



Phylogenomic Insights into Diversity and Evolution of Nonpathogenic *Xanthomonas* Strains Associated with Citrus

Kanika Bansal,^a Sanjeet Kumar,^{a*} Prabhu B. Patil^a

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

ABSTRACT *Xanthomonas* species are primarily known as a group of phytopathogens infecting diverse plants. Recent molecular studies reveal the existence of potential novel species and strains of *Xanthomonas* following a nonpathogenic lifestyle. In the present study, we report whole-genome sequences of four nonpathogenic strains from citrus (NPXc). Taxonogenomics revealed the surprising diversity, as each of these three isolates were found to be potential novel species that together form a citrus-associated nonpathogenic *Xanthomonas* species complex (NPXc complex). Interestingly, this NPXc complex is related to another nonpathogenic species, *Xanthomonas sontii*, from rice (NPXr). On the other hand, the fourth NPXc isolate was found to be related to nonpathogenic isolates from walnut (NPXw); altogether, they form a potential taxonomic outlier of pathogenic *Xanthomonas arboricola* species. Furthermore, genomic investigation of well-characterized pathogenicity clusters in NPXc isolates revealed lifestyle-specific gene content dynamics. Primarily, genes essential for virulence (i.e., type 1 secretion system [T1SS], T2SS and its effectors, T3SS and its effectors, T4SS, T6SS, adhesins, and *rpf* gene cluster) and adaptation (i.e., *gum*, iron uptake and utilization, xanthomonadin, and two-component systems) were depicted by comparative genomics of a *Xanthomonas* community comprising diverse lifestyles. Overall, the present analysis confers that nonpathogenic isolates of diverse hosts phylogenomically converge and are evolving in parallel with their pathogenic counterparts. Hence, there is a need to understand the world of nonpathogenic isolates from diverse and economically important hosts. Genomic knowledge and resources of nonpathogenic strains will be invaluable in both basic and applied research of the genus *Xanthomonas*.

IMPORTANCE *Xanthomonas citri* is one of the top phytopathogenic bacteria and is the causal agent of citrus canker. Interestingly, *Xanthomonas* is also reported to be associated with healthy citrus plants. The advent of the genomic era enabled us to carry out a detailed evolutionary study of a *Xanthomonas* community associated with citrus and other plants. Our genome-based investigations have revealed hidden and extreme interstrain diversity of nonpathogenic *Xanthomonas* strains from citrus plants, warranting further large-scale studies. This indicates an unexplored world of *Xanthomonas* from healthy citrus plant species that may be coevolving as a species complex with the host, unlike the variant pathogenic species. The knowledge and genomic resources will be valuable in evolutionary studies exploring its hidden potential and management of pathogenic species.

KEYWORDS nonpathogenic *Xanthomonas*, evolution, ecology, pathogenicity, comparative genomics, virulence- and plant adaptation-related gene clusters, speciation

Xanthomonas is a complex phytopathogen with an array of host ranges, including 268 dicots and 124 monocots (1–3). The genus comprises 34 different species with diversified pathogenic potential. *Citrus* spp. are widely known to be infected with

Citation Bansal K, Kumar S, Patil PB. 2020. Phylogenomic insights into diversity and evolution of nonpathogenic *Xanthomonas* strains associated with citrus. *mSphere* 5: e00087-20. <https://doi.org/10.1128/mSphere.00087-20>.

Editor Hideyuki Tamaki, National Institute of Advanced Industrial Science and Technology

Copyright © 2020 Bansal et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Prabhu B. Patil, pbpatil@imtech.res.in.

* Present address: Sanjeet Kumar, Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany.

Received 30 January 2020

Accepted 18 March 2020

Published 15 April 2020

Xanthomonas citri, a quarantine pathogen (4–7). Owing to its economic importance as one of the top phytopathogenic bacteria, its evolution into successful species is of interest to researchers, policy makers, etc. Moreover, it is the most successful *Xanthomonas* species, harboring 22 pathovars known to be host specific to diverse plants (8).

Interestingly, of these 34 species, only two species, i.e., *Xanthomonas sontii* and *Xanthomonas maliensis*, are exclusively nonpathogenic *Xanthomonas* strains associated with rice (NPXr) (9–12). *Xanthomonas arboricola* is known to comprise pathogenic strains of different pathovars affecting stone and nut fruit trees (13); nonpathogenic *X. arboricola* strains are also isolated from walnut trees (11, 12). Another well-studied nonpathogenic strain, *Xanthomonas sacchari* R1, was isolated from rice seeds and is known to be antagonistic to rice pathogens, which was later identified as a misclassified strain of *X. sontii* (14, 15). In addition to these, there are various unclassified and unexplored nonpathogenic isolates reported from a diversity of host plants (16). These host plants are already associated with successful pathogenic species of *Xanthomonas* and reported to cause devastating diseases worldwide. Although nonpathogenic strains have a significant impact on the evolution of *Xanthomonas* genus as a whole, they are widely overlooked due to their lower economic importance.

The genomic era has radically transformed our basic understanding of bacterial evolution. Phylogenomics and taxonogenomics are revealing the evolutionary processes that lead to genome diversification and speciation of the order *Lysobacterales*, comprising the genus *Xanthomonas* (17). *Xanthomonas* community analysis by comparative genomics of diverse lifestyles of bacteria is crucial for our understanding. The lifestyle and adaptability of a microbe to a new host or environment depends on its genomic content. In the genus *Xanthomonas*, there are widely known virulence-related and adaptation-related genes or gene clusters, such as type secretion systems and their effectors (15, 18–21), the *rpf* (regulation of pathogenicity factors) gene cluster (22, 23), iron uptake-, utilization-, and siderophore-related genes (24–26), global regulators and two-component systems (27–31), etc. The status of these genes or gene clusters among nonpathogenic strains of *Xanthomonas* will be of great significance. A genome-wide comparison of pathogenic and nonpathogenic isolates of *X. arboricola* reflects various genomic determinants, such as type 3 secretion systems (T3SS), chemoreceptors, adhesins, etc. (11). In this context, such an analysis of rice *Xanthomonas* community from our group has demonstrated the gene(s) or gene cluster(s) essential for virulence and adaptation to its host (15). Similarly, such a study on nonpathogenic *Xanthomonas* from diverse hosts of economically important citrus plants is lacking and is urgently needed.

