Xanthomonas is pathologically diverse group of bacteria with the controversial taxonomic status of its members. This has hindered systematic studies to understand their origin and evolution. The advent of genomics era has revolutionized the field of bacteriology. Now we can generate and access complete genotype of an organism at unprecedented rate and scale. Hence, whole genome sequence based studies of strains, pathovars and species, particularly from their putative centre of origin can fill the gap and provide much finest insights on their diversity.

5.1 Phylogenomic based insights into origin and evolution of fruit pathovars

India is one of the largest cultivators of fruits in the world and is a putative centre of origin of major fruit crops. Interestingly, the diseases caused by *Xanthomonas* on many fruits were first noticed in India. These are *X. axonopodis* pv. citri (*Xac*), the causal agent of canker in citrus, *X. citri* pv. mangiferaindicae (*Xmi*), the causal agent of black spot in mango, *X. axonopodis* pv. punicae (*Xap*), the causal agent of oily spot disease in pomegranate and *X. campestris* pv. viticola (*Xvt*), the causal agent of bacterial leaf spot of grapes. However, the taxonomic status and relationship of these pathovars are highly controversial and confusing. *Xac* was one of the first plant pathogenic bacterial genomes to be sequenced Hence, I worked towards obtaining whole genome sequences of remaining three pathovars. Further we carried out annotation and published the genomes in public domain.

Analysis of the genomes revealed their comparable sizes, referring to the absence of any reductive evolution. Phylogenomic analysis using conserved phylogenomic marker genes showed that they belong to a monophyletic group and suggested their emergence from a common ancestor in recent past. Further, ANI analysis showed that the four fruit pathovars belong to same species with high genome identity of nearly 99% in core genome. Recombinational analysis revealed the substantial impact of recombination on the evolution of these strains. Hence, genomes of the *Xanthomonas* pathovars infecting fruit crops were compared to identify the genes shared by all and the genes present uniquely in each pathovar. Each of the four pathovars contains ~300-500 unique genes that could have a putative role in providing the host specificity. The analysis also revealed the presence of a ~90 kb cluster in *Xmi*, which is absent from all others being compared. Further

investigation led to the identification of the cluster as a Non-Ribosomal Peptide/Polyketide Synthesis (NRPS/PKS) pathway, which could be a potential source of a novel antibiotic.

Xanthomonas typically is yellow in color, due to the production of the pigment xanthomonadin encoded by a specific gene cluster. The atypical white color of the pathovars *Xmi* and *Xvt* led us to further investigation in this regard. The analysis revealed that the gene cluster encoding for the pigment xanthomonadin is disrupted by the presence of transposase in *Xmi* and have a frameshift mutation in a gene in *Xvt*, which is imperative for pigment production. Previously published comparative studies have revealed that some gene clusters in *Xanthomonas* are hyper variable. Hence, I was interested in loci that are hypervariable or variations that occur at specific regions of the genome. Such clusters were analyzed in detail, specifically integron encoding cassette, LPS and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR). Comparison of Integron cassette in the fruit pathovars revealed a high level of variability in the gene content and high transpositional activity in the cassette.

All of the fruit pathovars compared showed the presence of a specific type of CRISPR repeats in their genomes while a second type of the CRISPR repeats were present specifically in the citrus and pomegranate pathogens. Genes involved in biosynthesis and transport of LPS are known to be involved in virulence and defense responses. Our analysis on this loci showed the presence of diverse cassette types in the members of these lineages, with type *Xac* being the most prevalent. Hence, recombination, horizontal gene transfer, point and deletion mutations along with hypervariable loci has played an important role in the diversification of the genome of these host specific pathovars.

5.2 Population genomics insights into origin and evolution of rice pathovars

Rice is a major staple crop of India and world. However, its cultivation is severely limited by two *Xanthomonas* pathovars. *X. oryzae* pv. oryzae (*Xoo*) infects xylem causing bacterial leaf bight while *X. oryzae* pv. oryzicola (*Xoc*) infects parenchyma causing bacterial leaf streak. However, the origin of these pathovars is still mysterious. *Xoo* is particularly devastating and many highly virulent strains are emerging in India. India is a major region of the diversity of the rice and also one of the largest cultivators of rice plant. However, the strains that have been sequenced are from other parts of Asia, Africa and the USA. Hence,

my focus was to sequence and analyse genome of a large number of isolates to throw light on the origin of these strains, pathovars and species *oryzae* itself.

