

# *In-Silico* Subcellular Localization and Functional Analysis of Computationally Predicted Virulent Proteins in *X. Oryzae* pv. *Oryzae* Strain PXO99A Causal Organism of Bacterial Leaf Blight (BLB) of Rice (*O. Sativa* L.)

Dharmendra Kashyap\* and Aafreen Khan

Department of Microbiology & Bioinformatics, Atal Bihari Vajpayee Vishwavidyalaya, Bilaspur University, Bilaspur, Chhattisgarh, India

## Abstract

Rice (*O. Sativa* L.) is an important cereal crop in the world and more than half of the human population depends on it as a staple diet. India ranks second among countries in the world as per rice production and area under cultivation. *X. oryzae* are a group of plant-specific yellow-pigmented microbes, which are important phytopathogens that infect crops such as rice and wheat and cause leaf blight disease with a yield loss of as much as 70%. Virulent proteins of pathogen and their subcellular localization play an important role in the invasion, colonization and survival of pathogens. In the present study, computational identification of pathogenic proteins and prediction of their sub-cellular has been done with an idea to facilitate the identification and understanding of their role in pathogenesis. The sequence of strain was downloaded from NCBI and various tools like MP3 software, VirulentPred server, etc. were used for the identification of pathogenic proteins. From analysis 116 proteins were predicted that may have a role in pathogenicity. The subcellular localization and Pfam/Interpro analysis using various online servers like BUSCA, Cello v.2.5, etc. for subcellular localization and Pfam server for Pfam/Interpro analysis were carried out for predicted 116 proteins. Of the 116 proteins most of the proteins were predicted as membrane, extracellular and cytoplasmic proteins, of the predicted proteins most have been predicted with multiple localization in the *X. oryzae*. The Pfam analysis revealed their role in pathogenesis/virulent associated functions too.

**Keywords:** *O. Sativa* L. • *X. Oryzae* pv. *Oryzae* • Bacterial leaf blight • Virulent proteins • Subcellular localization prediction • Pfam/Interpro

## Introduction

Rice (*O. Sativa* L.) is a major source of food for the world's population with an estimated global production of 508 million tons in the year 2020 [1]. The production of rice in India for the year 2020 is estimated to be 102.36 million tonnes [2]. For Indians rice is an important part of the diet with more than 70% of people directly dependent on it as a major source of calories. In India rice is cultivated in an area of 32.179 million hectares with a productivity of 3.18 metric tons per hectare [3]. Rice is a semi-aquatic plant and is infected by a large no of pathogens like fungi, bacteria, nematodes and viruses. These pathogens affect the productivity and quality of rice from moderate to severe levels. The bacterial leaf blight disease is one of the most severe diseases caused by bacteria *X. oryzae* and is observed throughout the Asian continent [4]. The losses due to disease may rise to 40% in susceptible varieties [5,6]. The *Xanthomonas* genus is plant-specific yellow-pigmented microbes, some of which are economically important phytopathogens that devastate crops such as citrus plants, rice, beans, grape and cotton. These organisms are almost exclusively found associated with their plant hosts [7,8]. The *X. oryzae*, a Gram's negative bacterial plant pathogen, enters through the natural opening like stomata or from wounds on the plant surface. After entry, the pathogens start the process of host recognition, colonization and reproduction [9]. The infection is driven by virulent proteins that

have a role in the initiation of infection and colonization of the host plant. The virulent proteins are crucial for the mechanism of infection [10]. The virulent proteins show a high level of expression in bacterial cells which undergoes a process of colonization and infection. Identification of such proteins is vital for the understanding of the pathogenicity of the pathogen. Due to advances in computational biological sciences, it has become easy to identify such proteins through computational methods. Various tools like MP3, VirulentPred server and VICM server are available free of cost for users to identify such pathogenic proteins. Virulence Factor Database in Bateria (VFDB) is another resource that can be used for the prediction of virulent protein which is also freely available for users. The subcellular location of proteins is quite significantly related to their biological functions [11]. The Gram's negative bacteria have five main subcellular localization sites it includes the inner membrane, the outer membrane, the periplasm, the cytoplasm and the extracellular space. Identification of such proteins, associated with the disease cycle and their subcellular prediction is crucial for the understanding of pathogenesis. Various online servers like BUSCA, CELLO, Gneg-mPLoc, PSORTb, PSL-Pred, SLP-Local, ngLOC, Gram-LocEN, CELLO2GO are available free to users for this purpose. These servers apply different techniques for the prediction of subcellular localization of proteins. Using multiple methods eliminates the chance of unilateral prediction of subcellular localization. The identified proteins were again subjected to Pfam/Interpro analysis for further functional analysis. The present research work is an *in-silico* approach for the prediction of virulent proteins, their subcellular localization and functional prediction of involved proteins in the virulence/pathogenicity process for *X. oryzae*, a potent pathogen of *O. Sativa* L.

\*Address for Correspondence: Dharmendra Kashyap, Department of Microbiology & Bioinformatics, Atal Bihari Vajpayee Vishwavidyalaya, Bilaspur University, Bilaspur, Chhattisgarh, India, Tel: 07752-220007; E-mail: dharmendra.kashyap@bilaspuruniversity.ac.in

**Copyright:** © 2023 Kashyap D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Received:** 04 May, 2023, Manuscript No. MBL-23-97803; **Editor assigned:** 06 May, 2023, PreQC No. P-97803; **Reviewed:** 22 May, 2023, QC No. Q-97803; **Revised:** 05 June, 2023, Manuscript No. R-97803; **Published:** 15 June, 2023, DOI: 10.37421/2168-9547.2023.12.377

## Materials and Methods

### Sequences

*X. oryzae* pv. *oryzae* PXO99A contains 5.24 Mbp of genome size (largest in *X. oryzae*) and contains 3907 proteins estimated through bio-project PRJNA131967. The Refseq no is NC\_010717.2. The necessary files containing protein sequences, nucleotide sequences were downloaded from the NCBI server. The files were used for all sorts of computational analyses.

