Revista Argentina de Clínica Psicológica
2019, Vol. XXVIII, N°5, 795-806
DOI: 10.53555/03276716.2019.01

Computational Prediction and Analysis of Signal Peptide-Bearing TALEs across *Xanthomonas oryzae* Strains Pathogenic to Rice.

Dharmendra Kashyap[a*]

Abstract

Rice (Oryza sativa L.) is a vital food crop for over half of the world's population, with global production reaching approximately 470.63 million metric tonnes. India, the second-largest producer after China, cultivates rice across 42.75 million hectares, achieving an average yield of 3.61 metric tonnes per hectare. However, rice is highly susceptible to diseases caused by fungi, bacteria, viruses, and nematodes, which significantly impact yields. Bacterial leaf blight, caused by *Xanthomonas oryzae pv. oryzae* (*Xoo*), can result in yield losses of up to 70%, while bacterial leaf streak, caused by *Xanthomonas oryzae pv. oryzicola* (*Xoc*), can lead to losses of 20%.

Transcription activator-like effectors (TALEs) are DNA-binding proteins employed by *Xanthomonas* bacteria to enhance virulence in host plants by inducing the expression of host susceptibility genes. TALEs are translocated into plant cells via the type III secretion system and possess a central domain with tandem repeats that determine DNA-binding specificity. Each repeat recognises a single base pair, with specific amino acids (RVDs) dictating base recognition. This research investigates the variability of TALEs in different *Xanthomonas* strains, focusing on *Xoo* and *Xoc*. Thirty-four strains were analysed using whole genome sequences obtained from NCBI. Tools such as AnnoTALE, MP3, UniProt, SMS Suite, Clustal X, PSORTb, SignalP, and T346Hunter were utilised for TALE identification, protein prediction, and functional analysis. AnnoTALE identified 592 TALEs across the strains, with further analysis revealing pathogenic proteins, molecular characteristics, and secretion mechanisms.

Examining transcription activator-like effectors (TALEs) in *Xanthomonas oryzae* has provided valuable insights into how they induce disease in various bacterial types. A total of 592 TALEs were identified from 34 strains, highlighting the genetic diversity of the pathogen. MP3 software predicted that 572 of these proteins are pathogenic, suggesting their involvement in virulence. UniProt analysis identified six homologous protein entries, emphasising their significance in host interactions.

The SMS Suite analysis indicated that the molecular weights of these TALEs range from 80 to 160 kDa, with isoelectric points between 6 and 8, which are crucial for understanding their biochemical properties. Clustal X clustering revealed major protein clusters, indicating conserved sequences that may be critical for their function. PSORTb predictions suggested that most proteins are cytoplasmic, aligning with their role in manipulating host gene expression. Additionally, SignalP identified nine proteins with signal peptides necessary for secretion into host cells, while T346Hunter confirmed the presence of type III secretion systems (T3SS), essential for TALE translocation into host cells. Overall, 310 non-redundant TALEs were identified for further analysis, providing insights into the virulence mechanisms of *Xanthomonas* species and guiding future strategies to combat rice diseases.

Keywords: Rice (Oryza sativa L.), Transcription Activator-Like Effectors (TALEs), pathogenicity, SignalP server, NCBI Blast.

Introduction

Rice (*Oryza sativa* L.) is a crucial food crop globally, sustaining over half of the world's population. The total global production of rice is approximately 470.63 million metric tonnes, cultivated over an area of 157.46 million hectares, yielding an average of 4.46 metric tonnes per hectare. India ranks second in rice production after China, where rice serves as a staple in the Indian diet, cultivated on 42.75 million hectares and achieving a productivity of 3.61 metric tonnes per hectare. However, rice is highly susceptible to various diseases caused by fungi, bacteria, viruses, and nematodes, which significantly impact both the quality and quantity of yields. Notably, bacterial leaf blight, caused by the bacterium *Xanthomonas oryzae*

[a*]**Assist. Professor, Department of Microbiology & Bioinformatics, Atal Bihari Vajpayee Vishwavidyalaya, formarly named as Bilaspur University, Bilaspur, Chhattisgarh, India Pin-495009**

***Corresponding Author: Dharmendra Kashyap**
***Assist Professor, Department of Microbiology & Bioinformatics, EmailID: kashyapdk97@gmail.com Atal Bihari Vajpayee Vishwavidyalaya, Bilaspur (C.G.) India Pin 495009**

pv. oryzae (*Xoo*), is one of the most severe diseases, potentially leading to yield losses of up to 70%. *Xoo* infects rice leaves through hydathodes and wounds, spreading through the plant's xylem. Another significant threat is bacterial leaf streak, caused by *Xanthomonas oryzae pv. oryzicola* (*Xoc*), a close relative of *Xoo*, which can cause yield losses of up to 20%. *Xoc* infects rice through stomata and colonises the intercellular spaces of leaf tissues. Resistance mechanisms against these pathogens include specific loci, such as the XOI locus, which is associated with qualitative resistance to *Xoc*. Multiple resistance quantitative trait loci (QTLs) have also been identified against *Xoc*, highlighting the ongoing efforts to combat these diseases in rice cultivation.[2,3]

Plant disease resistance is a complex system involving a two-tier innate immune response. There are two main types of immunity: pattern-triggered immunity (PTI) and effector-triggered immunity (ETI). PTI is initiated when host pattern recognition receptors (PRRs) on the plasma membrane recognise pathogen-associated molecular patterns (PAMPs), resulting in a basal, qualitative resistance. ETI is triggered when cytoplasmic nucleotide-binding leucine-rich repeat (NB-LRR) proteins recognise pathogen effectors, leading to a strong, race-specific resistance. This type of resistance is controlled by major resistance (R) genes and is often quantitative.[4,5,6]