In the present study, we have focused on the evolution of a citrus-associated *Xanthomonas* community known as a global devastating pathogen. Along with that, we also included genomes of nonpathogenic isolates from walnut and rice available from NCBI. Here, we sequenced four citrus-associated nonpathogenic *Xanthomonas* species complex (NPXc) strains and carried out phylogenomics, taxonogenomics, and comparative genomics studies. Primarily, we tried to address several important aspects, such as whether all these nonpathogenic counterparts from diverse hosts belong to a single species, irrespective of their host of isolation, or whether they belong to diverse species. What is their relation with two of the already known nonpathogenic species among those in the *Xanthomonas* genus? What is genomic content and flux of a gene(s)/gene cluster(s) during the course of evolution. The present study will be a genomic investigation towards understanding of widely unexplored nonpathogenic isolates and the evolution of successfully established pathogenic *Xanthomonas* species.

RESULTS

Whole-genome sequencing and assembly of NPXc isolates. The genomes of four NPXc isolates were sequenced using an Illumina MiSeq platform. The assemblies obtained had coverage ranging from 73X to 146X with N_{50} values of 94 kb to 213 kb. Genome assembly assessment according to CheckM revealed all genomes to be

approximately 99.9% complete with not more than 1.5% contamination. Genome size and GC content of all the strains are approximately 5 Mb and 68%, respectively. Annotation by NCBI resulted in 3,744 to 4,366 coding sequences with single copies each of 5S, 16S, and 23S in the rRNA operon. Genome statistics for the NPXc isolates are summarized in Table 1.

Phylogenomic and taxonogenomic diversity of nonpathogenic isolates. Phylogenomic analysis of all available nonpathogenic isolates (see Table S1 in the supplemental material) along with their pathogenic counterparts and other type or representative strains is shown in Fig. 1. The majority of the nonpathogenic strains, such as NPXc (LMG8989, LMG8992, and LMG9002) and NPXr (*X. sontii* PPL1, *X. sontii* PPL2, *X. sontii* PPL3, LMG12459, LMG12460, LMG12461, LMG12462, SHU166, SHU199, and SHU308) belong to clade II of the *Xanthomonas* genus. Interestingly, *X. sacchari* R1 was also with the NPXr strains in *X. sontii* and not in *X. sacchari*, whereas nonpathogenic isolates from walnut (NPXw; *X. arboricola* CFBP7634 and *X. arboricola* CFBP7651) and one of the NPXc (LMG8993) isolates belong to clade I, with *X. arboricola* as their closest phylogenomic relative. *X. maliensis* (NPXr) is more related to clade I but does not belong to any of the clades (Fig. 1).

To confirm the species status of these isolates, taxonogenomic analysis, including average nucleotide identity (OrthoANI) and digital DNA-DNA hybridization (dDDH), was performed. Here, NPXc, i.e., LMG8989, LMG8992, and LMG9002, had approximately 93% and 50% OrthoANI and dDDH, respectively, with *X. sontii* strains, and approximately 94% and 55%, respectively, with *X. sacchari* CFBP4641 (T). This suggests that none of the NPXc isolates belong to *X. sontii* or *X. sacchari*. Moreover, these three strains of NPXc had OrthoANI ranging from 92.7% to 93.7% and dDDH ranging from 48.9% to 52.7%. Therefore, these three strains represent three different potential novel species, which we refer to as the NPXc complex. The NPXc complex is phylogenomically related to *X. sontii*, which was reported earlier as a nonpathogenic species from rice (9). Furthermore, *X. sacchari* R1 has 96.6% OrthoANI and 69.7% dDDH with *X. sontii*, whereas these values are 94% and 54% with *X. sacchari*. This clearly depicts that *X. sacchari* R1 is a misclassified strain and that it belongs to *X. sontii*.

Interestingly, the fourth NPXc isolate, LMG8993, and walnut-associated NPXw strains (CFBP7634 and CFBP7651) had around 96% OrthoANI but 67% dDDH among themselves. Hence, according to multiple genome similarity parameters, these three nonpathogenic isolates form a potential taxonomic outlier of pathogenic *X. arboricola* (Fig. 2). While NPXc and NPXr isolates are in both clades I and II, the NPXw isolates are only in clade I.

***Xanthomonas*-associated gene cluster(s)/repertoire among bacteria with different lifestyles.** We further analyzed the status of a well-characterized *Xanthomonas*-associated gene cluster(s)/repertoire in these nonpathogenic isolates/species and their counterparts in the genus *Xanthomonas* (Fig. 3 and 4). Among the type secretion systems, T1SS, *xps* T2SS, T4SS, and *rpf* gene cluster were present in all the strains belonging to both clades (with exception of T4SS, which was absent from *X. populi*, *X. oryzae*, *X. bromi*, *X. axonopodis* of clade I, and *X. maliensis*). However, the remaining secretion systems were more prevalent in clade I than in clade II. For instance, *xcs* T2SS and T6SS were present in *X. maliensis* and most of the strains of clade I and absent from clade II (except for T6SS in *X. translucens*). Similarly, T3SS and its effectors were widely present in all strains of clade I except for *X. floridensis*, *X. melonis*, *X. pisi*, and nonpathogenic strains LMG8993 (NPXc) and CFBP7634 (NPXw). In case of the type III effectors, a diverse repertoire was observed for the strains isolated from diversified hosts. However, in clade II, T3SS and its effectors were present only in *X. theicola*, *X. translucens*, and *X. hyacinthi*. Furthermore, adhesins were also analyzed for the strains under study. Here, only *pilQ* was present in all the strains under study, and the remaining three adhesins, i.e., *xadA*, *xadB*, and *yapH*, were present in clade I, but their presence varied from species to species. Cell wall-degrading enzymes were present in all the isolates but with a diversified repertoire.