## Prediction of virulence factor

**MP3 standalone software:** It is a UNIX-based software tool used for the prediction of pathogenic proteins in Genomic and Metagenomic data. MP3 Software works on two approaches Support Vector Machines (SVM) and Hidden Markov Model (HMM) for analysis of protein sequences. The standalone version of the software works on Linux operating systems and has been developed at Meta-informatics Laboratory, Metagenomics and Systems Biology Group, Department of Biological Sciences Indian Institute of Science Education and Research, Bhopal, M.P, India.

**Virulentpred server:** VirulentPred server is a bacterial virulent protein prediction server based on a double-layer cascade support vector machine. The first layer of SVM classifiers is trained with various sequence characteristics such as amino acid composition, dipeptide composition, composition of higher order dipeptides and remote and optimizes evolutionary relationships with Position Specific Iterated BLAST (PSIBLAST). The second layer received the binary score output generated by the top 5 modules of the first layer in order to train the SVM model of the second layer.

**VICMpred server:** VICMpred is a web-server for the broad functional classification of proteins of Gram's negative bacteria into virulence factors, information molecules, cellular processes and metabolism molecules. The VICMpred server applies SVM-based methods for patterns, amino acid and dipeptide composition of protein sequences of bacteria. VICMpred server allows the users to predict the function of a protein as virulence factors, information molecules, cellular process and metabolism from its amino acid sequences.

**Local blast with VFDB:** The selected sequences passing the above criteria will be locally BLASTed against the Virulence Factor Database in Bateria (VFDB). The VFDB has been developed by Jin Qi and co-workers at the State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing, China, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK and National Center of Human Genome Research, Beijing, China.

## Servers for prediction of subcellular localization of identified pathogenic proteins

**BUSCA server:** The Bologna Unified Subcellular Component Annotator (BUSCA) web server, available at <http://busca.biocomp.unibo.it/>, integrates several established methods such as DeepSig, TPpred3, PredGPI, BetAware and ENSEMBLE3.0 and has been developed by the Bologna Biocomputing Group for the Prediction of the specific subcellular localization based on the protein sequence.

**CELLO v.2.5 server:** CELLO is a multiclass SVM-based subcellular localization prediction tool designed for Gram-negative, Gram-positive and Eukaryotic cells are available at <http://www.csbio.sjtu.edu.cn/bioinf/Gneg-multi/>. The server predicts subcellular localization by considering the amino acid composition, the di-peptide composition, the partitioned amino acid composition and the sequence composition of given protein sequences.

**Gneg-mPLOC:** The Gneg-mPLOC tool, available at <http://www.csbio.sjtu.edu.cn/bioinf/Gneg-multi/>, is part of Cell-PLOC 2.0 webserver and is specialized for prediction of subcellular localization of Gram's negative bacteria. The Gneg-mPLOC uses information of gene ontology, functional domain and sequential evolution for the purpose.

**PSORTb v.3.0 server:** PSORTb server available at PSORTb Subcellular Localization Prediction Tool - version 3.0 is developed and made available by The Brinkman Laboratory. PSORTb v.3.0 consists of several modules for analysis, each of which analyzes one biological feature known to influence or be characteristic of subcellular localization.

**PSL-Pred server:** The PSLpred server, available at <http://www.imtech.res.in/raghava/pslpred/>, developed for prediction of subcellular localization of bacterial proteins in Gram's negative bacteria is a hybrid approach-based method integrating PSI-BLAST and three SVM modules based on compositions of residues, dipeptides and the physicochemical properties.

**SLP-Local server:** The SLP-Local server, available at <https://sunflower.kuicr.kyoto-u.ac.jp/~smatsuda/slplocal.html>, predicts the subcellular location of proteins from the amino acid sequence. The method employed by the server is the local compositions of amino acids and twin amino acids and local frequencies of the distance between successive amino acids like basic, hydrophobic and others for prediction of subcellular localization.

**ngLOC server:** The ngLOC server is primarily an n-gram based Bayesian classifier on-line server for the prediction of subcellular localization of proteins each in prokaryotes and eukaryotes from the protein sequences the usage of bayesian type approach that fashions the density distribution of amino acids. The distributions are decided from datasets of experimentally annotated subcellular localization of proteins from diverse organism.

**Gram-LocEN server:** Gram-LocEN is an interpretable multi-label online predictor which makes use of unified capabilities to yield sparse and interpretable answers for large-scale prediction of proteins of various species, along with Gram's negative and Gram-positive microorganisms. Given a question protein series in a selected species, a hard and fast of GO phrases are retrieved from newly created compact databases, specifically ProSeq-GO. The frequencies of GO occurrences are used to formulate frequency vectors with a dimensionality of 8,000+. Based on the chosen essential GO terms, the user of Gram-LocEN determines where a protein resides is located.

**CELLO2GO server:** CELLO2GO server, available at <http://cello.life.nctu.edu.tw/cello2go/>, is a publicly accessible web system for the detection of various properties of a target protein and its subcellular location. The server can be used for detailed Gene Ontology (GO) annotations, including subcellular locations, for queried proteins by combining the CELLO location prediction and BLAST homology search approaches.

