Revista Argentina de Clínica Psicológica
2020, Vol. XXIX, N°1, 150-160
DOI: 10.53555/03276716.2020.20

Pathogenic Protein Identification and Localization Prediction in *Pseudomonas fuscovaginae*: A Study on Sheath Brown Rot in Rice.

Dharmendra Kashyap[1*], Aafreen Khan[2] and Bharti Lahare[3]

Abstract

P. fuscovaginae poses a concern to rice cropping systems since it is readily found on asymptomatic seeds.The fluorescent, Gram-negative *P. fuscovaginae* (Pfv) belongs to the Gamma proteobacteria class of bacteria. *P. fuscovaginae* LMG 2158's genome is a single circular chromosome with a 6592354 bp total GC content. The bacterial disease known as brown sheath rot affects rice fields almost everywhere from sea level to 1200–1700 m above sea level, low temperature (20–22°C), and high humidity, in both temperate and tropical climates. Symptoms of *P. fuscovaginae* can be found at various crop cycle stages. The most typical signs include unfilled grains, kernel spotting, poor panicle emergence, sterility, dark necrotic spots on the flag leaf's sheath that vary in length, or as necrosis spreads on the sheaths. The disease's severity varies by region, with some strains showing biochemical and physiological variability.

Pathogenic proteins plays a vital role in host-pathogen interaction. To computationally identify such pathogenic proteins, MP3 software were used to analyse the proteome in the Fasta file that was downloaded from the NCBI. We analysed the 5778 proteins using the MP3 software, The predictions found 670 proteins as pathogenic by the SVM approach, 441 proteins by the HMM method, and 880 proteins by the HYBRID method. We selected the top 500 proteins for our further investigation. Nine distinct sub-cellular localisation prediction servers are used to forecast the sub-cellular locations of 500 proteins. BUSCA, Cello v.2.5, Cell-PLoc v.2.0, PSORTb v.3.0, PSL-Pred, SLP-Local, ngLOC, gram-LocEN, and CELLO2GO are the servers were employed. The results of the servers used to prediction of the sub-cellular location of proteins vary. Our findings were roughly categorised into six groups: extracellular space proteins, plasma membrane proteins, and cytoplasmic proteins and a combination of them. Out of 500 proteins, we were able to predict 32 proteins as exclusively Extra-cellular, 370 as exclusively Plasma membrane associated, 62 as exclusively Cytoplasmic proteins, 7 as both Extra-cellular and Plasma membrane associated, 28 as Plasma membrane associated and Cytoplasmic protein and 01 with Extra-cellular and Cytoplasmic in nature.
VirulentPred and VICM-pred were utilised for 500 proteins for further virulence prediction. By using VirulentPred, 337 out of 500 proteins were identified to be virulent, by VICM 52 to be virulent. We identified 107 proteins as pathogenic out of 500 proteins based on predictions from 3 servers/tools by vote of majority. We also identified the sub-cellular localisation of such proteins. 80 of the proteins were projected to be linked with the plasma membrane, 18 to be extracellular, and 19 to be cytoplasmic. A few proteins were predicted to have multiple sites.

KEYWORDS*: -**P. fuscovaginae*** LMG - 2158, Oryza sativa L. , MP3, PSI BLAST, HMM, BUSCA.

Introduction

Rice (*Oryza sativa* L.) is a fundamental global food crop, sustaining over half of the world's population. The global rice production stands at approximately 470.63 million metric tons, cultivated over 157.46 million hectares, yielding an average of 4.46 metric tons per hectare. India is a major rice producer, ranking just behind China in both production and area under cultivation. Rice is a staple for over 70% of the Indian population, with a cultivation area of 42.75 million hectares and a productivity of 3.61 metric tons per hectare. The genus Oryza encompasses around 24 species, with *Oryza sativa* being the most widely consumed. This species includes three primary cultivars: *japonica*, *indica*, and *javanica*. *Oryza sativa* L. is cultivated across various climates, with *Oryza sativa* L. ssp. *japonica* thriving in subtropical and

[1*]Assistant Professor, Department of Microbiology & Bioinformatics, U.T.D. of Atal Bihari Vajpayee University, Bilaspur, Chhattisgarh, India. Pin 495009
[2,3]M.Sc. in Microbiology and Bioinformatics, Department of Microbiology & Bioinformatics, U.T.D. of Atal Bihari Vajpayee University, Bilaspur, Chhattisgarh, India. Pin 495009

***Corresponding Author Department of Microbiology & Bioinformatics,U.T.D. of Atal Bihari Vajpayee University, Bilaspur, Chhattisgarh, India. Pin 495009, kashyapdk97@gmail.com**

temperate regions. It prefers heavy, flood-tolerant soils. Rice faces numerous diseases caused by fungi, bacteria, viruses, and nematodes, impacting both its quality and productivity.[1]

Rice has historically supported dense populations and state systems in Asia. The domestication and spread of O. *indica* and O. *japonica* in India offer a compelling story of agricultural evolution. *O. rufipogon* and *Oryza nivara*, both native to India, were present since the Pleistocene. Archaeological evidence from the Ganges River valley indicates the presence of rice as early as 9000 B.C. It is believed that India was an independent center of rice cultivation, with *indica* rice likely originating in the Ganges plains. By 5000 B.C., rice had become a staple food in the region. [3,4]

In the 2014-15 fiscal year, rice cultivation in India covered 4.3855 million hectares. The leading states in rice cultivation were Uttar Pradesh, West Bengal, Orissa, Chhattisgarh, and Telangana, with areas of 58.6, 53.8, 41.8, 38.09, and 38.09 lakh hectares, respectively. The total rice production was 104.798 million tonnes, with West Bengal, Uttar Pradesh, Telangana, Punjab, and Orissa being the top producers. The average yield was 2390 kg/hectare.[1]

Rice was the first crop plant to have its genome sequenced. The *Oryza sativa japonica* genome (RefSeq ID NC_029256.1) contains 374.423 megabase pairs (Mbp), with 33,185 genes and 41,070 proteins. This data is accessible through NCBI's Bioproject ID PRJNA13141 and Genome Assembly ID 22512.[2]