Phylogenomic markers based analysis of the genome of 101 *X. oryzae* strains, collected from India in last two decades, revealed that Indian *Xoo* strains along with other Asian strains form a cluster that is distinct from the all *Xoc* strains, *Xoo* strains from Africa and *X. oryzae* strains from the USA. It also suggested that *Xoc* is probably an accidental and variant lineage that has emerged from a diverse *Xoo* population. Whole genome based analysis suggested that Indian *Xoo* population exhibits epidemiological population structure with a major clonal lineage and four minor but diverse lineages with a high rate of recombination and mutation. The lineages are also diverse with respect to the geographical location of their isolation, where lineage II and V are restricted to eastern and northern regions of the country, rest of the lineages are more widespread.

Rice has been domesticated in Asia first and India is one of the highest rice cultivating countries. Hence, I carried out molecular dating of Indian lineages of *Xoo*. SNP analysis has allowed us to estimate the evolutionary divergence time of the *Xoo* population using SNP rate of $1x10^{-7}$ SNPs per site per year. The molecular clock estimate suggested the ancestor of Indian *Xoo* lineages started diversifying ~14,000 years ago, the estimated age of Neolithic revolution and demographic change and probable domestication of rice. It also suggested the emergence of the major clonal lineage of *Xoo* ~2000-3000 years ago, that could be due to the advancements in agriculture technologies. Further SNP analysis on the highly pathogenic strains of pathotype 11 (IXO1088 and IXO1104) and three strains (IXO651, IXO685 and IXO1221) that can breakdown the most effective resistance gene (*Xa21*) in the population suggested a recent emergence of their high virulence.

Unraveling relationship of *X. oryzae* pathovars and the population structure of *Xoo* allowed us to systematically study the variation at gene(s) whose product is a potential pathogen/damage associated molecular pattern (PAMP/DAMP). Interestingly, the well-known PAMP gene coding for flagellin (*fliC*) is under purifying selection in *Xoo* population, suggesting it to a potential drug target. Another recently recognized PAMP, a secreted peptide encoded by *raxX* gene harbors variant alleles in isolates that can breakdown *Xa21* gene that encodes a receptor for RaxX peptide. Interestingly, the cellobiosidase (*cbsA*) gene that is known as DAMP, shows evidence of positive selection in *Xoo* population. I also looked at variation in genes that encode effector proteins that are

secreted into the plant cell by Type 3 secretion system. The analysis revealed that out of 24 T3Es, five (AvrBs2, XopI, XopQ, XopR and XopV) are present universally in all *Xoo* and can be further targeted for durable resistance in rice against *Xoo*.

The study also provided an opportunity to look at hypervariable loci like CRISPR and LPS. Surprisingly CRISPR locus is absent in the USA and African lineages including *Xoc*, but seems to have acquired in the ancestor that gave rise to lineages in India and Asia. The number of repeats is very high in clonal lineage, L-I that explains the low rate of recombination, horizontal gene transfer and variation. CRSIPR repeats may have also protected the clonal lineage from bacteriophage attack allowing it to dominate the population of *Xoo* in India. Interestingly, the whole of *Xoo* population in India has only two type of LPS cassettes i.e. BXO1 and BXO8 type. The latter seems to be ancestral while former is of recent origin through horizontal gene transfer. While the diverse lineages harbor either of the cassettes, the isolates in L-I lineage have the only BXO1 type of LPS cassette. The variant LPS in isolates of L-I lineage, might have played a role in its success by protecting recognition by bacteriophages and defense responses of host plant. Overall, the inter-pathovar variation is more random as seen in the case of fruit pathovars that infect distinct fruit plants while intra-pathovar variation is more systematic as seen in the case of *Xoo* population and lineages.