## Methods

The total number of protein sequences encoded by *X. Oryzae pv. Oryzae* PXO99A in FASTA format were subjected to MP3 software analysis. The software predicts the virulence of sequence for a range of 0 to 1. The number 0 and 1 denotes the to which extent the predicted protein has a probability of being virulent protein. The value 1 was chosen to predict virulence as we intended for those proteins only that have a very high chance of involvement in the pathogenesis process. An only a positive value of Support Vector Machine (SVM) prediction was taken as a criterion for the determination of virulent protein as negative values reflect non- virulent proteins only. The second criterion used is the detection of protein as pathogenic by Pfam database. The selected protein from this analysis was subjected to the next analysis of the VirulentPred server and VICM server. The proteins which were predicted by a vote of a majority were taken as virulent proteins and were selected for further analysis and prediction for subcellular localization. The selected sequences were BLASTed locally against the Virulence Factor DataBase (VFDB) with an E-value less than 0.05 and query coverage of more than 70% and more for selection as virulent sequences. The selected sequences passing all of the above selection criteria were then subjected for analysis on various web servers for prediction of the subcellular localization. The selected sequences were subjected to Pfam analysis using webserver of European Bioinformatics Institute and is available at <https://www.ebi.ac.uk/Tools/hmmer/>. The HMMER web server is a fast and responsive homology search engine. This site is designed to provide a near interactive search for most searches along with intuitive and interactive display of results.

## Results and Discussion

### Prediction of virulent proteins

Using MP3 software, we were able to predict 499 protein sequences to be virulent with the assignment of the Pfam domain as a pathogenic protein. The highest SVM predicted score was 7.26 and the lowest was 1.38. The predicted pathogenic proteins were subjected to VirulentPred server analysis and 406 of the 499 proteins were predicted to be virulent. The VICM server predicted that only 62 of the 406 proteins were virulent. The MP3 software and both servers predicted 116 proteins as pathogens by majority vote. Table 1 contains the information of Predicted 116 proteins from MP3, VirulentPred and VICM server.

The proteins identified here were annotated as membrane proteins like transporter protein, membrane proteins, extracellular protein. These proteins have a role in the transportation of vital elements, signalling and passing of signals and secretory proteins, having a role in modulation of the infectious process and countering the defense mechanism of the host.

### The VFDB local BLAST analysis

The VFDB local BLAST analysis with E-values of 0.05 and query coverage

**Table 1.** Predicted 116 proteins from MP3, VirulentPred and VICM.

S. No.	NCBI accession No.	Description of Protein
1	ACD59217.1	Glutamate synthase domain 2
2	ACD59410.1	Glutamate synthase domain 2
3	ACD58782.1	YapH protein
4	ACD59161.1	Flagellar protein
5	ACD59353.1	Flagellar protein
6	ACD58786.1	Cation efflux system protein
7	ACD57443.1	Zinc metalloprotease
8	ACD59162.1	Flagellar protein
9	ACD59354.1	Flagellar protein
10	ACD61157.1	<i>Xanthomonas</i> adhesin-like protein B
11	ACD56969.1	Hemagglutinin-like protein
12	ACD59553.1	Protein U
13	ACD60995.1	<i>Xanthomonas</i> adhesin-like proteinA
14	ACD59556.1	Spore Coat Protein U domain family
15	ACD61350.1	ECF sigma factor
16	ACD59147.1	Flagellar protein
17	ACD56741.1	HrcS
18	ACD59159.1	Flagellar hook- associated protein FlgK
19	ACD59351.1	Flagellar hook-associated protein FlgK
20	ACD57416.1	Hypothetical protein PXO04232
21	ACD59190.1	Flagellar protein
22	ACD59382.1	Flagellar protein
23	ACD61484.1	Toluene tolerance protein
24	ACD56748.1	Type III secretion protein HrpB2
25	ACD59776.1	Hypothetical protein PXO01615
26	ACD60863.1	Hypothetical protein PXO02579
27	ACD58790.1	Serine protease
28	ACD61244.1	Filamentous hemagglutinin; hemagglutination activity domain protein
29	ACD60175.1	TonB family C-terminal domain protein
30	ACD56865.1	Hypothetical protein PXO03662
31	ACD60387.1	Hypothetical protein PXO01985
32	ACD59153.1	Flagellar biosynthesis hook protein
33	ACD59344.1	Flagellar biosynthesis hook protein
34	ACD61468.1	Poly vinyl alcohol dehydrogenase
35	ACD57184.1	Hypothetical protein PXO05451
36	ACD59651.1	Serine protease
37	ACD59554.1	Pili assembly chaperone
38	ACD59066.1	Heme exporter protein CcmB
39	ACD59257.1	Heme exporter protein CcmB
40	ACD57446.1	Lytic enzyme
41	ACD57847.1	Extracellular protease
42	ACD57315.1	Hypothetical protein PXO04040
43	ACD59152.1	Flagellar protein
44	ACD57366.1	Hypothetical protein PXO04091
45	ACD58506.1	Outer membrane protein OprN
46	ACD59186.1	Flagellar protein
47	ACD59378.1	Flagellar protein
48	ACD56921.1	Hypothetical protein PXO03603
49	ACD60040.1	gpU
50	ACD59155.1	Flagellar basal bodyrod protein FlgG
51	ACD59347.1	Flagellar basal bodyrod protein FlgG
52	ACD60061.1	Pre-pilin like leader sequence
53	ACD60006.1	Fimbrial biogenesis protein
54	ACD61119.1	Hypothetical protein PXO03102
55	ACD61033.1	ABC transporter substrate binding protein
56	ACD57711.1	Protease
57	ACD61409.1	TonB-dependent receptor
58	ACD57444.1	Neutral protease A
59	ACD60986.1	General secretion pathway protein I
60	ACD60805.1	Hypothetical protein PXO02514
61	ACD59148.1	Flagella basal body P-ring formation protein FlgA
62	ACD59339.1	Flagella basal body P-ring formation protein FlgA
63	ACD60882.1	TonB-dependent receptor
64	ACD59213.1	Rhs family protein
65	ACD59406.1	Rhs family protein
66	ACD60762.1	Murein hydrolaseD
67	ACD60057.1	PIIY1
68	ACD58523.1	Soluble lytic murein transglycosylase
69	ACD58061.1	OmpA family protein
70	ACD58915.1	RhsD protein
71	ACD60904.1	Outer membrane protein required for AvrXa21 activity C (raxC)
72	ACD60174.1	TonB-dependent receptor /Oar-like
73	ACD60980.1	General secretion pathway protein D
74	ACD61242.1	Filamentous hemagglutinin
75	ACD60993.1	Extracellular protease
76	ACD59750.1	TonB-dependent receptor
77	ACD59555.1	Outer-membrane usher protein FasD
78	ACD58522.1	Repressor
79	ACD59216.1	RHS Repeat family
80	ACD59409.1	RHS Repeat family
81	ACD56829.1	TonB-dependent receptor
82	ACD58825.1	Hypothetical protein PXO00667
83	ACD60628.1	TonB-dependent receptor
84	ACD57576.1	Surface antigen protein
85	ACD61616.1	Outer-membrane efflux protein
86	ACD60059.1	prepilin-typeN-terminal cleavage/methylation domain protein
87	ACD57752.1	Oxidoreductase
88	ACD60279.1	Hypothetical protein PXO02098
89	ACD57988.1	RhsD protein
90	ACD61123.1	Outer membrane protein
91	ACD59535.1	Outer-membrane efflux protein
92	ACD57008.1	TonB-dependent receptor
93	ACD61333.1	Hypothetical protein PXO02882
94	ACD56750.1	Type III secretion protein HrpB4
95	ACD59684.1	Regulatory protein
96	ACD60436.1	TonB-dependent outermembrane Receptor
97	ACD60577.1	TonB-dependent outermembrane Receptor
98	ACD60838.1	TonB-dependent outermembrane Receptor
99	ACD61158.1	Protease
100	ACD57183.1	Proteinase
101	ACD59052.1	TonB-dependent outermembrane Receptor
102	ACD56914.1	TonB-dependent outermembrane Receptor
103	ACD56913.1	TonB-dependent outermembrane Receptor
104	ACD60033.1	Phage-related baseplate protein
105	ACD59463.1	Hypothetical protein PXO01227
106	ACD57738.1	Putative secreted protein
107	ACD60540.1	TonB-dependent receptor
108	ACD56756.1	HrpA type III secretion outermembrane pore, YscC/Hrc family
109	ACD61597.1	Outer membrane protein
110	ACD60064.1	Multiple antibiotic resistance protein MarC
111	ACD57730.1	TonB-dependent receptor
112	ACD60250.1	TonB-dependent receptor
113	ACD56749.1	Type III secretion apparatus lipoprotein, YscJ/HrcJfamily
114	ACD60636.1	Tetratricopeptide repeat domain protein
115	ACD59540.1	Penicillinacylase II
116	ACD60703.1	TonB-dependent receptor

