# Extensive Genomic Rearrangements along with Distinct Mobilome and TALome are Associated with Extreme Pathotypes of a Rice Pathogen

Amandeep Kaur, Kanika Bansal, and Prabhu B. Patil*

Bacterial Genomics and Evolution Laboratory, CSIR-Institute of Microbial Technology, Chandigarh, India

*Corresponding author: E-mail:pbpatil@imtech.res.in.

**Accepted:** January 31, 2020

**Data deposition:** This project has been deposited at National Centre for Biotechnology Information (NCBI) under the accession numbers CP040603, CP040604 and CP040687.

## Abstract

*Xanthomonas oryzae* pv. oryzae (Xoo) is a serious pathogen of rice which displays tremendous interstrain variation. The emergence of highly-virulent strains of Xoo is a major threat to rice cultivation. Evolutionary insights into genome dynamics of highly virulent strains as compared with the less-virulent ones are crucial for understanding the molecular basis of exceptional success of Xoo as a highly evolved plant pathogen. In the present study, we report complete genome sequence of Xoo strains with extreme-virulent pathotypes (XVPs) characterized based on their reaction toward ten resistance (*Xa*) genes. One strain, IXO1088, can overcome resistance mediated by all the ten resistance genes while the other strain IXO704 cannot overcome any of them. Interestingly, our investigation revealed that XVPs display dramatic variation in the genome structure with numerous rearrangements/inversions. Moreover, XVPs also possess distinct transposon content and prophage elements that may provide genomic flux required for the acquisition of novel gene cassettes and structural changes in the genome. Interestingly, analysis of transcription activator-like effector proteins, which are major virulence determinants of *Xanthomonas* pathogen show marked variation in the transcription activator-like effector content and DNA binding domain of *tal* genes. Overall, the present study indicates the possible role of mobilomes and repetitive elements in major structural and sequence alterations, which may be leading to the emergence of novel and extreme pathotypes. The knowledge and resource of XVPs will be invaluable in the further systematic understanding of evolution and management of variant pathotypes of Xoo.

**Key words:** *Xanthomonas oryzae*, pathotype, TALEs, genome dynamics, mobilomes.

## Introduction

*Xanthomonas oryzae* pv. oryzae is a model plant-pathogenic bacterium that causes bacterial blight disease in rice plants and leads to notable reduction in rice yield (Niño-Liu et al. 2006; Noh et al. 2007). For successful invasion, Xoo secretes various virulence factors such as transcription activator-like effectors (TALEs) into the host cells modulating host cell machinery for its own benefit (Boch et al. 2009; Moscou and Bogdanove 2009). These effectors bind to the promoter elements and regulate expression of either susceptible (*S*) genes, that make environment suitable for bacterial growth (Bogdanove et al. 2010; Römer et al. 2010) or resistance (*R*) genes that limit bacterial infection by generating host defense response (Bogdanove et al. 2010; Schornack et al. 2013; Zhang et al. 2015). The TALEs are highly specific toward their host targets and this specificity is provided by nearly identical, tandem repeats of 30–35 amino acids (aa) in the central repeat region. The number of repeats varies between different TALEs and polymorphism lies at the 12th and 13th positions of each repeat (Boch et al. 2009; Moscou and Bogdanove 2009). The 13th residue is a base-specifying residue that interacts with nucleotide, whereas 12th residue plays a role in stabilization of loop structure (Deng et al. 2012). Variation at 12th and 13th positions determines the nucleotide specificity of each repeat and together these residues are known as repeat variable diresidues (RVDs) (Boch et al. 2009; Moscou and Bogdanove 2009).

The host plant resistance is deployed as an effective approach for the development of resistant rice cultivars. To date, different resistance (R) genes also called as *Xa* genes have been identified from rice and deployed for the development of resistant rice cultivars (Iyer and McCouch 2004; Kim et al. 2015; Wang et al. 2015; Busungu et al. 2016). Such as, sequence variation in EBEs (effector binding sites) of "*S*" genes can inhibit binding of TALEs thereby inhibiting activation and expression of various susceptible genes (Li et al. 2012). Similarly, engineering "*R*" genes to include EBEs for multiple TALEs can generate a defense response against a broad spectrum of strains (Tian et al. 2014). Despite the exploitation of diverse repertoire of resistance (R) genes, Xoo strains are rapidly evolving to overcome the resistance conferred by these resistance (R) genes by acquiring TALEs that can utilize alternate target genes. For example, some Xoo strains are able to overcome *xa*13 mediated resistance by expressing a TALE that can activate expression of an alternate *SWEET* gene of the host (Antony et al. 2010; Carpenter et al. 2018). Similarly, strains that can express PthXo1 TALE can overcome *xa*5-mediated resistance by activating a host-susceptible gene called *SWEET*11 (Sugio et al. 2007).