Rice, being a semi-aquatic plant, is susceptible to various pathogens, including fungi, bacteria, nematodes, and viruses. These diseases can drastically reduce both the quality and quantity of rice yield. Bacterial infections are particularly problematic due to their rapid multiplication and dissemination. Such diseases can lead to significant economic losses, with potential yield reductions exceeding 60%. *Pseudomonas fuscovaginae* is a seed-borne, Gram-negative bacterium responsible for sheath brown rot disease and grain discoloration in rice. First identified in Hokkaido, Japan, in 1976, it exhibits both epiphytic and endophytic characteristics. Although it is known to colonize rice seeds, its endophytic role remains unclear. Sheath brown rot, caused by *Pseudomonas fuscovaginae*, affects rice in various climates, including high altitudes (1200-1700 m) and low temperatures (20-22°C). Symptoms include brown necrotic spots on sheaths, poor panicle emergence, grain staining, and sterility. The disease's severity varies by region, with some strains showing biochemical and physiological.[3] Sheath brown rot, caused by *Pseudomonas fuscovaginae*, affects rice in various climates, including high altitudes (1200-1700 m) and low temperatures (20-22°C). Symptoms include brown necrotic spots on sheaths, poor panicle emergence, grain staining, and sterility. The disease's severity varies by region, with some strains showing biochemical and physiological variability. [5,6,7] The pathogen is seed-borne and can be transmitted through infected seeds, leading to severe crop losses in various regions, including Asia, South America, and Africa. Disease Mechanism and Symptoms *Pseudomonas fuscovaginae* colonizes the rice sheath, resulting in necrotic lesions that can appear brown to reddish-brown. If environmental conditions are conducive, these lesions may extend towards the panicle, causing seed discoloration and potentially leading to grain sterility. The pathogen's ability to survive as an epiphyte on seed surfaces and endophytically in plant tissues (roots, stems, and leaves) suggests a versatile metabolic capability, enabling it to thrive in diverse environments. Impact on Rice Production The disease caused by Pfv can lead to significant yield losses. Reports indicate that under severe conditions, losses can reach up to 72.2% in Indonesia, with total yield losses observed in Madagascar during extreme outbreaks. Symptoms typically manifest as yellow to brown discoloration on the leaf sheath, progressing to dark, water-soaked lesions that can lead to the death of the entire leaf sheath and withering of the.[9,10,11]

***Pseudomonas fuscovaginae* LMG 2158**

Pseudomonas fuscovaginae LMG 2158 is a Gram-negative fluorescent *Pseudomonas* that was first identified as a rice pathogen in Japan in 1976. It is primarily associated with causing bacterial sheath brown rot disease in rice (*Oryza sativa*) and has also been implicated in similar diseases in other cereals, including maize (*Zea mays*), sorghum (*Sorghum bicolor*), and wheat (*Triticum aestivum*). This pathogen is one of the 18 recognised plant-pathogenic species within the *Pseudomonas* genus that are oxidase-positive.

Pathogenic Mechanism

P. fuscovaginae produces three key phytotoxic metabolites: syringotoxin, fuscopeptin A (FP-A), and fuscopeptin B (FP-B). These metabolites are crucial in the development of disease symptoms, particularly the necrotic lesions observed on the rice sheath. The pathogen's virulence and the specific mechanisms by which it induces disease symptoms remain largely unexplored, with few studies focusing on its virulence factors and subcellular localisation.[12]

Disease Symptoms and Effects

The disease manifests through typical symptoms such as necrotic lesions on the sheath, which can lead to a reduction in panicle emission and grain filling. The presence of these lesions can significantly impact crop yield, highlighting the importance of understanding the pathogen's biology and the mechanisms behind its

pathogenicity. Rice sheath rot disease presents symptoms at all growth stages of rice plants. At the seedling stage, the symptoms begin with yellow to brown discolouration on the lower leaf sheath, which progresses to grey-brown to dark-brown, ultimately causing the infected seedlings to rot and die. In mature rice plants, symptoms of *Pseudomonas fuscovaginae* (Pfv) can be observed on flag leaf sheaths, other leaf sheaths, and panicles. Under severe infection conditions, the entire leaf sheath becomes necrotic and desiccates. Emerging panicle spikelets may exhibit dis-colouration, sterility, or be symptom less, except for small brown spots. The disease is characterised by rotting and discolouration of the sheath, leading to chaffiness and sterility of the resulting grains. Crop intensification practices, such as increased plant density, high rates of nitrogen fertilisers, and the use of semi-dwarf and photoperiod-insensitive cultivars, favour the susceptibility of rice to sheath brown rot. Rice sheath rot is caused by various fungal and bacterial pathogens. The major pathogens associated with this disease include fungi such as *Sarocladium oryzae* and *Fusarium* species belonging to the *Fusarium fujikuroi* complex, as well as the bacterial pathogen *Pseudomonas fuscovaginae*.[9]

Virulence factors

Virulence traditionally refers to the extent of host mortality caused by a pathogen. However, this definition may not fully encompass all aspects of virulence. A model of virulence evolution focuses on the traits contributing to pathogen fitness, which can vary across different pathogens and host interactions. Pathogen secrets various kind of proteins to start the process of colonization of its host. The process of secretion starts with accumulation of certain type of proteins around organelles of the bacterium. These proteins are specifically up-regulated during the process of host colonization and infection. In Gram-negative bacteria, proteins can localize to the cytoplasm, periplasm, inner membrane, outer membrane, or extracellular space Understanding a protein's sub-cellular localization is crucial for its functional characterization and tracing its relationship with the process of infection and colonization. [7,8]
With the advancements in the databases, algorithmic and computational capacity of computers it is possible to trace the sub-cellular localization of proteins with virulent nature. Keeping this in consideration the present research work has been undertaken to identify the sub-cellular localization prediction of proteins potentially involved in disease processes of bacterium *Pseudomonas fuscovaginae*, a potent pathogen of rice and causal organism of Sheath brown rot disease of Rice *(Oryza sativa* L.**)**

Materials and methods

Sequence

The genome size of *Pseudomonas fuscovaginae* pathovar LMG 2158 is 6,592,354 base pairs, including 5869 genes and 5778 proteins with estimated sequences based on bioproject PRJNA224116.The taxonomy ID is 50340, and the reference number is GCF_900108595.1.I obtained the fasta file from NCBI that contained protein sequences.[2]

For Virulence factor prediction

MP3 software

It is a software program that predicts pathogenic proteins from both genomic and transcriptome data. This software was used to forecast a protein's pathogenicity. Two methods are used by MP3 software to operate: SVM and HMM. On Linux operating systems, we will utilise the software's standalone version to identify the pathogenic proteins in the provided sequences. The Fasta file downloaded from NCBI will be analyzed using MP3 software on Linux operating system. We use the Threshold value of 0.9 for the prediction of the proteins because we want to have only the best proteins full fulling the criteria of pathogenicity. The results are divided into three categories: the HMM Result lists proteins that the HMM module has classified as pathogenic or non-pathogenic; the SVM Result lists proteins that the SVM module has classified similarly, along with their prediction values; and the Hybrid Result lists proteins that the Hybrid module has classified, along with information on the SVM prediction value, the kind of domain that is present, and the final classification. Based on their SVM results, we will choose the top 500 proteins for more investigation.[13]