of 70% also confirmed that 116 proteins are a good candidate to be treated as virulent proteins. We can interpret that most of the proteins from the Table 2 we can interpret that a wide variety of proteins have been identified as virulent.

### Prediction of subcellular localization of protein

The subcellular localization of selected 116 proteins through various web server analysis yielded diverse results. A summarized result is given here. We can clearly analyze that most of the identified proteins belong to extracellular,

plasma membrane and cytoplasmic proteins, extracellular proteins are mostly secreted proteins, plasma membrane proteins (internal or external) play a role in channeling of metabolites across the membrane. They play a role in the transmission of signals across the cell membrane as well. Cytoplasmic proteins are important metabolites that play a role in modulating the information required for pathogenesis. Therefore, we can confirm that through *in silico* analysis could find some of the proteins which may play a defined role in the pathogenesis of *X. oryzae* in causing bacterial leaf blight in rice (*O. Sativa* L.) (Table 3).

**Table 2.** Prediction of subcellular localization of selected 116 proteins through various web server analysis.

S. No.	Accession No.	Description	Subcellular Location
1	ACD59217.1	Glutamate synthase domain 2	Extracellular/outermembrane/periplasmic
2	ACD59410.1	Glutamate synthase domain 2	Extracellular
3	ACD58782.1	YapH protein	Extracellular/Plasma membrane
4	ACD59161.1	Flagellar protein	Extracellular/Periplasm
5	ACD59353.1	Flagellar protein	Extracellular
6	ACD58786.1	Cation efflux system protein	Plasma membrane
7	ACD57443.1	Zinc metalloprotease	Extracellular
8	ACD59162.1	Flagellar protein	Extracellular
9	ACD59354.1	Flagellar protein	Extracellular
10	ACD61157.1	<i>Xanthomonas</i> adhesin-like protein B	Plasma membrane
11	ACD56969.1	Hemagglutinin-like protein	Cytoplasmic/Extracellular
12	ACD59553.1	Protein U	Extracellular
13	ACD60995.1	<i>Xanthomonas</i> adhesin-like proteinA	Extracellular
14	ACD59556.1	Spore Coat Protein U domain family	Extracellular
15	ACD61350.1	ECF sigma factor	Cytoplasmic
16	ACD59147.1	Flagellar protein	Extracellular /Plasma membrane
17	ACD56741.1	HrcS	Cytoplasmic /Plasma membrane
18	ACD59159.1	Flagellar hook- associated protein FlgK	Extracellular
19	ACD59351.1	Flagellar hook-associated protein FlgK	Extracellular /Plasma membrane
20	ACD57416.1	Hypothetical protein PXO04232	Plasma membrane/cytoplasmic
21	ACD59190.1	Flagellar protein	Plasma membrane
22	ACD59382.1	Flagellar protein	Plasma membrane
23	ACD61484.1	Toluene tolerance protein	Plasma membrane
24	ACD56748.1	Type III secretion protein HrpB2	Plasma membrane
25	ACD59776.1	Hypothetical protein PXO01615	Plasma membrane
26	ACD60863.1	Hypothetical protein PXO02579	Extracellular/Plasma membrane
27	ACD58790.1	Serine protease	Extracellular
28	ACD61244.1	Filamentous hemagglutinin; hemagglutination activity domain protein	Plasma membrane/cytoplasmic
29	ACD60175.1	TonB family C-terminal domain protein	Plasma membrane
30	ACD56865.1	Hypothetical protein PXO03662	Cytoplasmic/innermembrane
31	ACD60387.1	Hypothetical protein PXO01985	Plasma membrane/cytoplasmic
32	ACD59153.1	Flagellar biosynthesis hook protein	Extracellular/periplasm
33	ACD59344.1	Flagellar biosynthesis hook protein	Extracellular/cell membrane
34	ACD61468.1	Poly vinyl alcohol dehydrogenase	Extracellular /Plasma membrane
35	ACD57184.1	Hypothetical protein PXO05451	Plasma membrane/cytoplasmic
36	ACD59651.1	Serine protease	Extracellular
37	ACD59554.1	Pili assembly chaperone	Plasma membrane
38	ACD59066.1	Heme exporter proteinCcmB	Plasma inner membrane
39	ACD59257.1	Heme exporter proteinCcmB	Plasma inner membrane
40	ACD57446.1	Lytic enzyme	Plasma membrane
41	ACD57847.1	Extracellular protease	Extracellular
42	ACD57315.1	Hypothetical protein PXO04040	Plasma membrane
43	ACD59152.1	Flagellar protein	Extracellular/plasma membrane
44	ACD57366.1	Hypothetical protein PXO04091	outer cell membrane
45	ACD58506.1	Outer membrane proteinOprN	outer cell membrane