Based on the effectiveness of ten *Xa* genes (*Xa*1, *Xa*3, *Xa*4, *xa*5, *Xa*7, *xa*8, *Xa*10, *Xa*11, *xa*13, *Xa*21), Xoo isolates have been characterized into 11 pathotypes (Mishra et al. 2013). In the present study, we report complete genome-based comparative analysis of two Xoo strains, IXO1088 and IXO704, belonging to pathotypes XI and X, respectively. Pathotype XI isolates are highly-virulent and can overcome resistance conferred by all the resistance genes, whereas pathotype X isolates are less-virulent and susceptible to them. Hence, complete genome-based investigation of diverse pathotypes could shed novel insights on their origin, evolution, and emergence.

Xoo is reported to be the second most complex genome known in the bacterial world due to presence of large number of repetitive elements such as TALEs encoding genes (Schmid et al. 2018). Emergence of affordable and long read-sequencing technologies is enabling investigation of TALEs repertoire, diversity, and their host targets toward evolution of virulent strains which is not possible using draft genomes (Bansal et al. 2018). Other than *tal* genes, the present study also highlights the possible role of mobile elements such as transposable elements, prophages, and plasmids in emergence of variant pathotypes in Xoo. These elements can result in genomic rearrangements, acquisition of novel genes, and expression level changes subsequently enhancing virulence potential of a pathogen.

## Materials and Methods

### DNA Isolation and Genome Sequencing

Strains were grown in nutrient broth media at 28 °C for 48 h. Genomic DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen). Quantity and quality of DNA were assessed using Nanodrop and Qubit 2.0 Fluorometric. For Nanopore sequencing, 3 µg of initial DNA was used for DNA end prep using NEBNext Ultra II End repair/dA-Tailing modules (NEB). Library was prepared using Ligation Sequencing Kit 1 D (SQK-LSK108) and Native Barcoding Kit 1 D (EXP-NBD103). The barcode and adapter ligation steps were performed using NEB ligase master mix module and NEB T4 DNA ligase module, respectively. All beads washing steps in the protocol were performed using AMPure beads (Beckman Coulter). Finally, 12 µl of prepared DNA library was loaded onto the flow cell according to manufacturer's instructions and sequenced using MinION (FLO-MIN-106 version R9.4) flow cell using MinKNOW software (http://community.nanoporetech.com, last accessed February 11, 2020) (Oxford Nanopore Technologies) (v.1.13.1) for 48 h. Raw reads were base called using Albacore v2.3.3 software (http://community.nanoporetech.com, last accessed February 11, 2020).

### Genome Assembly and Annotation Using ONT and Illumina Reads

Reads obtained after demutiplexing were hybrid (both ONT and Illumina reads) assembled using Unicycler v.0.4.8-beta (Wick et al. 2017) in bold mode. The assembled genomes were then error corrected for multiple rounds with short reads generated by Illumina using pilon v1.22 (Walker et al. 2014). The assembled genomes were then checked for the completeness and presence of contamination using CheckM v1.0.13 (Parks et al. 2015). Genome coverage was checked using BBmap tool v38.42 (Bushnell 2014). Genomes were annotated using NCBI PGAP (Prokaryotic Genome Annotation Pipeline) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/, last accessed February 11, 2020). The IXO704 and IXO1088 genomes were submitted to NCBI WGS portal with accession numbers CP040604 and CP040687, respectively. IXO704 strain contains 25-kb plasmid for which the sequence was submitted with accession number CP040603.

### Genome Comparison

Genome comparison between IXO1088 and IXO704 strains was performed using progressive MAUVE tool with default parameters (Darling et al. 2004). Pan-genome analysis was performed using Roary (Page et al. 2015). Functional categorization of unique genes into cluster of orthologus classes was obtained using EggNOG (Jensen et al. 2007). The pictorial representation of IXO1088 and IXO704 genomes was drawn using DNA plotter (Carver et al. 2009).