Virulent-pred server

A approach for predicting virulent proteins in bacteria is called VirulentPred. The bi-layer cascade Support Vector Machine (SVM) serves as its foundation. By using the Position-Specific Iterated BLAST (PSI-BLAST) generated Position Specific Scoring Matrix (PSSM), the first layer SVM classifiers were trained and optimised with various individual protein sequence features like amino acid composition, dipeptide composition, higher order dipeptide composition, and remote evolutionary relationships. To train the second layer SVM model, the five best modules built in the first layer produced binary scores that were sent to the second layer. The pathogenic proteins in the provided sequences was predicted using an online service. To forecast pathogenic proteins, the VirulentPred server makes use of five separate modules: PSI-BLAST-generated PSSM profiles, higher-order dipeptide composition, amino acid composition, similarity-based using PSI-BLAST, and dipeptide composition. For every sequence, each module offers

similarity-search based information and SVM prediction scores. After compiling the data from these five modules, a second layer of SVM classifiers and PSI-BLAST are used to generate the final output, which includes a final prediction of each protein's virulence.[14]

VICM-pred server

Webserver tool called VICMpred helps classify gram-negative bacteria's proteins into categories such as virulence factors, information molecules, cellular processes, and metabolism molecules. Using patterns, amino acid content, and dipeptide composition of bacterial protein sequences, the VICMpred server employs an SVM-based approach. VICMpred, which lets users infer from a protein's amino acid sequences the function of the protein (virulence factors, information molecules, cellular process, and metabolism). The pathogenic proteins in the provided sequences was predicted using an online service. Protein sequences entered into the VICMpred system are categorised into four functional groups: information molecules, virulence factors, metabolism molecules, and cellular processes. The final result that *indica*tes the projected functional class is provided by the server based on the scores awarded to each sequence. Sequences with fewer than 100 amino acid residues are rejected by VICMpred, though, since they do not have enough information for a precise classification.[15]

Servers for prediction of sub-cellular localization of pathogenic proteins

The BUSCA server

The Bologna Biocomputing Group created the Bologna Unified sub-cellular Component Annotator (BUSCA), a special web server that combines multiple existing techniques to predict a particular sub-cellular location based only on protein sequence. Protein sequences in FASTA format are accepted as input by the BUSCA web server. An analysis of up to 500 protein sequences can be done at a time on the server. The user is additionally prompted to designate the taxonomic origin of the input protein sequences prior to submission, selecting from five possible categories: fungi, plants, animals, and both Gram-positive and Gram-negative bacteria.[16]

The server CELLO v.2.5

CELLO is a basic, uncomplicated implementation of a single module support vector machine (SVM) for sub-cellular localisation prediction that is based on multiple n-peptide composition. Specialised methods or specific input vectors are not required for every sub-cellular localisation site. Better prediction performances are attained using the single module technique CELLO. The server accepts up to 500 protein sequences per-submission. The CELLO v.2.5 predicts proteins in two classes one Cello prediction and second SVM classifier prediction. [17]

The server PSORTb v.3.0

The development of PSORTb, an improved PSORT algorithm with noticeably greater accuracy, was led by the Brinkman Laboratory. It prioritises accuracy above recall in order to provide precise predictions, even if this means producing fewer predictions overall than alternative approaches. PSORTb v.3.0 is composed of several analytical modules, each of which examines a single biological attribute or factor that is known to affect sub-cellular localisation. The server accepts up to 500 protein sequences per-submission. The server predicts the location of the given proteins and provides score also.[18]

The server PSL-Pred

A web service called PSLpred is used to forecast the sub-cellular location of proteins produced by gram-negative bacteria. PSLpred is a hybrid approach-based method that combines three SVM modules based on physico-chemical parameters, residue compositions, and dipeptides with PSI-BLAST. The server accepts up to 500 protein sequences per-submission. The server accepts sequences which are more than 70 residues. The server predicts the location, length of the sequence, reliability index and expected accuracy of the given proteins.[19]

The SLP-Local server

Local amino acid compositions, twin amino acid compositions, and local frequencies of the separation between successive amino acids (basic, hydrophobic, and other) are all included in this server. The N-terminal, middle, and C-terminal portions of each sequence are divided for the purpose of determining the local properties. To account for any ambiguity in the length and position of signal sequences, the N-terminal portion is further split into four parts. It is concluded that the sub-cellular position can be predicted by taking into account the corresponding properties in the N-terminal, middle, and C-terminal sections. The server accepts up to 500 protein sequences per-submission. The server predicts the location of the given proteins and provides SVM score also.[20]

The server for ngLOC

A web server and software program for ngLOC-based protein sequence sub-cellular localisation prediction. Prokaryotic and eukaryotic proteins' sub-cellular locations are predicted by the n-gram-based Bayesian classifier ngLOC. ngLOC is a general technique for protein sub-cellular localisation prediction that may be trained using data from multiple species or classes.

Under the GNU GPL, the standalone program is freely accessible for academic usage. The server accepts up to 500 protein sequences per-submission. ngLOC web server reject those sequences which is less then 70 amino acid residues in sequence length.[21]

The server Gram-LocEN

In order to produce clear and interpretable results for the large-scale prediction of both single-label and multi-label proteins of various species, including Gram-positive and Gram-negative bacteria, Gram-LocEN is an interpretable multi-label predictor that leverages unified characteristics. The server accepts up to 500 protein sequences per-submission. The server predicts the location of the given proteins and provides BLAST E-value score also. [22]

The server CELLO2GO

A web-based tool called CELLO2GO used to screen different aspects of a targeted protein, including its sub-cellular location. By merging the BLAST homology-searching and CELLO localization-predicting techniques for the questioned proteins. Since a protein's environment contributes to some of the pertinent context required for function, it is widely accepted that a protein's sub-cellular localisation affects its function. The server accepts up to 500 protein sequences per-submission. CELLO2GO web server reject those sequences which is less then 70 amino acid residues in sequence length. So, we will submit those sequences which are more then 70 residues length. CELLO2GO server predicts location of given protein and provide molecular function also.[23]

Results and Discussion

Result from MP3 software

The MP3 software predicts results across several categories, including Pfam Domain, HMM prediction, SVM score, SVM prediction, Hybrid prediction, and Assignment. Out of a total of 5,778 proteins which is the whole proteome of organism, analyzed, the software identified 670 proteins as pathogenic using the SVM method, 441 proteins as pathogenic using the HMM method, and 880 proteins as pathogenic using the Hybrid method. For our study, we used the SVM score as the primary criterion to sort the proteins. After sorting, we selected the top 500 proteins, which yielded a highest SVM score of 6.88 and a lowest score of 1.067 among the selected proteins.