46	ACD59186.1	Flagellar protein	Plasma membrane
47	ACD59378.1	Flagellar protein	Plasma membrane
48	ACD56921.1	Hypothetical protein PXO03603	Plasma membrane
49	ACD60040.1	gpU	Cytoplasmic
50	ACD59155.1	Flagellar basal bodyrod protein FlgG	Extracellular
51	ACD59347.1	Flagellar basal bodyrod protein FlgG	Extracellular

52	ACD60061.1	Pre-pilin like leader sequence	Extracellular
53	ACD60006.1	Fimbrial biogenesis protein	Cytoplasmic
54	ACD61119.1	Hypothetical protein PXO03102	Plasma membrane
55	ACD61033.1	ABC transporter substrate binding protein	Plasma membrane
56	ACD57711.1	Protease	Extracellular/periplasmic
57	ACD61409.1	TonB-dependent receptor	Plasma membrane/cytoplasmic
58	ACD57444.1	Neutral protease A	Extracellular
59	ACD60986.1	General secretion pathway protein I	Plasma inner membrane
60	ACD60805.1	Hypothetical protein PXO02514	Plasma outer membrane
61	ACD59148.1	Flagella basal body P-ring formation protein FlgA	Plasma membrane
62	ACD59339.1	Flagella basal body P-ring formation protein FlgA	Plasma membrane
63	ACD60882.1	TonB-dependent receptor	Plasma membrane
64	ACD59213.1	Rhs family protein	Plasma membrane
65	ACD59406.1	Rhs family protein	Plasma membrane
66	ACD60762.1	Murein hydrolaseD	Plasma membrane/periplasmic
67	ACD60057.1	PilY1	Extracellular
68	ACD58523.1	Soluble lytic murein transglycosylase	Plasma membrane/periplasmic
69	ACD58061.1	OmpA family protein	Plasma membrane
70	ACD58915.1	RhsD protein	Extracellular
71	ACD60904.1	Outer membrane proteinrequiredforAvrXa21activityC(rax)	Plasma membrane
72	ACD60174.1	TonB-dependent receptor /Oar-like	Plasma membrane
73	ACD60980.1	General secretion pathway protein D	Plasma membrane
74	ACD61242.1	Filamentous hemagglutinin	Extracellular
75	ACD60993.1	Extracellular protease	Extracellular
76	ACD59750.1	TonB-dependent receptor	Plasma outer membrane
77	ACD59555.1	Outer-membrane usher protein FasD	Plasma outer membrane
78	ACD58522.1	Repressor	Cytoplasmic
79	ACD59216.1	RHS Repeat family	Plasma membrane
80	ACD59409.1	RHS Repeat family	Plasma membrane
81	ACD56829.1	TonB-dependent receptor	Plasma outer membrane
82	ACD58825.1	Hypothetical protein PXO00667	Plasma membrane/periplasmic
83	ACD60628.1	TonB-dependent receptor	Plasma outer membrane
84	ACD57576.1	Surface antigen protein	Plasma inner membrane
85	ACD61616.1	Outer-membrane efflux protein	Cytoplasmic /Plasma membrane
86	ACD60059.1	prepilin-typeN-terminalcleavage/methylationdomainprotein	Extracellular /Plasma membrane
87	ACD57752.1	Oxidoreductase	Plasma membrane
88	ACD60279.1	Hypothetical protein PXO02098	Plasma membrane
89	ACD57988.1	RhsD protein	Plasma membrane
90	ACD61123.1	Outer membrane protein	Plasma membrane/periplasmic
91	ACD59535.1	Outer-membrane efflux protein	Plasma membrane
92	ACD57008.1	TonB-dependent receptor	Plasma outer membrane
93	ACD61333.1	Hypothetical protein PXO02882	Plasma membrane/periplasmic
94	ACD56750.1	Type III secretion protein HrpB4	Plasma membrane
95	ACD59684.1	Regulatory protein	Plasma membrane
96	ACD60436.1	TonB-dependent outermembrane Receptor	Plasma membrane
97	ACD60577.1	TonB-dependent outermembrane Receptor	Plasma outer membrane
98	ACD60838.1	TonB-dependent outermembrane Receptor	Plasma outer membrane
99	ACD61158.1	Protease	Extracellular
100	ACD57183.1	Proteinase	Plasma membrane
101	ACD59052.1	TonB-dependent outermembrane Receptor	Plasma outermembrane
102	ACD56914.1	TonB-dependent outermembrane Receptor	Plasma membrane
103	ACD56913.1	TonB-dependent outermembrane Receptor	Plasma outermembrane
104	ACD60033.1	Phage-related baseplate protein	Extracellular
105	ACD59463.1	Hypothetical protein PXO01227	Extracellular
106	ACD57738.1	Putative secreted protein	Extracellular
107	ACD60540.1	TonB-dependent receptor	Plasma outermembrane
108	ACD56756.1	HrpA type III secretion outermembrane pore, YscC/HrcC family	Plasma outermembrane
109	ACD61597.1	Outer membrane protein	Plasma outermembrane
110	ACD60064.1	Multiple antibiotic resistance protein MarC	Plasma innermembrane
111	ACD57730.1	TonB-dependent receptor	Plasma outermembrane
112	ACD60250.1	TonB-dependent receptor	Plasma outermembrane