### TALEs Analysis

All *tal* genes were identified using AnnoTALE software version 1.2 (Grau et al. 2016). Identified TALEs were assigned to

different families using AnnoTALE class builder file. In order to find out orthologs of IXO704 and IXO1088 TALEs, both AnnoTALE and FuncTAL (Pérez-Quintero et al. 2015) software were used. For FuncTAL, RVD sequence was used as input.

### Identification of Mobile Elements and Prophage Regions

Insertion elements (IS) elements were identified using ISsaga tool (Varani et al. 2011). Putative genomic islands containing prophage regions were identified using PHASTER tool (Arndt et al. 2016).

## Results and Discussion

### Complete Genome Characteristics of IXO1088 and IXO704 Genomes

Complete genome size of IXO704 and IXO1088 is 4,994,377 and 5,093,052 bp, respectively. IXO704 strain contains a plasmid of 25,634 bp named as pIXO-704, whereas IXO1088 does not contain any plasmid. The pIXO-704 plasmid is 99.97% identical to plasmid, pBXO1-2 (CP033203) from another Xoo strain BXO1 (Kaur et al. 2019). A total 4,762 CDS for IXO704 and 4,821 CDS for IXO1088 were found in the genome. Genome coverages for IXO704 and IXO1088 were 96× and 138×, respectively, with average GC content of 63.7% for both the genomes. Further, genome completeness and contamination were found to be 100% and 0%, respectively, for both the genomes. The pictorial representation of both the genomes is provided in supplementary figure 1, Supplementary Material online.

### Extensive Structural Variation in the Genome Architecture of Extreme-Virulent Pathotypes

Xoo genomes are known to be one of the most complex genomes (Schmid et al. 2018). Large-scale evolutionary events such as rearrangements, indels, inversions, and duplications can contribute to the evolutionary success of a major pathogen. Therefore, to understand genome-wide evolutionary dynamics of extreme-virulent pathotypes (XVPs), we carried out comparative genome analysis using progressive MAUVE. The genomes of XVPs displayed dramatic variation in genome structure as seen by numerous large-scale rearrangements and inversions (fig. 1*A*). Strain-specific features observed including some regions, which are present in IXO1088 but completely absent from IXO704 strain. Most of the strain-specific unique regions encode phage proteins pointing toward a role of phages as anchor regions in driving evolution of extreme pathotypes through gene duplication, gene disruption, and chromosomal rearrangements.

### Role of Mobilomes in Driving Evolution of XVPs

To further look at unique gene pool content contributing to variations between XVPs, we performed pan-genome analysis. The total size of the pan-genome is 5,507 genes with 4,139 core genes, which are present in both the strains, whereas 722 genes were unique to IXO1088 strain and 646 genes were unique to IXO704 strain. Further, we calculated the average GC content of unique genes. Overall proportion of unique genes with atypical GC content (Bansal et al. 2017) that is not in the range of 63.7% (± 2.5%) was 50.3% and 48.2% for IXO704 and IXO1088, respectively. The average GC content for Xoo chromosomal genome is 63.7%, whereas unique genes with significantly lower or higher GC content point toward their acquisition through horizontal gene transfer events. Unique genes were then functionally annotated into cluster of orthologus classes (fig. 1*B*). Genes belonging to category "L" (replication, recombination, and repair) and category "S" (with unknown function) are highly represented and their number varies in both the strains.

Mobile elements can play a major role in pathogens diversity and enhancing virulence profile through transfer of virulence genes. Analysis also revealed high number of unique genes with unknown functions in IXO1088. Hence, genes belonging to categories "L" and "S" may be contributing toward genome dynamics such as ISs, transposases, and prophages. Therefore, we investigated the mobilomes of two genomes. In IXO1088, ten prophage regions were identified with three intact regions of 16.6, 43.8, and 47.5 kb, whereas, in IXO704, only three prophage regions were identified with one intact region of 21.3 kb (fig. 1*A*). Detailed features of prophage regions with their location, size, and number of proteins are given in supplementary table 1, Supplementary Material online.

As IS elements are also known to play key role in genomic alterations, we studied IS element content and differences in XVPs using ISSaga. As can be seen in (fig. 1*C*), there is variation in number and type of IS elements in both pathotypes. About 747 IS elements were observed for IXO704 classified into 20 families. Whereas in IXO1088, total 725 IS elements were found classified into 21 families (fig. 1*C*). Interestingly, there is an addition of a new family Tn3 family transposase in IXO1088.