Table no. 1:: Result of output of MP3 Predictions

S. no	Method for Detection of Pathogenic proteins	Threshold	No of proteins detected as Pathogenic proteins
1	SVM	0.9	670
2	HMM	0.9	441
3	HYBRID	0.9	880

Result from VirulentPred server

The VirulentPred server was used to analyze the 500 proteins that passed the initial screening by the MP3 software. The server predicts the pathogenicity of proteins based on six different criteria: amino acid composition, dipeptide composition, higher-order dipeptide composition, similarity-based using PSI-BLAST, PSI-BLAST-created PSSM profiles, and a cascade of SVMs and PSI-BLAST. Out of the 500 proteins analyzed, the VirulentPred server predicted 337 as virulent and 163 as non-virulent.

Result from VICM Pred Server

The VICM-Pred server was utilized to analyze the 500 proteins that successfully passed the MP3 software's initial analysis. This server is designed to predict the pathogenicity of proteins. Among the 500 proteins evaluated, 52 were predicted to be virulent, while 436 were classified as non-virulent.

Results from sub-cellular Localization Servers

The prediction of sub-cellular localization of proteins was conducted using various servers, yielding diverse results. The findings were broadly classified into three categories: cytoplasmic proteins, plasma membrane proteins, and extracellular space proteins.

(a) BUSCA Server

The BUSCA server predicted the sub-cellular localization of proteins in several categories: plasma and outer membrane, cytoplasmic, extracellular space, and, for some proteins, predictions could not be made due to the protein length being shorter than the prescribed limit. Out of the 500 proteins analyzed, 189 were identified as plasma membrane proteins, 157 as cytoplasmic proteins, 132 as extracellular proteins, and 22 were not identified because their lengths fell below the required threshold.

(b) CELLO v.2.5 Server

The CELLO v.2.5 server categorized the sub-cellular localization of proteins into plasma membrane (outer membrane and inner membrane), cytoplasmic, and extracellular. From the 500 proteins, 366 were identified as plasma membrane proteins, 62 as cytoplasmic proteins, and 72 as extracellular proteins.

(c) PSORTb v.3.0 Server

The PSORTb v.3.0 server predicted sub-cellular localization in categories including plasma membrane (outer membrane and periplasm), cytoplasmic, and extracellular, with some proteins remaining unclassified. Among the 500 proteins, 77 were identified as plasma membrane proteins, 213 as cytoplasmic proteins, 30 as extracellular proteins, and 180 were not identified.

(d) PSL-Pred Server

The PSL-Pred server provided predictions for sub-cellular localization in the following categories: plasma membrane (outer membrane, inner membrane, periplasm), cytoplasmic, and extracellular. For some proteins, predictions could not be made due to insufficient length. Out of 500 proteins, 360 were identified as plasma membrane proteins, 49 as cytoplasmic proteins, 80 as extracellular proteins, and 11 were not identified because their lengths were shorter than the prescribed limit.

(e) SLP-Local Server

The SLP-Local server predicted sub-cellular localization in categories such as plasma membrane (periplasm), cytoplasmic, and extracellular. Of the 500 proteins analyzed, 148 were identified as plasma membrane proteins, 242 as cytoplasmic proteins, and 110 as extracellular proteins.

(f) ngLOC Server

The ngLOC server categorized the sub-cellular localization of proteins into plasma membrane (outer membrane, periplasm, inner membrane), cytoplasmic, and extracellular. Among the 500 proteins, 207 were identified as plasma membrane proteins, 285 as cytoplasmic proteins, and 8 as extracellular proteins.

(g) Gram-LocEN Server

The Gram-LocEN server predicted sub-cellular localization in categories including plasma membrane (outer membrane, periplasm, inner membrane, fimbrium, nucleoid), cytoplasmic, and extracellular. Out of the 500 proteins, 339 were identified as plasma membrane proteins, 133 as cytoplasmic proteins, and 28 as extracellular proteins.

(h) CELLO2GO Server

The CELLO2GO server provided predictions for sub-cellular localization in categories such as plasma membrane (outer membrane, periplasm, inner membrane), extracellular, and, for some proteins, cytoplasmic. out the 500 proteins analyzed, 316 were identified as plasma membrane proteins, 177 as extracellular proteins, and 7 were predicted as cytoplasmic.

Table no. 2 :: Sub-cellular localization prediction of 500 proteins

S no.	Prediction of Protein's localization	NCBI Accession no.
1	Extra-cellular	WP_010445499.1, WP_029530467.1, WP_010449940.1, WP_019361858.1, WP_010449972.1, WP_010449974.1, WP_010449945.1, WP_019361854.1, WP_010447993.1, WP_010447659.1, WP_010448372.1, WP_081354367.1, WP_019361260.1, WP_019361394.1, WP_010448376.1, WP_081354670.1, WP_019361004.1, WP_019362083.1, WP_019360510.1, WP_081354677.1, WP_010449983.1, WP_081354464.1, WP_026007646.1, WP_019362073.1, WP_019361855.1, WP_010453046.1, WP_019362482.1, WP_019363755.1, WP_081354628.1, WP_019362926.1, WP_010445549.1, WP_010449516.1.
2	Plasma membrane	WP_019360344.1, WP_010444007.1, WP_081354652.1, WP_019362126.1, WP_081354298.1, WP_019361857.1, WP_010444433.1, WP_010446010.1, WP_019363340.1, WP_010449061.1, WP_019361337.1, WP_010444008.1, WP_019361813.1, WP_010449981.1, WP_019361856.1, WP_010452368.1, WP_010453537.1, WP_026007833.1, WP_029531009.1, WP_029529929.1, WP_019362956.1, WP_081354566.1, WP_081354473.1, WP_019360837.1, WP_019360401.1, WP_019362258.1, WP_010449913.1, WP_029529951.1, WP_010446015.1, WP_019360276.1, WP_081354648.1, WP_010445511.1, WP_019363565.1, WP_029531028.1, WP_010453541.1, WP_010449786.1, WP_029530390.1, WP_010451299.1, WP_010448323.1, WP_026007617.1, WP_010448715.1, WP_010444322.1, WP_019363601.1, WP_010449677.1, WP_019362405.1, WP_010443994.1, WP_010443712.1, WP_019360267.1, WP_019363511.1, WP_010448391.1, WP_010450321.1, WP_081354302.1, WP_026007559.1, WP_019361709.1, WP_029529825.1, WP_019362422.1, WP_010448173.1, WP_019363616.1, WP_019360607.1, WP_019363315.1, WP_019361283.1, WP_010452053.1, WP_029530492.1, WP_019361022.1, WP_010447001.1, WP_026007631.1, WP_010452649.1, WP_019363573.1, WP_019361525.1, WP_010448617.1, WP_081354279.1, WP_019361915.1, WP_080589888.1, WP_019360563.1, WP_026007413.1, WP_019362709.1, WP_010444725.1, WP_010449373.1, WP_019360789.1, WP_010448416.1, WP_081354491.1, WP_019362662.1, WP_019360802.1, WP_010449909.1, WP_010451530.1, WP_019360525.1, WP_019363352.1, WP_010446719.1, WP_019362674.1, WP_019363045.1, WP_019362425.1, WP_080589843.1, WP_019362558.1, WP_010450107.1, WP_026007489.1, WP_019362788.1, WP_019361920.1, WP_026007673.1, WP_010447744.1, WP_081354322.1, WP_010447068.1, WP_026007602.1, WP_019363763.1, WP_010448854.1, WP_010451265.1, WP_010452064.1, WP_010445599.1, WP_019363473.1, WP_010444759.1, WP_010450629.1, WP_081354575.1, WP_010450771.1, WP_010446103.1, WP_026007421.1, WP_019360998.1, WP_019363241.1, WP_029530644.1, WP_010444335.1, WP_019363302.1, WP_019363510.1

