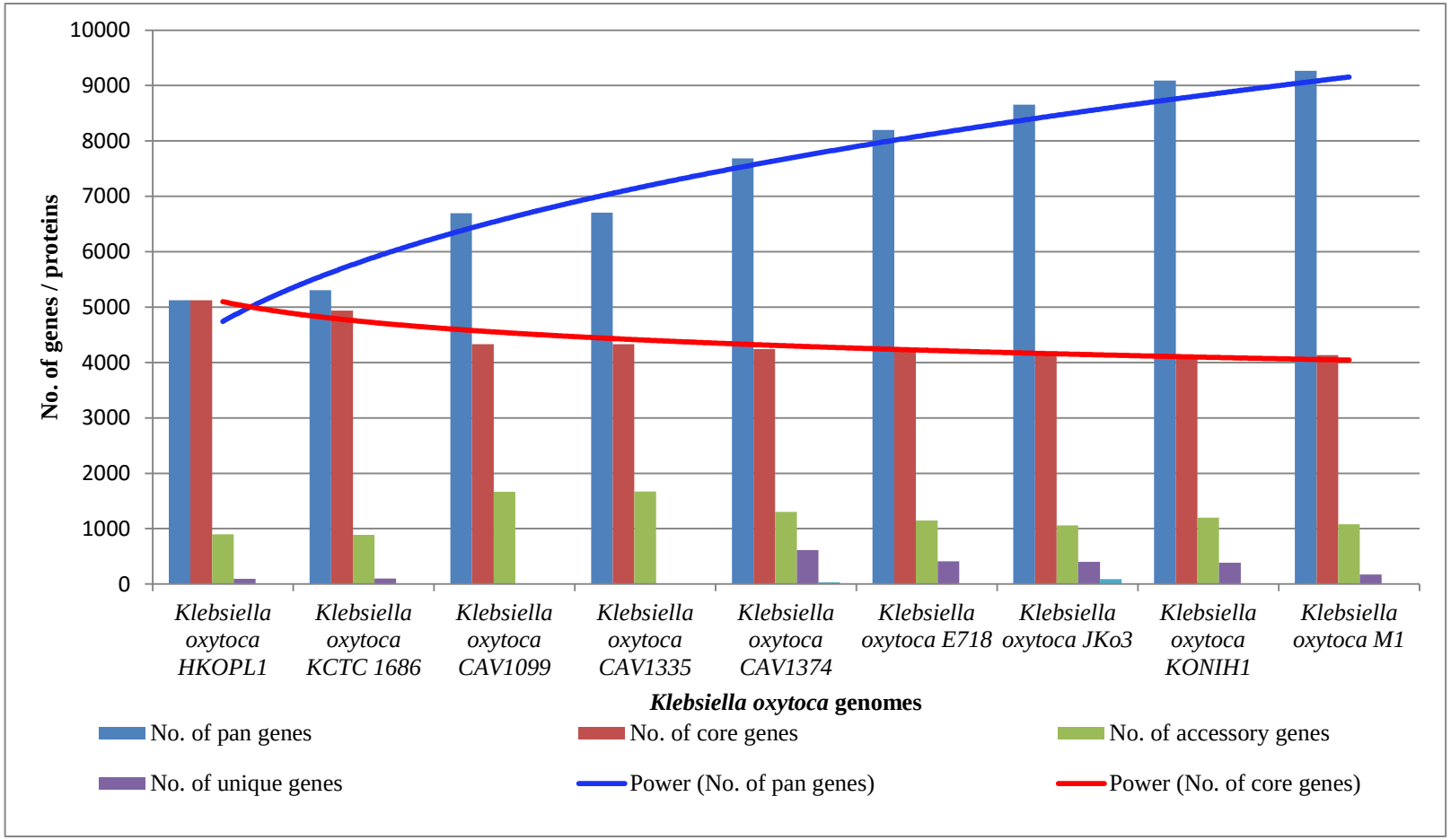


Figure 21: **Pan-genome Plot of *K. oxytocca* completely sequenced genomes.** The X-axis in the graphs represents *K. oxytocca* completely sequenced genomes and Y-axsia represents the number of genes expressed in each genome. The blue power line represents the persistant pan genome persistant increase and stability indicating an open pan-genome. The red power line represents the core genome conservation and stability sugesting possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the numer of genes in pan-genome, core-genome, accessory-genome, unique-genome and exclusively abscent genes for each strain.

d. Pan-genome estimation of *P. aroginosa* genomes

The interspecies pan-genome analysis was carried out on 59 complete drafts available at the time of analysis to comprehend *P.aroginosa* genetic land scape. Pan-genome analysis accesses us in portioning the conserved and dispensable genomes. The variation in pan-genomic groups corroborates the notion of frequent lateral gene transfer in the bacterium. In case of *P. aroginosa* total 3,60,559 genes were analyzed. The highest numbers of genes 7,071 were found in *P. aroginosa Carb01 63* and the lowest number of genes 5,629 was found in *P.aroginosa Cu1510.* The average genome is estimated to comprise 6,111 genes for a single strain.

The pan-genome analysis was performed utilizing the previously mentioned strategy.The pan-genes are estimated to be comprising a huge consortium of 2,91,789, illustrated in blue bars in the Figure 23, indicating *P.aroginosa* to be a highly complex organism. The pan-genome is found to increase persistently with the addition of each strain in the analysis due to the addition of dispensable genes brought by each strain. This persistent increase is highlighted in the Figure 23 with a blue power line indicative of the open pan-genome.

The conserved genome estimated to carry 4,151 genes, illustrated with red bars in the Figure 23. The conserved genomes show a persistently decreasing conserved genome, indicating the stability of the conserved genome, illustrated in red power line in the following figure. Therefore the numbers of conserved genes in a single strain are calculated to be 67%, further fortifying the stability of the genome. The dispensable genome analysis revealed a huge dispensable genome with 78,838 genes comprising accessory and 2,148 genes comprising unique genome. The highest numbers of accessory genes were found in case of *Pseudomonas aeruginosa Carb0163* comprising about 2,033 genes. Whereas the highest number of unique genes 277 genes were found in Pseudomonas aeruginosa PA7.

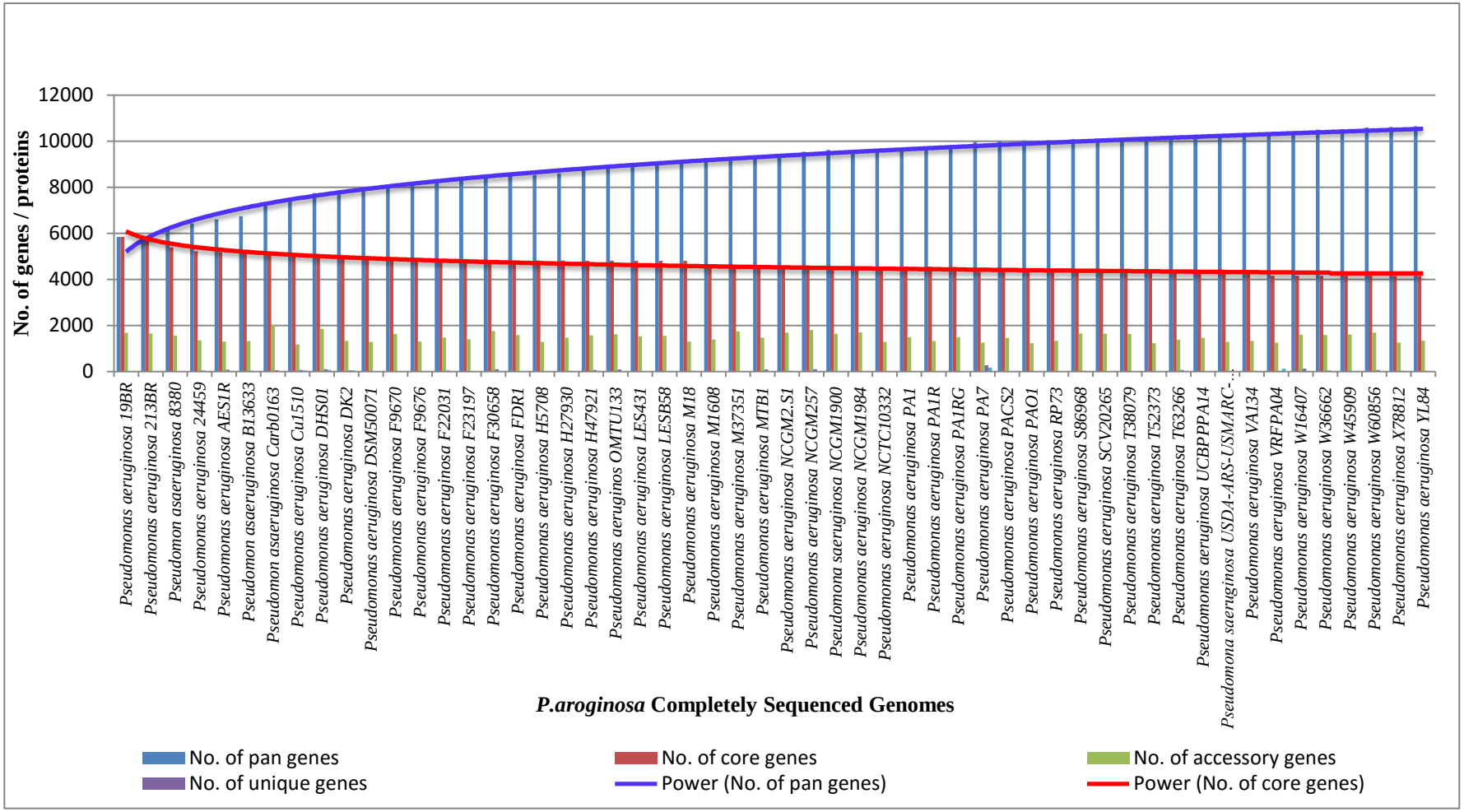


Figure 22: Pan-genome Plot of *P.aroginosa* completely sequenced genomes. The X-axis in the graphs represents *P.aroginosa* completely sequenced genomes and Y-axis represents the number of genes expressed in each genome. The blue power line represents the persistent pan genome persistent increase and stability indicating an open pan-genome. The red power line represents the core genome conservation and stability suggesting possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the number of genes in pan-genome, core-genome, accessory-genome, unique-genome and exclusively absent genes for each strain.

e. Pan-genome estimation of *E. faecalis* genomes

The pan-genome estimation of ***E. faecalis*** was applied on 7 complete drafts available at NCBI at the initiation of the analysis. Total about 21090 genes were analyzed from these genomes. The highest numbers of genes 3412 genes were estimated for *Enterococcus faecalis V583.* The lowest numbers of genes were estimated for *Enterococcus faecalis OG1RF with 2,636.* The average genome for a single genome or stain was calculated to carry 3012 genes.

The functional genome obtained from prodigal was accessed for pa-genome estimation by using BPGA as previously explained. The pan-genome comprising 4,031 genes was obtained, shown in blue bar in *E. faecalis* pan-genome plot in the figure 24. The pan-genome is found to persistently increase, this gradual increase with the addition of new strain in the analysis is indicative of an open pan-genome.

Later on, the availability of more draft genomes and the subsequent pan-genome analysis might give a larger pan-genome. The conserved genome comprised 2,169 genes with of huge genomic conservation at a strain level of 72% conserved average genome. The accessory genome indicative of genes present in only a few strains consisted of 2845 genes with the highest number of accessory genes, 549 genes present in *Enterococcus faecalis V583*.The unique genome indicative unique genes consisted of total 1001 genes. The highest numbers of genes 320 genes were also found in Enterococcus faecalis V583.

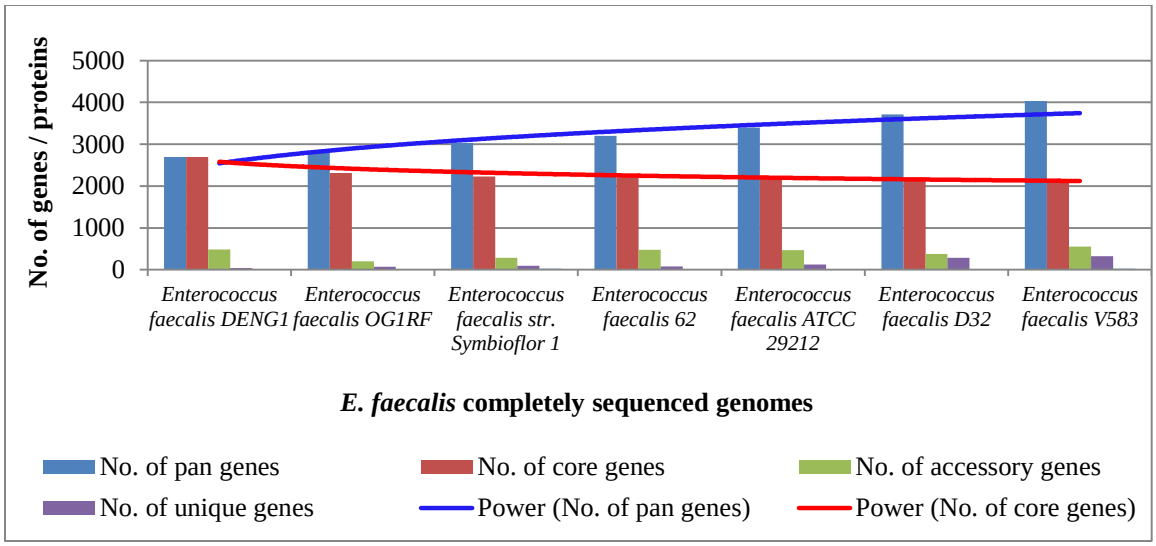


Figure 23 Pan-genome Plot of *E. faecalis* completely sequenced genomes. The X-axis in the graphs represents *E. faecalis* completely sequenced genomes and Y-axis represents the number of genes expressed in each genome. The blue power line represents the persistent pan genome persistent increase and stability indicating an open pan-genome. The red power line represents the core genome conservation and stability suggesting possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the number of genes in pan-genome, core-genome, accessory-genome, unique-genome and exclusively absent genes for each strain.

f. Pan-genome estimation of *P. mirabilis* genomes

The pan-genome estimation in case of *P. mirabilis*was applied on only three strains as at the start of the analysis only three complete drafts of *P. mirabilis* were available at NCBI. The pan-genome analysis was performed by applying the pre-established strategy. The pan-genome comprised 3,924 genes from the total 10,628 analyzed genes and an average genome of 3012 genes Figure 25. The conserved genome carried 2, 969 with average conserved genome at strain level estimated to be 50%, indicative of diversity even at the strain level. Total 342 accessory and 784 unique genes were found.