### Genome Sequence Reveals Variations in TALome in XVPs

TALEs are key pathogenicity factors of Xoo. Due to repetitive nature of TALEs, their diversity cannot be assessed using draft genomes. Availability of complete genomes allowed us to the exploration of TALEs. The TALEs analysis revealed rearrangements and inversions of *tal* genes with addition of some new *tal* genes. IXO704 and IXO1088 genomes contain 16 and 17 TALEs, respectively, with two pseudo TALEs in the both the genomes (fig. 2*A*). The RVD sequences of some TALEs vary in both the genomes with variations at both 12th and 13th positions (fig. 2*B*). In IXO704, 13 TALEs were orthologs of IXO1088. Out of these 13, 3 TALEs (TalAE13, TalAB15, and TalAG14) were identical to IXO1088 with no RVD variation,

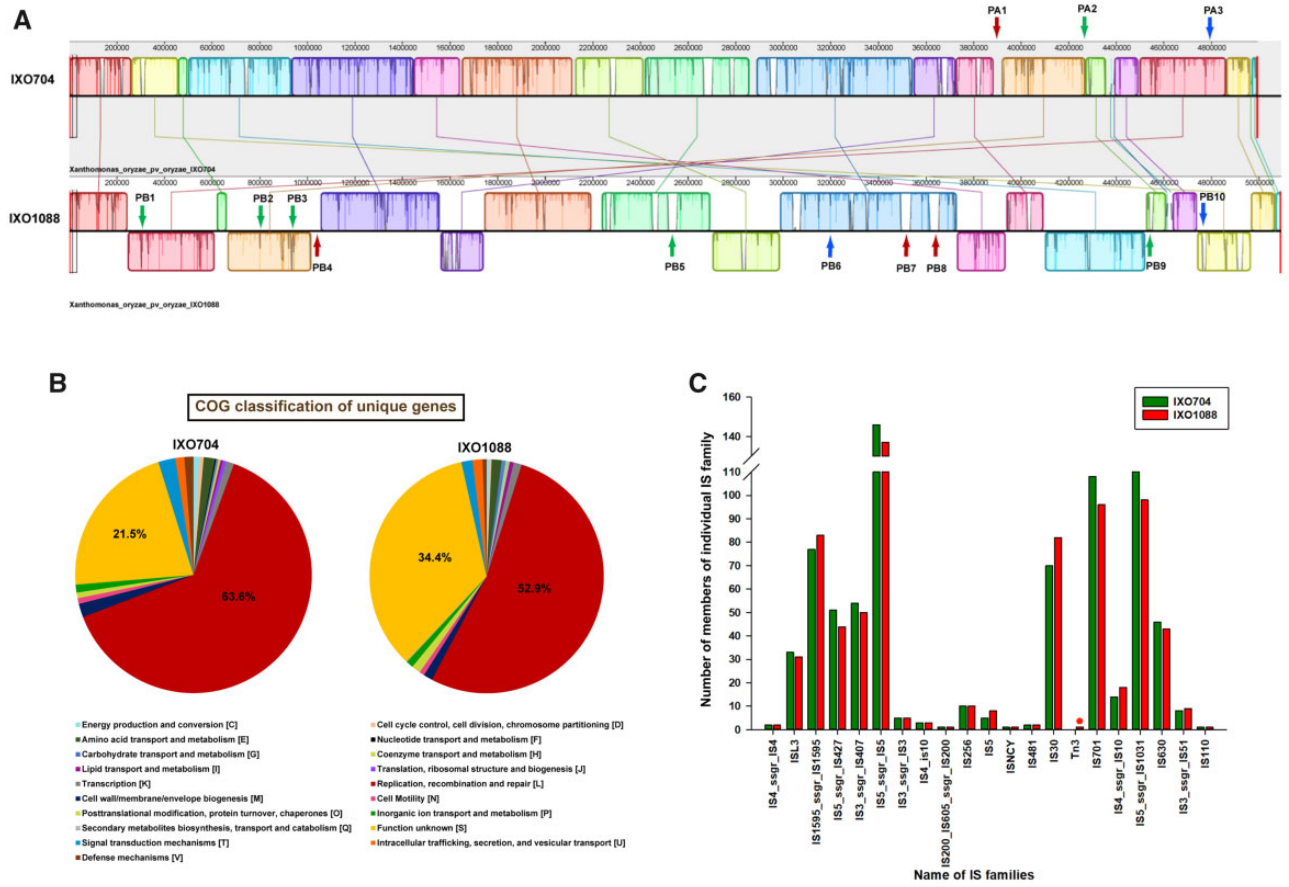


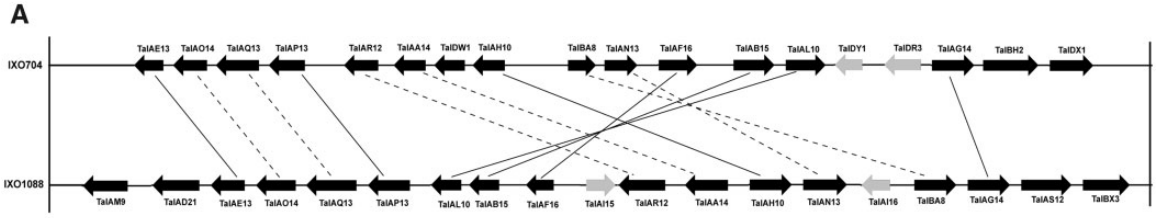

**Fig. 1.**—(*A*) Complete genome alignment of IXO704 and IXO1088 performed using progressive MAUVE. The scale represents coordinates of each genome. Different color blocks represent LCBs (Local Collinear Blocks) which are conserved segments in both the genomes. Within LCBs, white area represents low similarity regions or regions unique to one genome but absent in another. LCBs above black horizontal central line are in forward orientation and below this are in reverse orientation. Colored lines show rearrangement of LCBs between two genomes. Arrows above the scale represent putative prophage regions identified using PHASTER tool. Color code represents: maroon for intact phages (score >90), blue: incomplete phage regions (score 70–90), green: questionable phage regions (score <70). (*B*) Pie chart showing distribution of unique genes classified into cluster of orthologous groups in both the genomes. (*C*) Distribution of IS elements into different IS families in both the genomes. The asterisk sign shows the presence of Tn3 family transposase in IXO1088 but its absence in IXO704.