		WP_010449657.1, WP_019361285.1, WP_010449098.1, WP_019362125.1, WP_019363119.1 WP_019362196.1, WP_019361295., WP_019362946.1, WP_026007874.1, WP_010450470.1 WP_019360586.1, WP_019361016.1, WP_019361259.1, WP_026007334.1, WP_010445865.1 WP_010448894.1, WP_019362964.1, WP_019361033.1, WP_081354524.1, WP_010452558.1 WP_026007766., WP_019360932.1, WP_010449228.1, WP_010451933.1, WP_010447651.1 WP_026007552.1, WP_019363179.1, WP_080589870.1, WP_019360594.1, WP_081354574.1 WP_010446375.1, WP_019361694.1, WP_081354465.1, WP_019361723.1, WP_026007868.1 WP_019363118.1, WP_010450859.1, WP_081354432.1, WP_010444159.1, WP_010451279.1 WP_026007663.1, WP_081354597.1, WP_019363379.1, WP_010450078.1, WP_010448947.1 WP_081354397.1, WP_010448951.1, WP_019362528.1, WP_010452857.1, WP_029379617.1 WP_019362751.1, WP_019361935.1, WP_019360940.1, WP_080579377.1, WP_019362256.1 WP_019362894.1, WP_010448274.1, WP_081354528.1, WP_010453368.1, WP_019363028.1 WP_010452517.1, WP_029533476.1, WP_029531320.1, WP_081354447.1, WP_019362086.1 WP_026007333.1, WP_019361699.1, WP_019360593.1, WP_019363705.1, WP_019363156.1 WP_019361185.1, WP_010452179.1, WP_019361942.1, WP_051356168.1, WP_010445540.1 WP_081354537.1, WP_026007825.1, WP_080579430.1, WP_019360510.1, WP_081354340.1 WP_019361818.1, WP_003448337.1, WP_010452138.1, WP_080589854.1, WP_019363008.1 WP_019363617.1, WP_010453662.1, WP_019360683.1, WP_019362858.1, WP_019363736.1 WP_081354642.1, WP_029530962.1, WP_010444411.1, WP_019363492.1, WP_010446627.1 WP_019361964.1, WP_029530063.1, WP_010449608.1, WP_019362552.1, WP_010444282.1 WP_019360509.1, WP_019362915.1, WP_019362544.1, WP_036987812.1, WP_010447320.1 WP_081354529.1, WP_026007462.1, WP_019361223.1, WP_029530129.1, WP_010446570.1 WP_019360880.1, WP_019360342.1, WP_010444332.1, WP_010451512.1, WP_019362282.1 WP_019360788.1, WP_019361483.1, WP_029530749.1, WP_010451510.1, WP_019362851.1 WP_010453000.1, WP_026007495.1, WP_010449710.1, WP_010449906.1, WP_029530046.1 WP_081354498.1, WP_010450202.1, WP_026007899.1, WP_081354346.1, WP_010448369.1 WP_010445575.1, WP_019363219.1, WP_029379313.1, WP_010449643.1, WP_019360853.1 WP_019362740.1, WP_010444822.1, WP_051026120.1, WP_010451954.1, WP_081354599.1 WP_026007633.1, WP_026007449.1, WP_081354649.1, WP_010444005.1, WP_010449874.1 WP_019362359.1, WP_019361547.1, WP_019363328.1, WP_081354640.1, WP_019363044.1 WP_010451705.1, WP_010451499.1, WP_081354475.1, WP_010452639.1, WP_080660510.1 WP_029530387.1, WP_029531013.1, WP_010448876.1, WP_081354577.1, WP_010452877.1 WP_019363132.1, WP_010449390.1, WP_080589928.1, WP_081354552.1, WP_081354474.1 WP_010453275.1, WP_019361366.1, WP_081354565.1, WP_019363581.1, WP_010446710.1 WP_019362421.1, WP_019363512.1, WP_081354304.1, WP_010443779.1, WP_019363295.1 WP_081354313.1, WP_019360338.1, WP_010446817.1, WP_019361725.1, WP_080589923.1 WP_081354557.1, WP_010447466.1, WP_081354495.1, WP_010444900.1, WP_029529879.1 WP_026007753.1, WP_019361624.1, WP_010445873.1, WP_081354667.1, WP_019361352.1 WP_019360612.1, WP_019361142.1, WP_010449845.1, WP_081354562.1, WP_019360328.1 WP_029530275.1, WP_029379575.1, WP_010443807.1, WP_019361695.1, WP_026007463.1 WP_010446704.1, WP_010450989.1, WP_019362789.1, WP_019361706.1, WP_019362285.1 WP_019360615.1, WP_026007359.1, WP_019362979.1, WP_019362447.1, WP_010448497.1 WP_026007523.1, WP_019362094.1, WP_010444219.1, WP_026007465.1, WP_010450830.1 WP_019363023.1, WP_019362567.1, WP_019360574.1, WP_019361092.1, WP_029530534.1 WP_010451565.1, WP_026007632.1, WP_026007618.1, WP_019360606.1, WP_019361705.1 WP_010453690.1, WP_081354538.1, WP_019361958.1, WP_081354400.1, WP_081354512.1 WP_019363047.1, WP_019360669.1, WP_010449907.1, WP_019361077.1, WP_019363518.1 WP_019363472.1, WP_081354357.1, WP_010445195.1, WP_019362947.1, WP_019361637.1 WP_019363677.1, WP_019361609.1, WP_029532310.1, WP_019362500.1, WP_019361502.1 WP_010449410.1, WP_081354470.1, WP_019360636.1, WP_026007450.1, WP_081354335.1.
3	**Cytoplasmic**	WP_010451831.1, WP_026007353.1, WP_010444772.1, WP_080589918.1, WP_010448035.1 WP_026007455.1, WP_029530783.1, WP_019362856.1, WP_010449756.1, WP_010444310.1 WP_029530278.1, WP_010444496.1, WP_081354663.1, WP_081354605.1, WP_019361713.1 WP_010447661.1, WP_019363128.1, WP_019362300.1, WP_010450758.1, WP_019360792.1 WP_019363251.1, WP_026007388.1, WP_010451581.1, WP_019360500.1, WP_019361404.1 WP_010451949.1, WP_010447773.1, WP_010449979.1, WP_010443844.1, WP_019361641.1 WP_019362243.1, WP_019362755.1, WP_019360506.1, WP_019360754.1, WP_010451710.1 WP_080579429.1, WP_080579396.1, WP_019363366.1, WP_029530798.1, WP_010449315.1 WP_019361079.1, WP_019360994.1, WP_010445870.1, WP_019363520.1, WP_010448189.1 WP_010448878.1, WP_019360791.1, WP_019363801.1, WP_019360900.1, WP_029379352.1 WP_010446753.1, WP_010451658.1, WP_019361412.1, WP_026007490.1, WP_019363446.1 WP_019362396.1, WP_026007517.1, WP_026007505.1, WP_010453742.1, WP_019361443.1 WP_019360275.1, WP_019361005.1.
4	**Extracellular protein and Plasma-membrane**	WP_010452147.1, WP_026007845.1, WP_019362980.1, WP_019362208.1, WP_019362085.1 WP_019363426.1, WP_019361544.1.
5	**Plasma-membrane and Cytoplasmic**	WP_010452028.1, WP_019361075.1, WP_010447623.1, WP_081354497.1, WP_080589886.1 WP_029529842.1, WP_081354603.1, WP_019361154.1, WP_010448177.1, WP_026007464.1 WP_010449172.1, WP_081354291.1, WP_010452875.1, WP_010453421.1, WP_019361729.1 WP_019361969.1, WP_019363106.1, WP_026007772.1, WP_010445011.1, WP_036998690.1 WP_019363568.1, WP_019360355.1, WP_010448541.1, WP_019360984.1, WP_019363075.1 WP_010446250.1, WP_010450891.1, WP_019363546.1.