In some instances, plants utilise induced immunity to prevent pathogen attacks following an initial local infection. This resistance is termed systemic acquired resistance (SAR) or induced systemic resistance (ISR), involving plant hormones such as salicylic acid (SA), jasmonic acid (JA), and ethylene in systemic defence signalling. In addition to receptor genes (PRRs and R genes), numerous defence-related (DR) genes respond to pathogen infection and are often characterised as quantitative trait loci (QTLs) in rice. Some resistance genes provide broad-spectrum resistance (BSR) to multiple pathogen species, regulated by single genes.[7,8]

Transcription activator-like effector (TALE) proteins secreted by pathogens play a crucial role in plant resistance engineering. TALEs bind to specific DNA sequences known as effector binding elements (EBEs) in the host nucleus, transcriptionally activating downstream susceptibility (S) genes that promote disease development. During host-pathogen co-evolution, plants have developed variations in EBEs and genes encoding TALE helper proteins, leading to the evolution of recessively inherited resistance (r) genes. Furthermore, plants have also evolved dominantly inherited, transcriptionally controlled R genes composed of a promoter-embedded EBE and a downstream-encoded executor R protein. Understanding TALE biology and recent genome editing techniques can facilitate the design of new crop plants resistant to TALEs or TALE-like proteins, providing a promising approach for enhancing disease resistance in agriculture.[9,10,11,12]

The present research aims to investigate the variability of transcription activator-like effectors (TALEs) among different species, pathovars, and strains of *Xanthomonas*, with a particular focus on the rice pathogens *Xanthomonas oryzae pv. oryzae* and *Xanthomonas oryzae pv. oryzicola*. Notably, certain strains of *Xanthomonas oryzae pv. oryzae* possess over 20 TALEs, with more than 100 TALE-encoding genes identified, which are associated with either resistance or susceptibility-inducing traits. Recent advancements have led to the development of several algorithms for *insilico* TALE target prediction. This research seeks to utilise these algorithms for the *insilico* identification, characterisation, modelling, and docking studies of TALEs from the aforementioned rice pathogens. By elucidating the mechanisms of TALE function and their interactions with host plant genes, this study aims to contribute to a deeper understanding of the pathogenicity of *Xanthomonas* and to inform strategies for enhancing disease resistance in rice. The variability of TALEs across different *Xanthomonas* strains is crucial for understanding their role in pathogenicity. Previous studies have shown that the TALE repertoires in *Xanthomonas oryzae* strains are highly diverse, with implications for their virulence and the development of resistant rice varieties. The identification and characterisation of these effectors will provide insights into their evolutionary adaptations and functional relationships, which are essential for devising effective disease management strategies. Furthermore, the integration of computational tools for TALE analysis will facilitate a more comprehensive understanding of plant-pathogen interactions. By employing these advanced methodologies, the research aims to uncover the intricate dynamics between TALEs and host susceptibility genes, thereby enhancing our ability to develop targeted approaches for improving disease resistance in rice cultivation. This study not only seeks to advance the field of plant pathology but also aims to contribute to global food security by addressing the challenges posed by *Xanthomonas* pathogens in rice production.

.

Materials and methods

Thirty four strains of *Xoo* and *Xoc* werew under investigation and their whole genome were downloaded from the from NCBI's FTP facility.

Table no 01 : Details of Xoo and Xoc strains under investigation

sno	Organism/Name	Strain	Size (Mb)	GC%	Replicons	Gene	Protein
1	*Xanthomonas oryzae pv. oryzae*	PXO99A	5.23855	63.6	chromosome: NC_010717.2/ CP000967.2	5058	4031
2	*Xanthomonas oryzae pv. oryzae*	MAFF 311018	4.94022	63.7	chromosome: NC_007705.1/ AP008229.1	4852	3940
3	*Xanthomonas oryzae pv. oryzicola*	BLS256	4.83175	64.1	chromosome: NC_017267.2/ CP003057.2	4493	3674
4	*Xanthomonas oryzae pv. oryzicola*	CFBP7342	5.0801	64	chromosome: NZ_CP007221.1/CP007221.1	4811	3914
5	*Xanthomonas oryzae pv. oryzae*	PXO86	5.01662	63.7	chromosome: NZ_CP007166.1/CP007166.1	4847	3950
6	*Xanthomonas oryzae pv. oryzicola*	CFBP2286	5.00599	63.98	chromosome: NZ_CP011962.1/CP011962.1	4709	3883
7	*Xanthomonas oryzae pv. oryzicola*	B8-12	4.79432	64.1	chromosome: NZ_CP011955.1/CP011955.1	4484	3694
8	*Xanthomonas oryzae pv. oryzicola*	BLS279	4.79062	64.1	chromosome: NZ_CP011956.1/CP011956.1	4485	3693
9	*Xanthomonas oryzae pv. oryzicola*	BXOR1	4.69259	64.1	chromosome: NZ_CP011957.1/CP011957.1	4367	3594
10	*Xanthomonas oryzae pv. oryzicola*	CFBP7331	5.00829	63.9	chromosome: NZ_CP011958.1/CP011958.1	4711	3811
11	*Xanthomonas oryzae pv. oryzicola*	CFBP7341	5.01777	63.9	chromosome: NZ_CP011959.1/CP011959.1	4716	3821
12	*Xanthomonas oryzae pv. oryzicola*	L8	4.79653	64.1	chromosome: NZ_CP011960.1/CP011960.1	4479	3683
13	*Xanthomonas oryzae pv. oryzicola*	RS105	4.77995	64.1	chromosome: NZ_CP011961.1/CP011961.1	4480	3678
14	*Xanthomonas oryzae pv. oryzae*	AXO1947	4.67497	63.9	chromosome: NZ_CP013666.1/CP013666.1	4392	3533
15	*Xanthomonas oryzae pv. oryzae*	PXO83	5.02543	63.7	chromosome: NZ_CP012947.1/CP012947.1	4854	3953
16	*Xanthomonas oryzae pv. oryzae*	PXO71	4.907	63.7	chromosome: NZ_CP013670.1/CP013670.1	4807	3914
17	*Xanthomonas oryzae pv. oryzae*	PXO145	5.03976	63.7	chromosome: NZ_CP013961.	4839	3852