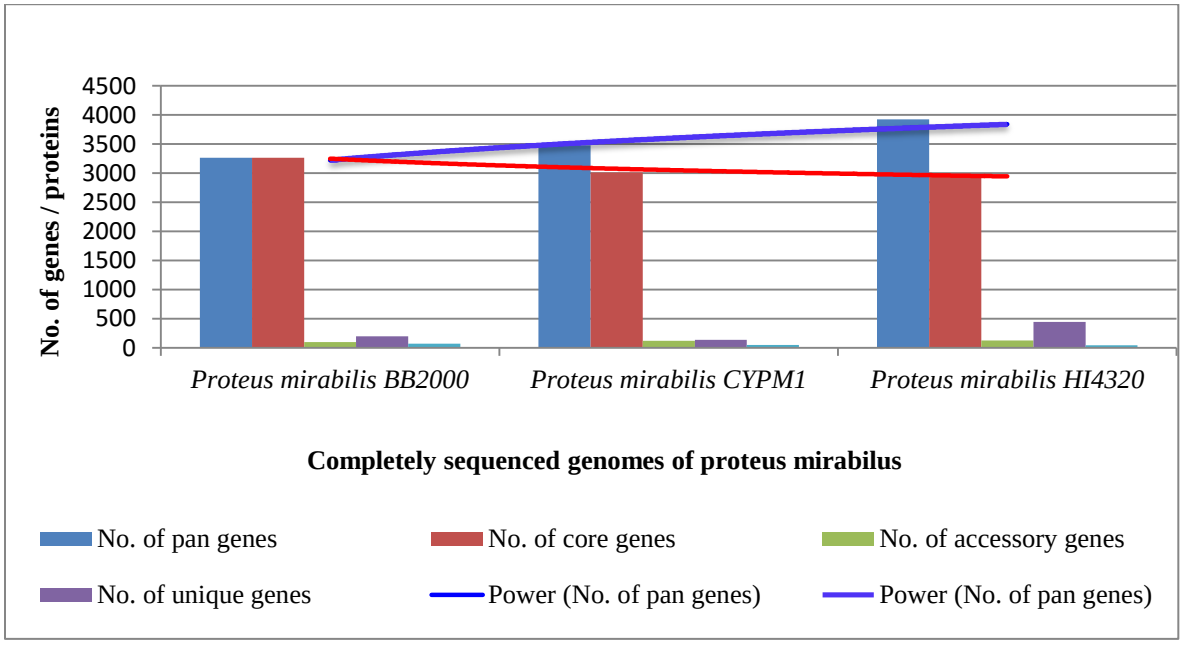


Figure 24: Pan-genome Plot of *Proteus mirabilis.* completely sequenced genomes. The X-axis in the graphs represents *P. mirabilus* completely sequenced genomes and Y-axsia represents the number of genes expressed in each genome. The blue power line represents the persistant pan genome persistant increase and stability indicating an open pan-genome. The red power line represents the core genome conservation and stability sugesting possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the numer of genes in pan-genome, core-genome, accessory-genome, unique-genome and exclusively abscent genes for each strain.

g. Pan-genome estimation of *E.faceium* genomes

The pan-core genome analysis of *E.faceium* with 12 completely sequenced genomes available at the time of analysis revealed an open pan-genome. The total number of pan-genes was observed to 4,031 for *E.faecalis*, Figure 26. The numbers of core genes were estimated to be 1,909. The pan-genome is illustrated in blue bars and the core genome is illustrated in red bars. The core genome to average genome ratio is estimated to be 66%. The accessory genome comprised 6,974 genes whereas the core-genome comprised 980 genes.

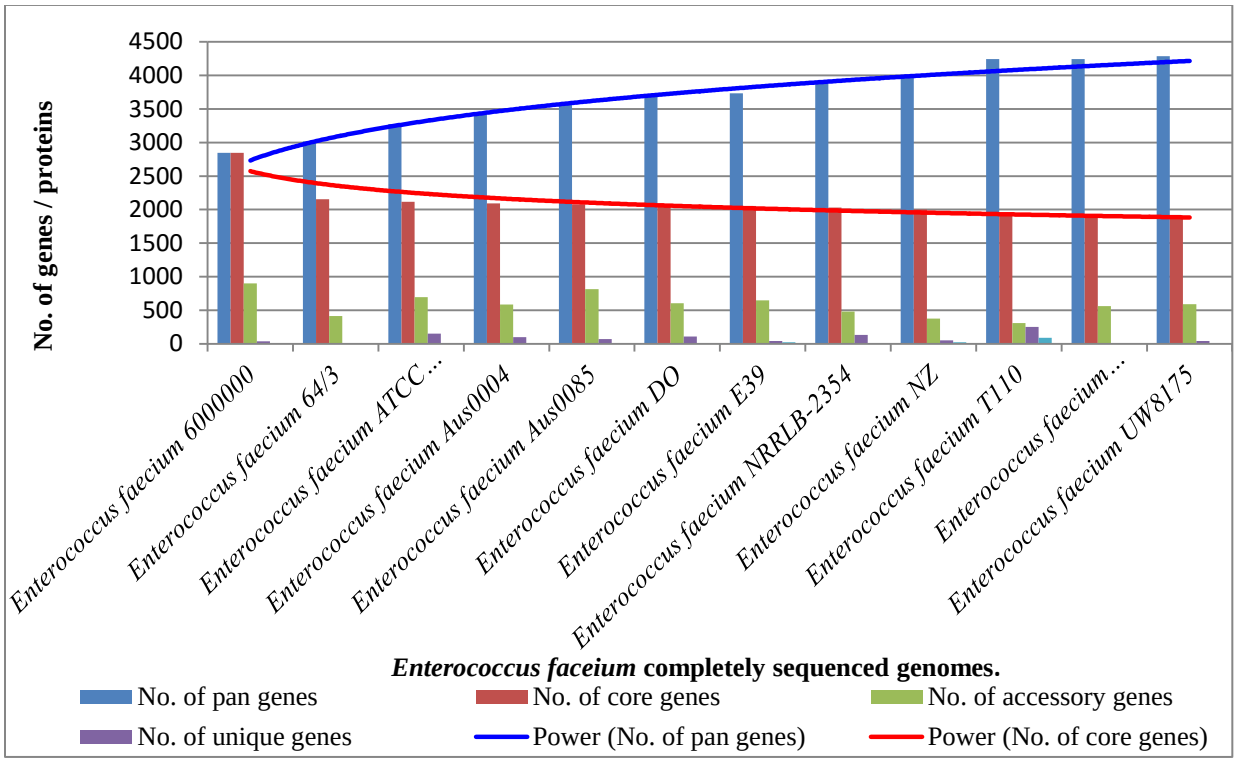


Figure 25: Pan-genome Plot of *E.faceium* completely sequenced genomes. The X-axis in the graphs represents *E.faceium* completely sequenced genomes and Y-axis represents the number of genes expressed in each genome. The blue power line represents the persistent pan genome persistent increase and stability indicating an open pan-genome. The red power line represents the core genome conservation and stability suggesting possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the number of genes in pan-genome, core-genome, accessory-genome and unique genome.

h. Pan-genome Plot of *S.aureus genomes*

The pan-genome analysis of *S.aureus* involved 109 complete draft genomes based on the previously determined strategy. The pan-genome which comprised all the genes at the specie level for *S.aureus* carries total 4,195 genes. The highest number of genes 3,170 were found in Staphylococcus aureus subsp. aureus Gv69 with a proteome estimated to carry 2,980 proteins. But the highest numbers of proteins 3,041 were observed for Staphylococcus aureus subsp. aureus TW20 proteome. The lowest number of genes and proteins were observed for Staphylococcus aureus RKI4, 2,551 and 2,414 respectively. There for the average number og genes for a single genome calculated to be 2,854.

The conserved genes were estimated to comprise 1,902 genes as shown in the Figure 27. The core-genome to average genome was estimated to be 66%. The accessory genome comprised 52,903 genes, whereas the unique genome comprised 438 genes. The pan-genome plot for the analysis is illustrated in the following, Figure 27. The plot indicates the *S. aureus* to comprise an open pan-genome with an increasing pan-genome indicated with a blue power line in Figure 27. The core-genome was found to decrease gradually indicating the stability of the genome, as illustrated with a red power line in the Figure 27.

The total accessory genes estimated to be 5,290 with the highest number of accessory genes 811 found in Staphylococcus aureus subsp. aureus TW20, which also estimated to constitute the largest proteome among *S.aureus* species. The total number of unique genes appraised to be 438 for *Staphylococcus aureus* pan-genome with the highest number of unique genes 44 genes, calculated for St*aphylococcus aureus XQ* in the current analysis.

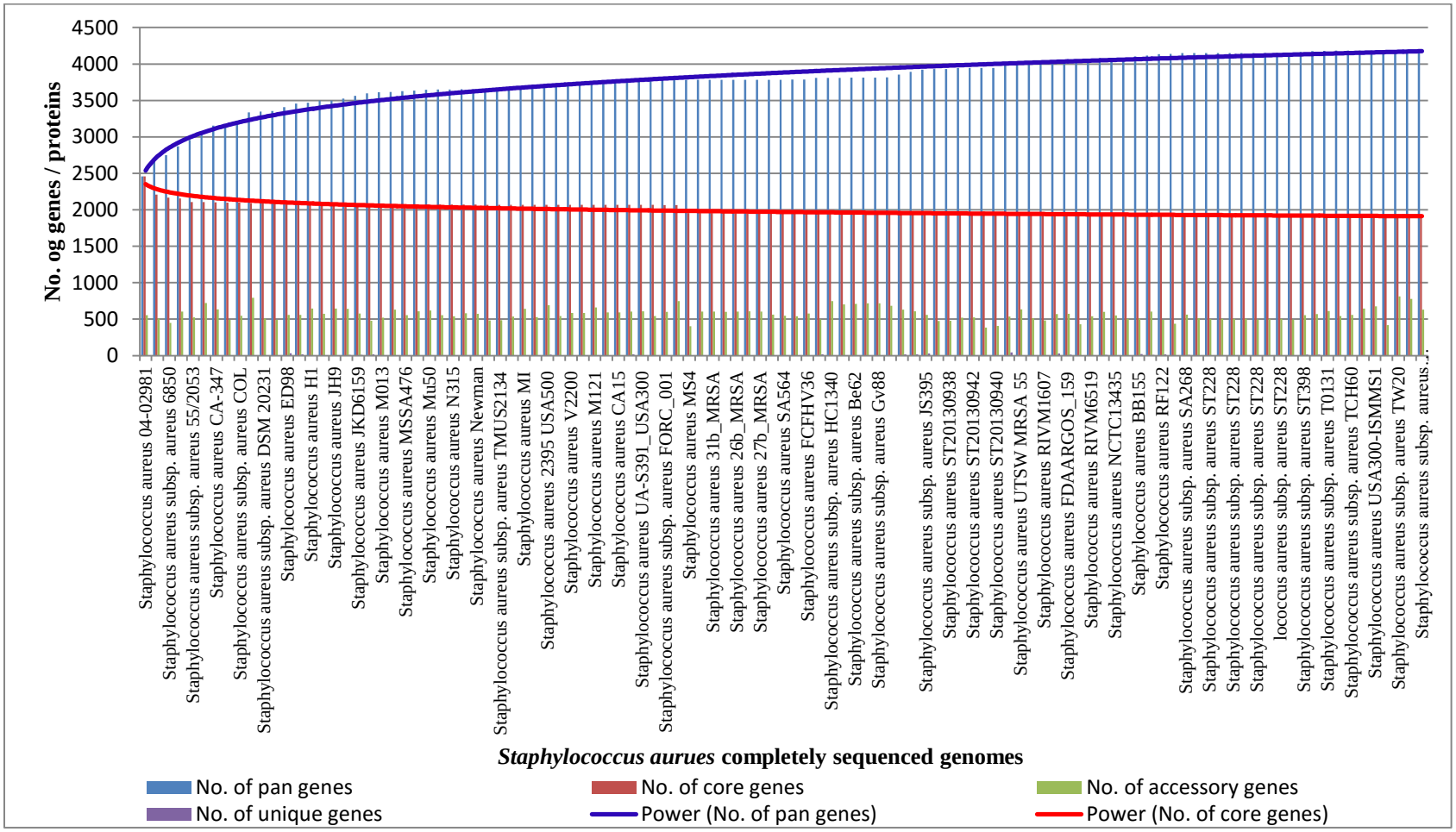


Figure 26: Pan-genome Plot of *S.aureus* completely sequenced genomes. The X-axis in the graphs represents *S.aureus* completely sequenced genomes and Y-axis represents the number of genes expressed in each genome. The blue power line represents the persistent pan genome persistent increase and stability indicating an open pan-genome. The red power line represents the core genome conservation and stability suggesting possibility of its utilization in vaccine candidate assortment. The blue, red, green, purple and light blue bars represents the number of genes in pan-genome, core-genome, accessory-genome and unique genome.

Table 2 Pan-core genome analysis of eight common UTI causing pathogens The table illustrates the number of genome species, universal gene suppositories, average genome and pan, core, accessory and unique genes of *E. coli, K. pneumoniae, K. oxytoca, P. aroginosa, E. faecalis, E. faecium, S.aureus* and *P. mirabilis.*

Organism	Genomes for species	Universal gene repository	Average genome	Pan-gene Repository	Core-gene repository	% core/av. Genome	Accessory genes for in all genomes	Unique genes in all genomes
Escherichia coli	159	655030	5000	13556	2595	52%	2096924	2627
Klebsiella pneumoniae	52	54988	6110	9266	4132	67%	104914	2187
Klebsiella oxytoca	9	284882	5585	10713	3534	63%	68549	2413
Pseudomonas aeruginosa	59	360559	6111	291789	4151	67%	80521	2148
Enterococcus faecalis	7	21090	3012	4031	2169	72%	2845	1001
Enterococcus faecium	12	34905	2909	4285	1909	66%	6974	980
Staphylococcus aureus	109	308199	2854	4195	1902	66%	52903	438
Proteus mirabilis	3	10628	5314	3924	2969	50%	342	784
Total	410	1730281	36895	341759	23361	-	2413972	12578

i. Comparative pan-genomic feature analysis

Average genome was calculated for each genome and core to average genome ratio was determined. Highest numbers of pan-genes, 2,01,789 genes were identified for *P. aroginosa* and lowest 3,924 genes for *P. mirabilis*. Others include *E.coli*, *K. pneumoniae, K. oxytoca, S. aureus*, *E. faecalis* with 13556, 10173, 4195 and 4031 genes respectively. Whereas, highest numbers of core-genes were observed in *P. aroginosa*, 4151 genes (67% of average genome) followed by *K. oxytoca*, *K. pneumoniae, E. faecalis, P. mirabilis, E. coli*, *E. faecium* and *S. aureu*s with 4132 (67% of average genome), 3534 (63% of average genome), 2969 (50% of average genome), 2595 (52% of average genome), 2169 (72% of average genome) and 1902 (66% of average genome) genes respectively. Total 410 genome sequences from 8 whole genomes of UTI causing pathogens were analyzed out of which 251759 functional or pan-genes were identified. Whereas 23362 genes, were conserved genes distributed among the analyzed uropathogens. A sum of 4, 29,882 accessory and 17,872 unique gene were also sorted out (Table 2).

