Extensive genomic rearrangements along with distinct mobilome and TALome is associated with extreme pathotypes of a rice pathogen

Amandeep Kaur¹, Kanika Bansal¹ and Prabhu B. Patil¹*

Affiliation

¹Bacterial Genomics and Evolution Laboratory, CSIR-Institute of Microbial Technology, Chandigarh, India.

*Author for Correspondence: Dr Prabhu B. Patil, Bacterial Genomics and Evolution Laboratory, CSIR-Institute of Microbial Technology, Chandigarh, India. pbpatil@imtech.res.in

Data deposition

The genomes of IXO704 and IXO1088 are submitted to NCBI with accession number CP040604 and CP040687 respectively. The plasmid pIXO-704 present in IXO704 strain is submitted with accession number CP040603.
Abstract

*Xanthomonas oryzae* pv. oryzae (Xoo) is a serious pathogen of rice which displays tremendous inter-strain 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 to 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 a complete genome sequence of Xoo strains with extreme virulent pathotypes (XVPs) characterized based on their reaction towards 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, XPVs 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 (TALE) proteins, which are major virulence determinants of *Xanthomonas* pathogen show marked variation in the TALE 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.

**Keywords:** *Xanthomonas oryzae*, pathotype, TALEs (Transcription activator-like effectors), 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, Ronald, & Bogdanove, 2006); (Noh et al., 2007). For successful invasion, Xoo secretes various virulence factors such as TALEs (transcription activator-like effectors) into the host cells modulating host cell machinery for its own benefit (Boch et al., 2009); (Moscou & Bogdanove, 2009). These effectors bind to the promoter elements and regulate expression of either susceptible (*S*) genes, that make environment suitable for bacterial growth (Römer et al., 2010); (Bogdanove, Schornack, & Lahaye, 2010) or resistance (*R*) genes that limit bacterial infection by generating host defence response (Bogdanove et al., 2010); (Schornack, Moscou, Ward, & Horvath, 2013); (Zhang, Yin, & White, 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 position of each repeat (Boch et al., 2009) (Moscou & 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 (D. Deng et al., 2012). Variation at 12th and 13th position determine the nucleotide specificity of each repeat and together these residues are known as repeat variable diresidues (RVDs) (Boch et al., 2009) (Moscou & 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 & McCouch, 2004) (Wang et al., 2015) (Kim et al., 2015a) (Busungu, Taura, Sakagami, & Ichitani, 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, Wei, Lin, & Chen, 2012). Similarly, engineering ‘R’ genes to include EBEs for multiple TALEs can generate a defence 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, Yang, Zhu, & White, 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 eleven 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 pathotype 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 are enabling investigation of TALEs repertoire, diversity and their host targets towards evolution of virulent strains which is not possible using draft genomes (Bansal, Kumar, & Patil, 2018). Other than \( tal \) genes, the present study also highlight 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 hrs. Genomic DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen). Quantity and quality of DNA was assessed using Nanodrop and Qubit 2.0 Fluorometric. For Nanopore sequencing, 3µg of initial DNA was used for DNA end prep using NEBNext Ultra II End repair/dA-Tailing modules (NEB,USA). Library was prepared using Ligation Sequencing Kit 1D (SQK-LSK108) and Native Barcoding Kit 1D (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) (Oxford Nanopore Technologies) (v.1.13.1) for 48hrs. Raw reads were base called using Albacore v2.3.3 software (http://community.nanoporetech.com).

**Genome assembly and annotation using ONT and Illumina reads**

Reads obtained after demultiplexing were hybrid (both ONT and Illumina reads) assembled using Unicycler v.0.4.8-beta (Wick, Judd, Gorrie, & Holt, 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, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015). Genome coverage was checked using BBmap tool v38.42 (Bushnell, 2014). Genomes were annotated using NCBI PGAP
Genome comparison

Genome comparison between IXO1088 and IXO704 strain was performed using progressive MAUVE tool with default parameters (Darling, Mau, Blattner, & Perna, 2004). Pan-genome analysis was performed using Roary (Page et al., 2015). Functional categorization of unique genes into cluster of orthologous (COG) classes was obtained using EggNOG (Jensen et al., 2007). The pictorial representation of IXO1088 and IXO704 genomes was drawn using DNA plotter (Carver, Thomson, Bleasby, Berriman, & Parkhill, 2008).

TALE analysis

All tal genes were identified using AnnoTALE software (Grau et al., 2016). Identified TALE’s 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

IS (Insertion elements) elements were identified using ISsaga tool (Varani, Siguier, Gourbeyre, Charneau, & Chandler, 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 bp 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, Bansal, Kumar, Sonti, & Patil, 2019). A total 4,762 CDS for IXO704 and 4,821 CDS for IXO1088 were found in the genome. Genome coverage for IXO704 and IXO1088 was 96x and 138x respectively with average GC content of 63.7% for both the genomes. Further, genome completeness and contamination was found to be 100% and 0% respectively for both the genomes. The pictorial representation of both the genomes is provided in Supplementary Figure 1.