TABLE 1 Genome assembly statistics and metadata of NPXc strains in the present study

Strain	Host	Geographical location	Yr	Genome size (Mb)	Coverage (x)	No. of contigs	GC content (%)	<i>N</i> ₅₀ (Kb)	No. of coding sequences	rRNA (5S, 16S, 23S)	tRNA	Completeness (%)	Contamination (%)	NCBI accession no.
<i>Xanthomonas</i> sp. LMG8989	Citrus sp. (orange)	USA	1989	4.8	125	72	69.26	213	3,744	1,1,1	36	100/0.34		QUX1000000000
<i>Xanthomonas</i> sp. LMG8992	Citrus sp. (orange)	USA	1989	4.5	146	253	68.53	163	4,366	1,1,1	51	99.9/1.5		QUX1000000000
<i>Xanthomonas</i> sp. LMG8993	Citrus sp. (orange)	USA	1989	4.9	93	78	65.82	139	4,141	1,1,1	51	99.9/0.5		QVIF0000000000
<i>Xanthomonas</i> sp. LMG9002	Citrus sp. (orange)	USA	1989	4.9	73	94	68.80	94	3,987	1,1,1	51	100/0.34		QJXK0000000000

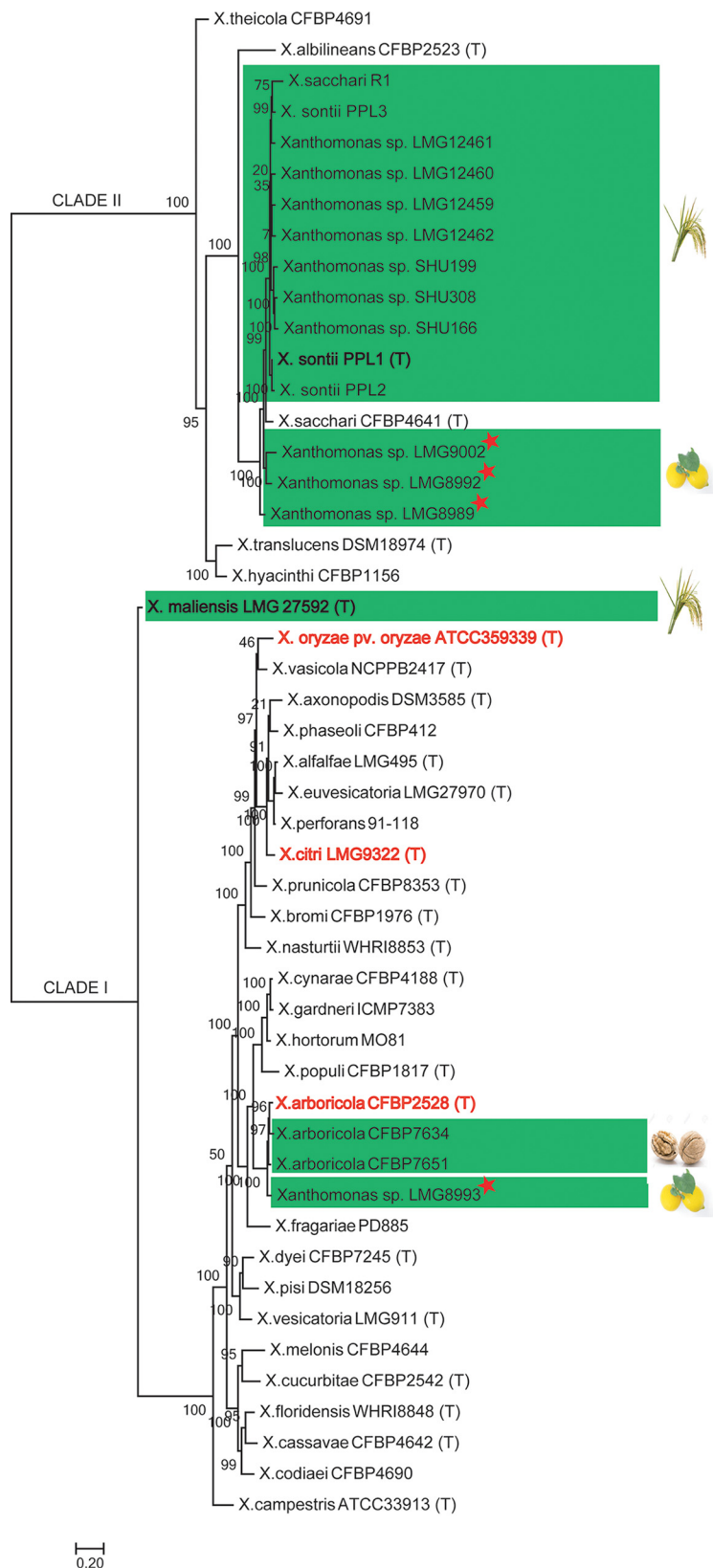


FIG 1 Phylogenomic tree of nonpathogenic and pathogenic *Xanthomonas* strains. Here, all the non-pathogenic strains are highlighted in green, and their respective hosts are shown. Pathogenic species *X. citri*, *X. oryzae*, and *X. arboricola* are highlighted in red. The type strains are designated by (T), and strains sequenced in the study are indicated by stars.