6	**Extracellular protein and Cytoplasmic**	WP_010452004.1

So we were able to predict 32 proteins as exclusively Extra-cellular, 370 as exclusively Plasma membrane associated, 62 as exclusively Cytoplasmic proteins, 7 as both Extra-cellular and Plasma membrane associated, 28 as Plasma membrane associated and Cytoplasmic protein and 1 as Extra-cellular and Cytoplasmic in nature.

Virulent proteins and their sub-cellular localization

Out of the 500 proteins analyzed, we identified 107 proteins that are predicted to have a role in pathogenicity based on the assessments from three different servers/tools. The determination of pathogenicity was made through a majority voting system; specifically, if at least two of the three servers predicted a protein to be pathogenic, we classified it as such; otherwise, it was considered non-pathogenic. Subsequently, we correlated these findings with the predictions of sub-cellular localization. Among the 107 proteins identified as pathogenic, 80 were predicted to be associated with the plasma membrane, 18 were classified as extracellular, and 9 were identified as cytoplasmic. Notably, some of these proteins were predicted to have multiple localization sites.

Table no. 3:: Virulent and sub-cellular localization prediction

S.NO.	NCBI Acession no.	Prediction for virulent	Final prediction	S.NO.	NCBI Accession no.	Prediction for virulent	Final prediction
1	WP_029530467.1	Pathogenic	Extracellular	55	WP_019363426.1	Pathogenic	Plasma-membrane ,Extracellular
2	WP_010449940.1	Pathogenic	Extracellular	56	WP_081354599.1	Pathogenic	Plasma-membrane
3	WP_019361858.1	Pathogenic	Extracellular	57	WP_081354649.1	Pathogenic	Plasma-membrane
4	WP_010449974.1	Pathogenic	Extracellular	58	WP_019362073.1	Pathogenic	Extracellular
5	WP_019361857.1	Pathogenic	Plasma-membrane	59	WP_010449874.1	Pathogenic	Plasma-membrane
6	WP_010444433.1	Pathogenic	Plasma-membrane	60	WP_019362359.1	Pathogenic	Plasma-membrane
7	WP_010449061.1	Pathogenic	Plasma-membrane	61	WP_019361547.1	Pathogenic	Plasma-membrane
8	WP_010444008.1	Pathogenic	Plasma-membrane	62	WP_019363328.1	Pathogenic	Plasma-membrane
9	WP_029529929.1	Pathogenic	Plasma-membrane	63	WP_019361855.1	Pathogenic	Extracellular
10	WP_081354566.1	Pathogenic	Plasma-membrane	64	WP_010453046.1	Pathogenic	Extracellular
11	WP_019360837.1	Pathogenic	Plasma-membrane	65	WP_019363044.1	Pathogenic	Plasma-membrane
12	WP_010449913.1	Pathogenic	Plasma-membrane	66	WP_019361920.1	Pathogenic	Plasma-membrane
13	WP_010452517.1	Pathogenic	Plasma-membrane	67	WP_010448376.1	Pathogenic	Extracellular
14	WP_026007333.1	Pathogenic	Plasma-membrane	68	WP_010446103.1	Pathogenic	Plasma-membrane
15	WP_019361699.1	Pathogenic	Plasma-membrane	69	WP_026007421.1	Pathogenic	Plasma-membrane
16	WP_010445011.1	Pathogenic	Plasma-membrane , cytoplasmic	70	WP_010444335.1	Pathogenic	Plasma-membrane
17	WP_019361942.1	Pathogenic	Plasma-membrane	71	WP_029529842.1	Pathogenic	Plasma-membrane , cytoplasmic
18	WP_010445540.1	Pathogenic	Plasma-membrane	72	WP_019363302.1	Pathogenic	Plasma-membrane
19	WP_026007825.1	Pathogenic	Plasma-membrane	73	WP_010451499.1	Pathogenic	Plasma-membrane
20	WP_019362083.1	Pathogenic	Extracellular	74	WP_019362482.1	Pathogenic	Extracellular
21	WP_080579430.1	Pathogenic	Plasma-membrane	75	WP_080660510.1	Pathogenic	Plasma-membrane
22	WP_019360510.1	Pathogenic	Extracellular	76	WP_010445870.1	Pathogenic	cytoplasm
23	WP_081354677.1	Pathogenic	Extracellular	77	WP_081354577.1	Pathogenic	Plasma-membrane
24	WP_019363008.1	Pathogenic	Plasma-membrane	78	WP_081354628.1	Pathogenic	Extracellular
25	WP_010453662.1	Pathogenic	Plasma-membrane	79	WP_019362926.1	Pathogenic	Extracellular
26	WP_029530962.1	Pathogenic	Plasma-membrane	80	WP_010447661.1	Pathogenic	cytoplasm
27	WP_029529951.1	Pathogenic	Plasma-membrane	81	WP_019361016.1	Pathogenic	Plasma-membrane
28	WP_026007353.1	Pathogenic	cytoplasm	82	WP_019362964.1	Pathogenic	Plasma-membrane
29	WP_081354367.1	Pathogenic	Extracellular	83	WP_019360792.1	Pathogenic	cytoplasm
30	WP_019361260.1	Pathogenic	Extracellular	84	WP_081354524.1	Pathogenic	Plasma-membrane
