OriginalArticle

Mining of novel drug/vaccine targets from the proteome of *Burkholderiacepacia*using computational tools through reverse vaccinology approach.

Vaishnavi Modupalli[1], John Arun[2*]

[1]Research Scholar,Department of BioinformaticsSaveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai,TamilNadu, India, Pincode:602105

[2]Project Guide, Corresponding author,Department of Bioinformatics,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai, TamilNadu, India, Pincode: 602105

Abstract

Aim : The present study anticipates prediction of novel drug/vaccine candidates from the group of uncharacterized proteins of *Burkholderiacepacia*using a series of computational tools. *B.cepacia* is an emerging pulmonary pathogen in immunocompromised patients and in patients with cystic fibrosis (CS). Infections caused by these bacteria are difficult to treat due to antibiotic resistance. So far there are no drugs/vaccines against *B.Cepacia*.**Materials and Methods:** A total of 50 hypothetical proteins were retrieved from NCBI and studied using computational tools for studying their localization, membrane helices, physicochemical properties, virulence factors, signal peptides, antigenicity, and epitopes. These proteins were then subjected to BLAST to find they are matching with any human homologs to avoid autoimmune reactions. **Results**: Out of 50, 3 were found to be highly virulent proteins that possess antigenic properties and they are not human homologs. **Conclusion**: Hence, these proteins could be putative drug targets against *Burkholderiacepacia*infections. Nevertheless, they require further *in-vitro* and *in-vivo* experiments for validation.

Keywords: Immunoinformatics, Reverse Vaccinology, Novel Drug Targets, *Burkholderiacepacia*, Proteome, Virulence Factors.

DOI: 10.47750/pnr.2022.13.S03.016

INTRODUCTION

Burkholderiacepacia, is a gram negative pathogenic bacteria, that often causes pneumonia in immunocompromised individuals with underlying lung diseases. *B.cepacia*causes emerging cases all over the world especially in patients with cystic fibrosis (CF) (Winkelstein et al. 2000). *B.cepacia* causes three major problems for the CF patients (1) Innate multi resistance to antimicrobial agents (2) Person-to-person transmission of epidemic strains through nosocomial or social contacts (3) Cepacia syndrome- a fatal pneumonia associated with septicaemia that occurs in approximately 20% of colonized patients including those with previously mild disease (Coenye and Vandamme 2007)(Shinoy et al. 2013).

Most of the *B.cepacia* strains are resistant to many of antibacterial agents that are commonly used for CF. No vaccines are currently available to treat *B.cepacia* infections separately (Ada 2001). With the advent of genomics, proteomics and bioinformatics techniques it is possible to discover antigens against *B.cepacia*(Shinoy et al. 2013); (Woods and Darie 2019); (Zhang et al. 2011)1; (Heinrich et al. 2017); (López-Macías and Cunningham 2016); (López-Macías and Cunningham 2016; Fulton and Twine 2020)(Puthucheary and Vadivelu 2002). Novel genomic techniques offer opportunities to effectively identify novel drug targets without growing pathogens *in-vitro*. One of such reliable strategies is Reverse Vaccinology, which clearly offers a platform to predict candidate proteins with vaccine potentials using computational tools (Rappuoli et al. 2016). About 750 studies related to proteome mining of novel drug targets against several pathogens have been found in Pubmed in the last 5 years. One such recent study is (Prabhu et al., 2020), where the uncharacterized proteins of the *Serratiamarcescens*were functionally characterized and reported as novel drug targets using computational tools. Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al.

2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021)

Nevertheless, the lacunae identified here is that, even though the genome and proteome of this organism is well studied, nearly one fourth of its proteome is annotated as hypothetical proteins. Hence, the present study aims to predict potential vaccine candidates with their epitopes from the uncharacterized protein pool of *B. cepacia*by applying RV and *immune-informatics* methods. These identified epitopes could be considered as promising candidates for effective protein-based vaccines against *B.cepacia*.

Materials and Methods

The proposed work is done in the Bioinformatics lab, Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, TamilNadu,India. There is no ethical approval as human samples are not involved. For each organism the number of groups is one. The sample size is 50 proteins per group.

The FASTA sequences of fifty uncharacterized proteins of *Burkholderiacepacia* were retrieved from NCBI. The physicochemical properties including the molecular weight, pI, Instability index, Aliphatic index, extinction coefficient and GRAVY of the hypothetical proteins were predicted using ProtParam(https://web.expasy.org/protparam/). VICM pred tool (http://www.imtech.res.in/raghava/vicmpred/) was used for functional classification of hypothetical proteins. Subcellular localization was predicted for the 50 sequences using CELLO2GO(http://cello.life.nctu.edu.tw/cello2go/)(Horton et al. 2007). Virulence and antigenicity properties of the fifty uncharacterized proteins were identified using VaxiJen ver. 2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). The combination of ABCpred, and VaxiJen servers (http://ailab-projects1.ist.psu.edu:8080/bcpred/predict.html) allowed the prediction of overlapping antigenic B cell epitopes (Rammensee, Bachmann, and Stevanovic 2013; Rammensee, Bachmann, and Stevanović 1997) from the uncharacterized proteins which is important for the development of immunodiagnostics (Doytchinova, Guan, and Flower 2006; Shirai et al. 2014).