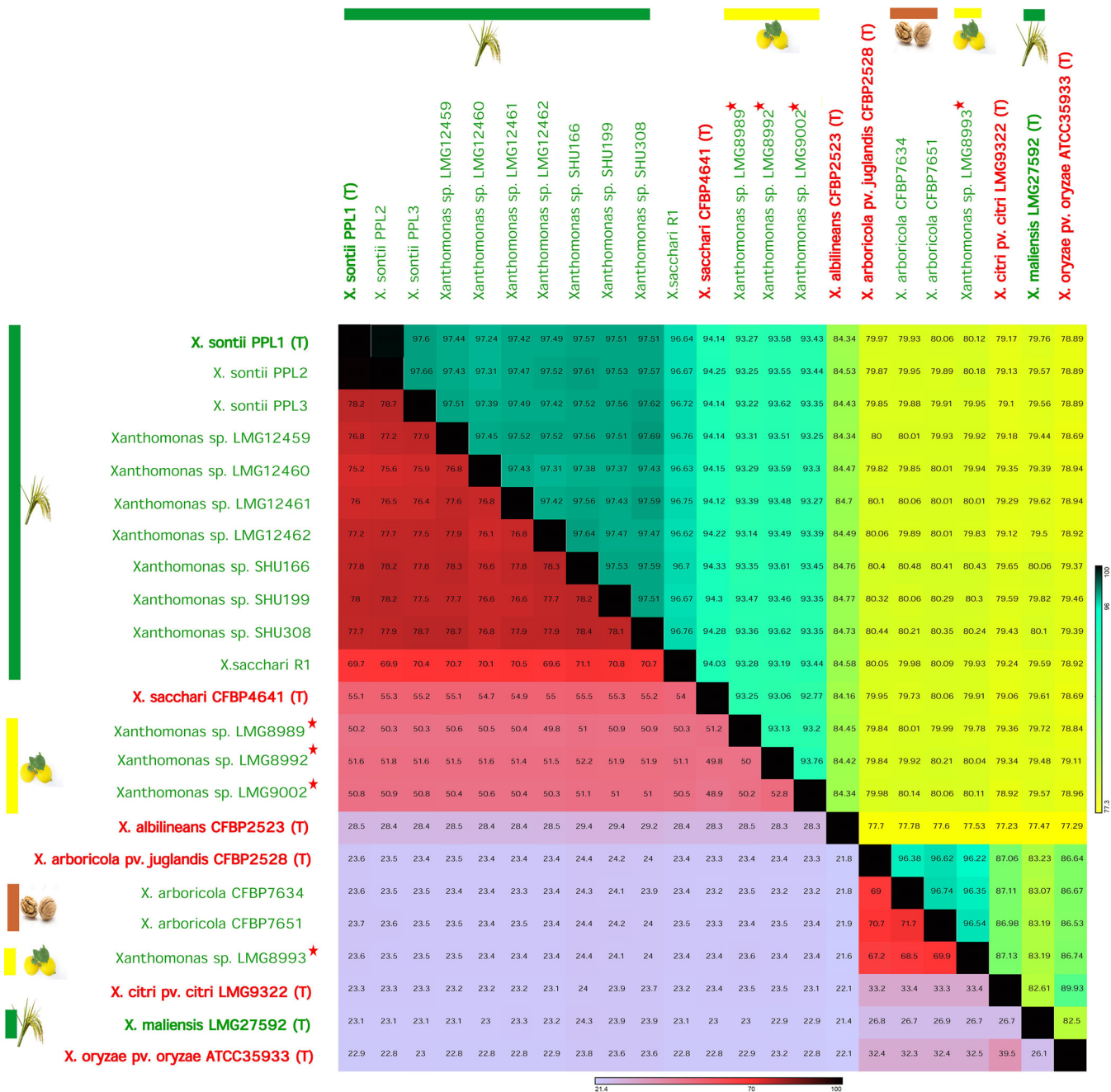


FIG 2 Taxonogenomic analysis of all nonpathogenic strains. Nonpathogenic and pathogenic strains are in red and green color, respectively. Nonpathogenic strains from rice, lemon, and walnut are represented by green, yellow, and brown bars, respectively. The type strains are designated by (T), and strains sequenced in the study are indicated by stars. Here, OrthoANI and dDDH values are shown in the upper and lower diagonal areas (separated by a line of black boxes), respectively.

Overall, the pathogenicity cluster status among nonpathogenic isolates, T1SS, *xps* T2SS, T4SS, and the *rpf* cluster were present in nonpathogenic isolates of both clades. T3SS and its effectors and T6SS were absent from all the nonpathogenic strains except for one of the NPXw (CFBP7651) isolates having T3SS. T2SS *xcs* was present in nonpathogenic isolates from clade I only. Cell wall-degrading enzymes were present in both clade I and clade II nonpathogenic isolates. Here, clade II nonpathogenic isolates had similar repertoires, whereas clade I nonpathogenic isolates had diversified repertoires. Similarly, among the adhesin genes, clade I isolates were rich compared to clade II isolates.