Analyzing pan-core genome not only signifies the genomic distribution of the functional genes with in sub-sets but also filters the conserved genome, in each case. The core-genome sub-set is further analyzed and rests of the subsets are excluded from the analysis to ensure that the protein priotoization (Phase II) involves only the conserved gene products of the respective genomes. This would guarantee the prioritized proteins to be broad-spectrum, thus can be able to provide immunogenic coverage against all the strains of the pathogen. Only those core genes are filtered which are present in the database of Clusters of Orthologous Groups of proteins (COGs) and Kyoto Encyclopedia of Genes and Genomes (KEGG).

B. Identification of resistome and its pan-genome dissemination

The Comprehensive Antibiotic Resistance Database analyzed the pan-genome subsets of 8 distinct bacterial pathogens. The data generated the results on the basis of homology analysis of FASTA files of each core, accessory and unique subsets with 1600 distinct resistance causing proteins in CARD under Perfect and Strict analysis categories of resistome analysis (McArthur, Waglechner et al. 2013). Presence of total 564 ABR genes was identified in eight bacterial pan-genomes 213 genes were found in core genome sub-sets, 208 in accessory genome sub-sets and 143 were located in unique genomes of eight analyzed uropathogens. Among these 546 genes 54 were found to be result of single-nucleotide polymorphism. Distribution of resistome identified in the individual pan-genome subsets of eight organisms is demonstrated in Table 3. This step signifies the presence of ABR genes at each sub-set level of bacterial genomes desperately requiring a vaccine based solution of ABR vulnerability.

Table 3 Distribution of resistance genes identified in pan-genome subsets of eight UTI causing bacterial organisms. Pan-genome subsets of common UTI causing organisms were individually analyzed for the identification of resistance genes through CARD screening..

Organism	ABR genes identified in core genome	SNPs among ABR core genes	ABR genes identified in accessory genome	SNPs among ABR accessory genes	ABR genes identified in unique genome	SNPs among ABR unique genes
Escherichia coli	28	3	18	1	8	0
Klebsiella pneumoniae	45	1	45	6	31	0
Klebsiella oxytoca	36	1	34	1	30	1
Pseudomonas aeruginosa	35	0	34	11	17	11
Enterococcus faecalis	17	3	5	0	5	0
Enterococcus faecium	12	2	9	0	11	0
Staphylococcus aureus	15	3	18	1	9	0
Proteus mirabilis	25	3	45	6	32	0
Total	213	16	208	26	143	12

4.2. Phase: II Priotoization of the core-proteome for the identification of suitable vaccine candidate.

A. Host non-homology exploration

A prospective vaccine target must not have homologs in the human proteome to prevent undesirable cross-reactivity of potential vaccine (Hassan, Naz et al. 2016). A host non-homology examination was executed to quarantine gene products that are not homologous to the human proteome. Out of total 23362 core proteins, 17393 (73%) did not show any similarity for the human proteome and were labeled as non-homologous i.e., exclusively residing in the examined organism and absent in the host.

P.aroginosa has highest number of non-host homologs with 3786 proteins, followed by *K.oxytocca* with 3776 non-host homolog proteins, *K. pneumoniae* with 3213 non-homologous proteins, *E.coli* with 2373 non-host homolog proteins, *E.faecalis* with 1968 non-homologous proteins, Enterococcus faecium with 1724 non-host homolog proteins and Staphylococcus aureus with 1714 non-host homolog proteins (Table 4).

A. Virulence factor (Vf) analysis

Virulent proteins are important source for the establishment and severity of infection if inhibited they can alter the pathogens virulence thus they can be good targets for vaccine candidates (Sannes, Kuskowski et al. 2004). Therefore virulent proteome was also prioritized among the core genome of each pathogen. Out of 23362 conserved proteins (extracted from 8 uropathogens) 3747 (16%) core proteins were identified to be virulent and optimum for a vaccine candidate. Highest proportion of virulence factors was found in *E. coli with 635 virulent factors* (24% of *E coli* core- genome), by means of 635 VFs (table3) and the lowest (0.2% of core genome) was estimated in *P.mirabilis.* Others include *K. oxytoca, K. pneumoniae, P. mirabilis, S. aureus, E. faecalis* and *E. faecium* with 841 (20%), 699(20%),

622 (20%), 324 (17%), 365 (16%) and 260 (16%) virulent genes in their individual core genome subset respectively.

Table 4 Phase: II protein mining for vaccine candidate assortment The table illustrates the number of proteins obtain in host non-homology analysis, virulence factor analysis, essential gene analysis, number of proteins having helices <2 and Exoproteomic proteins obtained in case of each priotoization step for analyzed uropathogens.

UTI causing organism	Host non-homolog analysis	Virulence factor analysis	Essential genes analysis	Protein helices < 2	Exoproteome
Escherichia coli	2373	635	1000	2066	189
Klebsiella pneumoniae	3213	699	1321	2722	93
Klebsiella oxytoca	3776	841	1422	3231	119
Pseudomonas aeruginosa	3786	10	10	10	8
Enterococcus faecalis	1968	356	648	1711	400
Enterococcus faecium	1724	260	495	1468	24
Staphylococcus aureus	1714	324	679	1495	431
Proteus mirabilis	2615	622	249	2330	224
Total	21169	3747	5824	15033	1488

B. Essential gene analysis

The inactivation of a targeted protein should be lethal for the pathogen there for a prospective vaccine target must be indispensible for the existence and growth of the pathogen (Jadhav, Shanmugham et al. 2014). The core-proteomes resulting from the pan-genome analysis were also conceded through homology search counter to the database of essential genes (DEG) (Zhang, Ou et al. 2004). Out of 23362 core proteins 5824 (24%) proteins distributed over eight core-genome sub-sets exhibited similarity to gene products in the DEG demonstrating that they are possible to be essential in the respective pathogen.

E. coli with 1000 essential genes (39% of its core genome) has comparatively the highest number of essential genes. Followed by *K. oxytoca* and *K. pneumoniae,* with 422 (34%) and 1,321 (34%) essential genes of each core-proteome essential genes respectively.

The lowest number of essential genes was identified in *P. aroginosa* carrying only 10 (0.24%) genes indispensible for core the genome. The number of essential genes in *S. aureus, E. faecalis, E. faecium and P. mirabilis* was 679 (37%), 648 (22%), 495 (26%) and 249 (8.3%) in the respective core-genome subset.

D. Protein helices analysis

A vaccine candidate protein should be able to easily colonize and express in the host cell so that it can play is pivotal role in host defense interactions and interplay. Therefore each core-genome sub-set was also screened through TMHMM data base (Fernando, Selvarani et al. 2004) and HMMTOP (Lu, Szafron et al. 2004). The group of proteins in each core-genome sub-set showing <2 helices were segregated to idyllic as vaccine candidate. Total out of 23,362 analyzed core proteins 15033 (64%) showed <2 helices in their structure, i.e., their ability to easily colonize and express in the host genome.

E. faecalis with its 90% of its core-genome with 1711 genes or gene products (table 3) having <2 helices in their structures. Other with exceptional ability to colonize and express include *K.oxytocca, K.pneumoniae , P.mirabilus, E.coli, S.aureus* and *E .faecium* with 2,331 (78% of *K.oxytocca* core-genome), 27,22 (77% of *K.pneumoniae* core-genome), 2,330 (79% of *P.mirabilus* of core-genome), 2,066 (77% of *E.coli* core-genome), 1,495 (79% of *S.aureus* core-genome) and 1,468 (68% of *E.faceium* core-genome) .

E. Sub-cellular localization analysis

In order to characterize the core proteins as vaccine targets, their propagation within the bacterial cell was determined using two prediction servers PSORTb and CELLO (Nancy, Wagner et al. 2010; Yu, Cheng et al. 2014). Target proteins localized in the extra cellular and outer membrane can be used as vaccine targets. Based on the localization score, PSORTb

categorization of each core-genome for each organism in cytoplasmic, cytoplasmic membrane, outer membrane, and periplasmic region, extra cellular and unknown categories is given in Table 4. Out of total 23,362 core proteins of eight core-genome subsets 10,057 proteins were cytoplasmic, 6,790 were localized on cytoplasmic membrane, 589 periplasmic, 456 outer-membranes, 261 extracellular and for 5,204 localization was unknown as presented in Table 5.

Table 5: Sub-cellular localization of analyzed UTI causing pathogens. The table illustrates subcellular location of core-genome of all analyzed uropathogens.

Organism	Cytoplasmic	Cytoplasmic membrane	Periplasmic	Outer-membrane	Extracellular	Unknown
Escherichia coli	1126	633	89	58	42	646
Klebsiella pneumoniae	1590	939	126	74	19	785
Klebsiella oxytoca	1873	1102	143	74	45	894
Pseudomonas aeruginosa	1536	904	112	118	58	1060
Enterococcus faecalis	1108	620	0	40	24	376
Enterococcus faecium	1005	458	12	7	5	424
Staphylococcus aureus	450	1420	14	7	15	395
Proteus mirabilis	1369	714	93	78	53	624
Total	10057	6790	589	456	261	5204

Endorsement of prioritized proteins mining parameters

Proteins with the attributes such as host non-homology, essentiality, virulence, exoproteome sub-cellular localization and <2 transmembrane helices in their structure have a higher probability to be a successful antigenic vaccine candidate. Proteins with satisfying

parameters for these attributes were extracted out from the core-genome sub-set of each analyzed uropathogen.

For *E. coli,* 8 vaccine candidates were identified including outer membrane N, outer membrane C, fimbrial, invasion plasmid antigen, type VII secretion system (T7SS) usher family, fimbrial assembly, gram-negative porin family and outer membrane porin ompD (Table 6). All these vaccine candidates are reported for their immunogenic character or as a vaccine candidate except OmpN (Granoff, Rathore et al. 1993; Luo, Zeng et al. 1999; Klemm and Schembri 2000; Gil-Cruz, Bobat et al. 2009; Simon, Wahid et al. 2009; Costa, Felisberto-Rodrigues et al. 2015).

OmpN belongs to porin family and is involved in transmembrane transport of <1000Da substances and having attributes desirable for a vaccine candidate, it can show a successful antigenic response against *E. coli.* OmpN is also prioritized as a vaccine candidate in *K. pneumoniae* and *K. oxytoca* along with three other proteins outer membrane C, fimbrial and fimbrial assembly which is also prioritized in *P. mirabilis.* All the 10 proteins prioritized vaccine candidates for Klebsiella species were common for *K. pneumoniae* and *K. oxytoca.*

The rest of the six vaccine candidates for Klebsiella species include family RND efflux system outer membrane lipoprotein, outer membrane channel, N-acetylneuraminic acid outer membrane channel, outer membrane F, outer membrane usher, and outer membrane A. The last three proteins are recognized as well known vaccine candidates and are in different phases of trials (Jeannin, Magistrelli et al. 2002; Telford, Barocchi et al. 2006).

In case of *P. aroginosa* channel, RND transporter, toluene efflux pump outer membrane precursor, outer membrane efflux and peptidoglycan-binding were prioritized as vaccine

candidates. Channel protein and toluene efflux pump outer membrane precursor are novel vaccine candidates were as rest reported to be successful vaccine candidates (Bina, Lavine et al. 2008; McConnell, Rumbo et al. 2011; Visweswaran, Leenhouts et al. 2014). For *E.faceium* only a novel immunogenic protein, zinc-binding protein is extracted out as a vaccine candidate.

In case *E. faecalis 50S ribosomal L13,* excinuclease ABC subunit A, assembly protein and decarboxylase were highlighted as vaccine aspirants. For *S. aureus* prophage derived single DNA binding protein a well-known *Salmonella enterica* immunogen and phospho ethyl pyramidine kinase were designated as vaccine candidates. Highest numbers of vaccine contenders, 11 vaccine candidates were identified for *P. mirabilis.*

Except Fe(3+)-binding periplasmic precursor all these 10 proteins fimbrial assembly, outer membrane channel, outer membrane usher, thiol peroxidase, ADP-heptose-LPS heptosyltransferase, metalloprotease, outer membrane lipo carrier, porin, peptidylprolyl isomerase, molecular chaperone and peptidoglycan-associated outer membrane lipoprotein are reported as vaccine candidates (Tommassen, Vermeij et al. 1990; Chakravarti, Fiske et al. 2000; Klemm and Schembri 2000; Sydenham, Douce et al. 2000; Pohl, Haller et al. 2003; Shin, Nam et al. 2003; Vermout, Brouta et al. 2004; Telford, Barocchi et al. 2006).

Table 6: Distribution of thirty-three distinct proteins in UTI causing organisms Protein extracted from the UTI causing organisms genome via susscessfully passing Phase I core genome extraction and Pahse II of priortoized protein mining drill.

Prioritized proteins	ece	Kp	ko	Pa	Efe	efs	sa	pm	Literature
50S ribosomal L13						✓			Novel
ADP-heptose--LPS heptosyltransferase								✓	Pohl, Ehmke, et al.
Decarboxylase						✓			Wherrett, Diane K., et al.
Excinuclease ABCsubunit A						✓			Crasta, Oswald R., et al.
family RND efflux system outer membrane lipoprotein		✓	✓						Novel
Fe(3+)-binding periplasmic precursor								✓	Novel
Fimbrial	✓	✓	✓						Moon, Harley W., and Thomas O. Bunn.
fimbrial assembly	✓	✓	✓					✓	Klemm, Per, and Mark A. Schembri.
gram-negative porin family	✓								Luo, Yugang, et al.
invasion plasmid antigen	✓								Simon, J. K., et al.
Metalloprotease								✓	Vermout, Sandy M., et al.
molecular chaperone								✓	Yang, Xinghong, et al.
N-acetylneuraminic acid outer membrane channel		✓	✓						Novel
outer membrane A		✓	✓						Makidon, P. E., et al.
outer membrane C	✓	✓	✓						Granoff, Dan M., et al.
outer membrane channel		✓	✓					✓	Tommassen, J., et al.
outer membrane efflux				✓					McConnell, Michael J., et al.
outer membrane F		✓	✓						Jeannin, Pascale, et al.
outer membrane lipo carrier								✓	Chakravarti, Deb N., et al.
outer membrane N	✓	✓	✓						Novel
outer membrane porin ompD	✓								Gil-Cruz, Cristina, et al.
outer membrane usher		✓	✓					✓	Telford, John L., et al.
peptidoglycan-binding				✓					Visweswaran, Ganesh Ram R., et al.

peptidylprolyl isomerase								✓	Sydenham, Mark, et al.
phospho ethyl pyramidine kinase							✓		Novel
Porin								✓	Tabaraie, Bahman, et al.
prophage derived single DNA binding							✓		Sydenham, Mark, et al.
RND transporter				✓					Bina, Xiaowen R., et al.
thiol peroxidase								✓	Shin, Ji-Hyun, et al.
toluene efflux pump outer membrane precursor				✓					Novel
type VII secretion system (T7SS) usher family	✓								Costa, Tiago RD, et al.
zinc-binding					✓				Novel
Total	8	10	10	4	1	3	2	11	

ece- *Escherichia coli*; kp- *Klebsiella pneumoniae;* ko- *Klebsiella oxytoca*; pa- *Pseudomonas aeruginosa;* efe- *Enterococcus faecium;* efs- *Enterococcus faecalis;* sa-*Staphylococcus aureus* ;pm- *Proteus mirabilis.* The symbol ✓represents the presence of a protein, in an organism.. Novel signifies the first analysis as a vaccine candidate.