CTLPred server (http://crdd.osdd.net/raghava/ctlpred/) predicted cytotoxic T cell epitopes from the uncharacterized proteins studied (Bhasin and Raghava 2004; “Website,” n.d.). A total of 2 cytotoxic T cell epitope regions were predicted as antigens by VaxiJen server. Of these 1 antigenic epitope is found in OHS92109.1 and another one found in OHS88591.1. Both the antigenic epitopes are allergens . Characteristics of transmembrane helices in uncharacterized proteins was predicted by TMHMM based on the hidden Markov model and HMMTOP (http://www.enzim.hu/hmmtop/html/submit.html) (Koski 2001); (Tusnady and Simon 2001; “Website,” n.d.). Avoiding interference against human immune mechanisms, these uncharacterized protein sequences are submitted to tBLASTn(https://blast.ncbi.nlm.nih.gov/) for non-human homologs (Heath and Ramakrishnan 2010).

Results

Table 1 shows the hypothetical protein sequences retrieved from NCBI. Table 2 Depicts the physio-chemical properties of the retrieved protein sequences predicted using PROTPARAM tools. Table 3 reveals the functional classification of hypothetical proteins found using the VICMpred tool. Table 4 shows the subcellular localization for all the proteins predicted using the CELLO2GO tool. Table 5 shows the antigenicity characteristics of the proteins revealed using VaxiJen ver. 2.0 tool. Table 6 depicts the presence of epitopes in the proteins studied using BCpred, ABCpred and VaxiJen tools. Table 7 reveals the presence of T- cell epitopes in the uncharacterized protein pool using CTLpred tool. Table 8 shows the number of transmembrane helices found in the 3 putative antigenic proteins analyzed using the HMMTOP tool. Table 9 depicts the tBLASTn results of the 3 putative novel antigenic proteins, confirming if they are homologous to any human proteins. From these results, we can conclude that these proteins are potential candidates for drug designing against *B.cepacia.*

Discussion

Out of 50, 3 were found to be highly virulent proteins that possess antigenic properties and they are not human homologs.Hence, these proteins could be putative drug targets against *Burkholderiacepacia*infections. Nevertheless, they require further *in-vitro* and *in-vivo* experiments for validation. Currently, over 50% of the *B.cepacia*proteins have no annotated function. Characterization of hypothetical proteins may be useful in better

understanding the organism's metabolic pathways, disease progression, drug development, and disease control strategies(Brüssow 2019)(Kaur et al. 2021)(Brüssow 2019). With a complete *B.cepacia*genome sequence (Nene and Kole 2008) and advancement in bioinformatics, it is now possible to identify potential vaccine candidates using reverse vaccinology which reduces the time and cost of designing and identifying vaccine candidates (Rappuoli et al. 2016).
This study utilized several bioinformatics and immunoinformatics tools for identification and characterization of hypothetical proteins of *B.cepacia*for vaccine development. Out of 50 hypothetical proteins studied, 2 (KML08942.1 and KML08444.1) were identified as potential drug targets *B.cepacia(Brüssow 2019)*. For prediction of localization with functional classification we have used a web server CELLO2GO (Yu et al. 2014)The BLAST results of the three selected HPs suggest that they could be used for drug development without causing autoimmunity. *In-silico* characterization of these three putative targets, affirms that they could be novel drug targets with potentials. VaxiJen server identified the potential antigens out of 50 hypothetical proteins. Out of 50, 3 proteins possessed virulence (Garg and Gupta 2008) and antigenic properties . Furthermore, these 3 were subjected to epitope prediction. The B cell epitopes were identified through ABCpred server while CTL **e**pitopes were predicted using CTLPred web server. This is based on the idea that the development of a peptide vaccine largely relies on identifying immunodominantepitopes(Sanchez-Trincado, Gomez-Perosanz, and Reche 2017)(Sanchez-Trincado, Gomez-Perosanz, and Reche 2017) that can induce specific immune responses without the need of involving whole microorganisms. From the two HPs, a number of antigenic B, cytotoxic and helper T cell epitopes were identified which could potentially be used for designing an epitope based vaccine against *B.cepacia*listed in Tables 7 and 8.

Therefore, information generated herein about the characteristics of novel drug targets identified could shed insight into bacterial pathogenesis and can assist in vaccine development. Nonetheless, the major limitation of this study is that the selected drug candidates along with their epitopes should be further validated for their immunogenicity and protective efficacy experimentally if they are to be used for future drug development against *B.cepacia*.