					1/CP013961.1		
18	*Xanthomonas oryzae pv. oryzae*	PXO211	5.03335	63.7	chromosome: NZ_CP013674.1/CP013674.1	4872	3951
19	*Xanthomonas oryzae pv. oryzae*	PXO236	4.96872	63.7	chromosome: NZ_CP013675.1/CP013675.1	4799	3907
20	*Xanthomonas oryzae pv. oryzae*	PXO524	4.9543	63.7	chromosome: NZ_CP013677.1/CP013677.1	4851	3960
21	*Xanthomonas oryzae pv. oryzae*	PXO563	4.93631	63.7	chromosome: NZ_CP013678.1/CP013678.1	4840	3933
22	*Xanthomonas oryzae pv. oryzae*	PXO602	4.95179	63.7	chromosome: NZ_CP013679.1/CP013679.1	4832	3932
23	*Xanthomonas oryzae pv. oryzae*	XF89b	4.96674	63.7	chromosome: NZ_CP011532.1/CP011532.1	4851	3947
24	*Xanthomonas oryzae pv. oryzae*	MAI73	4.70398	63.9	chromosome: NZ_CP019086.1/CP019086.1	4488	3598
25	*Xanthomonas oryzae pv. oryzae*	MAI145	4.70398	63.9	chromosome: NZ_CP019092.1/CP019092.1	4487	3607
26	*Xanthomonas oryzae pv. oryzae*	MAI68	4.70378	63.9	chromosome: NZ_CP019085.1/CP019085.1	4483	3605
27	*Xanthomonas oryzae pv. oryzae*	MAI106	4.70545	63.9	chromosome: NZ_CP019089.1/CP019089.1	4492	3586
28	*Xanthomonas oryzae pv. oryzae*	MAI129	4.70396	63.9	chromosome: NZ_CP019090.1/CP019090.1	4489	3609
29	*Xanthomonas oryzae pv. oryzae*	MAI134	4.73014	63.9	chromosome: NZ_CP019091.1/CP019091.1	4541	3639
30	*Xanthomonas oryzae pv. oryzae*	MAI95	4.70504	63.9	chromosome: NZ_CP019087.1/CP019087.1	4489	3598
31	*Xanthomonas oryzae pv. oryzae*	MAI99	4.69882	63.9	chromosome: NZ_CP019088.1/CP019088.1	4485	3611
32	*Xanthomonas oryzae pv. oryzae*	MAI1	4.73521	63.9	chromosome: NZ_CP025609.1/CP025609.1	4521	3652
33	*Xanthomonas oryzae pv. oryzae*	BAI3	4.72388	63.9	chromosome: NZ_CP025610.1/CP025610.1	4520	3640
34	*Xanthomonas oryzae pv. oryzae*	PXO61	4.99936	63.7	chromosome: NZ_CP020942.1/CP020942.1	4957	3830

The analysis was conducted on a PC with both Windows and Linux operating systems installed. The minimum hardware requirements include: At least 8 GB of RAM A CPU with a clock speed of at least 2.5 GHz The following software was utilized: Microsoft Office or Open Office for document processing Notepad++ for text editing Active Perl (version 5.26.3 or higher) for Windows, available at https://www.activestate.com/activeperl Java Runtime Environment (JRE) version 1.7 or higher, NCBI BLAST+ executable version 2.7.1 for local BLAST searches

The AnnoTALE Software Suite

The AnnoTALE software suite is specifically designed for the identification and analysis of transcription activator-like effectors (TALEs) within the genomes of *Xanthomonas*. This suite facilitates the clustering of TALEs based on their repeat variable di-residue (RVD) sequences, assigns novel TALEs to existing classes, proposes standardised nomenclature for TALEs, and predicts the targets of individual TALEs and their respective classes. AnnoTALE is available as a JavaFX-based standalone application, featuring a graphical user interface that supports interactive analysis sessions. For this study, AnnoTALE version 1.2 was utilised to identify TALEs across 34 strains of *Xanthomonas oryzae pv. oryzae* (*Xoo*) and *Xanthomonas oryzae pv. oryzicola* (*Xoc*). The specific version employed, AnnoTALE 1.2 (6GB), is capable of utilising 6GB of RAM for computational tasks and is distributed as a JAR file requiring a Java Runtime Environment (JRE) on both Linux and Windows operating systems. An installable version is also available for ease of installation. The genome sequences for the 34 strains were downloaded from the NCBI database for analysis.[14]

The MP3 Software Suite

The MP3 software suite is employed for predicting pathogenic proteins in genomic and metagenomic datasets. This tool utilises two methodologies: Support Vector Machines (SVM) and Hidden Markov Models (HMM). The standalone version of MP3 was implemented on a Linux operating system to identify proteins predicted by Pfam.[15]

UniProt Analysis

The UniProt database serves as a primary resource for protein sequences and is accessible at http://www.uniprot.org/. To identify homologous sequences, NCBI BLAST was performed on all 592 sequences against the UniProt database, employing a low E-value threshold of 0.005 to ensure the identification of nearest homologous sequences. Subsequent detailed analyses were conducted on the identified sequences.