whereas 10 TALEs (TalAA14, TalAF16, TalAO14, TalAQ13, TalAP13, TalAR12, TalAL10, TalAH10, TalBA8, TalAN13) showed variation in at least one or more RVDs. TalDW1, TalDX1, and TalDY1 of IXO704 were assigned to new class with no orthologs in *Xanthomonas* family (fig. 2*B*).

In comparison to other strains, IXO1088 contains 16 TALEs, which are orthologs of PXO99A TALEs, with 9 TALEs, which are identical and 7 TALEs with variations in RVD sequences, whereas 3 TALEs, TalAL10, TalAS12, TalAM9 have no orthologs in PXO99A. Interestingly, in IXO704, out of 18 TALEs, 12 TALEs are orthologs of PXO99A with only 2 TALEs that are identical, whereas 10 TALEs possess variations in the RVDs and 6 TALEs (TalAL10, TalBH2, TalDR3, TalDW1, TalDX1, and TalDY1) have no orthologs. As TALEs are highly specific for their targets, variation in the RVDs can disrupt their target specificity. Similarly, presence of TALEs with no orthologs or new family of TALE can target new or alternative host genes thus eliminating success of particular resistance gene.

## Conclusion

Emergence of highly resistant strains that can overcome resistance mediated by all major resistance genes is a grave threat to rice cultivation. Complete genome studies of Xoo strains with XVP surprisingly revealed distinct variation in both content and sequence of TALome. Interestingly, acquisition of novel genes and numerous genomic rearrangements mediated by unique and large number of IS elements indicates non-TALE origin in variant and new pathotypes. Presence of large number of prophage elements that may be leading to genomic flux that can affect structure and function of genome resulting in origin of XVPs. Spread of phages within or between species can even lead to clonal diversification of