Conclusion

Reverse vaccinology is a promising strategy for the screening and identification of antigenic antigens with potential capacity to elicit cellular and humoral immune responses against *B.cepacia*infection. In this study, two hypothetical proteins were selected through computational methods and verified as potential drug candidates against *B.cepacia*. We therefore recommend further in-depth immunoinformatics and structural biology approaches together with in vitro and in vivo experiments to validate their immunogenicity and protective efficacy to completely decipher the vaccine targets against *B.cepacia.*

Declarations

Conflict of Interest

The authors of this paper declare no conflict of interest.

Author Contribution

Author VM was involved in data collection, data analysis, manuscript writing. Author JA was involved in conceptualization, guidance and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Qbiogen, Chennai
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Ada, Gordon. 2001. "Vaccines and Vaccination." *New England Journal of Medicine*. https://doi.org/10.1056/nejmra011223.

2. Adhinarayanan, Rajesh, AravindhRamakrishnan, GopalKaliyaperumal, Melvinvíctor De Poures, Rajesh Kumar Babu, and DamodharanDillikannan. 2020. "Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
3. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. "Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre–reinforced Epoxy Composite." *Biomass Conversion and Biorefinery*, August. https://doi.org/10.1007/s13399-020-00938-0.
4. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. "Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique." *Biomass Conversion and Biorefinery*, May. https://doi.org/10.1007/s13399-021-01551-5.
5. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik R. Bhagat. 2021. "Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for PhotocatalyticTransesterification of Castor Oil." *Fuel* 304 (November): 121490.
6. Bhasin, Manoj, and G. P. S. Raghava. 2004. "Prediction of CTL Epitopes Using QM, SVM and ANN Techniques." *Vaccine*. https://doi.org/10.1016/j.vaccine.2004.02.005.
7. Brüssow, Harald. 2019. *Hurdles for Phage Therapy (PT) to Become a Reality*. MDPI.
8. Coenye, Tom, and Peter Vandamme. 2007. *Burkholderia: Molecular Microbiology and Genomics*. Taylor & Francis.
9. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. "Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement." *Biomass Conversion and Biorefinery*, February. https://doi.org/10.1007/s13399-021-01312-4.
10. Doytchinova, Irini A., Pingping Guan, and Darren R. Flower. 2006. "EpiJen: A Server for Multistep T Cell Epitope Prediction." *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-7-131.
11. Fulton, Kelly M., and Susan M. Twine. 2020. *Immunoproteomics: Methods and Protocols*. Humana.
12. Garg, Aarti, and Dinesh Gupta. 2008. "VirulentPred: A SVM Based Prediction Method for Virulent Proteins in Bacterial Pathogens." *BMC Bioinformatics* 9 (January): 62.
13. Heath, Lenwood S., and NarenRamakrishnan. 2010. *Problem Solving Handbook in Computational Biology and Bioinformatics*. Springer Science & Business Media.
14. Heinrich, Michael, Elizabeth M. Williamson, Simon Gibbons, Joanne Barnes, and Jose Prieto-Garcia. 2017. *Fundamentals of Pharmacognosy and Phytotherapy E-Book*. Elsevier Health Sciences.
15. Horton, P., K-J Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai. 2007. "WoLF PSORT: Protein Localization Predictor." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkm259.
16. Jayanth, BellappuVenkat, Melvin Victor Depoures, GopalKaliyaperumal, DamodharanDillikannan, DilipsinghJawahar, KumaranPalani, and Ganesha Prasad MeravanigeeShivappa. 2021. "A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
17. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and ShantanuPatil. 2020. "Melatonin Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration." *Process Biochemistry* 99 (December): 36–47.
18. Kaur, Harpreet, Vikram Singh, ManmohitKalia, Balvinder Mohan, and NeelamTaneja. 2021. "Identification and Functional Annotation of Hypothetical Proteins of Uropathogenic Strain CFT073 towards Designing Antimicrobial Drug Targets." *Journal of Biomolecular Structure & Dynamics*, November, 1–12.
19. Koski, T. 2001. *Hidden Markov Models for Bioinformatics*. Springer Science & Business Media.
20. López-Macías, Constantino, and Adam Cunningham. 2016. *How Salmonella Infection Can Inform on Mechanisms of Immune Function and Homeostasis*. Frontiers Media SA.
21. Nene, Vishvanath, and ChittaranjanKole. 2008. *Genome Mapping and Genomics in Animal-Associated Microbes*. Springer Science & Business Media.
22. Puthucheary, S. D., and J. Vadivelu. 2002. *Human Melioidosis*. NUS Press.
23. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. "Collective Influence of 1-Decanol Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends." *Fuel* 277 (October): 118166.
24. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. "Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications." *Fuel* 278 (October): 118315.
25. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. "Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour." *Biomass Conversion and Biorefinery*, March. https://doi.org/10.1007/s13399-021-01427-8.
26. Rammensee, Hans-Georg, Jutta Bachmann, and Stefan Stevanović. 1997. "The MHC Genes." *MHC Ligands and Peptide Motifs*. https://doi.org/10.1007/978-3-662-22162-4_2.
27. Rammensee, Hans-Georg, Jutta Bachmann, and Stefan Stevanovic. 2013. *MHC Ligands and Peptide Motifs*. Springer Science & Business Media.
28. Rappuoli, Rino, Matthew J. Bottomley, Ugo D'Oro, Oretta Finco, and Ennio De Gregorio. 2016. "Reverse Vaccinology 2.0: Human Immunology Instructs Vaccine Antigen Design." *Journal of Experimental Medicine*. https://doi.org/10.1084/jem.20151960.
29. Sanchez-Trincado, Jose L., Marta Gomez-Perosanz, and Pedro A. Reche. 2017. "Fundamentals and Methods for T- and B-Cell Epitope Prediction." *Journal of Immunology Research*. https://doi.org/10.1155/2017/2680160.
30. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). https://doi.org/10.1002/ep.13696.
31. Shanmugam, Rajasekaran, DamodharanDillikannan, GopalKaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 43 (23): 3064–81.
32. Shinoy, Minu, Ruth Dennehy, Lorraine Coleman, Stephen Carberry, Kirsten Schaffer, Máire Callaghan, Sean Doyle, and Siobhán McClean. 2013. "Immunoproteomic Analysis of Proteins Expressed by Two Related Pathogens, Burkholderia Multivorans and Burkholderia Cenocepacia, during Human Infection." *PLoS ONE*. https://doi.org/10.1371/journal.pone.0080796.