The SMS Suite

The Sequence Manipulation Suite (SMS) comprises a collection of JavaScript programmes designed for generating, formatting, and analysing short DNA and protein sequences. It is widely used in molecular biology for educational purposes and for testing programmes and algorithms. The SMS is available at http://www.bioinformatics.org/sms2/, enabling various analyses related to the biophysical characteristics of proteins.[16]

The Clustal X Software

Clustal X was employed for clustering protein sequences. It comes in two versions: Clustal W, which operates via the command line, and Clustal X, which provides a graphical interface. The latest version, Clustal 2.1, is available for download as precompiled executables for Linux, Mac OS X, and Windows operating systems, with the source code also accessible. [17]

The PSORTb Server

The PSORTb server is a computational tool used for predicting the subcellular localization of proteins, which is essential for genome analysis and annotation. Understanding a protein's localization can provide insights into its function, particularly for bacterial pathogens, where surface proteins may serve as potential drug or vaccine targets. The prediction of localization is influenced by features in the protein's primary structure, such as signal peptides and membrane-spanning alpha-helices. For this analysis, PSORTb version 3.0 was utilised, accessible at http://www.psort.org/downloads/index.html. [18]

The SignalP Server

The SignalP server, available at www.cbs.dtu.dk/services/SignalP/, consists of three components, two of which were employed for this analysis:

SecretomeP 2.0: This server predicts non-classical protein secretion, which does not rely on signal peptides. It integrates data from multiple feature prediction servers to provide comprehensive predictions regarding post-translational modifications and localization aspects of proteins. [19]

***SignalP 3.0*:** This server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from various organisms, including Gram-positive and Gram-negative prokaryotes as well as eukaryotes. It employs a combination of artificial neural networks and hidden Markov models for accurate predictions. [20]

The T346Hunter Server

T346Hunter is a web application available at http://bacterial-virulence-factors.cbgp.upm.es/

T346Hunter designed to identify type III, type IV, and type VI secretion systems (T3SS, T4SS, and T6SS) in bacterial genomes. These secretion systems are critical for delivering effector proteins, influencing the pathogenesis of bacterial strains. T346Hunter utilises a database of hidden Markov model (HMM) protein profiles to perform BLASTp and HMMER searches against user-provided bacterial sequences, identifying enriched regions associated with these secretion systems. [21]

Local BLAST of 34 strains

Following the identification of TALEs in *Xanthomonas oryzae pv. oryzae (Xoo)* using AnnoTALE, a local BLAST analysis was conducted to identify annotated proteins in 34 strains of *Xoo* from the NCBI database. This analysis will utilise the 592 TALEs predicted by AnnoTALE as a database to find corresponding sequences in each strain. Proteome for all 34 strains were downloaded for further analysis, with the BLAST executables available at https://blast.ncbi.nlm.nih.gov/. [22,23]

UniProt Analysis of Identified Proteins in 34 Strains

A UniProt analysis was performed using NCBI BLAST to identify sequences from the previously mentioned criteria. This analysis aims to locate experimentally validated sequences within the predicted sequences from the 34 strains of *Xoo*. The identification of experimentally proven sequences is crucial for accurately predicting various parameters in the analysed sequences.

RESULTS AND DISCUSSION

The AnnoTALE suite results analysis

After downloading the FastA file for the Genome for all 34 strains, the AnnoTALEs software was employed for identification of TALEs. For all 34 strains of *Xanthomonas oryzae* collectively 592 TALEs were computationally identified. The files were available in fastA format for further analysis.

Table no 02:: List of computationally identified TALEs for 24 strains of *Xoo* and 10 strains of *Xoc.*

S no	*Xanthomonas oryzae oryzae strains*	No of TALEs predicted	S no	*Xanthomonas oryzae pv oryzae strains*	No of TALEs predicted	S no	*Xanthomonas oryzae pv oryzaecola strains*	No of TALEs predicted
1	PXO99A	**19**	**13**	PXO602	**20**	**25**	BLS256	28
2	MAFF 311018	**17**	**14**	XF89b	**17**	**26**	CFBP7342	24
3	PXO61	**18**	**15**	MAI73	**9**	**27**	CFBP2286	28
4	PXO86	**18**	**16**	MAI145	**9**	**28**	B8-12	28
5	AXO1947	**09**	**17**	MAI68	**9**	**29**	BLS279	26
6	PXO83	**18**	**18**	MAI106	**9**	**30**	BXOR1	27
7	PXO71	**20**	**19**	MAI129	**9**	**31**	CFBP7331	22
8	PXO145	**18**	**20**	MAI134	**9**	**32**	CFBP7341	22
9	PXO211	**17**	**21**	MAI95	**9**	**33**	L8	29
10	PXO236	**16**	**22**	MAI99	**9**	**34**	RS105	24
11	PXO524	**19**	**23**	MAI1	**9**			
12	PXO563	**18**	**24**	BAI3	**9**			

Results from MP3 software

The MP3 software suite is a computational tool employed for predicting pathogenic proteins in genomic and metagenomic data. In this study, the standalone version of MP3 running on a Linux operating system was used to detect pathogenic proteins among the 592 TALEs predicted by the AnnoTALE software suite across 34 strains of *Xanthomonas oryzae pv. oryzae (Xoo)* and *Xanthomonas oryzae pv. oryzicola* (*Xoc*). A threshold value of 1.0, which is the maximum possible score for predicting pathogenicity, was employed to ensure the identification of only the most likely pathogenic proteins. To further investigate the "exclusive pathogenic" nature of the predicted TALEs, Pfam domain analysis was performed on all 592 proteins. The results showed that 572 of the predicted proteins contained domains associated with pathogenicity, while the remaining 20 proteins did not possess these specific domains according to the MP3 software's predictions. The high-confidence predictions, along with the Pfam domain analysis, provide valuable

insights into the pathogenic potential of the TALEs and their role in the virulence of these bacterial pathogens.