113	ACD56749.1	Type III secretion apparatus lipoprotein, YscJ/HrcJ family	Plasma outer membrane/periplasm
114	ACD60636.1	Tetratricopeptide repeat domain protein	Plasma membrane/periplasmic
115	ACD59540.1	Penicillinase II	Periplasm
116	ACD60703.1	TonB-dependent receptor	Plasma outer membrane

Table 3. Result of Pfam/Interpro analysis.

S. No.	Accession No.	Description	Pfam & Interpro Analysis
1	ACD59217.1	Glutamate synthase domain 2	No information available
2	ACD59410.1	Glutamate synthase domain 2	No information available
3	ACD58782.1	YapH protein	Role in the efficient transport of autotransporter virulence factors to the bacterial surface during growth and infection.
4	ACD59161.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
5	ACD59353.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
6	ACD58786.1	Cation efflux system protein	The OEP family (Outer membrane efflux protein) form trimeric channels that allow export of a variety of substrates in Gram's negative bacteria.
7	ACD57443.1	Zinc metalloprotease	This group of metalloproteases belong to the MEROPS peptidase family and includes a number of enzymes from a number of pathogenic bacteria.
8	ACD59162.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
9	ACD59354.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
10	ACD61157.1	<i>Xanthomonas</i> adhesin-like protein B	Responsible for the autotransport function in bacteria.
11	ACD56969.1	Hemagglutinin-like protein	Secreted protein considered virulence factors in many Gram's negative pathogens.
12	ACD59553.1	Protein U	A secretory precursor, pro-protein U, which is then secreted across the membrane.
13	ACD60995.1	<i>Xanthomonas</i> adhesin-like protein A	Class of a proteins found on the outer membrane used to infect their hosts.
14	ACD59556.1	Spore Coat Protein U domain family	A family of secreted pili proteins involved in motility and biofilm formation.
15	ACD61350.1	ECF sigma factor	Bacterial core RNA polymerase complex, responsible for transcription elongation and termination.
16	ACD59147.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
17	ACD56741.1	HrcS	Exporter proteins.
18	ACD59159.1	Flagellar hook-associated protein FlgK	Flagellar hook protein role in bacterial movement.
19	ACD59351.1	Flagellar hook-associated protein FlgK	Flagellar hook protein role in bacterial movement.
20	ACD57416.1	Hypothetical protein PXO04232	Functionally uncharacterised.
21	ACD59190.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
22	ACD59382.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
23	ACD61484.1	Toluene tolerance protein	Component of the Mla pathway, functions to maintain the asymmetry of the outer membrane.
24	ACD56748.1	Type III secretion protein HrpB2	Type III secretion operons in a narrow group of species including <i>Xanthomonas</i> .
25	ACD59776.1	Hypothetical protein PXO01615	Functionally uncharacterised.
26	ACD60863.1	Hypothetical protein PXO02579	Functionally uncharacterised.
27	ACD58790.1	Serine protease	Proteolytic enzymes.
28	ACD61244.1	Filamentous hemagglutinin; hemagglutination activity domain protein	Bacterial exotoxins.
29	ACD60175.1	TonB family C-terminal domain protein	Inner membrane-bound transporter protein with role in iron transport.
30	ACD56865.1	Hypothetical protein PXO03662	Functionally uncharacterised.
31	ACD60387.1	Hypothetical protein PXO01985	Functionally uncharacterised.
32	ACD59153.1	Flagellar biosynthesis hook protein	Flagellar hook proteins responsible in movement.
33	ACD59344.1	Flagellar biosynthesis hook protein	Flagellar hook proteins responsible in movement.
34	ACD61468.1	Poly vinyl alcohol dehydrogenase	Co-factor for a number of enzymes (quinoproteins) and particularly for some bacterial dehydrogenase.
35	ACD57184.1	Hypothetical protein PXO05451	Functionally uncharacterised.
36	ACD59651.1	Serine protease	Responsible for the autotransport function in bacteria.
37	ACD59554.1	Pili assembly chaperone	Required to assemble pili, but their assembly only takes place after translocation across the cytoplasmic membrane
38	ACD59066.1	Heme exporter protein CcmB	Necessary for cytochrome c biosynthesis in eubacteria.
39	ACD59257.1	Heme exporter protein CcmB	Necessary for cytochrome c biosynthesis in eubacteria.
40	ACD57446.1	Lytic enzyme	No information available.
41	ACD57847.1	Extracellular protease	Proteolytic enzymes having role in protein degradation.
42	ACD57315.1	Hypothetical protein PXO04040	Functionally uncharacterised.
43	ACD59152.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
44	ACD57366.1	Hypothetical protein PXO04091	Functionally uncharacterised.
45	ACD58506.1	Outer membrane protein OprN	Outer membrane efflux protein form trimeric (three-piece) channels that allow export of a variety of substrates
46	ACD59186.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
47	ACD59378.1	Flagellar protein	Responsible for flagellin's ability to polymerise into a filament.
48	ACD56921.1	Hypothetical protein PXO03603	Functionally uncharacterised.
49	ACD60040.1	gpU	Involved in bacterial locomotion.
50	ACD59155.1	Flagellar basal body rod protein FlgG	Involved in bacterial locomotion.