33. Shirai, Hiroki, Catherine Prades, Randi Vita, Paolo Marcatili, Bojana Popovic, Jianqing Xu, John P. Overington, et al. 2014. “Antibody Informatics for Drug Discovery.” *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. https://doi.org/10.1016/j.bbapap.2014.07.006.
34. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. “Pretreatment and Process Optimization of Bioethanol Production from Spent Biomass of GanodermaLucidum Using Saccharomyces Cerevisiae.” *Fuel* 306 (December): 121680.
35. Tusnady, G. E., and I. Simon. 2001. “The HMMTOP Transmembrane Topology Prediction Server.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/17.9.849.
36. “Website.” n.d. Bhasin, Manoj, and G. P. S. Raghava. 2004. “Prediction of CTL Epitopes Using QM, SVM and ANN Techniques.” Vaccine. https://doi.org/10.1016/j.vaccine.2004.02.005.
37. ———. n.d. Tusnady, G. E., and I. Simon. 2001. “The HMMTOP Transmembrane Topology Prediction Server.” Bioinformatics. https://doi.org/10.1093/bioinformatics/17.9.849.
38. Winkelstein, Jerry A., Mary C. Marino, Richard B. Johnston, John Boyle, John Curnutte, John I. Gallin, Harry L. Malech, et al. 2000. “Chronic Granulomatous Disease: Report on a National Registry of 368 Patients.” *Medicine*. https://doi.org/10.1097/00005792-200005000-00003.
39. Woods, Alisa G., and Costel C. Darie. 2019. *Advancements of Mass Spectrometry in Biomedical Research*. Springer.
40. Yu, Chin-Sheng, Chih-Wen Cheng, Wen-Chi Su, Kuei-Chung Chang, Shao-Wei Huang, Jenn-Kang Hwang, and Chih-Hao Lu. 2014. “CELLO2GO: A Web Server for Protein subCELlularLOcalization Prediction with Functional Gene Ontology Annotation.” *PloS One* 9 (6): e99368.
41. Zhang, Wei, Guangjin Liu, Fang Tang, Jing Shao, Yan Lu, YinliBao, Huochun Yao, and Chengping Lu. 2011. “Pre-Absorbed Immunoproteomics: A Novel Method for the Detection of Streptococcus Suis Surface Proteins.” *PloS One* 6 (6): e21234.