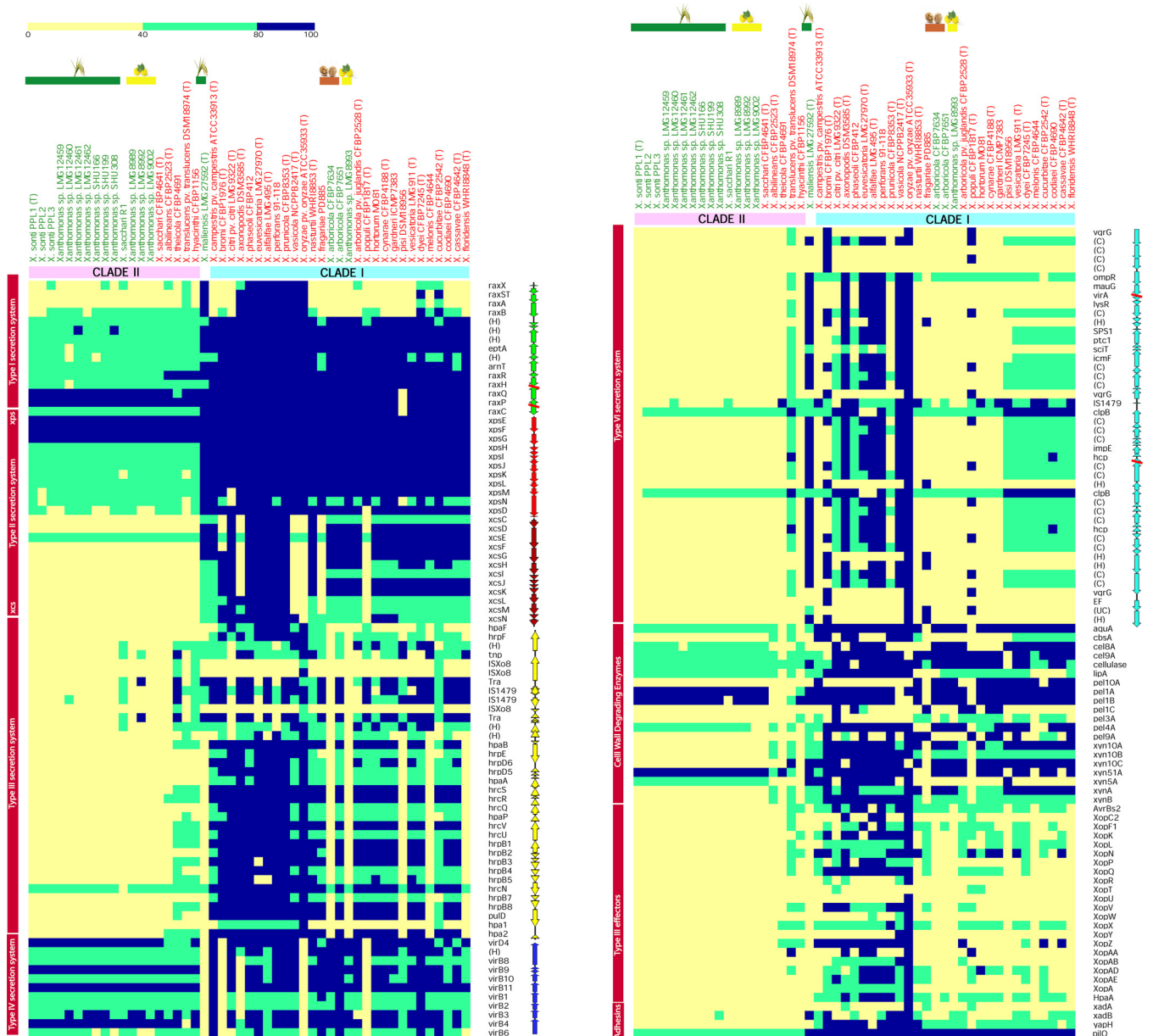


FIG 3 Heat maps showing status of pathogenicity-associated gene clusters (names indicated in vertical boxes, and vertical arrows indicate gene organization in the cluster). Here, contig breaks are represented by red lines for nonpathogenic strains (represented by green color) in the study and other species of the genus *Xanthomonas* (represented by red color). The intensity of the color indicates the level of identity with the query as depicted by the scale. (H), hypothetical gene; (C), conserved gene; (UC), uncharacterized gene; (T), transposase.

Apart from studying if the status of gene clusters was virulence related, we further looked at the status of genes related to adaptation (Fig. 4). Accordingly, iron uptake and utilization, *gum* cluster, xanthomonadin pigment production cluster, and two-component system genes were present in all of the strains analyzed with few exceptions. For instance, *hrpG* and *hrpX* were absent from *X. pisi* and most of the strains of clade II (except for *X. theicola*, *X. translucens*, and *X. hyacinthi*). In the case of the *gum* cluster, it was missing from *X. albilineans* and *X. theicola* but present in all the strains of both clades and lifestyles.

Gene content among nonpathogenic and pathogenic species of *Xanthomonas*.

Genome-wide studies of these ecological and phylogenomic relatives with diverse lifestyles enabled us to understand the evolution of highly successful lineages. To evaluate gene content differences among nonpathogenic strains (NPXc, NPXr, and