4.3. Phase III: Novel prioritized proteins quantitative analysis

Only novel proteins among the prioritized vaccine candidates are further analyzed. Outer membrane N , family RND efflux system outer membrane lipoprotein, N-acetylneuraminic acid outer membrane channel, channel, Toluene efflux pump outer membrane precursor, Zinc-binding, phospho ethyl pyramidine kinase, 50S ribosomal L13 and Fe(3+)-binding periplasmic precursor.

A. Molecular weight estimation

It is essential for vaccine candidates to be of small molecular size because large peptides cannot pass through glomerulus filtrate and are excreted. This would prevent the circulation of the prioritized proteins in the blood circulation for generating an effective immune response. Small sized proteins less than 110kD are required to limit the amount of acrylamide a carcinogenic neurotoxin required during purification process. The highest molecular weight, 57.41kDa (Table 7) was estimated for Zinc binding protein and lowest

16.26 k Da was estimated for 50 S ribosomal L13. Others include ompN, toluene efflux pump outer membrane precursor, N-acetylneuraminic acid outer membrane channel, phosphor enyl pyrimidine kinase, Fe3+ binding periplasmic precursor and family RND efflux system outer membrane lipoprotein with 2.88kDa, 25.99kDa, 26.2kDa, 2.99kDa 35.75kDa and 49.73kDa respectively. Thus these proteins have suitable molecular weight as a vaccine candidate.

B. Interactome analysis

Protein interactome analysis was performed to analyze the functional importance of short listed vaccine targets in the core-genome subsets. Protein interaction network of each these targets was constructed at a high confidence thresh hold of 0.700.. All of these proteins showed interactome of more than five interactors (Table 7). Prioritized protein with additional interactors is considered metabolically imperative active protein which can act as fitting vaccine or drug target.

C. Broad spectrum antibiotic-resistance activity analysis

Sequence alignment comparison of these vaccine candidates with additional medically significant pathogens expedites the evolution of the proposed targets as ideal broad spectrum targets. BLASTp homology search in contrast to entire proteome of each of 102 urogenital pathogens resulted in the identification of promising broad-spectrum vaccine targets. The analysis was carried out cut off value was of >100 bit score and expect e-value of 0.005. The analysis also revealed that these vaccine targets exhibit homology to proteome of medically important non-UTI causing pathogens. Few of such examples include *Megasphaera sp.*, *Haemophilus parainfluenzae*, *Sphingobacterium spiritivorum*, *Corynebacterium striatum*,

Finegoldia magna, Prevotella oralis, Prevotella bivia, Veillonella parvula and *Staphylococcus simulans.*

OmpN showed homology with total 28 targets of different urogenital pathogens including *Acinetobacter baumanni.* Toulene efflux pump outer membrane precursor has overall 213 homologs in urogenital pathogens, subsequently followed by zinc-binding protein, phospho enyl pyrimidine kinase, 50 S ribosomal L13 Fe, family RND efflux system outer membrane lipo and N-acetylneuraminic acid outer membrane channel binding periplasmic precursor with 158, 145, 157, 212, 43 and 6 homolog targets. In addition COG pursuit furthermore shown that all these vaccine targets have homologs in other bacterial pathogens.

D. Functionality analysis

The novel vaccine targets were screened utilizing INTERPROSCAN and STRING database in order to predict their function. Outer membrane N and Toluene efflux pump Omp predicted to be an integral component of the outer membrane. Omp N found to be involved in small pore formation which allows transmembrane transport via diffusion. Toluene efflux pump Omp predicted to be responsible for transport activity against the concentration gradient and also represents lipid binding domains.

Zinc binding protein Interproscan envisaged it as zinc ion and metal ion binding protein curtail in cell adhesion and metal ion transport. Phosphoenyl pyrimidine kinase was predicted to be participant in phosphomethyl pyrimidine kinase activity and found responsible for thiamine biosynthetic process. The interproscan result predicted 50S L13 protein to be an integral part of ribonucleoprotein complex involved in translation. Fe^{3+} binding periplasmic precursor was found responsible for ferric transport, ATPase activity and iron homeostasis.

N-acetylneuraminic acid outer membrane channel functionality analysis revealed it to be involved in pathogenesis; transmembrane transporter activity and family RND efflux system outer membrane lipo was found to be responsible for family RND efflux system outer membrane lipo activity.

Table 7: Prioritized proteins quantitative and functional annotation. The table illustrates the length, molecular weight, number of protein interactors, broad spectrum analysis, functionality analysis and number of pat homologs.

Selected protein	Length (a.a)	Molecular weight kDal	Interac tome analysis	Broad Spectrum Activity	Functionality analysis	Pat Homology analysis
Omp N	377	2.88kDa	20	28	porin activity, transport	100
Toluene efflux pump outer membrane precursor	487	25.99kDa	31	213	lipid binding , transport, protein assembly	22
zinc binding	507	16.26kDa	10	158	zinc binding, cell adhesion, ion transport	45
Phosphor ethyl pyramidine kinase	276	2.99kDa	8	145	kinase activity, thiamine biosynthesis	31
50S ribosomal L13	147	57.41kDa	50	157	structural component, translation biosynthesis	45
Fe binding precursor	343	37.57kDa	9	212	iron homeostasis, ferric transport	9
N-acetylneuraminic acid outer membrane channel	231	26.2kDa	9	6	pathogenesis, transmembrane transporter activity	36
family RND efflux system outer membrane lipo	459	49.73kDa	24	43	transport, lipid binding activity	8

E. Patent homologs analysis

Patent homology analysis revealed the homology of the selected proteins with the patented protein sequences pat to determine the worth of similar proteins in R&D sector. OmpN has more than 100 patented protein homologs in NCBI pat database with 22 patented proteins showing more than 80% homology. Toluene efflux pump Omp pat homology analysis revealed 16 pat homologs with sequences patent US 7517684 (*Pseudomonas aeruginosa* derived polypeptide for therapeutic and diagnostic interventions) and US 7700729 (*Pseudomonas aeruginosa* derived bacteriocins, bactericidal substances) showed 99% homology.

Zinc binding protein has 45 pat homologs relating to *Streptococcus uberi, Escherichia coli,* Shigella sp. Flexner sp. and *Bacillus licheniformis,* among these sequences US 6583275 (*Enterococcus faecium* derived antibodies) patent showed 100% homology. Phosphoenyl pyrimidine kinase 31 pat homolog sequences were identified where sequence 3318 from patent US 7608276 (amino acid and corresponding nucleotide sequence derived from *S. aureus* beneficial for vaccines, immunogenic structures, diagnostics, enzymatic analysis and as targets for antibiotics) exhibited 100% homology.

50S ribosomal L13 pat homology analysis revealed 100 pat homologs with 95% including similarity for sequence 4787 from patent US 6583275. Sequences were derived from *Streptococcus pyogenes, Streptococcus mutans, Streptococcus sp., Listeria monocytogenes, Bacillus sp., Symbiobacterium thermophilum, Propionibacterium acne, Leifsonia s, Oceanobacillus iheyensis, Prochlorococcus marinus* and *Clostridium perfringens*

pat peptides showed more than 60% homology. Generated five homologs under the preset criteria where four were showing 100% similarity.

Fe(3+)-binding periplasmic precursor pat database homology analysis results showed 8 shomologs, including patents of peptide derived from *Haemophilus influenza, Streptococcus pneumoniae,* sequence 6024 from patent US 66057 (*Proteus fragilis* derived polypeptides for antibody based prevention and treatment strategy against Proteus species) showed 99% homology. N-acetylneuraminic acid outer membrane channel found to have 36 unnamed patent homologs with three showing >80% homology. Family RND efflux system outer membrane lipo has 6 homologs in pat database. The patent protein homologs novel prioritized proteins are illustrated in the Table 7.

4.4. Phase: IV Epitope analysis and 3D structure prediction

A. 3D Structure prediction

The 3D structures of novel vaccine candidates were predicted by Phyre2 . Accessibility of the 3D structure generally delivers an enhanced mode to visualize, apprehend and infer the associations with other proteins. Homology modeling was utilized for predicting the 3D structures from scratch as these are the potentially immunogenic genome conserved structures not belonging to a single genome strain. B-cell and T-cell epitope exposure was also depicted for each protein.

OmpN structure homology analysis generated its 3D (Figure 28) structure utilizing four PDB residues d2jnaa1, d2zfga1, d1phoa and d1osma. The structure was generated with 100% confidence via 345 residues (94%) modelled at >90% accuracy Phosphor ethyl pyramidine kinase 3D structure (Figure 28) prediction is based on three PBD residues utilizing c4c5lC the structure of the pyridoxal kinase from Staphylococcus aureus in complex with pyridoxal

as template. The homology analysis was carried out at 100% confidence with 270 residues (98%) modelled at >90% accuracy.

Zinc-binding vaccine candidate protein 3D (Figure 28) structure spawned from five templates d1txla, c1txlA crystal structure of metal-binding protein yoda from E. coli, Pfam duf149, d1psza "Helical backbone" metal receptor, c5hx7A a metal abc transporter from *Listeria monocytogenes* and c2o1eB crystal structure of the metal-dependent lipoprotein ycdh2 from *Bacillus subtilis*. Phyre 2 generated the structure with 476 residues (94%) modelled at >90% accuracy at 100% confidence.

Fe(3+)-binding periplasmic precursor 3D structure was generated on c4r73B template structure of the periplasmic binding protein afua a abc-type fe3+ transport system, periplasmic component from *Actinobacillus pleuropneumoniae* with 318 residues (93%) modelled at >90% accuracy and 100% confidence.

The family RND efflux system outer membrane lipo protein prioritized vaccine candidate was procreated by Phyre 2 using five templates c5azsC, c5azpA crystal structures of a membrane protein from pseudomonas aeruginosa , c3pikA outer membrane protein cusc, c4mt0A the crystal structure of the open state of the *Neisseria gonorrhoeae mtre2* outer membrane channel, c4mt4C and d1wp1a outer membrane efflux proteins (OEP). The 3D structure of family RND efflux system outer membrane lipo protein was generated with 445 residues (97%) modelled at >90% accuracy.

Phyre 2 generated N-acetylneuraminic acid outer membrane channel 3D structure via two templates c4fqeA a porin protein and c2wjqA nanc porin structure in hexagonal crystal form. The 3D structure (Figure 28) was generated at 100% confidence with 209 residues (91%) modelled at >90% accuracy. Toluene efflux omp precursor 3D structure (figure7) was

modelled utilizing c5azpA crystal structure of the *Campylobacter jejuni cmec* outer membrane2 channel as template. Phyre 2 carried the analysis with 448 residues (92%) modelled at >90% accuracy at 100% confidence.

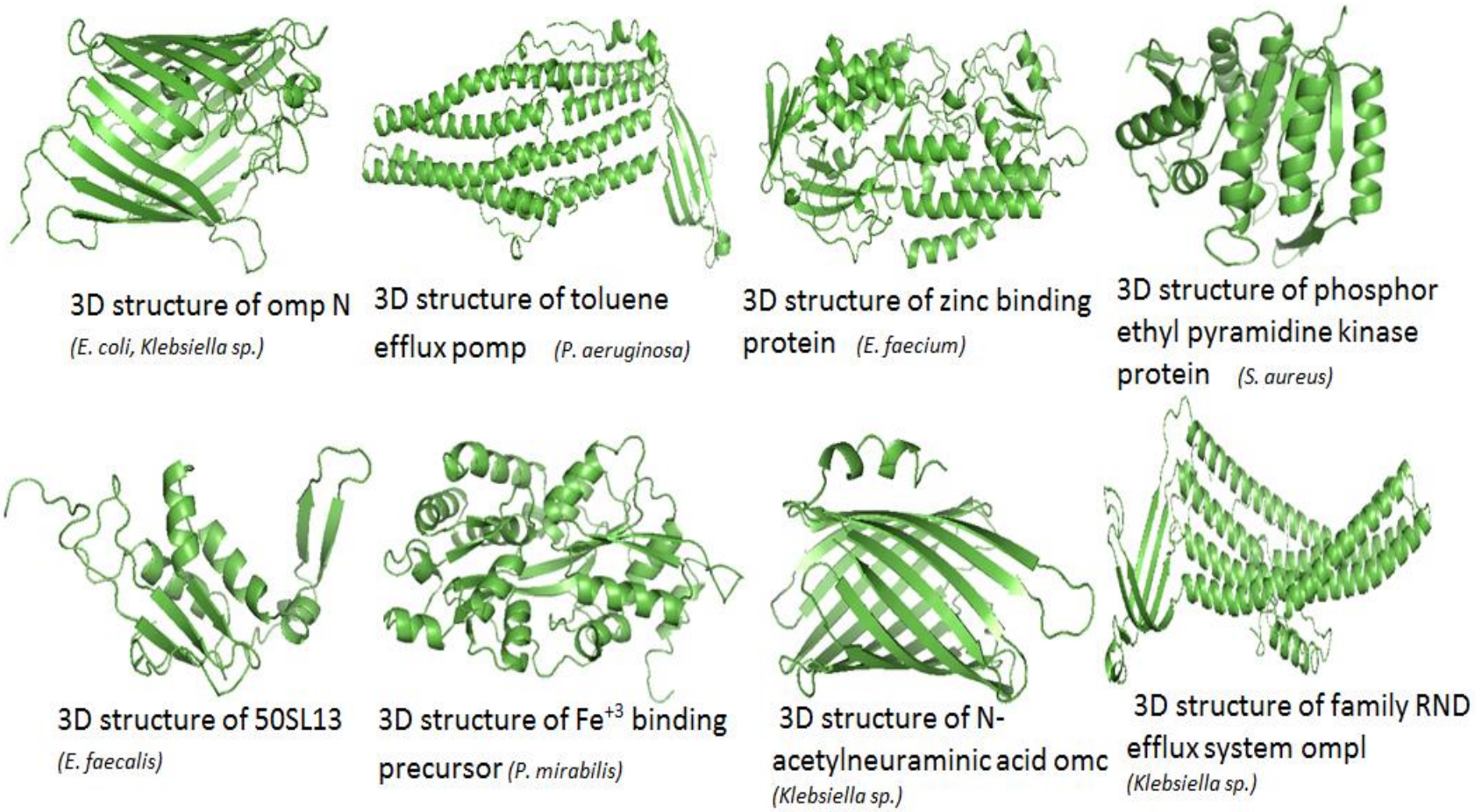


Figure 27: Novel prioritized proteins 3D structures Outer membrane protein N, Phospho ethyl pyramidine kinase, Toluene efflux pump outer membrane precursor, zinc binding , Fe binding precursor, family RND efflux system outer membrane lipo protein, N-acetylneuraminic acid outer membrane channel and 50 S ribosomal L13 3D structure obtained by homology modeling.