List of Tables

Table 1. Retrieval of hypothetical protein sequences from NCBI

S.No	Accession Number	Protein
1.	SES73574.1	Hypothetical
2.	KML 14202.1	Hypothetical
3.	KML06707.1	Hypothetical
4.	KML 21601.1	Hypothetical
5.	KML20291.1	Hypothetical
6.	KML19256.1	Hypothetical
7.	KML15370.1	Hypothetical
8.	KML14433.1	Hypothetical
9.	KMK97725.1	Hypothetical
10.	KML22375.1	Hypothetical
11.	KML22249.1	Hypothetical
12.	KML21985.1	Hypothetical
13.	KML21650.1	Hypothetical
14.	KML21640.1	Hypothetical
15.	KML21092.1	Hypothetical
16.	KML21022.1	Hypothetical
17.	KML20341.1	Hypothetical

18.	KML19501.1	Hypothetical
19.	KML18866.1	Hypothetical
20.	KML18863.1	Hypothetical
21.	KML17458.1	Hypothetical
22.	KML17380.1	Hypothetical
23.	KML17359.1	Hypothetical
24.	KML17040.1	Hypothetical
25.	KML16936.1	Hypothetical
26.	KML16332.1	Hypothetical
27.	KML15872.1	Hypothetical
28.	KML15222.1	Hypothetical
29.	KML14666.1	Hypothetical
30.	KML14459.1	Hypothetical
31.	KML14250.1	Hypothetical
32.	KML12027.1	Hypothetical
33.	KML12003.1	Hypothetical
34.	KML10949.1	Hypothetical
35.	KML10146.1	Hypothetical
36.	KML09187.1	Hypothetical
37.	KML08942.1	Hypothetical
38.	KML08715.1	Hypothetical
39.	KML08444.1	Hypothetical
40.	KML03670.1	Hypothetical
41.	KML02576.1	Hypothetical
42.	KML02316.1	Hypothetical
43.	KML02308.1	Hypothetical
44.	KML01017.1	Hypothetical

45.	KML00665.1	Hypothetical
46.	KML00027.1	Hypothetical
47.	KMK99432.1	Hypothetical
48.	KMK98973.1	Hypothetical
49.	KMK98856.1	Hypothetical
50.	KMK92235.1	Hypothetical

Table 2. Physio-chemical properties of the retrieved protein sequences using PROTPARAM

S.NO	**Protein ID**	**Molecular Weight**	**P_I**	**GRAVY**	**Instability Index**	**Aliphatic Index**	**Extinction Coefficient**
1.	SES73574.1	58139.83	5.68	-0.046	28.45	93.10	64315
2.	KML14202.1	13527.65	6.05	0.127	33.15	98.39	9970
3.	KML06707.1	47101.74	8.42	-0.004	38.34	96.55	51700
4.	KML21601.1	51296.28	6.77	-0.033	38.95	100.46	38390
5.	KML20291.1	84429.68	6.05	-0.058	42.82	88.78	93320
6.	KML19256.1	50331.72	8.77	-0.145	36.60	83.85	90995
7.	KML15370.1	89562.78	8.70	-0.542	53.25	60.85	160895
8.	KML14433.1	46515.66	5.89	-0.087	27.45	76.27	83770
9.	KMK97725.1	23011.43	7.18	-0.050	31.63	94.98	24980
10.	KML22375.1	84358.49	5.94	-0.102	31.91	91.88	76375

11.	KML22249.1	49005.76	6.11	0.133	30.59	93.15	32110
12.	KML21985.1	40540.24	6.72	0.012	41.50	95.84	41940
13.	KML21650.1	47225.76	5.77	0.005	33.68	91.61	55850
14.	KML21640.1	35292.46	10.29	-0.065	62.15	108.00	9970
15.	KML21092.1	28394.75	9.62	-0.228	44.22	98.02	17085
16.	KML21022.1	29557.62	10.22	-0.379	49.41	74.44	42190
17.	KML20341.1	17101.49	11.70	-0.257	42.49	72.47	16625
18.	KML19501.1	8632.51	10.16	1.114	17.78	143.81	13980
19.	KML18866.1	29253.01	6.31	-0.350	27.27	81.45	64985
20.	KML18863.1	21370.75	5.47	-0.615	29.63	69.37	55920
21.	KML17458.1	13404.75	10.79	-0.009	42.05	102.35	16500
22.	KML17380.1	21251.15	6.10	-0.135	42.46	92.67	23045
23.	KML17359.1	36753.72	5.58	-0.269	39.08	83.12	38305
24.	KML17040.1	26055.04	11.16	-0.164	55.85	86.52	25230
25.	KML16936.1	37571.95	9.88	-0.018	45.08	91.26	52940

26.	KML16332.1	17592.87	6.36	-0.035	58.40	94.70	15470
27.	KML15872.1	15139.07	9.03	0.187	24.58	93.33	4595
28.	KML15222.1	57929.46	6.02	0.220	37.62	100.94	38640
29.	KML14666.1	35113.71	6.61	-0.517	48.95	72.00	76110
30.	KML14459.1	48956.91	5.71	-0.117	34.10	90.11	55725
31.	KML14250.1	57545.66	5.57	-0.258	37.56	80.75	54235
32.	KML12027.1	55231.92	5.30	0.202	33.99	98.27	48610
33.	KML12003.1	47761.38	6.19	-0.025	33.96	94.14	29715
34.	KML10949.1	59047.43	5.81	0.055	26.86	96.72	41745
35.	KML10146.1	49975.41	5.88	-0.085	37.69	89.28	69690
36.	KML09187.1	36389.61	7.91	-0.119	41.98	87.20	54555
37.	KML08942.1	31919.53	6.86	0.229	24.26	76.84	42775
38.	KML08715.1	22588.71	6.75	-0.038	42.36	95.89	12490
39.	KML08444.1	11362.01	5.31	-0.245	31.74	88.82	2980
40.	KML03670.1	66831.96	6.59	-0.448	36.71	89.46	64290