5.3 Evolutionary genomics of *X. axonopodis* - connecting the evolution of fruit and rice pathovars

Phylogenomics studies clearly established that fruit pathovars belonged to one species. Further to clarify the exact specie status and phylogeny of the fruit pathovars, we selected the all type strains of species that were historically assigned to fruit pathovars and their phylogenetic relatives. In this direction, I carried out sequencing of type strains of *X. axonopodis* DSM3585 (*Xaxn3585*) *X. citri* subsp. citri LMG 9322 (*Xc9322*), *X. euvesicatoria* LMG 27970 (*Xe27970*), *X. alfalfae* subsp. alfalfae LMG 495 (*Xaf495*) along with pathovar reference strain *Xanthomonas axonopodis* pv. manihotis LMG 784 (*Xam784*). The genome of 4 type strains of species *X. fuscans* subsp. fuscans strain NCPPB 381 (*Xf381*), *X. perforans* 91-118 (*Xp91-118*), *X. vasicola* strain NCPPB 2417 (*Xv2417*) and *X. oryzae* ATCC 35933 (*Xo35933*) were accessed through NCBI.

Surprisingly, phylogenomic analysis with type strains revealed that all the four fruit pathovars belong to *X. citri* and not to *X. axonopodis* or *X. campestris* to which these fruit pathovars were originally classified. Further analysis using type strains of representative members of the both lineages with other sequenced genomes showed that lineage *X. axonopodis* further consists of at least five species, where only one strain *X. axonopodis* pv. vasculorum strain NCPPB 900 belongs to species *X. axonopodis*, while all the other pathovars previously named as *X. axonopodis* are misclassified. Interestingly, *X. axonopodis* was found to be phylogenomically equidistant to *X. oryzae* and *X. citri*, the species whose pathovars are the focus of my work.

Phylogenomic tree based on 31 phylogenomic marker genes confirmed that *X. axonopodis* is indeed a phylogenetic intermediate to *Xoo* and *Xac*. Because of such intermediate relationship, traditional classification methods like rep-PCR, DNA-DNA hybridization and or limited gene sequences failed to resolve the relationship.

It is pertinent to note that only *X. oryzae* and *X. citri* have a putative origin in Asia, while *X. axonopodis* has a putative origin in South America. Interestingly, *X. oryzae* and *X. axonopodis* infect monocots while *X. citri* infects dicots. Hence, these three related species can be ideal for ecological and evolutionary genomic studies. For this I extended the study by comparing the genomes of all other taxonomic relatives and strains of *X. axonopodis*, *X. citri* and *X. oryzae* available in public domain for in depth comparative genomics.

Xanthomonas contains various pathogenicity islands that are known to be core to this genus. Genomic analysis revealed that the cluster encoding for biosynthesis of xanthan gum (*gum*), hypersensitive reaction and pathogenicity (*hrp*) gene cluster, regulation of pathogenicity factors (*rpf*) gene cluster, lipopolysaccharide (LPS) and one of the type II secretory pathway (*xps*) gene cluster are core to *X. axonopodis*, *X. citri* and *X. oryzae* and they represent the ancestral acquisition, while the other type II secretory pathway (*xcs*) gene cluster is either absent or has most of the part deleted in vascular pathogens of monocotyledonous plants suggesting the dispensability of *xcs* cluster in them.

Our analysis on LPS locus revealed the presence of diverse cassette types in the members of these lineages. Interestingly despite the presence of this variability, one of the cassette type (*Xac*) has its homolog in several members of these lineages. This cassette type is shared by citrus and pomegranate pathovars, *X. axonopodis*, *X. alfalfae* and *X. oryzae* type strain, pathovar manihotis, some of the other pathogenic strains of *Xoo* including the

ancestral strains and also partially by *X. perforans*, *X. vasicola*, *Xoc*, *Xmi* and *Xvt*, again suggesting that *Xac* cassette type could be the ancestral cassette.

Overall my work has shown the power of whole genome sequence in resolving the relationship and differences in a controversial group of bacteria like *Xanthomonas*. My work also showed that such genomic studies will be more meaningful if they involve systematic study of pathovars and their population from their putative centre of origin or eco-evo relatives.