51	ACD59347.1	Flagellar basal body rod protein FlgG	Involved in bacterial locomotion.
52	ACD60061.1	Pre-pilin like leader sequence	Act as initiators or regulators of the biogenesis and dynamics of the pilus.
53	ACD60006.1	Fimbrial biogenesis protein	Involved in type II bacterial export systems
54	ACD61119.1	Hypothetical protein PXO03102	Functionally uncharacterised.
55	ACD61033.1	ABC transporter substrate binding protein	Part of the Mla pathway, it acts to maintain the asymmetry of the outer membrane.
56	ACD57711.1	Protease	Proteolytic enzymes having role in protein degradation.
57	ACD61409.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
58	ACD57444.1	Neutral protease A	Secreted eubacterial endopeptidases.
59	ACD60986.1	General secretion pathway protein I	Role in pilus biogenesis, competence proteins, and type II secretion proteins.
60	ACD60805.1	Hypothetical protein PXO02514	Functionally uncharacterised
61	ACD59148.1	Flagella basal body P-ring formation protein FlgA	Chaperones for flagellar basal-body proteins and pilus-assembly proteins.
62	ACD59339.1	Flagella basal body P-ring formation protein FlgA	Chaperones for flagellar basal-body proteins and pilus-assembly proteins.
63	ACD60882.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
64	ACD59213.1	Rhs family protein	RHS repeats form an extended strip of beta-sheet that spirals around to form a hollow shell, role in flagella formation.
65	ACD59406.1	Rhs family protein	RHS repeats form an extended strip of beta-sheet that spirals around to form a hollow shell, role in flagella formation.
66	ACD60762.1	Murein hydrolase D	Part of type II, type III and type IV secretion systems.
67	ACD60057.1	PIIY1	Role in mobility.
68	ACD58523.1	Soluble lytic murein transglycosylase	Part of type II, type III and type IV secretion systems.
69	ACD58061.1	OmpA family protein	Interact with host receptor molecules.
70	ACD58915.1	RhsD protein	Part of ABC toxin complexes.
71	ACD60904.1	Outer membrane protein required for AvrXa21 activity (raxC)	Allow export of a variety of substrates in Gram's negative bacteria.
72	ACD60174.1	TonB-dependent receptor /Oar-like	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
73	ACD60980.1	General secretion pathway protein D	Involvement in type II secretion pathway.
74	ACD61242.1	Filamentous hemagglutinin	Bacterial exotoxins.
75	ACD60993.1	Extracellular protease	Proteolytic enzymes.
76	ACD59750.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
77	ACD59555.1	Outer-membrane usher protein FasD	Involvement in biogenesis of the pilus in Gram's negative bacteria.
78	ACD58522.1	Repressor	Role in transcriptional regulation response.
79	ACD59216.1	RHS Repeat family	RHS repeats form an extended strip of beta-sheet that spirals around to form a hollow shell, role in flagella formation.
80	ACD59409.1	RHS Repeat family	RHS repeats form an extended strip of beta-sheet that spirals around to form a hollow shell, role in flagella formation.
81	ACD56829.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
82	ACD58825.1	Hypothetical protein PXO00667	Functionally uncharacterised.
83	ACD60628.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
84	ACD57576.1	Surface antigen protein	Includes a transmembrane alpha-helical region that contains glycine zipper motifs.
85	ACD61616.1	Outer-membrane efflux protein	The OEP family (Outer membrane efflux protein) form trimeric channels that allow export of a variety of substrates in Gram's negative bacteria.
86	ACD60059.1	prepilin-type N-terminal cleavage/ methylation domain protein	Role in type IV pilus biogenesis process.
87	ACD57752.1	Oxidoreductase	Belongs to the family of transporter proteins.
88	ACD60279.1	Hypothetical protein PXO02098	Functionally uncharacterised
89	ACD57988.1	RhsD protein	RHS repeats form an extended strip of beta-sheet that spirals around to form a hollow shell, role in flagella formation.
90	ACD61123.1	Outer membrane protein	Outer membrane protein.
91	ACD59535.1	Outer-membrane efflux protein	Part of outer membrane efflux protein channel, allow export of a variety of substrates in Gram's negative bacteria.
92	ACD57008.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
93	ACD61333.1	Hypothetical protein PXO02882	Functionally uncharacterised.
94	ACD56750.1	Type III secretion protein HrpB4	Type III secretion operons in a narrow group of species including <i>Xanthomonas</i> .
95	ACD59684.1	Regulatory protein	Proteins involved in heat shock response.
96	ACD60436.1	TonB-dependent outer membrane Receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
97	ACD60577.1	TonB-dependent outer membrane Receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.

