International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023
DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

1,4 - ß cellobiosidase (CbsA) in *Xanthomonas* bacteria involved in switch to non vascular infection phenotype shows large scale deletions and structural changes

Varshini A[1] , Shreya S[1], Kruthika T R[1],Shilpa S K[1], Keerthi B[1], Manohar G M[2]

[1]*Department of PG studies in Biotechnology, Nrupathunga University, Bangalore -01.*
[2]*halliLabs, Ragihalli, Anekal Tq, Bangalore-83*
OrcidID : https://orcid.org/0009-0003-2087-6384

Abstract :
Vascular plant pathogens spread through the veins of the host causing systemic infections whereas non vascular pathogens are confined to infection sites and cause localized symptom development. CbsA -1,4 beta cellobiosidase is a cell wall-degrading enzyme responsible for causing vascular infection and a mutant version of it is found in non vascular infection. Here, we investigate the sequence variation and structural changes accompanying a mutant CbsA gene present in some pathovars exhibiting non vascuar infection by *Xanthomonas* bacteria and also chalk out the evolutionary history of the non vascular phenotype. Among *X.oryaze pv oryaze* and *X. oryzae pv oryzicola , X.oryzicola* showed a number of large scale deletions and amino acid substitutions. Protein structure of *X. oryaze pv oryaze* showed 12 helices and *X. oryzae pv oryzicola* showed missing 2 helices. *X. citri pv vignicola* did not show the same large scale deletions however the sequence had accumulated large variation. The non vascular phenotype may have evolved in the ancestor of *X. citri, X.oryzae and X. oryzicola* pathovars, but may have reverted back in *X. oryzae.* We elaborate that sequence and structural changes accompanying just one gene might have had a major role in the phenotypic swift from vascular to nonvascular infection. This may have implications in plant disease because vascular infecting pathogens are efficient in invading the whole body of the plant while non vascuar infection is localized.
Keywords : CbsA gene, pathovar, phenotypic, *Xanthomonas,* transposons, substitutions, clade.

1. INTRODUCTION

Bacteria that infect Vascular tissues of a plant like xylem and phloem can travel long distances through them thus spreading the infection[1] . The genus *Xanthomonas* and *Xylella* belonging to the family *Xanthomonadacea*e are the most common infection causing bacteria in plants [2].These are gram negative bacteria that cause vascular and nonvascular disease in different plant hosts[3]. Vascular *Xanthomonas* invade the water-transporting xylem whereas nonvascular *Xanthomonas* pathovars affect the mesophyll thereby showing visible symptoms [4]. *Xanthomonas* cannot directly penetrate through any plant tissue but through stoma or any natural wound a plant can be infected [5]. Some of the *Xanthomonas species* causing infection in plants are *Xanthomonas translucens pv translucens , Xanthomonas translucens pv. undulosa and Xanthomonas translucens pv. cerealis* causing bacterial leaf streak in barley, wheat & grasses [6] . In case of barley, at the beginning translucent spots of light green color appear on the leaf then these spots are expanded to yellow and turn to brown and even black and sticky slime is observed which forms yellowish film at drying on the spots. Affected plants give yellow stripes and are stunted or they fail to produce flowers or fruits [7]. Bacterial blight is the most common disease caused by *Xanthomonas oryzae pv oryzae* species in rice and the same disease is caused by *Xanthomonas campestris pv.* in barley [8]. *Xanthomonas*

DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

campestris pv. phaseoli cause blight of beans and *Xanthomonas campestris pv. malvacearum* cause angular leaf spot in cotton [9,10] . CbsA (1,4-beta cellobiosidase) gene plays a major role in vascular pathogenesis. It is found to be mutated in some strains of *Xanthomonas* pathovars which infect non vascular tissue [4].

We wanted to characterize what kind of mutations in CbsA gene lead specifically to non vascular infection phenotypic switch.We analysed the sequences of available *Xanthomonas* sequences and found that non vascular pathovars shared a large deletion in CbsA gene. Other pathovars showed a significant variation in CbsA gene. Phylogenetic analysis suggested that the vascular infection phenotype is the ancestral phenotype. Structural analysis of the Alphafold2 model of CbsA protein showed two missing helices in the pathovars involved in non vascular infection. Overall our work characterizes the mutation at both sequence and structural level and underlines how mutation in a single gene can cause a big switch in the phenotype like vascular and nonvascular mode of infection [4].