**Extensive structural variation in the genome architecture of XVPs**

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 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, inversions etc. (Figure 1A). 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 towards 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 5507 genes with 4139 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 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 points towards their acquisition through horizontal gene transfer events. Unique genes were then functionally annotated into COG (cluster of orthologous groups) classes (Figure 1B). 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 category ‘L’ and ‘S’ may be contributing towards genome dynamics such as insertion elements, transposases, prophages etc. Therefore, we investigated the mobilomes of two genomes. In IXO1088, 10 prophage regions were identified with three intact regions of 16.6kb, 43.8kb and 47.5kb whereas, in IXO704, only three prophage regions were identified with one intact region of 21.3kb (Figure 1A). Detailed features of prophage regions with their location, size and number of proteins are given in Supplementary Table 1.

Since 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 (Figure 1C), there is variation in number and type of IS elements in both pathotypes. 747 IS elements were observed for IXO704 classified into 20 families. Whereas in IXO1088, total 725 IS elements were found classified into 21 families (Figure 1C). 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 genome contains 16 and 17 TALEs respectively with two pseudo TALEs in the both the genomes (Figure 2A). The RVD sequences of some TALEs varies in both the genomes with variations at both 12th and 13th position (Figure 2B). 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 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 (Figure 2B).

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 two TALEs that are identical whereas 10 TALEs possess variations in the RVDs. and six 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 a Xoo strains with extreme virulent pathotype 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 indicate 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 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.

**Authors Contribution:**

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

**Acknowledgements**

This work was supported by a project entitled “Mega-genomic insights into co-evolution of rice and its microbiome (MICRA)” MLP0020 of Council of Scientific and Industrial Research (CSIR) and “High throughput and integrative genomic approaches to understand adaptation of probiotic and pathogenic bacterium” (OLP148). AK is supported by DST-INSPIRE fellowship. The authors declare that the research was conducted in the absence of any commercial or financial support that could be considered as a potential conflict of interest.

**References:**

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., & Yang, B. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. *The Plant Cell, 22*(11), 3864-3876.
Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*, 44(W1), W16-W21.

Bansal, K., Kumar, S., & Patil, P. B. (2018). Complete genome sequence reveals evolutionary dynamics of an emerging and variant pathovar of Xanthomonas euvesicatoria. *Genome biology and evolution*.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., . . . Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959), 1509-1512.

Bogdanove, A. J., Schornack, S., & Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology*, 13(4), 394-401.

Bushnell, B. (2014). BBMap: a fast, accurate, splice-aware aligner.

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. *Breeding science*, 66(4), 636-645.

Carpenter, S. C., Mishra, P., Ghoshal, C., Dash, P., Wang, L., Midha, S., . . . Singh, N. K. (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. *Frontiers in microbiology*, 9, 2703.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., & Parkhill, J. (2008). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*, 25(1), 119-120.

Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*, 14(7), 1394-1403.
Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., . . . Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science, 335*(6069), 720-723.

Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G., . . . Boch, J. (2016). AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci Rep, 6*.

Iyer, A. S., & McCouch, S. R. (2004). The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Mol Plant Microbe Interact, 17*(12), 1348-1354.

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., & Bork, P. (2007). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic acids research, 36*(suppl_1), D250-D254.

Kaur, A., Bansal, K., Kumar, S., Sonti, R. V., & Patil, P. B. (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 Microbiology, acmi000063*.

Kim, S.-M., Suh, J.-P., Qin, Y., Noh, T.-H., Reinke, R. F., & Jena, K. K. (2015a). 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.

Kim, S.-M., Suh, J.-P., Qin, Y., Noh, T.-H., Reinke, R. F., & Jena, K. K. (2015b). Identification and fine-mapping of a new resistance gene, *Xa40*, conferring resistance to bacterial blight races in rice (*Oryza sativa* L.). *Theoretical and applied genetics, 128*(10), 1933-1943.

http://mc.manuscriptcentral.com/gbe
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 reports, 31*(5), 851-862.

Mishra, D., Vishnupriya, M. R., Anil, M. G., Konda, K., Raj, Y., & Sonti, R. V. (2013). Pathotype and genetic diversity amongst Indian isolates of *Xanthomonas oryzae* pv. oryzae. *PLoS One, 8*(11), e81996.

Moscou, M. J., & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science, 326*(5959), 1501-1501.

NIÑO-LIU, D. O., Ronald, P. C., & Bogdanove, A. J. (2006). *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Molecular Plant Pathology, 7*(5), 303-324.