B

| TALE | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TalAA14 | NI | HG | NI | NG | HG | HD | NS | NG | HD | NN | NG | HG | NG | HD | HG | HD | HD | NI | NN | NG | | | | | | | | | |
| TalAA14 | NI | HG | NS | HG | HG | HD | NS | NG | HD | NN | NG | HG | NG | HD | HG | HD | HD | NI | NN | NG | | | | | | | | | |
| TalAB15 | NI | HG | NI | NI | NI | NN | HD | NS | NN | NS | NN | HD | NN | NI | HD | NN | NS | NG | | | | | | | | | | | |
| TalAB15 | NI | HG | NI | NI | NI | NN | HD | NS | NN | NS | NN | HD | NN | NI | HD | NN | NS | NG | | | | | | | | | | | |
| TalAE13 | NI | NN | NI | HG | HG | HD | NG | HD | HG | HD | HD | HD | HD | NG | | | | | | | | | | | | | | | |
| TalAE13 | NI | NN | NI | HG | HG | HD | NG | HD | HG | HD | HD | HD | HD | NG | | | | | | | | | | | | | | | |
| TalAF16 | NI | NN | NN | NI | NI | NI | HD | NS | HG | NN | NN | NN | NI | NI | NG | HD | | | | | | | | | | | | | |
| TalAF16 | NI | NN | NN | NI | NI | NI | HD | NS | HG | NN | NN | NN | NI | NI | HG | HD | | | | | | | | | | | | | |
| TalAG14 | NI | NG | NN | NG | NK | NG | NI | NN | NI | NN | NI | NN | NS | NG | NS | NN | NI | N* | NS | NG | | | | | | | | | |
| TalAG14 | NI | NG | NN | NG | NK | NG | NI | NN | NI | NN | NI | NN | NS | NG | NS | NN | NI | N* | NS | NG | | | | | | | | | |
| TalAH10 | NI | N* | NI | NS | NN | NG | NN | NS | N* | NS | NN | NS | N* | NN | HG | HD | NI | HD | HD | NG | | | | | | | | | |
| TalAH10 | NI | N* | NI | NS | NN | NG | NN | NS | N* | NS | NN | NS | N* | HD | HG | HD | NI | HD | HD | NG | | | | | | | | | |
| TalAL10 | NI | NS | HD | NG | NS | NN | HD | H* | NN | NN | HI | NN | HD | NG | HD | HD | HD | NG | | | | | | | | | | | |
| TalAL10 | NI | NS | HD | NG | NS | NN | HD | NG | NN | NN | HI | NN | HD | NG | HD | HD | HD | NG | | | | | | | | | | | |
| TalAN13 | NI | HG | NI | HG | NI | NI | NI | HD | NN | HD | NS | NG | NS | HD | NI | NI | NN | NI | NN | SS | NG | | | | | | | | |
| TalAN13 | NI | HG | NI | HG | NI | NI | NI | HD | NN | HD | NS | NG | SS | HD | NI | NI | NN | NI | NN | NI | NG | | | | | | | | |
| TalAO14 | NI | NN | N* | NG | NS | NN | NN | NN | NI | NN | NI | NG | HD | HD | NI | HG | N* | | | | | | | | | | | | |
| TalAO14 | NI | NN | N* | NG | NS | NN | NN | NN | NI | NN | NI | N* | HD | HD | NI | NG | NG | | | | | | | | | | | | |
| TalAP13 | HD | HD | HD | NG | N* | NN | HD | HD | N* | NI | NI | NN | HD | HI | ND | HD | NI | HD | NG | NG | | | | | | | | | |
| TalAP13 | HD | HD | HD | NG | N* | NN | HD | HD | N* | NI | NI | NN | HD | HI | ND | HD | NI | HD | NG | N* | | | | | | | | | |
| TalAQ13 | HD | HD | NN | NN | NI | NG | HD | S* | HG | HD | NG | N* | NG | HD | HD | N* | NI | NI | NN | HD | HI | ND | HD | NG | NN | HG | N* | | |
| TalAQ13 | HD | HD | NN | NN | NG | NG | HD | NS | HG | HD | NG | N* | HD | HD | HD | N* | NN | NI | NN | HD | HI | ND | HD | HG | NN | HG | N* | | |
| TalAR12 | NI | H* | NI | NN | NN | NN | NN | NN | HD | NI | NS | HG | HD | NI | N* | NS | NI | NI | HD | N* | NS | NI | N* | | | | | | |
| TalAR12 | NI | H* | NI | NN | NN | NN | NN | NN | HD | NI | HD | HG | HD | NI | N* | NS | NI | NI | HG | HD | NS | NS | NG | | | | | | |
| TalBA8 | NI | NS | HD | HG | NS | NN | HD | H* | NG | NN | NN | HD | NI | HD | HD | NG | | | | | | | | | | | | | |
| TalBA8 | NI | NS | HD | HG | NS | NN | HD | H* | NG | NN | NN | HD | HD | NG | HD | NI | NG | NS | NG | | | | | | | | | | |
| TalBH2 | NI | HG | NI | HG | NI | NI | NS | HD | NN | HD | HD | HD | NG | HD | N* | NI | HD | HD | NN | NS | HD | NN | NN | NG | NN | HD | NG | NS | N* |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| TalDW1 | NI | HG | N* | NI | NN | HG | HG | NN | HD | HD | HD | NG | HD | NS | NI | N* | NS | NG | | | | | | | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | | | | | | | | | | | |
| TalDX1 | NI | NG | NI | NI | NI | NN | HD | NI | HD | HD | NG | NI | NI | NS | NI | HD | NS | NN | HD | HD | NG | N* | | | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | | | | | | | | |
| TalDY1 (Pseudo) | NS | HD | NG | NG | NG | HD | HD | NI | NG | HD | NG | NG | NG | HD | HD | H* | | | | | | | | | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | | | | | | | | | | | | | |
| TalDR3 (Pseudo) | NS | HD | NG | NG | NG | NG | HD | HD | NG | HD | NI | HD | NG | HD | NI | NI | N* | | | | | | | | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | | | | | | | | | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| TalAM9 | NI | HG | NI | NN | NN | NI | NN | HD | NS | HD | NS | NS | NS | HD | NN | HD | NG | HD | HD | HD | NG | NG | | | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| TalAS12 | NI | HG | NI | NN | NS | HD | NN | HD | HG | HD | NI | NI | NN | NI | HD | HD | HD | HG | NN | NN | HD | NS | NN | HD | N* | NS | N* | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| TalBX3 | NN | HD | NI | HG | HD | NG | N* | HD | HD | NI | NG | NG | NI | HD | NG | NN | NG | NI | NI | NI | NI | N* | NS | N* | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| TalAD21 | NN | HD | NS | NG | HD | NN | N* | NI | HD | NS | HD | NN | HD | NN | HD | NN | NN | NN | NN | NN | NN | NN | HD | NG | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| TalAI16 (Pseudo) | NS | NG | NG | NG | NG | NG | NG | HD | HD | HD | NN | HD | NG | HD | HD | HD | HD | H* | | | | | | | | | | | |
| – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | | | | | | | | | | | |
| TalAI15 (Pseudo) | NS | HD | NG | NG | NG | NG | NG | HD | HD | HD | NN | HD | NG | HD | NI | HD | NN | N* | | | | | | | | | | | |