The UniProt database result analysis:

NCBI BLAST was conducted on 592 sequences to identify homologous proteins in the UniProt database, utilising a low E-value threshold of 0.005. The analysis revealed that all sequences showed homology to six protein entries: P14727.2, B2SU53.2, Q56830.1, Q8XYE3.1, E5AW45.1, and E5AV36.1.

P14727.2: This entry corresponds to the Avirulence protein AvrBs3 from *Xanthomonas* euvesicatoria, which acts as a transcription factor in *Capsicum annuum*. In susceptible plants lacking the Bs3 resistance gene, AvrBs3 induces the expression of several genes, including those homologous to auxin-induced genes and pectate lyase, leading to plant hypertrophy. In resistant plants, it triggers a hypersensitive response by inducing transcription of the Bs3 gene.

***B2SU53.2*:** This protein is also an avirulence factor that acts as a transcription factor in rice, inducing the expression of host genes such as SWEET11 in susceptible plants with the Xa13 allele. Plants with the xa13 allele do not induce SWEET11 and exhibit resistance to bacterial blight, eliciting an atypical hypersensitive response.

***Q56830.1*:** Another avirulence protein, this entry induces a hypersensitive response in rice plants carrying the resistance gene *Xa10*. Its activity relies on the presence of core repeat domains, and it likely functions as a transcription factor to promote plant resistance.

***Q8XYE3.1*:** This protein is exported into plant cells and targeted to the nucleus, where it likely acts as a transcription factor, binding DNA in a sequence-specific manner and contributing to plant pathogenicity.

***E5AW45.1*:** Similar to Q8XYE3.1, this protein is exported into plant cells, targets the nucleus, and functions as a transcription factor, binding DNA in a sequence-specific manner, potentially contributing to pathogenicity.

***E5AV36.1*:** This protein binds to double-stranded DNA in a sequence-specific manner and can also bind to DNA/RNA duplexes. Each tandem core repeat recognises a single nucleotide in the target DNA, with specificity determined by the base-specifying residue.

Overall, the results from the UniProt analysis provide significant insights into the functional roles of these proteins in the pathogenicity of *Xanthomonas* species, highlighting their importance as transcription factors that manipulate host plant responses during infection.

SMS Suite Results Analysis

An analysis of 592 proteins was conducted, yielding significant insights into their molecular characteristics. The molecular weight analysis indicated that the majority of the proteins exhibited molecular weights ranging from 80 to 160 kDa. This suggests a relatively uniform size distribution among the analysed proteins, which may reflect their functional roles and structural properties. The isoelectric point (pI) analysis revealed that most sequences had pI values within the range of 6 to 8. This range is indicative of proteins that are likely to be soluble in physiological conditions, which is important for their biological functions and interactions within the cellular environment. The sequence length analysis showed that the majority of the proteins had lengths varying between 850 and 1450 residues. This length range is typical for proteins involved in complex biological processes, as it allows for the necessary structural diversity and functionality. Overall, the analysis conducted using the SMS suite provides valuable information regarding the molecular weight, pI, and sequence length of the proteins, contributing to the understanding of their potential roles in biological systems. These findings may assist in further investigations into the functional implications of these proteins in the context of plant-pathogen interactions and other related studies.

The Clustal X Result analysis

The Clustal X version 2.1 were used for clustering of proteins. This clustering will help to identify the similar proteins and cluster them in a group of interest. All 592 proteins were subjected to clustering analysis with default values. 3 major clusters and near about 22 sub-clusters were visually identified and can be analyzed further for other interest.

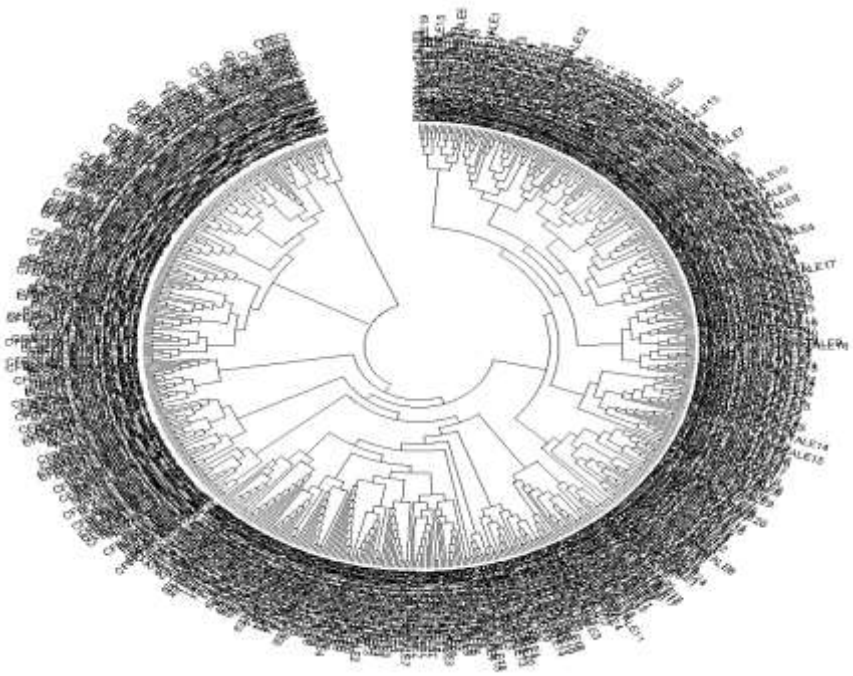

Figure 01: Result of Clustal X analysis for 592 proteins

The PSORTb server analysis

In this study, all 592 protein sequences in FASTA format, derived from Gram-negative bacterial strains, were submitted to PSORTb v3.0.2 for localization prediction. The results showed that almost all the sequences were predicted to have a "cytoplasmic" subcellular localization, with a high confidence score of 8.96 out of 10.