98	ACD60838.1	TonB-dependent outermembrane Receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
99	ACD61158.1	Protease	Proteolytic enzymes having role in protein degradation.
100	ACD57183.1	Proteinase	No results.
101	ACD59052.1	TonB-dependent outermembrane Receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
102	ACD56914.1	TonB-dependent outermembrane Receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
103	ACD56913.1	TonB-dependent outermembrane Receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
104	ACD60033.1	Phage-related baseplate protein	Type VI secretion system.
105	ACD59463.1	Hypothetical protein PXO01227	Functionally uncharacterised.
106	ACD57738.1	Putative secreted protein	Hydrolyse the glycosidic bond between two or more carbohydrates, role in energy driven processes.
107	ACD60540.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
108	ACD56756.1	HrpA type III secretion outermembrane pore, YscC/HrcC family	Involved in the type II and type III secretion pathway (GSP) in Gram's negative bacteria, responsible for protein export.
109	ACD61597.1	Outer membrane protein	Outer membrane protein.
110	ACD60064.1	Multiple antibiotic resistance protein MarC	The function of this family is not clear
111	ACD57730.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
112	ACD60250.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.
113	ACD56749.1	Type III secretion apparatus lipoprotein, YscJ/HrcJ family	Member of the YscJ family, involved in secretion of several proteins.
114	ACD60636.1	Tetratricopeptide repeat domain protein	Part of plasma membrane protein other extracellular proteins.
115	ACD59540.1	Penicillinacylase II	Role in penicillin acylase activity.
116	ACD60703.1	TonB-dependent receptor	TonB protein interacts with receptor proteins of the outer membrane, carry out a high-affinity binding and energy-dependent absorption of specific substrates in the periplasmic space.

## Conclusion

Results from Pfam/Interpro analysis using webserver of European Bioinformatics Institute is given here in Table 3. The function of these proteins was referenced by Pfam database and Interpro server analysis. From the above Table 3 it can be seen that most of the identified proteins play a role in host recognition as a part of the outer membrane protein, in the movement of bacteria as a part of the pathways responsible for formation and assembly of flagella or cilia, in the energy regulation as part of cytochrome proteins, part of the secretion system of type II, III, IV, transporters of various secreted proteins, as exotoxins and proteolytic enzymes play a role in the breakdown of the host cell wall, as heat shock protein plays a role in combating the fluctuations in ambient temperature. The proteins identified here are good candidates for further computational and wet laboratory analysis for analysis of various pathways responsible for infection and survival processes of the pathogen.

## Acknowledgement

Above work was carried out in the Bioinformatics lab of Dept. of Microbiology and Bioinformatics of Atal Bihari Vajpayee Vishwavidyalaya, Bilaspur, Chhattisgarh, India.

## Conflict of Interest

None.

## References

1. Ben Hassen, Tarek, Hamid El Bilali and Mohammad S. Allahyari. "Impact of COVID-19 on food behavior and consumption in Qatar." *Sustainability* 12 (2020): 6973.
2. Syed-Ab-Rahman, Sharifah Farhana, Lilia Costa Carvalhais and Dzulkhifli Omar. "Development of plant-based emulsion formulations to control bacterial leaf blight and sheath brown rot of rice." *Heliyon* 6 (2020).
3. Shaheen, Rozina, Muhammad Zahid Sharif, Luqman Amrao and Aiping Zheng, et al. "Investigation of bacterial leaf blight of rice through various detection tools

and its impact on crop yield in Punjab, Pakistan." *Pak J Bot* 51 (2019): 307-312.

4. Joshi, Johnson Beslin, Loganathan Arul, Jegadeesan Ramalingam and Sivakumar Uthandi. "Advances in the Xoo-rice pathosystem interaction and its exploitation in disease management." *J Biosci* 45 (2020): 1-17.
5. Shen, Yuwei and Pamela Ronald. "Molecular determinants of disease and resistance in interactions of *X. oryzae pv. oryzae* and rice." *Microbes Infect* 4 (2002): 1361-1367.
6. Furutani, Ayako, Minako Takaoka, Harumi Sanada and Yukari Noguchi, et al. "Identification of novel type III secretion effectors in *X. oryzae pv. oryzae*." *Mol Plant-Microbe* 22 (2009): 96-106.
7. Jensen, L. Juhl, Ramneek Gupta, Nikolaj Blom and D. Devos, et al. "Prediction of human protein function from post-translational modifications and localization features." *J Mol Biol* 319 (2002): 1257-1265.
8. Salzberg, Steven L., Daniel D. Sommer, Michael C. Schatz and Adam M. Phillippy, et al. "Genome sequence and rapid evolution of the rice pathogen *X. oryzae pv. oryzae* PXO99 A." *BMC Genom* 9 (2008): 1-16.
9. Gupta, Ankit, Rohan Kapil, Darshan B. Dhakan and Vineet K. Sharma. "MP3: A software tool for the prediction of pathogenic proteins in genomic and metagenomic data." *PLoS one* 9 (2014): e93907.
10. Garg, Aarti and Dinesh Gupta. "VirulentPred: A SVM based prediction method for virulent proteins in bacterial pathogens." *BMC Bioinform* 9 (2008): 1-12.
11. Saha, Sudipto and G. P. S. Raghava. "VICMpred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition." *Genom Proteom Bioinform* 4 (2006): 42-47.

**How to cite this article:** Kashyap, Dharmendra and Aafreen Khan. "In-Silico Subcellular Localization and Functional Analysis of Computationally Predicted Virulent Proteins in *X. Oryzae pv. Oryzae* Strain PXO99A Causal Organism of Bacterial Leaf Blight (BLB) of Rice (*O. Sativa* L.)." *Mol Bio* 12 (2023): 377.