FIG. 2.—(*A*) Map of *tal* genes of IXO704 and IXO1088 genome. Black arrows represent full length *tal* genes and gray arrows represent pseudo *tal* genes. Solid lines show *tal* genes which are identical between both the genomes or have single variation in the RVD sequence, whereas dotted lines show orthologous *tal* genes with two or more variations in the RVD sequence. (*B*) Alignment of RVD sequences of TALEs encoded in the IXO704 and IXO1088 genomes. Green color shows TALEs encoded in the IXO704 genome, whereas red color shows TALEs encoded in the IXO1088 genome. RVD sequences of each repeat are shown in black and variation in the RVD sequence between TALEs of both the genomes is highlighted in orange color. Blank "–" represents no orthologous of TALEs between the genomes. The asterisk sign indicates the absence of amino acid at 13th position of a repeat.

pathogens. Moreover, finding large number of genes with atypical GC content indicates role of horizontal gene transfer mediated by mobile elements in XVPs. The resource and findings from this study will aid in further molecular studies and management of pathotypes in Xoo.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This work was supported by projects entitled "Mega-genomic insights into coevolution of rice and its microbiome (MICRA)" MLP0020 and "High throughput and integrative genomic approaches to understand adaptation of probiotic and pathogenic bacterium" (OLP148) of Council of Scientific and Industrial Research (CSIR), India. A.K. is supported by Department of Science and Technology (INSPIRE-Fellowship). Authors declare that the research was conducted

in absence of any commercial or financial support that could be considered as a potential conflict of interest.

## Authors Contribution

A.K. and K.B. carried out complete genome sequencing and NCBI submission. A.K. carried out downstream analysis and drafted the manuscript with inputs from K.B. and P.B.P. P.B.P. conceived the study and participated in its design and interpretation of data with inputs from A.K. and K.B. All authors have read and approved the manuscript.

## Literature Cited


Antony G, et al. 2010. Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. Plant Cell 22(11):3864–3876.

Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44(W1):W16–W21.

Bansal K, Midha S, Kumar S, Patil PB. 2017. Ecological and evolutionary insights into *Xanthomonas citri* pathovar diversity. Appl. Environ. Microbiol. 83(9):e02993–16.

Bansal K, Kumar S, Patil PB. 2018. Complete genome sequence reveals evolutionary dynamics of an emerging and variant pathovar of *Xanthomonas euvesicatoria*. Genome Biol Evol. 10(11):3104–3109.

Boch J, et al. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326(5959):1509–1512.

Bogdanove AJ, Schornack S, Lahaye T. 2010. TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol. 13(4):394–401.

Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner (No. LBNL-7065E). Berkeley (CA): Lawrence Berkeley National Lab (LBNL).

Busungu C, Taura S, Sakagami J-I, Ichitani K. 2016. Identification and linkage analysis of a new rice bacterial blight resistance gene from XM14, a mutant line from IR24. Breed Sci. 66(4):636–645.

Carpenter SC, et al. 2018. A strain of an emerging Indian *Xanthomonas oryzae* pv. oryzae pathotype defeats the rice bacterial blight resistance gene xa13 without inducing a clade III SWEET gene and is nearly identical to a recent Thai isolate. Front Microbiol. 9:2703.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25(1):119–120.

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14(7):1394–1403.

Deng D, et al. 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. Science 335(6069):720–723.

Grau J, et al. 2016. AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. Sci Rep. 6(1):21077.

Iyer AS, McCouch SR. 2004. The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. Mol Plant Microbe Interact. 17(12):1348–1354.

Jensen LJ, et al. 2007. eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. 36(Database):D250–D254.

Kaur A, Bansal K, Kumar S, Sonti RV, Patil PB. 2019. Complete genome dynamics of a dominant-lineage strain of *Xanthomonas oryzae* pv. oryzae harbouring a novel plasmid encoding a type IV secretion system. Access Microbiol. 1(9):acmi.0.000063.

Kim S-M, et al. 2015. Identification and fine-mapping of a new resistance gene, *Xa40*, conferring resistance to bacterial blight races in rice (*Oryza sativa* L.). Theor Appl Genet. 128(10):1933–1943.

Li C, Wei J, Lin Y, Chen H. 2012. Gene silencing using the recessive rice bacterial blight resistance gene xa13 as a new paradigm in plant breeding. Plant Cell Rep. 31(5):851–862.

Mishra D, et al. 2013. Pathotype and genetic diversity amongst Indian isolates of *Xanthomonas oryzae* pv. oryzae. PLoS One 8(11):e81996.

Moscou MJ, Bogdanove AJ. 2009. A simple cipher governs DNA recognition by TAL effectors. Science 326(5959):1501.

Niño-Liu DO, Ronald PC, Bogdanove AJ. 2006. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. Mol Plant Pathol. 7(5):303–324.

Noh T-H, et al. 2007. Effects of bacterial leaf blight occurrence on rice yield and grain quality in different rice growth stage. Res Plant Dis. 13(1):20–23.

Page AJ, et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31(22):3691–3693.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25(7):1043–1055.

Pérez-Quintero AL, et al. 2015. QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. Front Plant Sci. 6:545.

Römer P, et al. 2010. Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. oryzae. N Phytol. 187(4):1048–1057.

Schmid M, et al. 2018. Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. Nucleic Acids Res. 46(17):8953–8965.

Schornack S, Moscou MJ, Ward ER, Horvath DM. 2013. Engineering plant disease resistance based on TAL effectors. Annu Rev Phytopathol. 51(1):383–406.

Sugio A, Yang B, Zhu T, White FF. 2007. Two type III effector genes of *Xanthomonas oryzae* pv. oryzae control the induction of the host genes OsTFIIA$_\gamma$1 and OsTFX1 during bacterial blight of rice. Proc Natl Acad Sci U S A. 104(25):10720–10725.

Tian D, et al. 2014. The rice TAL effector–dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. Plant Cell 26(1):497–515.

Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 12(3):R30.

Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.

Wang C, et al. 2015. XA23 is an executor R protein and confers broad-spectrum disease resistance in rice. Mol Plant. 8(2):290–302.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 13(6):e1005595.

Zhang J, Yin Z, White F. 2015. TAL effectors and the executor R genes. Front Plant Sci. 6:641.


**Associate editor**: Rachel Whitaker