B. Epitope analysis

Antigenicity analysis signified that all the novel vaccine targets are antigenic with a VaxiJen score of >0.4 (Table 8). Among them N-acetylneuraminic acid outer membrane channel showed the highest antigenic score of 2.2738 and family RND efflux system outer membrane lipoprotein showed the lowest antigenicity score of 0.4917. Other target proteins include OmpN, zinc-binding, phospho ethyl pyramidine kinase, 50S ribosomal L13 with antigenicity score, Toluene efflux pump outer membrane precursor, and N-acetylneuraminic acid outer membrane channel of 1.252, 0.8178, 0.6288 ,0.8178, 0.5446 a respectively.

These proteins were subjected to B-cell epitope prediction via ABC pred. Further MHCI and MHCII molecules binding with T-cell epitopes were predicted by employing Preopred and Propred I. IC50 values were determined by exploiting DRB1*0101 as it's the most commonly occurring allele, in human population worldwide Table 8. Highest affinity among the selected epitopes 1.66nM is shown by 9-mere FMTGRANGV T-cell epitope of OmpN with total 18 MHCI and MHCII binding alleles and virulence score of 42.72.

The 9-mer T-cell epitope YNRIEHIDV of Phosphor ethyl pyramidine kinase calculated to have 35.4nM IC50 value along with 25 total MHCI and MHCII binding molecules along with 116.87 virulence score. Toluene efflux pump outer membrane precursor 9-mer YRLDVLRGE T-cell epitope exhibited an IC50 value of 17.74nM, total 26 MHCI and MHCII binding molecules and virulence score of 158.34. ILNYEKGNR a 9-mer T-cell epitope of zinc binding protein have 77.09nm IC50 value and aggregate of 26 MHCI and MHCII molecules with a 1.2 lowest most virulence score.

Family RND efflux system outer membrane lipoprotein 9-mer YVDQALRYN T-cell epitope IC50 value of 4.58nM and total 7 MHCI and MHCII molecules with a huge virulence score of 150.92.

With IC50 value of 9nM YKPNLRVQY a 9-mer T-cell epitope of N-acetylneuraminic acid outer membrane channel has an combined 17 MHCI and MHCII molecules with a huge virulence score of 500. All the T-cell epitopes were passed through NetSuefP Exposure database for surface exposure analysis, all T-cell epitopes illustrated ≤ 8 NetSurefP score representing maximum surface exposure.

The analysis implied that all the epitopes were showing desirable scores for antigenicity, virulence and stability and these can be categorized as potential vaccine targets. Their incorporation in a single vaccine would generate an efficacious formulation and would ensure a broad-spectrum protection and therapeutic advantage against UTI causing pathogens. All the novel antigenic epitopes along with their scores are illustrated in Table 8. The epitopes with the protein 3D structure were illustrated in Figure 29.

Table 8 Phase IV epitope analysis of the novel vaccine candidates Antigenic epitopes from the novel predicted proteins along with their B-cell and T-cell epitopes, surface exposure, IC50,location, virulence score, MHC I & II allele count and antigenicity score is illustrated in the following table.

Selected vaccine candidate protein	B-Cell Epitope (Thresh hold 0.8)	T-Cell Epitope (Thresh hold 0.8)	NetSurfP Exposure (Cut-off 7)	IC50 (nM) (Cut-off 125)	Location	Virulence Score (Thresh hold 0.5)	MHC-I & II Allele Count	VaxiJen Antigenicity Score (Cut-off 0.4)
Omp N	FGGDSYTNADNFMTGRANGV	FMTGRANGV	9	1.66	144-153	42.72	18	1.2572
Phosphor ethyl pyramidine kinase	GPVDHGAYNRIEHIDVEVTE	YNRIEHIDV	8	35.4	264-255	116.87	25	0.6288
Toluene efflux omp precursor	TALESARYRLDVLRGEAPGS	YRLDVLRGE	9	17.74	250-259	158.34	26	0.5446
zinc binding	YKILNYEKGNRGVRFNFETD	ILNYEKGNR	9	77.09	427-436	1.2	26	0.8178
Fe3 binding precursor	AESSPMALKLSDLKLINYDM	LKLSDLKLI	9	123.03	308	5.7	29	1.0138
family RND efflux system outer membrane lipoprotein	NFHDSTLNLYVDQALRYNSD	YVDQALRYN	9	4.58	61	150.92	7	0.4917
N-acetylneuraminic acid outer membrane channel	FNIESNDSRSIYKPNLRVQY	YKPNLRVQY	8	9	112	500	17	2.2738

.

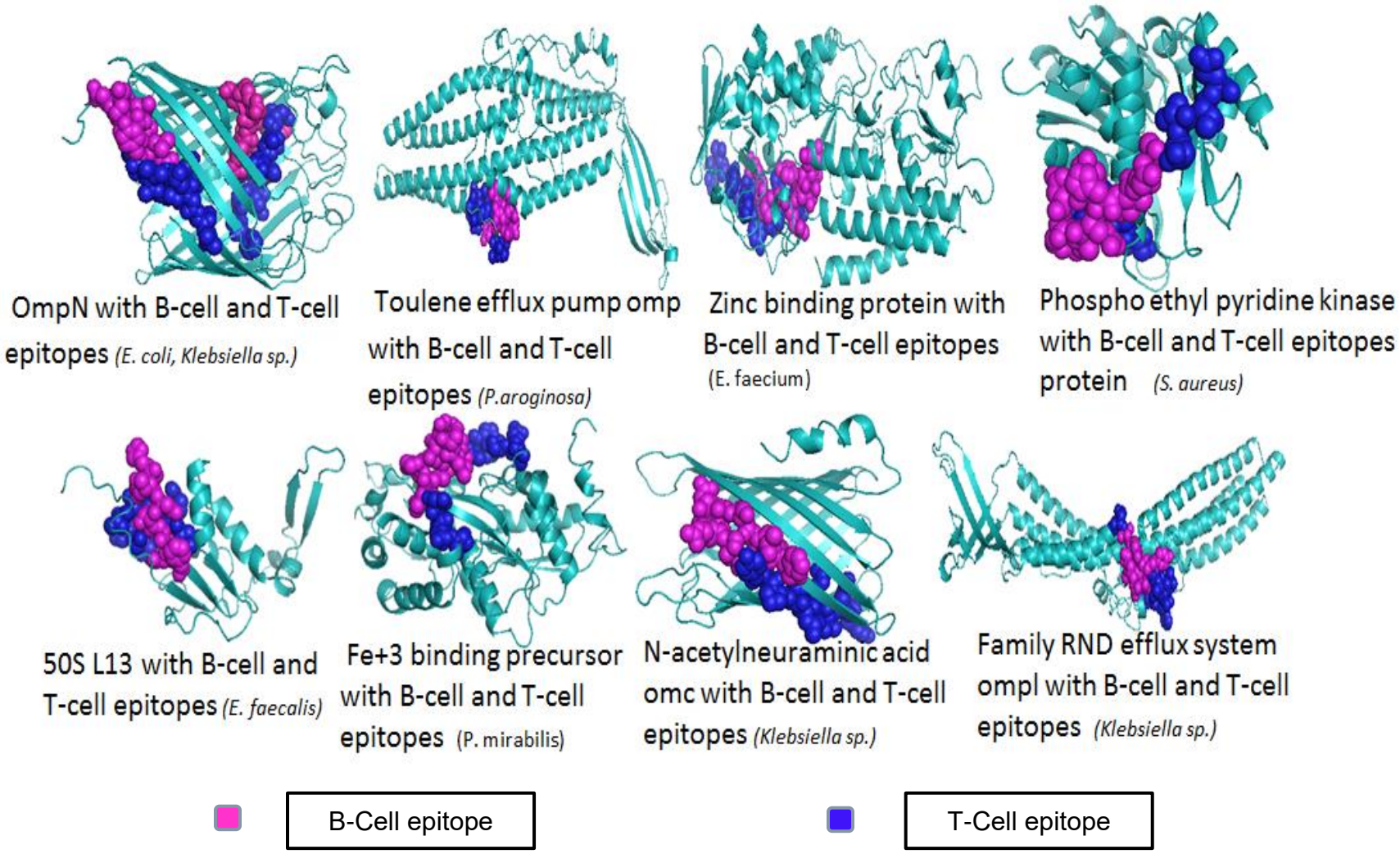


Figure 28: Novel prioritized proteins 3D structures with epitopes. Outer membrane protein N, Phospho ethyl pyramidine kinase, Toluene efflux pump outer membrane precursor, zinc binding , Fe binding precursor, family RND efflux system outer membrane lipo protein, N-acetylneuraminic acid outer membrane channel and 50 S ribosomal L13 3D structures with B-cell and T-cell epitopes.

CHAPTER 5

DISCUSSION

CHAPTER 5 DISCUSSION

The lack of a broadly protective vaccine against antibiotic resistant UTI pathogens is a foremost problem for the present society (Russo and Johnson 2003). Moreover, UTI's are associated with high costs to health care systems due to medical issues, such as sepsis, renal failure, premature birth and complications in women in gestational phase (Foxman 2002). UTI cases are on the rise in susceptible subpopulations, such as patients with diabetes, HIV/AIDS, catheters, spinal cord injuries, multiple sclerosis, pregnant women, children and the elderly (Foxman 2002).

The overall problem is aggravated by accumulative antibiotic resistance and recurrent infections (Blango and Mulvey 2010). The efforts to develop a broadly protective and safe vaccine against uropathogens have not yet been fruitful. The large antigenic and genetic variability of the strains of these uropathogens (both gram positive and gram negative bacteria) is a major obstacle to find conserved antigens (Moriel, Bertoldi et al. 2010).

Conventional approaches for vaccine development against uropathogens have not been successful. Reverse vaccinology and pan-genomics approaches for identifying vaccine targets in uropathogens is considered most cost effective and efficient method for targeting conserved proteins (Hassan, Tiwari et al. 2014). The 33 prioritized vaccine candidates fulfilled the essential parameters for a vaccine candidate, including host non-homology (Roy, Shakil et al. 2013), virulence (Shanmugham and Pan 2013) , essentiality (Malipatil, Madagi

et al. 2013), conservation (Ali, Soares et al. 2012) and functional importance (Hassan, Naz et al. 2016).

The novel vaccine candidates also provided encouraging results for their analysis regarding interactions in metabolically active proteins (Shanmugham and Pan 2013), molecular weight (Naz, Awan et al. 2015), similar proteins, broad spectrum activity (Jadhav, Shanmugham et al. 2014) and epitope analysis (Hassan, Naz et al. 2016).

In this study, 410 complete genomes of 8 uropathogens were analyzed within the framework of pan-genomics, comparative genomics and proteomic approaches. Previously, *E. coli* pan-core genome analysis of 69 genomes and 20 genomes revealed a conserved genome of 6% and 42% of average genome respectively (Rasko, Rosovitz et al. 2008; Lukjancenko, Wassenaar et al. 2010).

In the current study, pan-core genome analysis of 159 *E. coli* genomes revealed a core genome 52% of average genome, indicating a stable reliable prediction. In case of *K. pneumoniae*, earlier a pan-genome analysis comprising 300 strains including only 40 publically available strains revealed just 1,888 core genes (Holt, Wertheim et al. 2015), whereas in the current study 3,5434 (63% of average genome) core genes were extracted from 52 publically available strains. Novel findings regarding *K. oxytoca* and *P. mirabilis* pan-genome analysis are reported.

Previously, not much information was available regarding it. *P. aroginosa* pan-genome analysis only comprised 21 completely sequenced genomes (Hilker, Munder et al. 2015), whereas in this study successful pan-genome analysis of 59 completely sequenced genomes

was carried out. *E. faecium* pan-genome analysis of 12 whereas prior pan-genome analysis comprised 7 genomes (van Schaik, Top et al. 2010).

Compiled analysis of the *E. faecalis* pan-genome analysis data from Aakra et al., McBride et al. and from Margrete Solheim et al. produced a core genome estimate of just 1722 genes in 11 isolates (Aakra, Nyquist et al. 2007; McBride, Fischetti et al. 2007).The present analysis was based on publically available *E. faecalis* genome strains. A comprehensive study regarding all the plasmids and the chromosome for a genome was conducted (Galardini, Mengoni et al. 2015).

A broad analysis established on all entirely sequenced genomes would provide an enhanced framework for comparison than few organisms. Pan-genome analysis of all the analyzed uropathogens revealed an open genome with increase in the pan-genome and decrease in the core genome addition of new genomes. This increase is suggestive of an open pan-genome accentuates on the presence of gene procurement and loss occasions in the evolution, adaptation and persistent pathogenic nature of these uropathogens (Tettelin, Riley et al. 2008; Hassan, Naz et al. 2016) .

Among all, *E. coli* is the major uropathogen. The commonly prescribed antibiotics for *E. coli* infection are mentioned in literature, although *E. coli* isolates are not uniformly susceptible to these drugs (Czaja and Hooton 2006; Nicolle 2008; Salvatore, Salvatore et al. 2011; Shepherd and Pottinger 2013). The situation is further aggravated by beta-lactamases and extended spectrum beta lactamases (Paterson and Bonomo 2005). The current study also revealed the presence of resistance genes in the pan-genome including the core-genome.

The *E. coli* antibiotic resistance was the driving force for vaccine development against it. FimH of type 1 fimbriae based vaccine was not successful as it was not covering all the strains of the *E. coli* (Moriel, Bertoldi et al. 2010). Whereas all *E. coli* vaccine candidates identified in this study, are conserved in all 159 analyzed *E. coli* strains. Also ompN expectedly potent T-cell epitope may solve, the problem of modest protection by Uro-Vaxom (Wagenlehner and Naber 2006).

Besides with no homolog or interactors in human proteome it may not induce any auto immune reaction, as in the host as occurs in penicillin or cephalosporin users (Salama 2009). Furthermore it would surpass any of the antibiotic side effects mentioned in the literature by preventing its exposure. With 28 homologs in other urogenital isolates it is expected to be ideal as broad- spectrum vaccine candidate more effective than FimH of type 1 fimbriae, Uro-Vaxom and Solco-Urovac (Mobley and Alteri 2015).

K. oxytoca and *K. pneumoniae* specific prioritized proteins are conserved in all analyzed Klebsiella specie strains. Vaccine based treatment targeting all strains of *K. pneumoniae* is the necessity as the ESBL-producing Klebsiella strains have become resistant to most of the penicillins, cephalosporins and fluoroquinolones and meropenem which are often prescribed in Klebsiella specie infections (Kurupati, Teh et al. 2006).