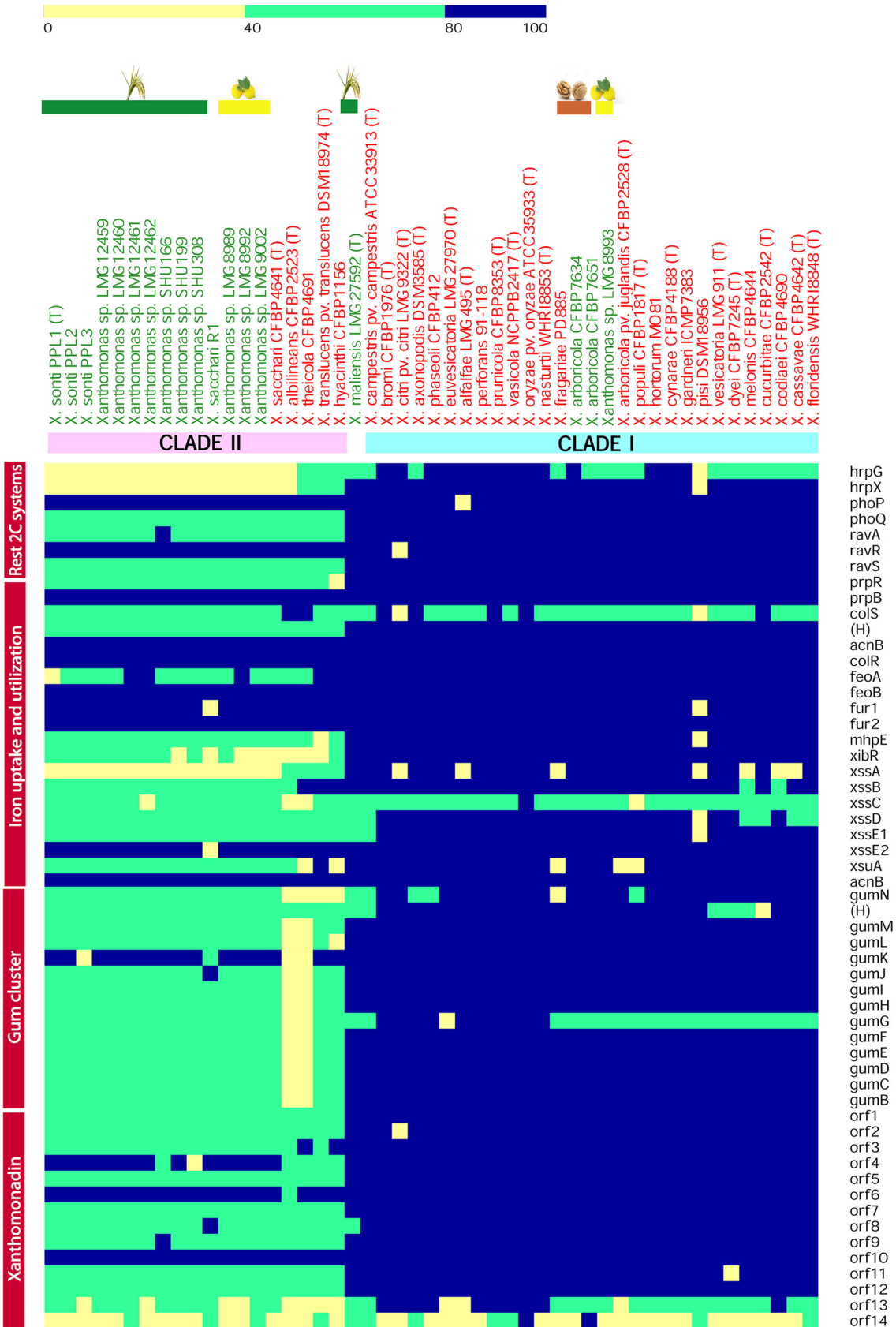


FIG 4 Heat map showing status of plant adaptation-related clusters (names indicated in vertical boxes) for nonpathogenic strains (represented by green color) in the study and other species of the genus *Xanthomonas* (represented by red color). The intensity of the color indicates the level of identity with the query as depicted by the scale. (H), hypothetical gene.

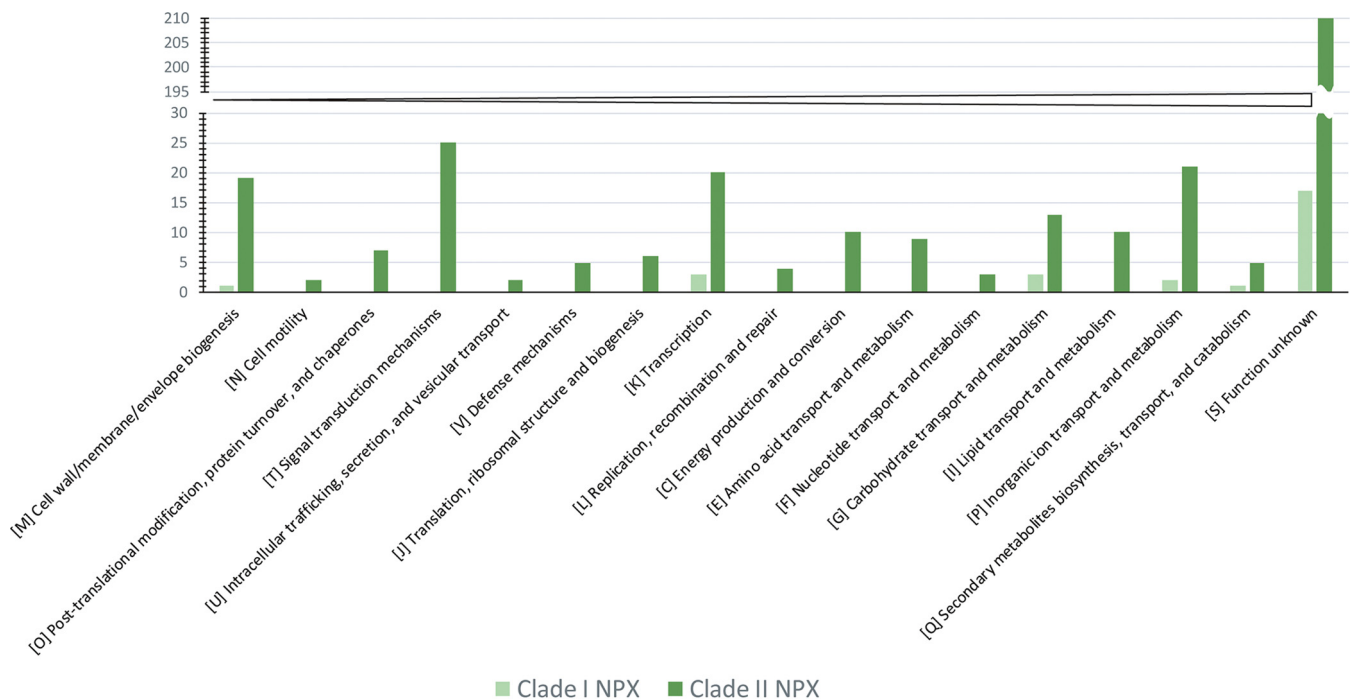


FIG 5 Distribution of COG-based functional categories of nonpathogenic isolates of clade I and clade II.

NPXw) and their pathogenic counterparts, i.e., those isolated from the host (*X. citri*, *X. oryzae*, and *X. arboricola*), we performed pan genome analysis. Here, 1,322 and 12,094 genes constituted core and pan genome, respectively, for all the strains from distinct lifestyles. However, only 1 and 12 genes were found to be present in all nonpathogenic and pathogenic strains, respectively. This low number of core genes can be explained by the fact that even though strains belong to same lifestyle, phylogenomically, they are diverse, i.e., they belong to distinct clades and hosts.