The predominance of cytoplasmic localization among the analysed proteins suggests that they are likely involved in various intracellular processes and metabolic pathways within the bacterial cells. Cytoplasmic proteins play crucial roles in cellular functions, such as enzymatic catalysis, signal transduction, and protein synthesis. While the majority of the proteins were predicted to be cytoplasmic, it is important to note that a small fraction may still be localised to other subcellular compartments, such as the cell membrane, periplasm, or extracellular space. These proteins may have specialised functions related to cell-cell communication, nutrient acquisition, or host-pathogen interactions.

In summary, the PSORTb analysis provides a high-confidence prediction of the subcellular localization of the 592 bacterial proteins, with the majority being cytoplasmic. This information can guide further investigations into the functional roles of these proteins within the bacterial cells and their potential interactions with host organisms.

SignalP Server Analysis Results

The analysis of protein secretion was conducted using two servers: SecretomeP 2.0 and SignalP 3.0, each serving distinct purposes in predicting protein localization and secretion mechanisms.

SecretomeP 2.0 Analysis: This server was utilised to predict non-classical protein secretion, which does not rely on signal peptides. SecretomeP 2.0 generates *ab initio* predictions by querying multiple feature prediction servers to gather information on various post-translational modifications and localization aspects of proteins. A score greater than 0.5 indicates a protein is likely secreted. Out of 592 proteins analysed, 302 proteins achieved scores above 0.5, confirming their secreted nature.

***SignalP 3.0 Analysis*:** This server predicts the presence and location of signal peptide cleavage sites in amino acid sequences across different organisms, including Gram-positive and Gram-negative prokaryotes and eukaryotes. SignalP 3.0 employs a combination of artificial neural networks and hidden Markov models to predict both signal peptides and the positions of signal peptidase I (SPase I) cleavage sites.

The graphical output from SignalP includes three primary scores: S, C, and Y. The S-score indicates the likelihood of an amino acid being part of a signal peptide, with higher scores suggesting a signal peptide presence. The C score assesses the likelihood of a cleavage site, with significant values expected only at the cleavage site. The Y-max score, a derivative of the C-score combined with the S-score, enhances cleavage site prediction by identifying the true cleavage site among multiple high C-scores. The results from both SecretomeP 2.0 and SignalP 3.0 provide critical insights into the secretion mechanisms of proteins, with a significant proportion of the analysed proteins predicted to be secreted, thereby aiding in the

understanding of their functional roles in bacterial pathogenesis.

The SignalP 3.0 analysis of 592 protein sequences revealed that only 9 proteins were predicted to contain signal peptides. The specific results for each of these 9 proteins are as follows:

The PXO145-tempTALE10:

Prediction: Signal peptide

Signal peptide probability: 0.763

Max cleavage site probability: 0.601 between positions 22 and 23

The PXO145-tempTALE13, PXO524-tempTALE7, PXO61-tempTALE1, PXO61-tempTALE5, PXO61-tempTALE7, PXO61-tempTALE14, PXO61-tempTALE15,

PXO61-tempTALE18:

Prediction: Signal peptide

Signal peptide probability: 0.557Max cleavage site probability: 0.273 between positions 17 and 18The presence of a signal peptide in these 9 proteins suggests that they are likely to be secreted or localized to the cell membrane, as signal peptides play a crucial role in targeting proteins for secretion or membrane integration. The signal peptide probability scores indicate the likelihood of a signal peptide being present, with higher scores indicating a higher probability. The max cleavage site probability identifies the most likely position where the signal peptidase will cleave the signal peptide from the mature protein.

These results provide valuable insights into the potential secretion mechanisms and localization of the analyzed proteins, which is particularly relevant for understanding their functional roles in bacterial pathogenesis and host-pathogen interactions.

The T346 Hunter server results analysis

All of 592 Coding sequences in fastA format for all 34 strains were subjected for Analysis to find the type of Secretion System responsible for their production. The server predicted the proteins to be the part of T3SS system. A summarized results are given below:

Table no 02:: T346 Hunter server results for *Xoo* and *Xoc* strains.

S no	*Xanthomonas oryzae strains*	No of T3SS predicted	S no	*Xanthomonas oryzae strains*	No of T3SS predicted
1	PXO99A	02	18	PXO211	03
2	MAFF 311018	05	19	PXO236	03
3	BLS256	04	20	PXO524	04
4	CFBP7342	04	21	PXO563	04
5	PXO86	03	22	PXO602	03
6	CFBP2286	04	23	XF89b	03
7	B8-12	03	24	MAI73	00
8	BLS279	03	25	MAI145	00
9	BXOR1	05	26	MAI68	00
10	CFBP7331	05	27	MAI106	00
11	CFBP7341	05	28	MAI129	00
12	L8	03	29	MAI134	00
13	RS105	03	30	MAI95	00
14	AXO1947	00	31	MAI99	00
15	PXO83	03	32	MAI1	00
16	PXO71	04	33	BAI3	00
17	PXO145	03	34	PXO61	03

Local BLAST Analysis of 34 Strains of Xoo

Local BLAST analysis was conducted on the proteomes of 34 strains of *Xanthomonas oryzae pv. oryzae (Xoo)* using the 592 predicted TALEs from the AnnoTALE software. The analysis employed a very low E-value threshold to identify the best-matched proteins among the strains. This step was crucial, as the initial predictions from AnnoTALE indicated potential protein sequences, but it was necessary to confirm their actual presence within the proteomes of the strains.