K. pneumoniae pan-genome resistome analysis in the current study also elucidates a broad spectrum resistance by the pathogen. Though a successful phage therapy has been developed for antibiotic resistant *K. pneumoniae* strains but they would aggravate the resistance problem (Bogovazova, Voroshilova et al. 1991; Chanishvili 2012). Therefore mostly meropenem is usually prescribed for resistant *K. pneumoniae* which confers GIT

adverse even thrombophlebitis, seizures and severe hypokalemia (Ellsworth, Witt et al. 2005).

Klebsiella specie specific vaccine candidates prioritized in this study lack any of these side effects, as they are not homologs to neither host proteome nor show any interaction with any host proteome at any level. Among Klebsiella specific prioritized proteins family RND efflux system outer membrane lipoprotein and N-acetylneuraminic acid outer membrane channel are novel vaccine candidates with each having immunogenic B-cell and T-cell epitopes.

These epitopes may effectively solve the major problem of adaptive and immune response invasion in fighting against *K. oxytoca* and *K. pneumoniae* pathogenesis which is the major concern in R&D for designing an efficacious treatment program against the pathogen (Kurupati, Teh et al. 2006).

Another uropathogen requiring immunotherapy based treatment is *P. aroginosa,* its pathogenesis involves bio-film formation impenetrable by antibiotics (Cornelis 2008). The major problem in vaccine based therapy against *P. aroginosa* is lack of availability of a vaccine covering all *P. aroginosa* strains and failure of previous conjugate vaccine to produce an efficient B-cell based response (Missiakas and Schneewind 2016). RND transporter, toluene efflux pump outer membrane precursor and outer membrane efflux peptidoglycan-binding our *P. aroginosa* specific prioritized vaccine targets are expected to provide coverage against all the *P. aroginosa* strains. They generated B-cell and T-cell epitopes, indicating to produce efficient humoral and cellular response with MHC I&II binding molecules, exclusive virulence and antigenicity score.

Enterococcus species including *E. faecalis* and *E. faecium* are also reported for UTI infections and are also common hospital and community acquired pathogens. Resistance by *E. faecalis* is becoming more common in hospital environment (Van Bambeke 2006; Amyes 2007). The current treatment options for vencomycin-resistant *E. faecalis* include nitrofurantoin linezolid, and daptomycin. Quinupristin/dalfopristin is prescribed to treat *E. faecium* but not *E. faecalis (Abdulla and Abdulla 2006; Fowler Jr, Boucher et al. 2006; Denyer, Hodges et al. 2008).*

Prophylactic and empirical treatment requires an efficient antibacterial targeting both *E. faecalis* and *E.faceium.* Also Quinupristin/dalfopristin are known to cause arthralgia, myalgia, GIT side effects, itching, headache and phlebitis(Denyer, Hodges et al. 2008) . Out of the analyzed core-proteome of *E.faceium* and *E.faecalis,* zinc-binding protein and 50S ribosomal L13, is not only non-homologous, lack any interactions to host proteome therefore not expected to show any side effects. A combination of zinc-binding protein and 50S ribosomal L13 vaccine would treat all the strains of both *E. faecalis* and *E.faceium* without inducing or escalating ABR problem.

In case of *P.mirabilus* treatment the problem is the inaccessibility of antibiotics to the *P. mirabilis,* hidden in urease crystal structure. Therefore a vaccine based treatment was proposed. No successful vaccine is available for *P.mirabilus* (Li and Mobley 2002; Nielubowicz, Smith et al. 2008). The current protein priotoization analysis prioritized twelve proteins for providing immunogenic protection against *P.mirabilus* and all of these belonged to the pathogens core-genome. The epitope novel vaccine candidate for *P.mirabilus* $Fe^{(3+)}$ binding with a conserved B-cell and T-cell having with potent immunogenic response predicted can be an ideal solution.

For *S.aureus* two proteins vaccine candidates, prophage derived single DNA binding and phospho ethyl pyramidine kinase were prioritized as vaccine candidates. Both of these are part of core-genome of 109 analyzed *S.aureus* species. Phospho ethyl pyramidine kinase being a novel vaccine candidate was also analyzed for its epitopes. With host proteome interaction, strong expected immunogenic response and conservation in all species these candidates do not indicate any disadvantages observed for Nabi's StaphVax and PentaStaph, Intercell's / Merck'sV710 and VRi's SA75 (Gaitas and Kim 2015; Torre, Bacconi et al. 2015; Giersing, Dastgheyb et al. 2016).

In case of *P.mirabilus,* 11 proteins were prioritized as vaccine candidate. 10 out of these 11 prioritized proteins were already reported as a vaccine candidate while ferric binding periplasmic precursor was prioritized as a novel vaccine candidate. The protein has also shown excellent result in epitope analysis.

The combination of these proteins either in the form of peptide vaccine or a recombinant vaccine would be an excellent broad-spectrum vaccine. These peptides can also be used in different combinations for narrow spectrum vaccines or even specie specific vaccines. The pipeline adopted for the analysis can also be used for other pathogens especially in the case of bio-hazardous pathogens for which wet-lab analysis is not possible.

CONCLUSION

The current study revealed 32 essential and virulent proteins among the core proteins distributed distinctively among eight common uropathogens. The analysis reported the immunogenic activity of few proteins for the first time including family RND efflux system outer membrane lipoprotein, Fe(3+)-binding periplasmic precursor, N-acetylneuraminic acid

outer membrane channel, outer membrane N, phospho ethyl pyramidine kinase, toluene efflux pump outer membrane precursor, 50S rL13 and zinc-binding. These 8 proteins on epitopes analysis resulted in immunogenic epitopes. The combination of these epitopes can be utilized for the development of a broad-spectrum multivalent, peptide or recombinant vaccine.

FUTURE PROSPECTS

The prioritized proteins can be used for specie specific vaccine development that would provide immunogenic coverage against all the strains for which the pan-genome analysis has been carried out. Similarly the strategy can be adapted to prioritized drug targets while adopting endoproteomic subcellular localization during mining of prioritized proteins. Vaccine or drug target identification for other mammalian hosts can be carried out while adopting non-host homology analysis against the concerned mammal, avian or any other desired host. In future the strategy can also be modified with utilization of enhanced algorithms for personalized vaccine or drug target identification.

REFERENCES

Aakra, Å., O. L. Nyquist, et al. (2007). "Survey of genomic diversity among Enterococcus faecalis strains by microarray-based comparative genomic hybridization." Applied and Environmental Microbiology **73**(7): 2207-2217.

Abdulla, F. E. and E. M. Abdulla (2006). "Antibiotic options for enterococcus faecalis infection." PAKISTAN JOURNAL OF MEDICAL SCIENCES **22**(3): 286.

Ali, A., S. C. Soares, et al. (2012). "Campylobacter fetus subspecies: Comparative genomics and prediction of potential virulence targets." Gene **508**(2): 145-156.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." Journal of molecular biology **215**(3): 403-410.

Amyes, S. G. (2007). "Enterococci and streptococci." International journal of antimicrobial agents **29**: S43-S52.

Apweiler, R., A. Bairoch, et al. (2004). "UniProt: the universal protein knowledgebase." Nucleic acids research **32**(suppl 1): D115-D119.

Arias, C. A. and B. E. Murray (2009). "Antibiotic-resistant bugs in the 21st century—a clinical super-challenge." New England Journal of Medicine **360**(5): 439-443.

Baghal, S. M. L., S. L. M. Gargari, et al. (2010). "Production and immunogenicity of recombinant ferric enterobactin protein (FepA)." International journal of infectious diseases **14**: e166-e170.

Bagshaw, S. M. and K. B. Laupland (2006). "Epidemiology of intensive care unit-acquired urinary tract infections." Current opinion in infectious diseases **19**(1): 67-71.

Barbosa-Cesnik, C., M. B. Brown, et al. (2011). "Cranberry juice fails to prevent recurrent urinary tract infection: results from a randomized placebo-controlled trial." Clinical Infectious Diseases **52**(1): 23-30.

Barh, D., S. Tiwari, et al. (2011). "In silico subtractive genomics for target identification in human bacterial pathogens." Drug Development Research **72**(2): 162-177.

Baron, C. and B. Coombes (2007). "Targeting bacterial secretion systems: benefits of disarmament in the microcosm." Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders) **7**(1): 19-27.

Bauer, H. W., V. W. Rahlfs, et al. (2002). "Prevention of recurrent urinary tract infections with immuno-active E. coli fractions: a meta-analysis of five placebo-controlled double-blind studies." International journal of antimicrobial agents **19**(6): 451-456.

Bénet, T., J. Haesebaert, et al. (2015). "Characterization of patients exposed to multiple devices but free of hospital-acquired infection at intensive care unit discharge." American journal of infection control **43**(2): 171-173.

Bijker, M. S., C. J. Melief, et al. (2007). "Design and development of synthetic peptide vaccines: past, present and future." Expert review of vaccines **6**(4): 591-603.

Bina, X. R., C. L. Lavine, et al. (2008). "The AcrAB RND efflux system from the live vaccine strain of Francisella tularensis is a multiple drug efflux system that is required for virulence in mice." FEMS microbiology letters **279**(2): 226-233.

Blair, J. M., V. N. Bavro, et al. (2015). "AcrB drug-binding pocket substitution confers clinically relevant resistance and altered substrate specificity." Proceedings of the National Academy of Sciences **112**(11): 3511-3516.

Blango, M. G. and M. A. Mulvey (2010). "Persistence of uropathogenic Escherichia coli in the face of multiple antibiotics." Antimicrobial agents and chemotherapy **54**(5): 1855-1863.

Bogovazova, G., N. Voroshilova, et al. (1991). "[The efficacy of Klebsiella pneumoniae bacteriophage in the therapy of experimental Klebsiella infection]." Zhurnal mikrobiologii, epidemiologii, i immunobiologii(4): 5-8.

Brotman, R. M., J. Ravel, et al. (2014). "Microbiome, sex hormones, and immune responses in the reproductive tract: challenges for vaccine development against sexually transmitted infections." Vaccine **32**(14): 1543-1552.

Brumbaugh, A. R. and H. L. Mobley (2012). "Preventing urinary tract infection: progress toward an effective Escherichia coli vaccine." Expert review of vaccines **11**(6): 663-676.

Burnie, J., T. Carter, et al. (2002). "Identification of ABC transporters in vancomycin-resistant Enterococcus faecium as potential targets for antibody therapy." FEMS Immunology & Medical Microbiology **33**(3): 179-189.

Cai, T., S. Mazzoli, et al. (2012). "188 The role of asymptomatic bacteriuria in young women with recurrent urinary tract infections: To treat or not to treat." European Urology Supplements **1**(11): e188, e188a.

Capecchi, B., D. Serruto, et al. (2004). "The genome revolution in vaccine research." Current issues in molecular biology **6**: 17-28.

Carnielli, C. M., F. V. Winck, et al. (2015). "Functional annotation and biological interpretation of proteomics data." Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics **1854**(1): 46-54.

Chakravarti, D. N., M. J. Fiske, et al. (2000). "Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates." Vaccine **19**(6): 601-612.

Chanishvili, N. (2012). A literature review of the practical application of bacteriophage research, Nova Biomedical Books.

Chaudhari, N. M., V. K. Gupta, et al. (2016). "BPGA-an ultra-fast pan-genome analysis pipeline." Scientific reports **6**.

Chen, P.-C., L.-Y. Chang, et al. (2014). "Drug susceptibility and treatment response of common urinary tract infection pathogens in children." Journal of Microbiology, Immunology and Infection **47**(6): 478-483.

Chong, Y., Y. Ito, et al. (2011). "Genetic evolution and clinical impact in extended-spectrum β-lactamase-producing Escherichia coli and Klebsiella pneumoniae." Infection, Genetics and Evolution **11**(7): 1499-1504.

Clements, A., A. W. Jenney, et al. (2008). "Targeting subcapsular antigens for prevention of Klebsiella pneumoniae infections." Vaccine **26**(44): 5649-5653.

Cornelis, P. (2008). Pseudomonas: genomics and molecular biology, Horizon Scientific Press.

Costa, T. R., C. Felisberto-Rodrigues, et al. (2015). "Secretion systems in Gram-negative bacteria: structural and mechanistic insights." Nature Reviews Microbiology **13**(6): 343-359.

Czaja, C. A. and T. M. Hooton (2006). "Update on acute uncomplicated urinary tract infection in women." Postgraduate medicine **119**(1): 39-45.

D'Costa, V. M., K. M. McGrann, et al. (2006). "Sampling the antibiotic resistome." Science **311**(5759): 374-377.

De Las Rivas, J. and C. Fontanillo (2010). "Protein–protein interactions essentials: key concepts to building and analyzing interactome networks." PLoS Comput Biol **6**(6): e1000807.

DeLano, W. L. (2002). "PyMOL." DeLano Scientific, San Carlos, CA **700**.

Delany, I., R. Rappuoli, et al. (2013). "Vaccines, reverse vaccinology, and bacterial pathogenesis." Cold Spring Harbor perspectives in medicine **3**(5): a012476.

Demetriou, E., S. J. Emans, et al. (1982). "Dysuria in adolescent girls: urinary tract infection or vaginitis?" Pediatrics **70**(2): 299-301.

Denyer, S. P., N. A. Hodges, et al. (2008). Hugo and Russell's pharmaceutical microbiology, John Wiley & Sons.

Dielubanza, E. J. and A. J. Schaeffer (2011). "Urinary tract infections in women." Medical Clinics of North America **95**(1): 27-41.

Do Thi, T. N., N. T. K. Chuc, et al. (2014). "Antibiotic sales in rural and urban pharmacies in northern Vietnam: an observational study." BMC Pharmacology and Toxicology **15**(1): 6.

Donati, C., N. L. Hiller, et al. (2010). "Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species." Genome biology **11**(10): 1.

Donati, C. and R. Rappuoli (2013). "Reverse vaccinology in the 21st century: improvements over the original design." Annals of the New York Academy of Sciences **1285**(1): 115-132.

Donlan, R. M. (2001). "Biofilms and device-associated infections." Emerging infectious diseases **7**(2): 277.

Doytchinova, I. A. and D. R. Flower (2007). "VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines." BMC bioinformatics **8**(1): 1.

Drenkard, E. (2003). "Antimicrobial resistance of Pseudomonas aeruginosa biofilms." Microbes and infection **5**(13): 1213-1219.

Ellsworth, A. J., D. M. Witt, et al. (2005). Mosby's Medical Drug Reference 2006, Mosby Limited.

Fernando, S., P. Selvarani, et al. (2004). "THGS: a web-based database of Transmembrane Helices in Genome Sequences." Nucleic acids research **32**(suppl 1): D125-D128.

Finley, R. L., P. Collignon, et al. (2013). "The scourge of antibiotic resistance: the important role of the environment." Clinical infectious diseases: cit355.

Flores-Mireles, A. L., J. N. Walker, et al. (2015). "Urinary tract infections: epidemiology, mechanisms of infection and treatment options." Nature Reviews Microbiology **13**(5): 269-284.