31	WP_019361075.1	Pathogenic	Plasma-membrane , cytoplasmic	85	WP_026007464.1	Pathogenic	Plasma-membrane , cytoplasmic
32	WP_081354648.1	Pathogenic	Plasma-membrane	86	WP_026007753.1	Pathogenic	Plasma-membrane
33	WP_010445511.1	Pathogenic	Plasma-membrane	87	WP_019360791.1	Pathogenic	cytoplasm
34	WP_010444772.1	Pathogenic	cytoplasm	88	WP_010445873.1	Pathogenic	Plasma-membrane
35	WP_019362405.1	Pathogenic	Plasma-membrane	89	WP_081354667.1	Pathogenic	Plasma-membrane
36	WP_081354302.1	Pathogenic	Plasma-membrane	90	WP_010449845.1	Pathogenic	Plasma-membrane
37	WP_029529825.1	Pathogenic	Plasma-membrane	91	WP_019363075.1	Pathogenic	Plasma-membrane , cytoplasmic
38	WP_019362422.1	Pathogenic	Plasma-membrane	92	WP_010448497.1	Pathogenic	Plasma-membrane
39	WP_010449983.1	Pathogenic	Extracellular	93	WP_080589870.1	Pathogenic	Plasma-membrane
40	WP_019361964.1	Pathogenic	Plasma-membrane	94	WP_019361404.1	Pathogenic	cytoplasm

41	WP_081354464.1	Pathogenic	Extracellular	95	WP_081354465.1	Pathogenic	Plasma-membrane
42	WP_010449608.1	Pathogenic	Plasma-membrane	96	WP_019361641.	Pathogenic	cytoplasm
43	WP_019360509.1	Pathogenic	Plasma-membrane	97	WP_019362243.1	Pathogenic	cytoplasm
44	WP_019362208.1	Pathogenic	Plasma-membrane ,Extracellular	98	WP_081354397.1	Pathogenic	Plasma-membrane
45	WP_019362915.1	Pathogenic	Plasma-membrane	99	WP_019362528.1	Pathogenic	Plasma-membrane
46	WP_036987812.1	Pathogenic	Plasma-membrane	100	WP_010452857.1	Pathogenic	Plasma-membrane
47	WP_019360342.1	Pathogenic	Plasma-membrane	101	WP_029379617.1	Pathogenic	Plasma-membrane
48	WP_019363568.1	Pathogenic	Plasma-membrane , cytoplasmic	102	WP_019360574.1	Pathogenic	Plasma-membrane
49	WP_019362282.1	Pathogenic	Plasma-membrane	103	WP_019361092.1	Pathogenic	Plasma-membrane
50	WP_019363616.1	Pathogenic	Plasma-membrane	104	WP_019363518.1	Pathogenic	Plasma-membrane
51	WP_029530492.1	Pathogenic	Plasma-membrane	105	WP_019362947.1	Pathogenic	Plasma-membrane
52	WP_026007631.1	Pathogenic	Plasma-membrane	106	WP_026007772.1	Pathogenic	Plasma-membrane , cytoplasmic
53	WP_080589888.1	Pathogenic	Plasma-membrane	107	WP_019360636.1	Pathogenic	Plasma-membrane
54	WP_081354491.1	Pathogenic	Plasma-membrane				

Summary, Conclusion and Suggestions for Future Research Work

Summary for Virulence Prediction

The analysis of 5,778 proteins from *Pseudomonas fuscovaginae* pv. LMG 2158 was conducted using the MP3 software to identify proteins potentially involved in pathogenicity. The MP3 standalone software, operating on a Linux OS, facilitated the detection of pathogenic proteins, with the total number of proteins derived from predictions by the NCBI Bioproject. The pathogenicity of proteins was assessed using three methods: the HMM method identified 441 proteins as pathogenic, the SVM method identified 670, and the HYBRID method identified 880. For our study, we prioritized the SVM prediction scores, which ranged from a highest score of 6.88 to a lowest of 1.067, ultimately selecting the top 500 proteins based on these scores. Subsequently, the VirulentPred server was employed to analyze the selected 500 proteins, predicting 337 as virulent and 163 as non-virulent. Additionally, the VICM-Pred server predicted 52 proteins as virulent and 436 as non-virulent.

Summary for sub-cellular localization prediction

sub-cellular localization predictions were performed using various servers, including BUSCA, CELLO v.2.5, PSORTb v.3.0, PSL-Pred, SLP-Local, ngLOC, Gram-LocEN, and CELLO2GO. From the 500 proteins analyzed, 405 were predicted to be associated with the plasma membrane, 91 as cytoplasmic, and 40 as extracellular, with some proteins predicted to have multiple localization sites. In conclusion, the *Pseudomonas fuscovaginae* LMG-2158 contains a total of 5,778 proteins, of which 670 were identified as pathogenic by the SVM method, 441 by the HMM method, and 880 by the HYBRID method. The VirulentPred and VICM-Pred servers further classified these proteins into virulent and non-virulent categories. The sub-cellular localization analysis *indica*ted that a significant proportion of the proteins were associated with the plasma membrane.