Hence, to get lifestyle-specific genes, we analyzed gene content common to all the nonpathogenic strains belonging to a particular clade. Interestingly, for the clade II strains, i.e., LMG8989, LMG8992, LMG9002, *X. sontii* PPL1, *X. sontii* PPL2, *X. sontii* PPL3, LMG12459, LMG12460, LMG12461, LMG12462, SHU166, SHU199, SHU308, and *X. sacchari* R1, there were 377 core genes (see Table S2). Similarly, there were 27 core genes for nonpathogenic strains from clade I, i.e., LMG8993, *X. arboricola* CFBP7634 and *X. arboricola* CFBP7651 (see Table S3). This clearly shows nonpathogenic strains of clade II are diversifying rapidly compared to clade I nonpathogenic strains. All these genes were further subjected to Clusters of Orthologous Groups (COG) functional classification. Of 25 COG classes, clade II nonpathogenic core genes were classified in 17 COG classes, and clade I nonpathogenic core genes were only in 6 COG classes (Fig. 5). Among clade II nonpathogenic isolates, genes were classified under “signal transduction mechanisms,” “transcription,” “cell wall/membrane/envelope biogenesis,” etc., whereas clade I nonpathogenic isolates belonged to “transcription,” “carbohydrate and inorganic ion transport and metabolism,” and “secondary metabolites biosynthesis, transport, and catabolism.”

To look for lifestyle-specific genes among citrus associated strains of *Xanthomonas*, we carried out pan genome analysis, including 4 NPXc (LMG8989, LMG8992, LMG8993, and LMG9002) and 4 *X. citri* pv. *citri* strains (LMG9322, XACB100, XACJJ10, and XALG98) (8, 32). Here, 2,344 and 7,736 genes constituted core genome and pan genome, respectively. Furthermore, 585 and 78 genes, respectively, were present in all pathogenic and nonpathogenic strains (see Tables S4 and S5). These genes were also then COG classified, and their distribution is shown in Fig. 6. Functional classification suggested that pathogenic isolates had a rich unique gene repertoire compared to that

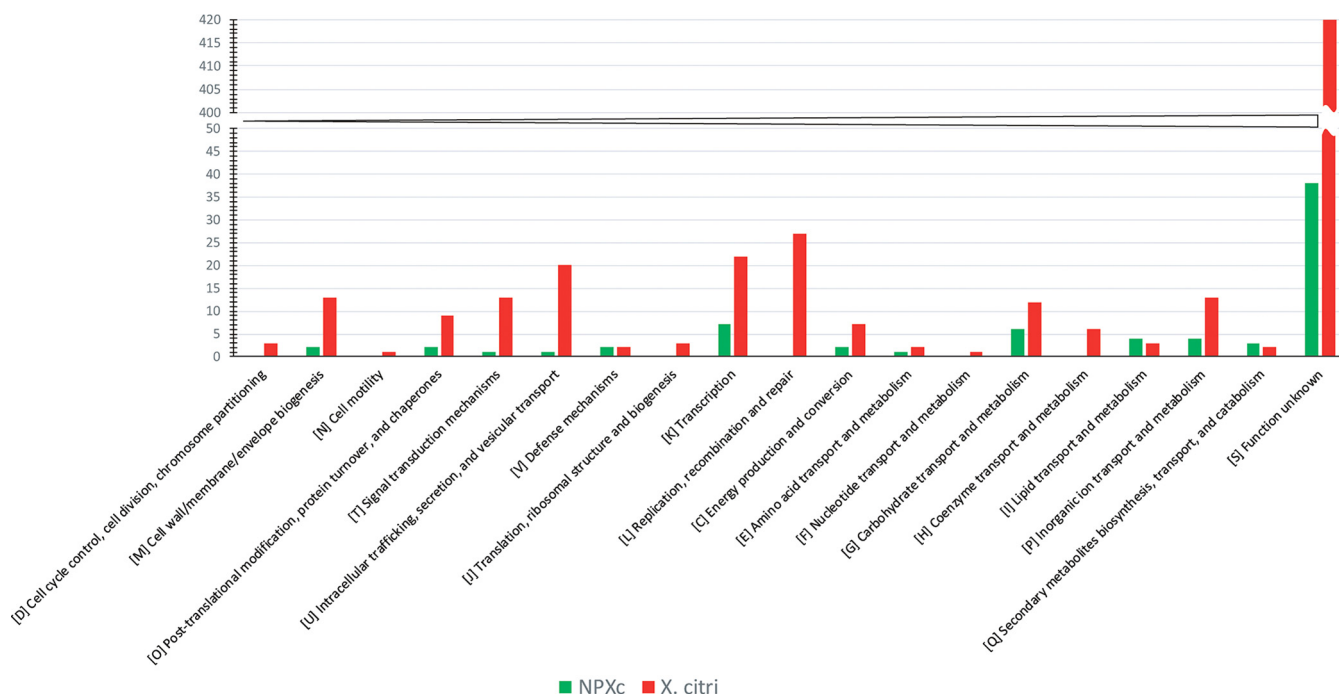


FIG 6 Distribution of COG-based functional categories of citrus-associated nonpathogenic and pathogenic isolates.

of nonpathogenic isolates. Pathogenic strains had unique genes related to “replication, recombination, and repair,” “transcription,” “intracellular trafficking, secretion, and vesicular transport,” and “cell wall/membrane/envelope biogenesis.” However, the genes unique to NPXc isolates belonged to “transcription” and “carbohydrate, lipid, and inorganic ion transport and metabolism.”