41.	KML02576.1	18530.11	7.64	-0.013	35.53	86.96	16640
42.	KML02316.1	8634.79	4.10	0.180	58.11	102.25	4470
43.	KML02308.1	16007.01	6.89	-0.534	46.69	62.14	27180
44.	KML01017.1	36466.21	5.67	0.865	19.63	124.94	43430
45.	KML00665.1	41895.55	8.71	-0.021	27.78	93.04	33585
46.	KML00027.1	34125.53	4.98	-0.379	40.68	70.37	82195
47.	KMK99432.1	29118.13	8.79	-0.270	34.15	86.04	37360
48.	KMK98973.1	7485.85	7.87	0.444	34.14	113.38	1615
49.	KMK98856.1	24128.87	10.18	0.294	28.06	98.68	16625
50.	KMK92235.1	50598.57	6.09	-0.022	34.23	92.13	43235

Table 3. Functional classification of hypothetical proteins using VICMpred.

S.No.	Accession Number	Functional Class	Score
1.	SES73574.1	Information and storage	1.947109
2.	KML14202.1	Cellular Process	3.2760795
3.	KML06707.1	Information and storage	-2.8863021
4.	KML21601.1	Metabolism Molecule	28.163698
5.	KML20291.1	Information and storage	12.64434
6.	KML19256.1	Information and storage	-3.0464891

7.	KML15370.1	Cellular Process	3.6625241
8.	KML14433.1	Cellular Process	2.1738985
9.	KMK97725.1	Cellular Process	1.7553734
10.	KML22375.1	Metabolism Molecule	6.2803919
11.	KML22249.1	Information and storage	-2.740058
12.	KML21985.1	Metabolism Molecule	-0.29996911
13.	KML21650.1	Virulence factors	-1.5581426
14.	KML21640.1	Cellular Process	1.820983
15.	KML21092.1	Cellular Process	1.675772
16.	KML21022.1	Cellular Process	3.3126758
17.	KML20341.1	Cellular Process	4.1958572
18.	KML19501.1	Metabolism Molecule	0.77439422
19.	KML18866.1	Cellular Process	0.40367169
20.	KML18863.1	Cellular Process	1.2318979
21.	KML17458.1	Cellular Process	2.802308
22.	KML17380.1	Cellular Process	2.7047251
23.	KML17359.1	Cellular Process	1.3048138
24.	KML17040.1	Cellular Process	2.7097199
25.	KML16936.1	Cellular Process	1.2581723
26.	KML16332.1	Metabolism Molecule	1.188392
27.	KML15872.1	Cellular Process	2.1691374
28.	KML15222.1	Information and storage	1.621023
29.	KML14666.1	Metabolism Molecule	0.93437648
30.	KML14459.1	Cellular Process	1.8255092
31.	KML14250.1	Cellular Process	6.1602703

32.	KML12027.1	Cellular Process	-0.24186409
33.	KML12003.1	Metabolism Molecule	5.8478392
34.	KML10949.1	Metabolism Molecule	-4.2839721
35.	KML10146.1	Metabolism Molecule	-0.72723536
36.	KML09187.1	Metabolism Molecule	0.96349173
37.	KML08942.1	Virulence factors	1.4292084
38.	KML08715.1	Cellular Process	4.2513801
39.	KML08444.1	Virulence factors	035750887
40.	KML03670.1	Information and storage	-5.3803569
41.	KML02576.1	Cellular Process	1.7964097
42.	KML02316.1	Cellular Process	2.0961271
43.	KML02308.1	Metabolism Molecule	3.3396351
44.	KML01017.1	Metabolism Molecule	6.5564375
45.	KML00665.1	Metabolism Molecule	-3.6310361
46.	KML00027.1	Cellular Process	2.088157
47.	KMK99432.1	Cellular Process	1.101841
48.	KMK98973.1	Cellular Process	0.49633792
49.	KMK98856.1	Cellular Process	1.117195
50.	KMK92235.1	Metabolism Molecule	1.5638527

Table 4. Subcellular localization prediction for the proteins using CELLO2GO

S.No	Accession Number	Localization
1.	SES73574.1	Cytoplasmic
2.	KML14202.1	Cytoplasmic
3.	KML06707.1	Cytoplasmic
4.	KML21601.1	Outer Membrane