The results revealed that 543 proteins across the 34 strains exhibited high homology to the TALEs predicted by AnnoTALE. These identified proteins were annotated within the respective proteomes and will serve as a foundation for further analyses. In total, the study successfully identified 543 TALEs distributed among 24 strains of *Xoo* and 10 strains of *Xanthomonas oryzae pv. oryzicola* (*Xoc*). This identification of TALEs is significant for understanding the functional roles of these proteins in the context of bacterial pathogenesis and host interactions. Further

analysis of these proteins will enhance insights into the mechanisms of virulence employed by *Xanthomonas* species.

Table 03: List of 543 proteins of 24 strains of Xoo and 10 of Xoc having high homology to the AnnoTALEs predicted proteins.

No	Organism/Name	Strain	No of homologous proteins
1	*Xanthomonas oryzae pv. oryzicola*	BLS256	27
2	*Xanthomonas oryzae pv. oryzicola*	CFBP7342	22
3	*Xanthomonas oryzae pv. oryzicola*	CFBP2286	27
4	*Xanthomonas oryzae pv. oryzicola*	B8-12	27
5	*Xanthomonas oryzae pv. oryzicola*	BLS279	26
6	*Xanthomonas oryzae pv. oryzicola*	BXOR1	26
7	*Xanthomonas oryzae pv. oryzicola*	CFBP7331	20
8	*Xanthomonas oryzae pv. oryzicola*	CFBP7341	19
9	*Xanthomonas oryzae pv. oryzicola*	L8	29
10	*Xanthomonas oryzae pv. oryzicola*	RS105	22
11	*Xanthomonas oryzae pv. oryzae*	PXO99A	16
12	*Xanthomonas oryzae pv. oryzae*	MAFF 311018	17
13	*Xanthomonas oryzae pv. oryzae*	PXO86	17
14	*Xanthomonas oryzae pv. oryzae*	AXO1947	9
15	*Xanthomonas oryzae pv. oryzae*	PXO83	17
16	*Xanthomonas oryzae pv. oryzae*	PXO71	16
17	*Xanthomonas oryzae pv. oryzae*	PXO145	14
18	*Xanthomonas oryzae pv. oryzae*	PXO211	16
19	*Xanthomonas oryzae pv. oryzae*	PXO236	17
20	*Xanthomonas oryzae pv. oryzae*	PXO524	15
21	*Xanthomonas oryzae pv. oryzae*	PXO563	17
22	*Xanthomonas oryzae pv. oryzae*	PXO602	16
23	*Xanthomonas oryzae pv. oryzae*	XF89b	24
24	*Xanthomonas oryzae pv. oryzae*	MAI73	9
25	*Xanthomonas oryzae pv. oryzae*	MAI145	9
26	*Xanthomonas oryzae pv. oryzae*	MAI68	9
27	*Xanthomonas oryzae pv. oryzae*	MAI106	9
28	*Xanthomonas oryzae pv. oryzae*	MAI129	9
29	*Xanthomonas oryzae pv. oryzae*	MAI134	9
30	*Xanthomonas oryzae pv. oryzae*	MAI95	9
31	*Xanthomonas oryzae pv. oryzae*	MAI99	9
32	*Xanthomonas oryzae pv. oryzae*	MAI1	9
33	*Xanthomonas oryzae pv. oryzae*	BAI3	9
34	*Xanthomonas oryzae pv. oryzae*	PXO61	9

Working on 34 strains of *Xanthomonas,* there will be duplicates in the sequences. Duplicate proteins were removed and only non-redundant proteins were kept for further analysis. Finally we were able to select **310** non-redundant proteins for all 34 strains of *Xoo.* These proteins will be further analyzed for other interesting topics. Thus finally we were able to identify 310 non-redundant proteins for all 34 strains of *Xoo.*

The Uniprot Analysis of identified proteins in 34 strains of Xoo

The UniProt analysis of 310 unique proteins identified across 34 strains of *Xanthomonas oryzae pv. oryzae (Xoo)* yielded 41 highly homologous entries from the UniProt database. These identifications were made using a very low E-value threshold, indicating a high degree of similarity between the predicted 310 sequences and the UniProt entries. These 41 UniProt proteins, which include B2SU53.2, Q56830.1, P14727.2, Q8XYE3.1, E5AW45.1, E5AV36.1, Q68A49.1, E5AW43.1, P69929.1, Q9FMG4.1, Q9LFF1.1, Q6CT49.1, Q8C6L5.1, Q73X75.1, P9WN44.1, A7I1G1.1, P59816.1,

A5U226.1, Q1B568.1, Q9INJ1.1, A3Q396.1, Q8WZ42.4, P54924.2, Q12IZ9.2, Q6GR21.1, A6NE01.3, Q47GX7.1, D3ZUI5.1, Q5M2T7.1, Q5LY80.1, Q03IZ8.1, Q5NXV7.1, Q9LFK0.1, Q59UH5.1, Q640Q5.3, Q69151.1, O15162.1, Q8XMQ5.1, Q9LFJ9.2, A7RR34.1, and A4STD5.2, will be valuable for further analysis of the predicted TALEs. These proteins exhibit good homology with the TALEs and can be investigated for additional interesting features and their potential roles in bacterial pathogenesis and host-pathogen interactions.