2. MATERIALS AND METHODS

1. Sequence retrieval : Nucleotide sequence of *Xanthomonas oryzae pv. oryzae* strain BXO43 cellobiosidase (CbsA) gene (MF521486.1) was retrieved from the NCBI Genbank database by using keywords Xanthomonas + cbsA . MF521486.1 was used as a query sequence in BLASTn. Results were filtered for an except threshold value of 0.05 and maximum target sequences to 5000 (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

From *Xanthomonas oryzae pv. oryzae strain BXO1* chromosome, complete genome (CP033201.1) 1,4 - ß cellobiosidase protein sequence was retrieved from BLASTn result of MF521486.1 and 1,4 - ß cellobiosidase protein sequence was used as query sequence in tblastn. Results were filtered for an of except threshold value of 0.05 and maximum target sequences to 5000. (https://blast.ncbi.nlm.nih.gov/Blast.cgi#dtr_1864553229).

2. Sequence Manipulation Suite : The reverse complement sequences of *Xanthomonas species* of cbsA gene were obtained using Reverse complement conversion tool in Sequence Manipulation Suite version 2 [11] (https://www.bioinformatics.org/sms2/index.html).

3. Multiple sequence alignment and Phylogenetic tree construction : The representative *Xanthomonas species* sequences were aligned in MEGA11 [12] [MEGA version 11.0.13] by using Align by Muscle.

Phylogenetic analysis was done by constructing Neighbour joining tree of the representative sequences, using Bootstrap method and number of Bootstrap Replications was taken as 5000 and set to Auto-size tree by using layout option. The tree was further edited to represent vascular and non vascular infections caused by *Xanthomonas species* in different colors using drawing options from subtree options [MEGA11].

The optimal tree is shown. Number of Bootstrap Replications was taken as 5000. The evolutionary distances were computed using the Maximum Likelihood method and are in the units of the number of base substitutions per site. This analysis involved 58 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1701 positions in the final dataset. Evolutionary analyses were conducted in MEGA11.

4. Interproscan : The domains of vascular and non vascular infection causing protein sequence were analyzed by using their protein fasta sequence [13].

5. Alphafold2 : AlphaFold is an AI system developed by Deepmind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment [14]. 3D protein structure of *Xanthomonas oryzae* and *Xanthomonas oryzae oryzicola* protein sequences was used as a query sequence.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023
DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

3. RESULTS

1,4 - ß cellobiosidase gene is majorly involved in vascular infection in plants. We obtained 300 nucleotide sequences of 1,4-ß cellobiosidase (cbsA) gene, Exoglucanase, Glycoside hydrolase family 6 protein of various *Xanthomonas species* through BLASTn using *Xanthomonas oryzae pv. oryzae* strain BXO43 cellobiosidase (CbsA) gene (MF521486.1) as a query. In order to collect more cbsA sequences of *Xanthomonas* species sequences, we performed tblastn using *Xanthomonas oryzae pv. oryzae* strain BXO1 (CP033201.1) of cbsA protein sequence as a query. As a result we retrieved 390 sequences including *Mycobacterium, Streptomyces* and many other species along with *Xanthomonas.* All the sequences obtained from BLASTn were also common to the list obtained through tblastn. From the raw data obtained in previous studies 57 *Xanthomonas species* sequence of 54 cbsA gene and 3 exoglucanase A gene were selected as they have 96% sequence similarity and from this 45 sequences from species known to cause vascular infection and 11 sequences from species known to cause non vascular infection along with query sequence of species that were not identified as either causing vascular or nonvascular infection were selected for further analysis.