Suggestions for Future Research Work

Future research could expand the identification of pathogenic proteins to include additional proteins identified by the MP3 software and replicate similar studies in other strains of the organism. Furthermore, analyzing the pathways involving these pathogenic proteins could provide deeper insights into their functions. Three-dimensional structure predictions could facilitate the identification of functional sites within these proteins, enabling the design of ligands aimed at blocking or altering their active sites to attenuate infection severity. Additionally, sub-cellular localization predictions could be refined to determine the precise accumulation of various protein types, aiding in further analysis. Finally, *insilico* identification and study of protein interactions with host cells could elucidate the molecular mechanisms at play, identifying key proteins for further investigation.

Acknowledgments

The authors thanks the Dept of Microbiology and Bioinformatics, Atal Bihari Vajpayee Vishwavidyalaya, Bilaspur, Chhattisgarh, India for providing all the necessary infrastructure to carryout the research work.

Conflict of Interest

No conflict of interest exist

Funding

No grants were received for carrying out the work.

Bibliography

1. http://www.mospi.gov.in/statistical-year-book-india/2016/177.
2. Patel HK, Passos da Silva D, Devescovi G, Maraite H, Paszkiewicz K, Studholme DJ, et al. Draft genome sequence of *Pseudomonas fuscovaginae*, a broad-host-range pathogen of plants. J Bacteriol [Internet]. 2012 [cited 2024

Aug 4];194(10):2765–6. Available from: http://dx.doi.org/10.1128/jb.00341-12.
3. Fuller DQ. Finding plant domestication in the Indian subcontinent. Curr Anthropol [Internet]. 2011;52(S4):S347–62. Available from: http://dx.doi.org/10.1086/658900.
4. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science [Internet]. 2002;296(5565):92–100. Available from: http://dx.doi.org/10.1126/science.1068275.
5. Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S, et al. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae pv. oryza*e PXO99A. BMC Genomics [Internet]. 2008;9(1):204. Available from: http://dx.doi.org/10.1186/1471-2164-9-204.
6. Adorada DL, Stodart BJ, Pangga IB, Ash GJ. Implications of bacterial contaminated seed lots and endophytic colonization by *Pseudomonas fuscovaginae* on rice establishment. Plant Pathol [Internet]. 2015;64(1):43–50. Available from: http://dx.doi.org/10.1111/ppa.12243.
7. Bigirimanav DP, Ha GKH, Nyamangyokuo I, Höfte M. Rice Sheath Rot: An Emerging Ubiquitous Destructive Disease Complex. 2015.
8. Wang J, Sung WK, Krishnan A, Li KB. Protein subcellular localization prediction for Gram-negative bacteria using amino acid subalphabets and a combination of multiple support vector machines. BMC Bioinformatics [Internet]. 2005;6(1):174. Available from: http://dx.doi.org/10.1186/1471-2105-6-174
9. Bacterial sheath brown rot - IRRI Rice Knowledge Bank. Available from: http://www.knowledgebank.irri.org/training/fact-sheets/pest-management/diseases/item/bacterial-sheath-brown-rot.
10. Tanii A, Miyajima K, Akita T. The Sheath Brown Rot Disease of Rice Plant and Its Causal Bacterium, *Pseudomonas fuscovaginae* A. Tanii, K. Miyajima et T. Akita sp. nov. Nippon Shokubutsu Byori Gakkaiho [Internet]. 1976;42(5):540–8. Available from: https://www.jstage.jst.go.jp/article/jjphytopath1918/42/5/42_5_540/_article.
11. Apsnet.org.[cited 2024 Aug 4]. Available from: https://www.apsnet.org/publications/plantdisease /backissues/Documents/1987Articles/PlantDisease71n07_592.pdf
12. Ferrarini E, Špacapan M, Lam VB, McCann A, Cesa-Luna C, Marahatta BP, et al. Versatile role of *Pseudomonas fuscovaginae* cyclic lipopeptides in plant and microbial interactions. Front Plant Sci [Internet]. 2022 [cited 2024 Aug 4];13. Available from: http://dx.doi.org/10.3389/fpls.2022.1008980
13. Gupta A, Kapil R, Dhakan DB, Sharma V. MP3: A Software Tool for the Prediction of Pathogenic Proteins in Genomic and Metagenomic Data Ankit Gupta, Rohan Kapil, Darshan B. Kapil R, Dhakan DB, Sharma VK, editors. PLoS One. 2017;9(4).
14. Garg A, Gupta D. VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. BMC Bioinformatics [Internet]. 2008;9(1). Available from: http://dx.doi.org/10.1186/1471-2105-9-62.
15. Saha S, Raghava GPS. VICMpred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition. Genomics Proteomics Bioinformatics [Internet]. 2006;4(1):42–7. Available from: http://dx.doi.org/10.1016/S1672-0229(06)60015-6.
16. Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acids Res [Internet]. 2018;46(W1):W459–66. Available from: http://dx.doi.org/10.1093/nar/gky320.
17. Yu CS, Lin CJ, Hwang JK. Hwang JK(2004): Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Science. 2004;13:1402–6.
18. Chou KC, Shen HB. Cell-PLoc: A package of web-servers for predicting subcellular localization of proteins in various organisms. Nature Protocols. 2008;3:153–62.
19. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics [Internet]. 2010;26(13):1608–15. Available from: http://dx.doi.org/10.1093/bioinformatics/btq249.
20. Bhasin M, Garg A, Raghava GPS. PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics [Internet]. 2005;21(10):2522–4. Available from: http://dx.doi.org/10.1093/bioinformatics/bti309.
21. Matsuda S, Vert JP, Saigo H, Ueda N, Toh H, Akutsu T. A novel representation of protein sequences for prediction of subcellular location using support vector machines: A NOVEL REPRESENTATION OF PROTEIN SEQUENCES FOR PREDICTION OF SUBCELLULAR LOCATION USING SUPPORT VECTOR MACHINES. Protein Sci

[Internet]. 2005;14(11):2804–13. Available from: http://dx.doi.org/10.1110/ps.051597405.

22. King BR, Vural S, Pandey S, Barteau A, Guda C. ngLOC: software and web server for predicting protein subcellular localization in prokaryotes and eukaryotes. BMC Res Notes [Internet]. 2012;5(1):351. Available from: http://dx.doi.org/10.1186/1756-0500-5-351.
23. Wan S, Mak MW, Kung S-Y. Gram-LocEN: Interpretable prediction of subcellular multi-localization of Gram-positive and Gram-negative bacterial proteins. Chemometr Intell Lab Syst [Internet]. 2017;162:1–9. Available from: http://dx.doi.org/10.1016/j.chemolab.2016.12.014.
24. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, et al. CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS One [Internet]. 2014;9(6):e99368. Available from: http://dx.doi.org/10.1371/journal.pone.0099368.