Summary, Conclusion and Suggestion for Future work

In this study, transcription activator-like effectors (TALEs) from *Xanthomonas oryzae* were computationally identified using the AnnoTALE software suite, resulting in the detection of 592 TALEs across 24 strains of *Xoo* and 10 strains of *Xoc*. Subsequent analyses were conducted using the MP3 suite, UniProt database, and SMS suite. The MP3 suite successfully identified regions classified as "Exclusively Pathogenic" based on predictions from the Pfam database. The UniProt analysis confirmed the presence of six reviewed protein sequences (P14727.2, B2SU53.2, Q56830.1, Q8XYE3.1, E5AW45.1, and E5AV36.1), validating the findings from the AnnoTALE suite. Additionally, the SMS suite analysis indicated a homogeneous nature of the TALEs concerning molecular weight, isoelectric point, and protein length. ClustalX analysis further revealed three major clusters and 22 minor clusters, suggesting avenues for future investigations into specific domains and functions. The PSORTb server analysis determined the cytoplasmic localization of the proteins, while the SignalP analysis identified the majority as secreted proteins, with nine proteins containing signal peptide domains. The T346Hunter server results indicated that these proteins are part of the Type 3 Secretion System. A Local BLAST analysis conducted on the proteomes of the 34 strains confirmed the presence of 310 unique proteins, with the UniProt analysis identifying 41 protein entries that corroborate the TALE predictions. These findings provide a robust framework for further domain-related analyses of the identified proteins, enhancing our understanding of their roles in bacterial pathogenesis and potential applications in crop protection.

insilico identification and analysis of interactions between host cell proteins and pathogen proteins could serve as a promising area for future research. This work could be broadened to include a diverse array of proteins from both hosts and pathogens. Additionally, the design of appropriate ligands that facilitate the recognition of pathogens by the host and directly interact with TALEs presents another intriguing research avenue. Furthermore, this research could be expanded to encompass other rice pathogens as well as various pathogens affecting other cereal crops.

Acknowledgments

The author 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.

REFERENCES

1. Anonymous. (2016). Foreign Agricultural Service / Office of Global Analysis. International Production. Assessment Division (IPAD / PECAD). USDA, World Agricultural Production (May 2016). Ag Box 1051, Room 4630, South Building, Washington, DC 20250-1051, p. 19.
2. Ou, S. H. (1985). Rice diseases (2nd ed.). Commonwealth Mycological Institute.
3. Kou, Y., & Wang, S. (2013). Bacterial blight resistance in rice. In R. Varshney & R. Tuberosa (Eds.), Genomics applications in plant breeding (pp. 11–30). Wiley-Blackwell Press.
4. Jones, J. D., & Dangl, J. L. (2006). The plant immune system. Nature, 444, 323–329. https://doi.org/10.1038/nature05286
5. Kou, Y., & Wang, S. (2010). Broad-spectrum and durability: understanding of quantitative disease resistance. Current Opinion in Plant Biology, 13, 181–185. https://doi.org/10.1016/j.pbi.2009.12.003
6. Thomma, B. P., Nurnberger, T., & Joosten, M. H. (2011). Of PAMPs and effectors: the blurred PTI-ETI dichotomy. Plant Cell, 23, 4–15. https://doi.org/10.1105/tpc.110.082602
7. Grant, M., & Lamb, C. (2006). Systemic immunity. Current Opinion in Plant Biology, 9, 414–420. https://doi.org/10.1016/j.pbi.2006.05.014
8. Shah, J. (2009). Plants under attack: systemic signals in defence. Current Opinion in Plant Biology, 12, 459–464. https://doi.org/10.1016/j.pbi.2009.05.003
9. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science, 326, 1509–1512. https://doi.org/10.1126/science.1178811
10. Boch, J., & Bonas, U. (2010). *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. Annual Review of Phytopathology, 48, 419–436. https://doi.org/10.1146/annurev-phyto-080508-081936
11. Bogdanove, A. J., Schornack, S., & Lahaye, T. (2010). TAL effectors: finding plant genes for

disease and defense. Current Opinion in Plant Biology, 13, 394–401. https://doi.org/10.1016/j.pbi.2010.04.010

12. Schornack, S., Moscou, M. J., Ward, E., & Horvath, D. (2013). Engineering plant disease resistance based on TAL effectors. Annual Review of Phytopathology, 51, 383–406. https://doi.org/10.1146/annurev-phyto-082712-102314
13. Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G., Koebnik, R., & Boch, J. (2016). AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. Scientific Reports, 6(1), 21077. https://doi.org/10.1038/srep21077
14. Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R., Wilson, G., Koebnik, R., & Boch, J. (2016). AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. Scientific Reports, 6(1). https://doi.org/10.1038/srep21077
15. Gupta, A., Kapil, R., Dhakan, D. B., & Sharma, V. K. (2014). MP3: A software tool for the prediction of pathogenic proteins in genomic and metagenomic data. PLoS ONE, 9(4), e93907. https://doi.org/10.1371/journal.pone.0093907
16. Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques, 28, 1102–1104. https://doi.org/10.2144/00286ir01
17. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 25(24), 4876–4882. https://doi.org/10.1093/nar/25.24.4876
18. Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 26(13), 1608–1615. https://doi.org/10.1093/bioinformatics/btq249
19. Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. Protein Engineering Design and Selection, 17(4), 349–356. https://doi.org/10.1093/protein/gzh037
20. Emanuelsson, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols, 2, 953–971. https://doi.org/10.1038/nprot.2007.131
21. Martínez-García, P. M., Ramos, C., & Rodríguez-Palenzuela, P. (2015). T346Hunter: A novel web-based tool for the prediction of type III, type IV and type VI secretion systems in bacterial genomes. PLoS ONE, 10(4), e0119317. https://doi.org/10.1371/journal.pone.0119317
22. Altschul, S., Gish, W., Miller, W., Myers, E., & Lipman, D. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
23. Altschul, S. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1006/jmbi.1990.9999