It turned out that 32 sequences were actually complementary sequences of CbsA gene. Therefore a reverse complement of these sequences was obtained from Sequence Manipulation Suite [2]. The multiple sequence alignment of these DNA sequences was translated into protein sequence by using the standard genetic code in MEGA. We wanted to see what residue changes are unique to species that cause non vascular infection as compared to vascular infection. We report the following differences in amino acid substitutions and large scale deletions that are unique to species that exhibit non vascular infection phenotype at different sites in *Xanthomonas oryzae* pv *oryzae* strains and *Xanthomonas oryzae* pv *oryzicola* at strains level.

Table 1 **:** Changes occurring in the amino acids when compared with *oryzae and oryzicola* CbsA gene

Site in MSA	From *X.oryzae Pv oryzae* strains	Electrical nature	Chemical nature	To *X.oryzae pv oryzicola* strains of	Electrical nature	Chemical nature
42	Glycine	Neutral	Hydrophilic	Aspartic acid	Acidic	Hydrophilic
286	Glycine	Neutral	Hydrophilic	Arginine	Basic	Hydrophilic
290	Serine	Neutral	Hydrophilic	Tryptophan	Neutral	Hydrophobic
471	Arginine	Basic	Hydrophilic	Glutamine	Neutral	Hydrophobic
490	Arginine	Basic	Hydrophilic	Proline	Neutral	Hydrophobic
535	Alanine	Neutral	Hydrophobic	Threonine	Neutral	Hydrophilic

As the table suggests almost all the changed amino acids are changed into radically different chemical or electrical nature. So either some of these or any one of these can be implicated in the switch between the vascular and nonvascular phenotype.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023
DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

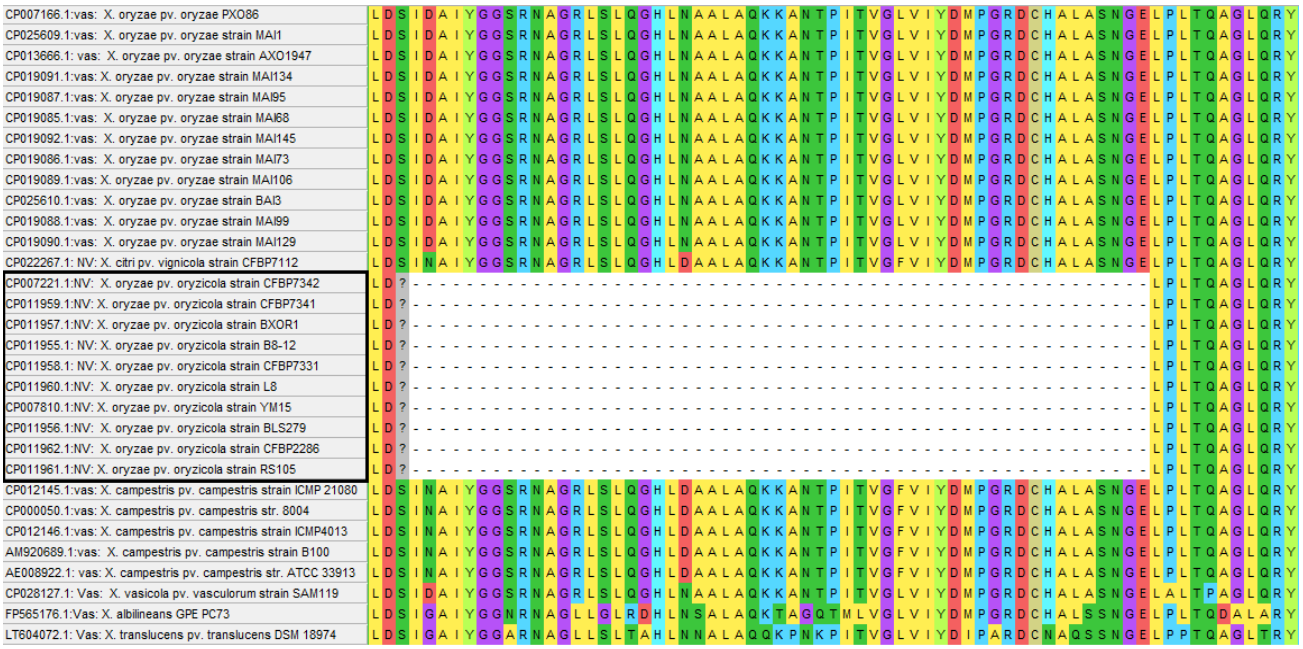

Fig. 1. Multiple sequence alignment of *Xanthomonas species* containing CbsA gene showing a large-scale deletion in *X oryzae* pv *oryzicola* which can be implicated in the swift from pathogenic to non pathogenic lifestyle .