Fowler Jr, V. G., H. W. Boucher, et al. (2006). "Daptomycin versus standard therapy for bacteremia and endocarditis caused by Staphylococcus aureus." New England Journal of Medicine **355**(7): 653-665.

Foxman, B. (2002). "Epidemiology of urinary tract infections: incidence, morbidity, and economic costs." The American journal of medicine **113**(1): 5-13.

Gaitas, A. and G. Kim (2015). "An introduction to antibiotic-free techniques to eliminate Staphylococcus aureus from blood." bioRxiv: 030072.

Galardini, M., A. Mengoni, et al. (2015). "Mapping contigs using CONTIGuator." Bacterial Pangenomics: Methods and Protocols: 163-176.

García-Quintanilla, M., M. R. Pulido, et al. (2016). "Vaccines for Antibiotic-Resistant Bacteria: Possibility or Pipe Dream?" Trends in pharmacological sciences **37**(2): 143-152.

Gardy, J. L., M. R. Laird, et al. (2005). "PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis." Bioinformatics **21**(5): 617-623.

Garg, A. and D. Gupta (2008). "VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens." BMC bioinformatics **9**(1): 1.

Gasteiger, E., C. Hoogland, et al. (2005). Protein identification and analysis tools on the ExPASy server, Springer.

Giersing, B. K., S. S. Dastgheyb, et al. (2016). "Status of vaccine research and development of vaccines for Staphylococcus aureus." Vaccine **34**(26): 2962-2966.

Gil-Cruz, C., S. Bobat, et al. (2009). "The porin OmpD from nontyphoidal Salmonella is a key target for a protective B1b cell antibody response." Proceedings of the National Academy of Sciences **106**(24): 9803-9808.

Girard, R., S. Gaujard, et al. (2015). "Controlling urinary tract infections associated with intermittent bladder catheterization in geriatric hospitals." Journal of Hospital Infection.

Gona, F., F. Barbera, et al. (2014). "In vivo multiclonal transfer of blaKPC-3 from Klebsiella pneumoniae to Escherichia coli in surgery patients." Clinical Microbiology and Infection **20**(10): O633-O635.

Gonzalez, R., J. S. Cullor, et al. (1989). "Prevention of clinical coliform mastitis in dairy cows by a mutant Escherichia coli vaccine." Canadian Journal of Veterinary Research **53**(3): 301.

Graham, D. W. (2015). "Antibiotic Resistance in the Environment: Not the Usual Suspects." Chemistry & biology **22**(7): 805-806.

Granoff, D. M., M. H. Rathore, et al. (1993). "Effect of immunity to the carrier protein on antibody responses to Haemophilus influenzae type b conjugate vaccines." Vaccine **11**: S46-S51.

Greenfield, S., G. Friedland, et al. (1974). "Protocol management of dysuria, urinary frequency, and vaginal discharge." Annals of internal medicine **81**(4): 452-457.

Guan, P., I. A. Doytchinova, et al. (2003). "MHCPred: a server for quantitative prediction of peptide–MHC binding." Nucleic acids research **31**(13): 3621-3624.

Gunther, N. W., V. Lockatell, et al. (2001). "In Vivo Dynamics of Type 1 Fimbria Regulation in UropathogenicEscherichia coli during Experimental Urinary Tract Infection." Infection and immunity **69**(5): 2838-2846.

Guzman, C., C. Pruzzo, et al. (1989). "Role of adherence in pathogenesis of Enterococcus faecalis urinary tract infection and endocarditis." Infection and immunity **57**(6): 1834-1838.

Hancock, R. E., A. Nijnik, et al. (2012). "Modulating immunity as a therapy for bacterial infections." Nature Reviews Microbiology **10**(4): 243-254.

Hassan, A., A. Naz, et al. (2016). "Pangenome and immuno-proteomics analysis of Acinetobacter baumannii strains revealed the core peptide vaccine targets." BMC genomics **17**(1): 732.

Hassan, S. S., S. Tiwari, et al. (2014). "Proteome scale comparative modeling for conserved drug and vaccine targets identification in Corynebacterium pseudotuberculosis." BMC genomics **15**(7): 1.

He, Y. (2012). "Analyses of Brucella pathogenesis, host immunity, and vaccine targets using systems biology and bioinformatics." Frontiers in cellular and infection microbiology **2**: 2.

Hilker, R., A. Munder, et al. (2015). "Interclonal gradient of virulence in the Pseudomonas aeruginosa pangenome from disease and environment." Environmental microbiology **17**(1): 29-46.

Hill, C., M. Pan, et al. (2015). "Presence of Virulence-Associated Genes and Ability to Form Biofilm among Clinical Isolates of Escherichia coli Causing Urinary Infection in Domestic Animals." Advances in Microbiology **5**(08): 573.

Hiller, N. L., B. Janto, et al. (2007). "Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome." Journal of bacteriology **189**(22): 8186-8195.

Högberg, L. D., A. Heddini, et al. (2010). "The global need for effective antibiotics: challenges and recent advances." Trends in pharmacological sciences **31**(11): 509-515.

Holt, K. E., H. Wertheim, et al. (2015). "Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health." Proceedings of the National Academy of Sciences **112**(27): E3574-E3581.

Hong, H., D. R. Patel, et al. (2006). "The outer membrane protein OmpW forms an eight-stranded β-barrel with a hydrophobic channel." Journal of biological chemistry **281**(11): 7568-7577.

Howett, M. K. and J. W. Kreider (2002). Human xenografts for microbicide testing and anatomical modeling, Google Patents.

Hyatt, D., G.-L. Chen, et al. (2010). "Prodigal: prokaryotic gene recognition and translation initiation site identification." BMC bioinformatics **11**(1): 1.

Hyyryläinen, H. L., B. C. Marciniak, et al. (2010). "Penicillin-binding protein folding is dependent on the PrsA peptidyl-prolyl cis-trans isomerase in Bacillus subtilis." Molecular microbiology **77**(1): 108-127.

Jadhav, A., B. Shanmugham, et al. (2014). "Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis." Infection, Genetics and Evolution **27**: 300-308.

Jaffe, A. B., J. Lerner, et al. (2001). Innovation policy and the economy, Mit Press.

Jarvis, W. R. and W. J. Martone (1992). "Predominant pathogens in hospital infections." Journal of Antimicrobial Chemotherapy **29**(suppl A): 19-24.

Jean, S.-S., G. Coombs, et al. (2016). "Epidemiology and antimicrobial susceptibility profiles of pathogens causing urinary tract infections in the Asia-Pacific region: Results from the Study for Monitoring Antimicrobial Resistance Trends (SMART), 2010–2013." International journal of antimicrobial agents **47**(4): 328-334.

Jeannin, P., G. Magistrelli, et al. (2002). "Outer membrane protein A (OmpA): a new pathogen-associated molecular pattern that interacts with antigen presenting cells—impact on vaccine strategies." Vaccine **20**: A23-A27.

Jeannin, P., T. Renno, et al. (2000). "OmpA targets dendritic cells, induces their maturation and delivers antigen into the MHC class I presentation pathway." Nature immunology **1**(6): 502-509.

Jensen, L. J., M. Kuhn, et al. (2009). "STRING 8—a global view on proteins and their functional interactions in 630 organisms." Nucleic acids research **37**(suppl 1): D412-D416.

Jha, N. and S. Bapat (2005). "A study of sensitivity and resistance of pathogenic micro organisms causing UTI in Kathmandu valley."

Johnson, J. R. (1991). "Virulence factors in Escherichia coli urinary tract infection." Clinical microbiology reviews **4**(1): 80-128.

Kanampalliwar, A. M., R. Soni, et al. (2013). "Reverse vaccinology: basics and applications." Journal of Vaccines & Vaccination **2013**.

Keshavjee, S. and P. E. Farmer (2012). "Tuberculosis, drug resistance, and the history of modern medicine." New England Journal of Medicine **367**(10): 931-936.

Kim, Y., S.-H. Wie, et al. (2014). "Comparison of the clinical characteristics of diabetic and non-diabetic women with community-acquired acute pyelonephritis: a multicenter study." Journal of Infection **69**(3): 244-251.

Klemm, P. and M. A. Schembri (2000). "Fimbrial surface display systems in bacteria: from vaccines to random libraries." Microbiology **146**(12): 3025-3032.

Koehn, F. E. and G. T. Carter (2005). "The evolving role of natural products in drug discovery." Nature reviews Drug discovery **4**(3): 206-220.

Kropec, A., I. Sava, et al. (2011). "Identification of SagA as a novel vaccine target for the prevention of Enterococcus faecium infections." Microbiology **157**(12): 3429-3434.

Kumar, D., A. K. Singh, et al. (2014). "Antimicrobial Susceptibility Profile of Extended Spectrum β-Lactamase (ESBL) Producing Escherichia coli from Various Clinical Samples." Infectious diseases **7**: 1.

Kunin, C. M. (1987). Detection, prevention, and management of urinary tract infections, Lea & Febiger.

Kurupati, P., B. K. Teh, et al. (2006). "Identification of vaccine candidate antigens of an ESBL producing Klebsiella pneumoniae clinical strain by immunoproteome analysis." Proteomics **6**(3): 836-844.

Langermann, S., R. Möllby, et al. (2000). "Vaccination with FimH Adhesin Protects Cynomolgus Monkeys from Colonization and Infection by Uropathogenic Eschevichia coli." Journal of Infectious Diseases **181**(2): 774-778.

Lapierre, P. and J. P. Gogarten (2009). "Estimating the size of the bacterial pan-genome." Trends in Genetics **25**(3): 107-110.

Laupland, K. B., S. M. Bagshaw, et al. (2005). "Intensive care unit-acquired urinary tract infections in a regional critical care system." Critical Care **9**(2): 1.

Laxminarayan, R., A. Duse, et al. (2013). "Antibiotic resistance—the need for global solutions." The Lancet infectious diseases **13**(12): 1057-1098.

Lehman, S. M. and R. M. Donlan (2015). "Bacteriophage-Mediated Control of a Two-Species Biofilm Formed by Microorganisms Causing Catheter-Associated Urinary Tract Infections in an In Vitro Urinary Catheter Model." Antimicrobial agents and chemotherapy **59**(2): 1127-1137.

Li, W., B. Kondratowicz, et al. (2013). "The annotation-enriched non-redundant patent sequence databases." Database **2013**: bat005.

Li, X., C. V. Lockatell, et al. (2004). "Development of an intranasal vaccine to prevent urinary tract infection by Proteus mirabilis." Infection and immunity **72**(1): 66-75.
Li, X. and H. L. Mobley (2002). "Vaccines for Proteus mirabilis in urinary tract infection." International journal of antimicrobial agents **19**(6): 461-465.
Lim, J. K., N. W. Gunther, et al. (1998). "In vivo phase variation of Escherichia coli type 1 fimbrial genes in women with urinary tract infection." Infection and immunity **66**(7): 3303-3310.
Linhares, I., T. Raposo, et al. (2013). "Frequency and antimicrobial resistance patterns of bacteria implicated in community urinary tract infections: a ten-year surveillance study (2000–2009)." BMC infectious diseases **13**(1): 19.
Litwin, M. S., C. S. Saigal, et al. (2005). "Urologic diseases in America Project: analytical methods and principal findings." The Journal of urology **173**(3): 933-937.
Llor, C. and L. Bjerrum (2014). "Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem." Therapeutic advances in drug safety **5**(6): 229-241.
Lu, Z., D. Szafron, et al. (2004). "Predicting subcellular localization of proteins using machine-learned classifiers." Bioinformatics **20**(4): 547-556.
Lukjancenko, O., T. M. Wassenaar, et al. (2010). "Comparison of 61 sequenced Escherichia coli genomes." Microbial ecology **60**(4): 708-720.
Luo, Y., Q. Zeng, et al. (1999). "Sequence analysis of Pasteurella multocida major outer membrane protein (OmpH) and application of synthetic peptides in vaccination of chickens against homologous strain challenge." Vaccine **17**(7): 821-831.
Malipatil, V., S. Madagi, et al. (2013). "Sexually transmitted diseases putative drug target database: A comprehensive database of putative drug targets of pathogens identified by comparative genomics." Indian journal of pharmacology **45**(5): 434.
Martínez, J. L., F. Baquero, et al. (2007). "Predicting antibiotic resistance." Nature Reviews Microbiology **5**(12): 958-965.
Mathers, A. J., H. L. Cox, et al. (2011). "Molecular dissection of an outbreak of carbapenem-resistant Enterobacteriaceae reveals intergenus KPC carbapenemase transmission through a promiscuous plasmid." MBio **2**(6): e00204-00211.
McArthur, A. G., N. Waglechner, et al. (2013). "The comprehensive antibiotic resistance database." Antimicrobial agents and chemotherapy **57**(7): 3348-3357.
McBride, S. M., V. A. Fischetti, et al. (2007). "Genetic diversity among Enterococcus faecalis." PLoS One **2**(7): e582.
McConnell, M. J., C. Rumbo, et al. (2011). "Outer membrane vesicles as an acellular vaccine against Acinetobacter baumannii." Vaccine **29**(34): 5705-5710.
McKenna, M. (2012). "Man vs MRSA." Nature **482**(7383): 23-25.
Medini, D., C. Donati, et al. (2005). "The microbial pan-genome." Current opinion in genetics & development **15**(6): 589-594.
Missiakas, D. and O. Schneewind (2016). "Staphylococcus aureus vaccines: Deviating from the carol." The Journal of Experimental Medicine **213**(9): 1645-1653.
Mittal, R., S. Aggarwal, et al. (2009). "Urinary tract infections caused by Pseudomonas aeruginosa: a minireview." Journal of infection and public health **2**(3): 101-111.
Mobley, H. L. and C. J. Alteri (2015). "Development of a vaccine against Escherichia coli urinary tract infections." Pathogens **5**(1): 1.

Moriel, D. G., I. Bertoldi, et al. (2010). "Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli." Proceedings of the National Academy of Sciences **107**(20): 9072-9077.

Mulder, N. and R. Apweiler (2007). "InterPro and InterProScan." Comparative genomics: 59-70.

Mullard, A. (2011). "2010 FDA drug approvals." Nature Reviews Drug Discovery **10**(2): 82-85.

Murray, B. E. (2000). "Vancomycin-resistant enterococcal infections." New England Journal of Medicine **342**(10): 710-721.

Muzzi, A., V. Masignani, et al. (2007). "The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials." Drug discovery today **12**(11): 429-439.

Nancy, Y. Y., J. R. Wagner, et al. (2010). "PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes." Bioinformatics **26**(13): 1608-1615.

Naz, A., F. M. Awan, et al. (2015). "Identification of putative vaccine candidates against Helicobacter pylori exploiting exoproteome and secretome: a reverse vaccinology based approach." Infection, Genetics and Evolution **32**: 280-291.

Nicolle, L. E. (2008). "Uncomplicated urinary tract infection in adults including uncomplicated pyelonephritis." Urologic Clinics of North America **35**(1): 1-12.