5.	KML20291.1	Cytoplasmic
6.	KML19256.1	Outer Membrane
7.	KML15370.1	Inner Membrane
8.	KML14433.1	Extracellular
9.	KMK97725.1	Cytoplasmic
10.	KML22375.1	Cytoplasmic
11.	KML22249.1	Cytoplasmic
12.	KML21985.1	Cytoplasmic
13.	KML21650.1	Cytoplasmic
14.	KML21640.1	Cytoplasmic
15.	KML21092.1	Cytoplasmic
16.	KML21022.1	Outer membrane
17.	KML20341.1	Cytoplasmic
18.	KML19501.1	Inner Membrane
19.	KML18866.1	Cytoplasmic
20.	KML18863.1	Periplasmic
21	KML17458.1	Cytoplasmic
22	KML17380.1	Cytoplasmic
23	KML17359.1	Cytoplasmic
24	KML17040.1	Cytoplasmic
25	KML16936.1	Cytoplasmic
26	KML16332.1	Cytoplasmic
27	KML15872.1	Extracellular
28	KML15222.1	Cytoplasmic
29	KML14666.1	Cytoplasmic
30	KML14459.1	Cytoplasmic
31	KML14250.1	Cytoplasmic
32	KML12027.1	Cytoplasmic
33	KML12003.1	Cytoplasmic
34	KML10949.1	Cytoplasmic

35	KML10146.1	Cytoplasmic
36	KML09187.1	Cytoplasmic
37	KML08942.1	Extracellular
38	KML08715.1	Inner Membrane
39	KML08444.1	Cytoplasmic
40	KML03670.1	Periplasmic
41	KML02576.1	Cytoplasmic
42	KML02316.1	Cytoplasmic
43	KML02308.1	Cytoplasmic
44	KML01017.1	Inner Membrane
45	KML00665.1	Inner Membrane
46	KML00027.1	Cytoplasmic
47	KMK99432.1	Cytoplasmic
48	KMK98973.1	Inner Membrane
49	KMK98856.1	Cytoplasmic
50	KMK92235.1	Cytoplasmic

Table 5. Evaluating the antigenicity of the proteins using VaxiJen ver. 2.0

S.No	Accession Number	Antigenicity score
1.	SES73574.1	0.4987
2.	KML14202.1	0.5441
3.	KML06707.1	0.4927
4.	KML21601.1	0.5338
5.	KML20291.1	0.4560
6.	KML19256.1	0.5761
7.	KML15370.1	0.5378
8.	KML14433.1	0.5811
9.	KMK97725.1	0.5565
10.	KML22375.1	0.4380

11.	KML22249.1	0.3752
12.	KML21985.1	0.4014
13.	KML21650.1	0.4109
14.	KML21640.1	0.4706
15.	KML21092.1	0.3634
16.	KML21022.1	0.4893
17.	KML20341.1	0.5647
18.	KML19501.1	0.1686
19.	KML18866.1	0.3392
20.	KML18863.1	0.4082
21.	KML17458.1	0.4094
22.	KML17380.1	0.4800
23.	KML17359.1	0.3381
24.	KML17040.1	0.2480
25.	KML16936.1	0.4212
26.	KML16332.1	0.5503
27.	KML15872.1	0.9457
28.	KML15222.1	0.4061
29.	KML14666.1	0.2096
30	KML14459.1	0.3882
31.	KML14250.1	0.4733
32.	KML12027.1	0.4169
33.	KML12003.1	0.4670
34.	KML10949.1	0.4746
35.	KML10146.1	0.5264
36.	KML09187.1	0.3529

37.	KML08942.1	0.9120
38.	KML08715.1	0.5860
39.	KML08444.1	0.6263
40.	KML03670.1	0.4363
41.	KML02576.1	0.4306
42.	KML02316.1	0.2902
43.	KML02308.1	0.4965
44.	KML01017.1	0.3873
45.	KML00665.1	0.4126
46.	KML00027.1	0.4718
47.	KMK99432.1	0.4421
48.	KMK98973.1	0.3237
49.	KMK98856.1	0.3934
50.	KMK92235.1	0.2981