From multiple sequence alignment (Fig. 1) , at the site 85-140, deletion of sequences was observed in *Xanthomonas oryzae pv oryzicola* strains which are represented in a black rectangular box. This could be due to transposon activity or mutation. It could be that this deletion could have forced the *oryzicola* strain to have adopted a non vascular infection lifestyle. *Xanthomonas campestris* causes vascular infection and we did not observe any deletion.
Xanthomonas citri pv vignicola strain CFBP7112 is also a non vascular disease causing species but deletion was not observed.
A phylogenetic tree (Fig.2), was constructed for all the strains using the Maximum likelihood method to infer the evolutionary history for the gene. The optimal tree is shown. Number of Bootstrap Replications was taken as 1000. The evolutionary distances were computed using the Maximum Likelihood method and are in the units of the number of base substitutions per site. This analysis involved 58 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1701 positions in the final dataset. Evolutionary analyses were conducted in MEGA11.
The phylogenetic tree suggests the following possibilities.

1. Vascular phenotype is ancestral.
2. There are two possibilities.

A. Non vascular phenotype could have originated in the ancestor of *X. citri*. In that case *X.citri*, *X. oryzae* and *X.oryzicola* will be non vascular. But as we know *X. oryzae* is exhibiting non vascular phenotype. It is a possibility that the *X.oryzae* clade reverted back to the vascular phenotype.
B. Mutations may have happened independently in *X. oryzae* ancestor and *X. citri* ancestor leading to a non vascular phenotype.
Interpro: In both vascular and nonvascular infection causing species, we observed that FN3_domain, Fibronectin type - III domain profile, FN3, FN3_2 domains were reported. Our InterPro analysis of both sequences causing vascular and nonvascular infection showed the same number and type of domains. No domains were found to be present in the 85-140 deleted region found in non vascular sequences. Even though no domains were found in the deleted region we cannot rule out the possibility of linker regions between the domains playing a role in the switch between vascular and nonvascular phenotype.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023
DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

Alpha fold2: In order to see the how the sequence variation in *Xanthomonas oryzae* which is a vascular pathogen and the *Xanthomonas oryzicola* which is a non vascular pathogen reflects in the structure of CbsA we used Alphafold2 to model both the protein structures. AlphaFold is an AI system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment.

ColabFold v1.5.2: AlphaFold2 using MMseqs2

AlphaFold2 and Alphafold2-multimer are used in colabfold, an intuitive programme for predicting protein structure and complexes. MMseqs2 and HHsearch are used to build sequence alignments and templates.

Structure interpretation

CbsA protein contains two beta barrels. One is found outside the protein and another which is inside is surrounded by alpha helices. Beta barrels are a common motifs found throughout the protein structures. Usually the region between the alpha helices and the beta barrel will be hydrophobic in nature. The outside beta barrel also closely resembles a Greek key motif another common motif found across alpha/beta protein structures.

CbsA protein structure **(Fig 3)** (wild type) of *Xanthomonas oryzae* which is a vascular pathogen consists of 12 alpha helices as predicted by alpha fold2. Interestingly the CbsA protein of the non vascular *Xanthomonas oryzicola* contains only 10 alpha helices. **(Fig 3)**