Noh, T.-H., Lee, D.-K., Park, J.-C., Shim, H.-K., Choi, M.-Y., Kang, M.-H., & Kim, J.-D. (2007). Effects of bacterial leaf blight occurrence on rice yield and grain quality in different rice growth stage. *Research in Plant Disease, 13*(1), 20-23.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., . . . Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics, 31*(22), 3691-3693.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research, 25*(7), 1043-1055.

Pérez-Quintero, A. L., Lamy, L., Gordon, J., Escalon, A., Cunnac, S., Szurek, B., & Gagnevin, L. (2015). QucTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. *Frontiers in plant science, 6*, 545.

Römer, P., Recht, S., Strauß, T., Elsaesser, J., Schornack, S., Boch, J., . . . Lahaye, T. (2010). Promoter elements of rice susceptibility genes are bound and activated by specific
TAL effectors from the bacterial blight pathogen, Xanthomonas oryzae pv. oryzae. New Phytologist, 187(4), 1048-1057.

Schmid, M., Frei, D., Patrignani, A., Schlapbach, R., Frey, J. E., Remus-Emsermann, M. N., & Ahrens, C. H. (2018). Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats. Nucleic acids research, 46(17), 8953-8965.

Schornack, S., Moscou, M. J., Ward, E. R., & Horvath, D. M. (2013). Engineering plant disease resistance based on TAL effectors. Annual Review of Phytopathology, 51, 383-406.

Sugio, A., Yang, B., Zhu, T., & White, F. F. (2007). Two type III effector genes of Xanthomonas oryzae pv. oryzae control the induction of the host genes OsTFIIAγ1 and OsTFX1 during bacterial blight of rice. Proceedings of the National Academy of Sciences, 104(25), 10720-10725.

Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., ... Murata-Hori, M. (2014). The rice TAL effector–dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. The Plant Cell, 26(1), 497-515.

Varani, A. M., 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 biology, 12(3), R30.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelli, A., Sakthikumar, S., ... Young, S. K. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One, 9(11), e112963.

http://mc.manuscriptcentral.com/gbe
Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., . . . Zhang, M. (2015). XA23 is an executor R protein and confers broad-spectrum disease resistance in rice. *Molecular plant*, 8(2), 290-302.

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, 13(6), e1005595.

Zhang, J., Yin, Z., & White, F. (2015). TAL effectors and the executor R genes. *Frontiers in plant science*, 6, 641.

**Figure legends:**

**Figure 1:** (A) Complete genome alignment of IXO704 and IXO1088 performed using progressive MAUVE. The scale represent coordinates of each genome. Different colour 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. Coloured lines shows rearrangement of LCBs between two genomes. Arrows above the scale represent putative prophage regions identified using PHASTER tool. Colour 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 different COG (cluster of orthologous groups) families in both the genomes. (C) Distribution of IS elements into different IS families in both the genomes. ‘*’ sign shows the presence of Tn3 family transposase in IXO1088 but its absence in IXO704.
**Figure 2:** (A) Map of *tal* genes of IXO704 and IXO1088 genome. Black arrows represent full length *tal* genes and grey arrows represent pseudo *tal* genes. Solid lines show *tal* genes which are identical between both the genomes whereas dotted lines shows 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 colour shows TALEs encoded in the IXO704 genome whereas red colour 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 are highlighted in orange colour. Blank ‘--‘represents no orthologous of TALEs between the genomes. The * asterisk indicates the absence of amino acid at 13\textsuperscript{th} position of a repeat.

**Supplementary material**

**Supplementary Figure 1:** Circular representation of IXO1088 and IXO704 genomes. Rings represents from outside to inside (i) protein coding genes (forward strand), (ii) protein coding genes (reverse strand), (iii) Phage protein coding genes (black), TAL effectors (red), (v) GC content, (vi) GC skew.

**Supplementary Table 1:** Putative prophage regions predicted using PHASTER tool with their size, number of phage proteins, chromosomal location and GC content.
Figure 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. Coloured lines show rearrangements of LCBs between two genomes. Arrows above the scale represent putative prophage regions identified using PHASTER tool. Colour 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 different COG (cluster of orthologous groups) families in both the genomes. (C) Distribution of IS elements into different IS families in both the genomes. '*' sign shows the presence of Tn3 family transposase in IXO1088 but its absence in IXO704.
Figure 2: (A) Map of tal genes of IXO704 and IXO1088 genome. Black arrows represent full length tal genes and grey arrows represent pseudo tal genes. Solid lines show tal genes which are identical between both the genomes 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 colour shows TALEs encoded in the IXO704 genome whereas red colour 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 are highlighted in orange colour. Blank ‘--’ represents no orthologous of TALEs between the genomes. The * asterisk indicates the absence of amino acid at 13th position of a repeat.

260x262mm (300 x 300 DPI)