Nielubowicz, G. R., S. N. Smith, et al. (2008). "Outer membrane antigens of the uropathogen Proteus mirabilis recognized by the humoral response during experimental murine urinary tract infection." Infection and immunity **76**(9): 4222-4231.

Nitzan, O., M. Elias, et al. (2015). "Urinary tract infections in patients with type 2 diabetes mellitus: review of prevalence, diagnosis, and management." Diabetes, metabolic syndrome and obesity: targets and therapy **8**: 129.

Organization, W. H. (2000). The world health report 2000: health systems: improving performance, World Health Organization.

Paterson, D. L. and R. A. Bonomo (2005). "Extended-spectrum β-lactamases: a clinical update." Clinical microbiology reviews **18**(4): 657-686.

Phimister, E. G., C. A. Arias, et al. (2015). "A New Antibiotic and the Evolution of Resistance." New England Journal of Medicine **372**(12): 1168-1170.

Pitts, S. R., R. W. Niska, et al. (2008). "National hospital ambulatory medical care survey: 2006 emergency department summary." Natl Health Stat Report **7**(7): 1-38.

Pizza, M., V. Scarlato, et al. (2000). "Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing." Science **287**(5459): 1816-1820.

Platt, R. (1982). "Adverse effects of third-generation cephalosporins." Journal of Antimicrobial Chemotherapy **10**(suppl C): 135-140.

Pohl, E., J. C. Haller, et al. (2003). "Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator." Molecular microbiology **47**(4): 903-915.

Poolman, J. T. and M. Wacker (2015). "Extra-intestinal Pathogenic Escherichia coli (ExPEC), a common human pathogen: challenges for vaccine development and progress in the field." Journal of Infectious Diseases: jiv429.

Pop, M. and D. Kosack (2004). "Using the TIGR assembler in shotgun sequencing projects." Bacterial Artificial Chromosomes: Volume 1 Library Construction, Physical Mapping, and Sequencing: 279-294.

Prevention, E. C. f. D., Control, et al. (2013). Point Prevalence Survey of Healthcare-associated Infections and Antimicrobial Use in European Acute Care Hospitals: 2011-2012, Publications Office of the European Union.

Raman, K., K. Yeturu, et al. (2008). "targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis." BMC systems biology **2**(1): 1.

Rasko, D. A., M. Rosovitz, et al. (2008). "The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates." Journal of bacteriology **190**(20): 6881-6893.

Raz, R., B. Chazan, et al. (2004). "Cranberry juice and urinary tract infection." Clinical Infectious Diseases **38**(10): 1413-1419.

Reinhardt, J. A., D. A. Baltrus, et al. (2009). "De novo assembly using low-coverage short read sequence data from the rice pathogen Pseudomonas syringae pv. oryzae." Genome research **19**(2): 294-305.

Reno, M. L., N. L. Held, et al. (2009). "Biogeography of the Sulfolobus islandicus pan-genome." Proceedings of the National Academy of Sciences **106**(21): 8605-8610.

Rinaudo, C. D., J. L. Telford, et al. (2009). "Vaccinology in the genome era." The Journal of clinical investigation **119**(9): 2515-2525.

Rodriguez-Valera, F. and D. W. Ussery (2012). "Is the pan-genome also a pan-selectome?" F1000Research **1**.

Roy, P. K., S. K. Shakil, et al. (2013). "Identification of putative drug targets of Listeria monocytogenes F2365 by subtractive genomics approach."

Rumbaugh, K. P., T. Sawa, et al. (2003). New perspectives on prevention and management of Pseudomonas aeruginosa infections. Severe infections caused by Pseudomonas aeruginosa, Springer**:** 183-199.

Russo, T. A. and J. R. Johnson (2003). "Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem." Microbes and infection **5**(5): 449-456.

Sader, H. S., M. A. Pfaller, et al. (1994). "Prevalence of important pathogens and the antimicrobial activity of parenteral drugs at numerous medical centers in the United States II. Study of the intra-and interlaboratory dissemination of extended-spectrum β-lactamase-producing Enterobacteriaceae." Diagnostic microbiology and infectious disease **20**(4): 203-208.

Saha, S. and G. Raghava (2006). "Prediction of continuous B-cell epitopes in an antigen using recurrent neural network." Proteins: Structure, Function, and Bioinformatics **65**(1): 40-48.

Sahebnasagh, R., H. Saderi, et al. (2015). "Evaluation of Antibiotic Susceptibility of Escherichia coli and Klebsiella spp. Isolated from Urinary Tract Infections of Patients Attending a University Hospital in Tehran." Journal of Basic and Clinical Pathophysiology **3**(2): 9-12.

Sakran, W., V. Smolkin, et al. (2015). "Community-Acquired Urinary Tract Infection in Hospitalized Children Etiology and Antimicrobial Resistance. A Comparison Between First Episode and Recurrent Infection." Clinical pediatrics **54**(5): 479-483.

Salama, A. (2009). "Drug-induced immune hemolytic anemia." Expert opinion on drug safety **8**(1): 73-79.

Salvatore, S., S. Salvatore, et al. (2011). "Urinary tract infections in women." European Journal of Obstetrics & Gynecology and Reproductive Biology **156**(2): 131-136.

Sannes, M. R., M. A. Kuskowski, et al. (2004). "Virulence factor profiles and phylogenetic background of Escherichia coli isolates from veterans with bacteremia and uninfected control subjects." Journal of Infectious Diseases **190**(12): 2121-2128.

Sbai, H., A. Mehta, et al. (2001). "Use of T cell epitopes for vaccine development." Current Drug Targets-Infectious Disorders **1**(3): 303-313.

Scavone, P., A. Miyoshi, et al. (2007). "Intranasal immunisation with recombinant Lactococcus lactis displaying either anchored or secreted forms of Proteus mirabilis MrpA fimbrial protein confers specific immune response and induces a significant reduction of kidney bacterial colonisation in mice." Microbes and infection **9**(7): 821-828.

Schäberle, T. F. and I. M. Hack (2014). "Overcoming the current deadlock in antibiotic research." Trends in microbiology **22**(4): 165-167.

Seeber, F. (2007). "Patent searches as a complement to literature searches in the life sciences—a'how-to'tutorial." Nature protocols **2**(10): 2418-2428.

Seixas-Mikelus, S. A., L. C. Jenkins, et al. (2009). "Pediatric pyeloplasty: comparison of literature meta-analysis of laparoscopic and open techniques with open surgery at a single institution." The Journal of urology **182**(5): 2428-2434.

Shahzada, H., S. Shehzada, et al. (2014). "EPIDMOLOGICAL AND SENSITIVITY PATTERNS OF SPECIFIED ANTIBIOTICS INIMICAL TO PATHOGENS CAUSING URINARY TRACT INFECTIONS (UTIs)." International Journal of Innovation and Applied Studies **7**(1): 115.

Shanmugham, B. and A. Pan (2013). "Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach." PLoS One **8**(3): e59126.

Sharma, A., A. Krause, et al. (2011). "Recent developments for Pseudomonas vaccines." Human vaccines **7**(10): 999-1011.

Sharma, G., T. Narwani, et al. (2016). "Complete Genome Sequence and Comparative Genomics of a Novel Myxobacterium Myxococcus hansupus." PLoS One **11**(2): e0148593.

Shepherd, A. K. and P. S. Pottinger (2013). "Management of urinary tract infections in the era of increasing antimicrobial resistance." Medical Clinics of North America **97**(4): 737-757.

Shigemura, K., K. Tanaka, et al. (2014). "Postoperative bacteriuria, pyuria and urinary tract infection in patients with an orthotopic sigmoid colon neobladder replacement." The Journal of antibiotics **67**(2): 143-145.

Shin, J.-H., S.-W. Nam, et al. (2003). "Identification of immunodominant Helicobacter pylori proteins with reactivity to H. pylori-specific egg-yolk immunoglobulin." Journal of medical microbiology **52**(3): 217-222.

Sievert, D. M., P. Ricks, et al. (2013). "Antimicrobial-resistant pathogens associated with healthcare-associated infections summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009–2010." Infection Control & Hospital Epidemiology **34**(01): 1-14.

Silverman, R. B. and M. W. Holladay (2014). The organic chemistry of drug design and drug action, Academic press.

Simon, J., R. Wahid, et al. (2009). "Antigen-specific B memory cell responses to lipopolysaccharide (LPS) and invasion plasmid antigen (Ipa) B elicited in volunteers vaccinated with live-attenuated Shigella flexneri 2a vaccine candidates." Vaccine **27**(4): 565-572.

Singhal, A., R. Sharma, et al. (2015). "Hospital and community isolates of uropathogens and their antibiotic sensitivity pattern from a tertiary care hospital in North West India." Annals of medical and health sciences research **4**(1): 51-56.

Sokurenko, E. V., V. Chesnokova, et al. (1998). "Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin." Proceedings of the National Academy of Sciences **95**(15): 8922-8926.

Sorichter, S., U. Baumann, et al. (2009). "Immune responses in the airways by nasal vaccination with systemic boosting against Pseudomonas aeruginosa in chronic lung disease." Vaccine **27**(21): 2755-2759.

Sydenham, M., G. Douce, et al. (2000). "Salmonella enterica Serovar TyphimuriumsurA Mutants Are Attenuated and Effective Live Oral Vaccines." Infection and immunity **68**(3): 1109-1115.

Szklarczyk, D., A. Franceschini, et al. (2011). "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored." Nucleic acids research **39**(suppl 1): D561-D568.

Szklarczyk, D., A. Franceschini, et al. (2014). "STRING v10: protein–protein interaction networks, integrated over the tree of life." Nucleic acids research: gku1003.

Tang, S. S., A. Apisarnthanarak, et al. (2014). "Mechanisms of β-lactam antimicrobial resistance and epidemiology of major community-and healthcare-associated multidrug-resistant bacteria." Advanced drug delivery reviews **78**: 3-13.

Telford, J. L., M. A. Barocchi, et al. (2006). "Pili in gram-positive pathogens." Nature Reviews Microbiology **4**(7): 509-519.

Tettelin, H., V. Masignani, et al. (2005). "Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome"." Proceedings of the National Academy of Sciences of the United States of America **102**(39): 13950-13955.

Tettelin, H., D. Riley, et al. (2008). "Comparative genomics: the bacterial pan-genome." Current opinion in microbiology **11**(5): 472-477.

Thomas, L. and C. R. Tracy (2015). "Treatment of Fungal Urinary Tract Infection." Urologic Clinics of North America **42**(4): 473-483.

Tomas, M. E., D. Getman, et al. (2015). "Overdiagnosis of Urinary Tract Infection and Underdiagnosis of Sexually Transmitted Infection in Adult Women Presenting to an Emergency Department." Journal of clinical microbiology **53**(8): 2686-2692.

Tommassen, J., P. Vermeij, et al. (1990). "Isolation of Neisseria meningitidis mutants deficient in class 1 (porA) and class 3 (porB) outer membrane proteins." Infection and immunity **58**(5): 1355-1359.

Torre, A., M. Bacconi, et al. (2015). "Four-component Staphylococcus aureus vaccine 4C-Staph enhances Fcγ receptor expression in neutrophils and monocytes and mitigates S. aureus infection in neutropenic mice." Infection and immunity **83**(8): 3157-3163.

Treitman, A. N., P. R. Yarnold, et al. (2005). "Emerging incidence of Enterococcus faecium among hospital isolates (1993 to 2002)." Journal of clinical microbiology **43**(1): 462-463.

Tsai, S.-S., J.-C. Huang, et al. (2010). "Characteristics of Klebsiella pneumoniae bacteremia in community-acquired and nosocomial infections in diabetic patients." Chang Gung Med J **33**(5): 532-539.

Tsay, R.-W., L. Siu, et al. (2002). "Characteristics of bacteremia between community-acquired and nosocomial Klebsiella pneumoniae infection: risk factor for mortality and the impact of capsular serotypes as a herald for community-acquired infection." Archives of internal medicine **162**(9): 1021-1027.

Van Bambeke, F. (2006). "Glycopeptides and glycodepsipeptides in clinical development: a comparative review of their antibacterial spectrum, pharmacokinetics and clinical efficacy." Current Opinion in Investigational Drugs **7**(8): 740.

van der Burg, S. H., M. S. Bijker, et al. (2006). "Improved peptide vaccine strategies, creating synthetic artificial infections to maximize immune efficacy." Advanced drug delivery reviews **58**(8): 916-930.

van Schaik, W., J. Top, et al. (2010). "Pyrosequencing-based comparative genome analysis of the nosocomial pathogen Enterococcus faecium and identification of a large transferable pathogenicity island." BMC genomics **11**(1): 1.

Vermout, S. M., F. D. Brouta, et al. (2004). "Evaluation of immunogenicity and protective efficacy of a Microsporum canis metalloprotease subunit vaccine in guinea pigs." FEMS Immunology & Medical Microbiology **40**(1): 75-80.

Vetrivel, U., G. Subramanian, et al. (2011). "A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens." The HUGO journal **5**(1): 25.

Vidal, E., J. Torre-Cisneros, et al. (2012). "Bacterial urinary tract infection after solid organ transplantation in the RESITRA cohort." Transplant Infectious Disease **14**(6): 595-603.

Visweswaran, G. R. R., K. Leenhouts, et al. (2014). "Exploiting the peptidoglycan-binding motif, LysM, for medical and industrial applications." Applied microbiology and biotechnology **98**(10): 4331-4345.

Wagenlehner, F. M. and K. G. Naber (2006). "Treatment of bacterial urinary tract infections: presence and future." European urology **49**(2): 235-244.

Wagner, V. E. and B. H. Iglewski (2008). "P. aeruginosa biofilms in CF infection." Clinical reviews in allergy & immunology **35**(3): 124-134.

Wan, J., S. J. Skoog, et al. (2012). "Section on Urology response to new guidelines for the diagnosis and management of UTI." Pediatrics **129**(4): e1051-e1053.

Werner, G., T. Coque, et al. (2008). "Emergence and spread of vancomycin resistance among enterococci in Europe."

Yu, C.-S., C.-W. Cheng, et al. (2014). "CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation." PLoS One **9**(6): e99368.

Yu, C. S., C. J. Lin, et al. (2004). "Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions." Protein Science **13**(5): 1402-1406.

Zarb, P., B. Coignard, et al. (2012). "The European Centre for Disease Prevention and Control (ECDC) pilot point prevalence survey of healthcare-associated infections and antimicrobial use." Euro Surveill **17**(46): 20316.

Zhang, R. and Y. Lin (2009). "DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes." Nucleic acids research **37**(suppl 1): D455-D458.

Zhang, R., H. Y. Ou, et al. (2004). "DEG: a database of essential genes." Nucleic acids research **32**(suppl 1): D271-D272.

Zhang, X., F. L. Paganelli, et al. (2012). "Genome-wide identification of ampicillin resistance determinants in Enterococcus faecium." PLoS Genet **8**(6): e1002804.