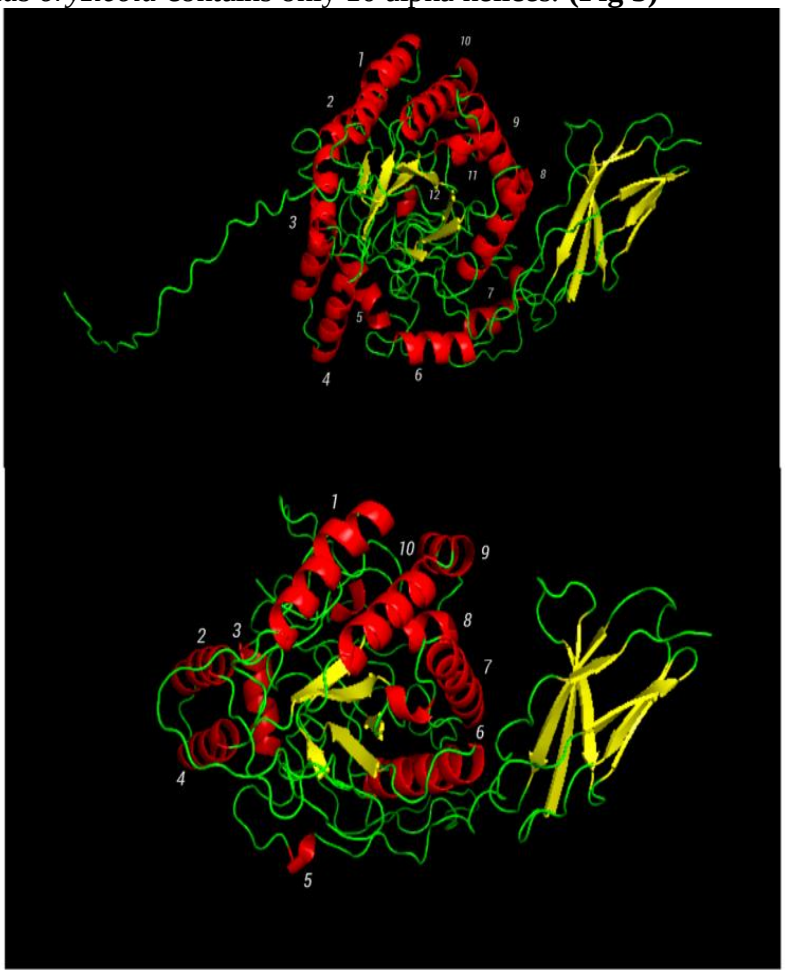

Fig. 3. 3D protein structure of *Xanthomonas oryzae* and *Xanthomonas oryzae oryzicola.* Unrelaxed rank 001 3D structure of CbsA from *Xanthomonas oryzae* (that causes vascular infection). It has 12 alpha helices distributed on either side of beta sheets whereas the unrelaxed rank 001 A 3D structure of CbsA from *Xanthomonas oryzicola* (causing non vascular infection). It has 10 alpha helices and missing 2 helix from its counter part CbsA of *Xanthomonas oryzae.*

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023
DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