Table 6. Epitope prediction using ABCpred and VaxiJen

Accession Number	**B cell epitope**	**ABC Pred**	**Vaxijen**	**Allergen FP**
KML08942.1	ASLTTSSGTFPFGATA			
	AHSIRWGDGTGGSATV	0.93	0.3495	Non-allergen
	GGTIGYGLYTDAAHSI			
	NAPVTATGTVSVTCTW			
	AVPAQTTPAPGSYSDT	0.91	2.1469	Allergen
	QITGNQPTVPTTGNST			
	TVTVYGQITGNQPTVP	0.91	0.7998	Allergen
	SVYSGTTPTSVTLAKP			
	GGHSVAWGSVYSGTTP	0.90	0.9789	Non-allergen
	NSTTTYSQQFGGNTTS			
	SCTATASTVSFGSVSP	0.90	0.7856	Non-allergen
	DGTGGSATVTGVGTGT			
	ALPRTEINVLRKHLDE	0.90	1.1880	Non-allergen
	SGTLAATGTITAQCTN			
	TGTITAQCTNGDAWKI	0.89	0.5213	Non-allergen
	TFPFGATADVINNCNI			
	GTVSVTCTWSAVSLVP	0.89	0.5424	Allergen
	AMQYDLYLDGGHSVAW			
	PTHETICALPRTEINV	0.89	0.5703	Non-allergen
	SQVVTVYGAVPAQTTP			
	GNTTSINTGFYLLVAP	0.88	1.0838	Allergen

	AFAAWCGVPHHARAES GTASILSGTLAATGTI NLGGTSPRSLVNGTNA DVTARQMKRSGGGGTI SSGGSGDVTARQMKRS AWKIALSSGGSGDVTA RSLVNGTNAMQYDLYL MKRSGGGGTIGYGLYT STVSFGSVSPISNAPV GLYTDAAHSIRWGDGT CTWSAVSLVPNVLVCL PAPGSYSDTITATISF IACEPACIAMRLALLL NCNINATNVSFGTASI TGFYLLVAPTCASLTT NVLRKHLDERIACEPA SVTLAKPALGTSASAT	0.87 0.86 0.86 0.85 0.84 0.84 0.82 0.82 0.80 0.80 0.80 0.79 0.79 0.79 0.78 0.78 0.78 0.77 0.76 0.74 0.70 0.69 0.69 0.68 0.63 0.61 0.61 0.60	0.9222 2.6435 0.2232 0.7065 1.5586 0.4000 0.6012 0.8593 0.1008 0.4196 1.1890 0.4423 0.3239 1.1432 2.0844 2.3773 2.1397 0.7043 1.9053 0.1950 0.5922 0.4634 0.9553 0.9657 0.8637 0.7397 0.9542 1.0588	Allergen Non-allergen Non-allergen Allergen Allergen Non-allergen Allergen Non-allergen Non-allergen Non-allergen Non-allergen Non-allergen Non-allergen Allergen Non-allergen Non-allergen Non-allergen Non-allergen Non-allergen Allergen Allergen Non-allergen Non-allergen Non-allergen Non-allergen Non-allergen Non-allergen Non-allergen

KML08444.1	SGGNYTGLTITVRATS	0.92	1.1382	Non-allergen
	ALPRTEINVLRKHLDE	0.86	0.2232	Non-allergen
	TKTIEVTGAIDTRKES	0.85	1.0495	Non-allergen
	PIKIMGKAHPEFKDTI	0.84	0.5598	Non-allergen
	RATSQEQLDNIYRALT	0.84	-0.0125	Non-allergen
	TGAIDTRKESLLEFPC	0.83	0.7122	Non-allergen
	EFPCDFPIKIMGKAHP	0.82	0.9264	Non-allergen
	DERIACEPACIAMSEP	0.81	0.7093	Non-allergen
	PTHETICALPRTEINV	0.80	0.1008	Non-allergen
	GLTITVRATSQEQLDN	0.79	0.9429	Non-allergen
	LDNIYRALTGHPMVKV	0.78	0.0175	Non-allergen
	VEKIEERASSGGNYTG	0.76	1.3776	Non-allergen
	KVVAVHDNEIDVEKIE	0.76	0.9199	Allergen
	ACIAMSEPTKTIEVTG	0.73	0.7850	Non-allergen
	HPEFKDTIFKVVAVHD	0.70	0.4093	Non-allergen
	NVLRKHLDERIACEPA	0.61	0.9542	Non-allergen

Table 7. T- cell epitope prediction of uncharacterized proteins using CTLpred

Accession Number	Peptide	Start	TC pred score	Antigenicity score	Allergenicity
KML08942.1	AKPALGTSA	155	1.000	1.1657	Allergen
	TVSFGSVSP	70	0.990	0.6200	Allergen
	SPISNAPVT	77	0.990	0.3530	Allergen
KML08444.1	PCDFPIKIM	63	0.990	1.5955	Allergen
	SGGNYTGLT	104	0.990	1.3661	Allergen
	LDNIYRALT	123	0.990	-1.2344	Allergen

Table 8. Transmembrane helices prediction for the 2 putative antigenic proteins using HMMTOP

Accession Number	TMHMM Score	HMMTOP Score
KML08942.1	0	1
KML08444.1	0	0

Table 9. Screening of the 2 putative antigenic proteins using tBLASTn for off targets

Accession Number	Non-human Homologous
KML08942.1	No significant similarity found
KML08444.1	No significant similarity found