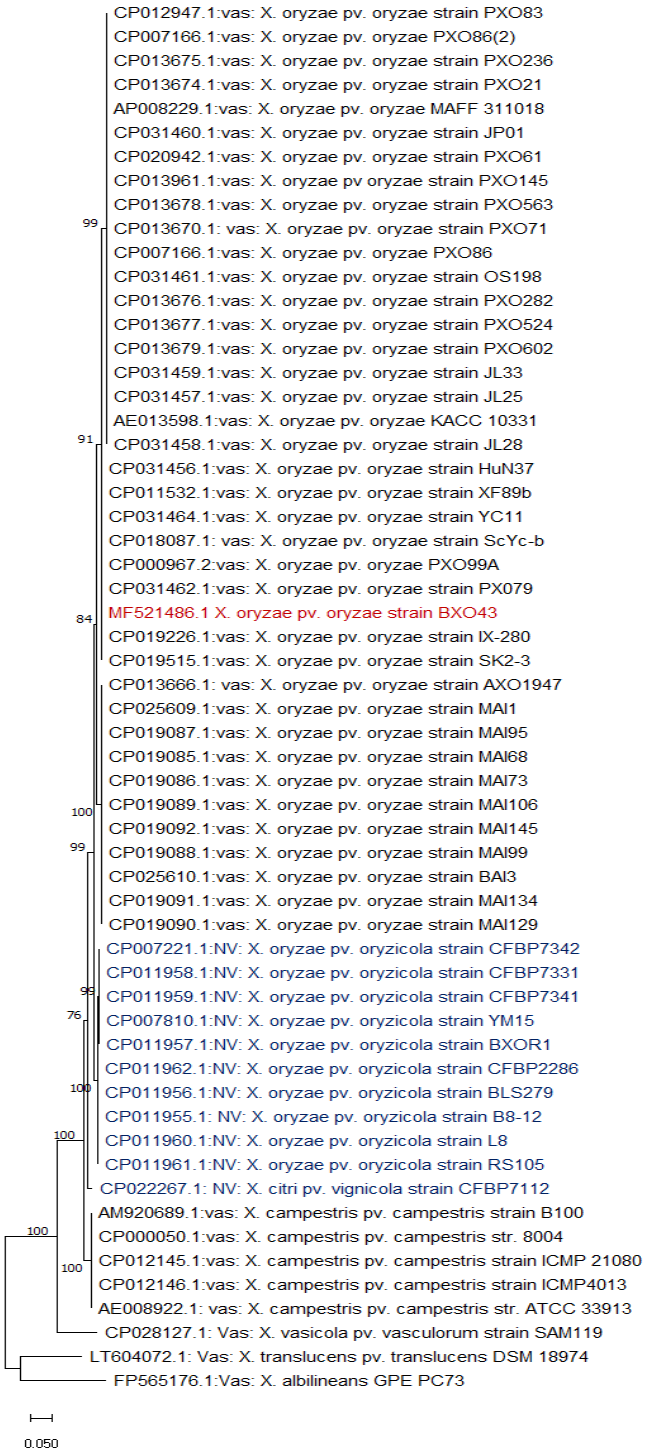

Fig 2. Maximum likelihood tree of CbsA gene of Xanthomonas bacteria. The blue colored are non vascular infection causing bacteria, black colored are vascular infection causing bacteria and red colored is not determined whether it causes vascular infection or non vascular infection , which is used as a query.

DISCUSSION

1,4 beta cellobiosidase of some bacteria is involved in vascular infection in plants. Multiple sequence alignment of 57 translated DNA sequences belonging to *Xanthomonas oryzae pv oryzae strains, Xanthomonas oryzae pv oryzicola* and *X.campestris along with other bacterial sequences* reported certain deletions and amino acid substitutions. All the amino acids had changed into

DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

amino acids that showed radically different chemical and electrical nature which may have contributed to phenotypic switch between vascular and nonvascular infection.

A phylogenetic tree using maximum likelihood tree was constructed taking bootstrap replications as 1000 which further suggested that vascular phenotype is ancestral to non vascular phenotype. Of many possibilities we suggest that non vascular phenotype may have evolved in the ancestor of *X. citri, X.oryzae and X. oryzicola* pathovars, but may have reverted back in *X. oryzae.*

The interpro analysis suggested that the vascular and non vascular causing species were observed in certain domains and no domains were noted in between 85-140 deleted regions found in non vascular sequences.

Structural analysis showed that CbsA protein was visualized to contain 2 beta barrels (common motifs found throughout the protein structures) and 12 Alpha helices. In the CbsA protein of *Xanthomonas oryzicola* which was a non vascular pathovar the protein contained only 10 alpha helices with 2 alpha helices missing as compared to its vascular CbsA counterpart.

CbsA gene acts as a phenotypic switch between vascular and nonvascular infection. Due to repeated gain and loss of a gene mutation occurs.

Some of the *Xanthomonas* species causing infection in plants are *Xanthomonas translucens pv. translucens, Xanthomonas translucens pv. undulosa* and *Xanthomonas translucens pv. cerealis* causing bacterial leaf streak in barley, wheat and grasses . Bacterial blight is the most common disease caused by *Xanthomonas oryzae pv. oryzae* species in rice and the same disease is caused by *Xanthomonas campestris pv.* in barley. *Xanthomonas campestries pv. phaseoli* cause blight of beans and *Xanthomonas campestris pv. malvacearum* cause angular leaf spot in cotton.

Inclusion of *Xylella* and *Ralstonia* species which also showed up in the blast results and other closely related species along with *Xanthomonas* can shed more light on the role of CbsA in the switch between the phenotypes.

Xanthomonas citri pv. vignicola showed a sequence variation and did not show the deletion but still causes nonvascular infection. It will be interesting to study what sequence variation other than the deletion can be implicated in nonvascular phenotype.

In protein structure prediction, the CbsA protein of the non vascular *Xanthomonas oryzicola* contains only 10 alpha helices. The way in which it might be involved in the non vascular phenotype can be further studied at the structural level.

A mutant CbsA gene results in confining the infecting bacteria to a specific tissue and not enabling it to spread throughout the plant body instead. This may have implications in the productivity of a particular crop plant.

ACKNOWLEDGEMENT

We deeply appreciate the help and mentoring from **Mr. Prakash Sivakumar**, Research scholar , CCMB , Hyderabad. **Dr. Shaik Naseer Pasha**, PhD from NCBS, Consultant for ZS associates for regular inputs for this project.

REFERENCES

1. Redkar, Amey, Mugdha Sabale, and Antonio Di Pietro. "A'Hydrolase Switch'for Vascular Specialization in Plant Pathogenic Bacteria." *Trends in Plant Science* 26.5 (2021): 427-429.
2. Costa, Joana, et al. "Integrating Science on Xanthomonas and Xylella for Integrated Plant Disease Management." *Microorganisms* 11.1 (2022): 6.
3. Ryan, Robert P., et al. "Pathogenomics of Xanthomonas: understanding bacterium–plant interactions." *Nature Reviews Microbiology* 9.5 (2011): 344-355.
4. Gluck-Thaler, Emile, et al. "Repeated gain and loss of a single gene modulates the evolution of vascular plant pathogen lifestyles." *Science advances* 6.46 (2020): eabc4516.
5. Cao, Jianbo, et al. "Different cell wall-degradation ability leads to tissue-specificity between Xanthomonas oryzae pv. oryzae and Xanthomonas oryzae pv. oryzicola." *Pathogens* 9.3 (2020): 187.

DOI:10.46647/ijetms.2023.v07i03.128 ISSN: 2581-4621

6. Roman-Reyna, Verónica, et al. "Genome resource of barley bacterial blight and leaf streak pathogen Xanthomonas translucens pv. translucens strain UPB886." *Plant disease* 104.1 (2020): 13-15.
7. Mew, T. W. "Xanthomonas oryzae pathovars on rice: cause of bacterial blight and bacterial leaf streak." *Xanthomonas. New York: Chapman and Hall* (1993): 30-40.
8. Dharmapuri, Sridhar, and Ramesh V. Sonti. "A transposon insertion in the gumG homologue of Xanthomonas oryzae pv. oryzae causes loss of extracellular polysaccharide production and virulence." *FEMS microbiology letters* 179.1 (1999): 53-59.
9. Tayi, Lavanya, et al. "A mutation in an exoglucanase of Xanthomonas oryzae pv. oryzae, which confers an endo mode of activity, affects bacterial virulence, but not the induction of immune responses, in rice." *Molecular plant pathology* 19.6 (2018): 1364-1376.
10. He, Ya-Wen, et al. "Rice bacterial blight pathogen Xanthomonas oryzae pv. oryzae produces multiple DSF-family signals in regulation of virulence factor production." *BMC microbiology* 10 (2010): 1-9.
11. Stothard, Paul. "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences." *Biotechniques* 28.6 (2000): 1102-1104.
12. Tamura, Koichiro, Glen Stecher, and Sudhir Kumar. "MEGA11: molecular evolutionary genetics analysis version 11." *Molecular biology and evolution* 38.7 (2021): 3022-3027.
13. Paysan-Lafosse, Typhaine, et al. "InterPro in 2022." *Nucleic Acids Research* 51.D1 (2023): D418-D427.
14. Mirdita, Milot, et al. "ColabFold: making protein folding accessible to all." *Nature methods* 19.6 (2022): 679-682.
15. Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." *Nature* 596.7873 (2021): 583-589.

SUPPLEMENTARY SECTION
https://docs.google.com/document/d/1KDDo482j84vQ4D5aw0xueVW1FW9g1UZPJLZOPyznsjs/edit