DISCUSSION

Xanthomonas pathogens of citrus and rice belong to *X. citri* and *X. oryzae*, respectively (7, 8, 33). In this context, it is interesting to find that nonpathogenic isolates from citrus are highly diverse, with each isolate belonging to a potentially novel species and altogether forming a species complex, the NPXc complex. This suggests the existence of a highly diverse and coevolving community of nonpathogenic *Xanthomonas* on citrus plants. This is unlike nonpathogenic isolates from rice that primarily belong to one species, *X. sontii*, suggesting distinct selection pressure and diversification in nonpathogenic isolates depending on their hosts. Owing to the diversity of NPXc, there is a need to expand such studies by isolating and investigating large numbers of nonpathogenic strains from citrus plants. Incidentally, unlike rice, citrus plants themselves belong to several highly diverse species. Hence, such studies are required to be expanded to all citrus species and not just one species to understand the origin, evolution, and adaptation of a nonpathogenic population on citrus hosts.

Most of the nonpathogenic isolates in both rice and citrus belong to clade II. At the same time, a minority of the strains in both rice (*X. maliensis*), citrus (LMG8993), and even *Xanthomonas*-associated with walnut (NPXw; CFBP7634 and CFBP7651) were found to be more related to pathogenic isolates. This suggests the existence of a minor group that may be ancestral to the pathogenic counterparts and ongoing dynamics toward diverse lifestyles. Also, pathogenic *X. citri* and *X. arboricola* species are known to cause similar types of symptoms in diverse hosts (34). One of the nonpathogenic isolates from citrus (LMG8993) is related to pathogenic and nonpathogenic isolates of walnut, indicating the basic functions required to adapt to fruit plants. Moreover, a previous study comparing pathogenic and nonpathogenic strains from walnut found that they belong to the same species (*X. arboricola*). Interestingly, whole-genome-based

comparisons have shown that NPXw isolate (CFBP7651, also included in the present study) had noncanonical T3SS (11). Similarly, among NPXr isolates as well, *X. maliensis* was found to be more related to pathogenic isolates and also had some of the key pathogenicity determinants which were otherwise absent from the majority of nonpathogenic isolates (15).

Phylogenetically pathogenic species of rice (*X. oryzae*) and citrus (*X. citri* pv. *citri*) are distantly related (35), and unlike nonpathogenic isolates, both of the pathogenic species fall in a single clade (clade I). This may be to do with the pathogenic lifestyle on respective hosts and the requirement of distinct set of virulence genes. Hence, a nonpathogenic lifestyle may have been the default or original lifestyle. During the course of evolution and with the advent of agriculture or anthropological factors, this might have led to the emergence of variant pathogenic species or strains. Such emergence might have been from a nonpathogenic *Xanthomonas* strain which is not successful, such as that seen in *X. maliensis* that is also intermediate to both nonpathogenic and pathogenic strains in relationship and gene content. Even though the nonpathogenic strains are from two diverse hosts (monocot versus dicot), in the phylogenomic tree, the majority of the isolates were closely related owing to their nonpathogenic lifestyle. This highlights that the origin of both communities is from a common ancestor in the recent past and that lifestyle is also important along with host specialization. There is parallel evolution and convergence of function related to this lifestyle. This also means there is requirement for a core set of functions to be a successful nonpathogenic *Xanthomonas*.

Earlier molecular genetic studies have identified many genes/gene clusters that are required for pathogenicity in *Xanthomonas*. Our genomic investigation studies also allowed us to demarcate genes/gene clusters that might have played important roles in the evolution of a pathogenic lifestyle, such as T3SS, *rpf*-based gene regulation, *gum* cluster for exopolysaccharide production, xanthomonadin pigment production, etc., and are core to both lifestyles of *Xanthomonas*. These findings suggest that the functions required for plant adaptation, primarily in nonpathogenic strains, is at the level of the epiphyte for their establishment. The pathogenic lifestyle is more aggressive and requires functions at all stages until plant death. This observation is also supported by pan genome analysis that indicates the need to acquire a large number of unique genes with unknown and diverse functions for a successful pathogen. The dominance of genes related to replication, repair, and recombination reiterates the importance of horizontal gene transfer in the acquisition of a large number of genes in *Xanthomonas* strains that are pathogenic to citrus. The dominance of transcription category genes also suggests that apart from diversification and acquisition of novel genes through horizontal gene transfer, regulation and on-going variation in the core genes is also important.

Hence, the present study reflects that phylogenomically and gene content wise, nonpathogenic strains have two waves of evolution. The first one is followed by a majority of the strains, forming a species complex with the absence of virulence-related clusters but basic adaptation functions. The second one is more diversified and related to pathogenic strains, although they are represented by a minority of the strains. Resource and knowledge pertaining to the population of nonpathogenic *Xanthomonas* from diverse hosts will shed light on the epidemiology and evolution of pathogenic isolates. Genomics has led to a new era in *Xanthomonas* research by enabling researchers to carry out large-scale comprehensive phylogenomic profiling to understand evolutionary differences in both minor and major groups of nonpathogenic isolates across diverse hosts, such as grasses, shrubs, trees, etc.

MATERIALS AND METHODS

Strain collection, genome sequencing, assembly and annotation. All four citrus-associated strains were procured from BCCM-LMG culture collection and were grown as per the instructions. Genomic DNA was extracted by using a ZR Fungal/Bacterial DNA MiniPrep kit (Zymo Research). DNA quality checking was carried out using a NanoDrop 1000 instrument (Thermo Fisher Scientific) and agarose gel electrophoresis. Quantitation of DNA was performed using a Qubit 2.0 fluorometer (Life Technologies). Illumina

paired-end sequencing libraries (read length, 2×250 bp) of genomic DNA were prepared using Nextera XT sample preparation kits (Illumina, Inc., San Diego, CA, USA) with dual indexing adapters. Illumina sequencing libraries were sequenced in-house using an Illumina MiSeq platform (Illumina, Inc., San Diego, CA, USA) and company-supplied paired-end sequencing kits. Adapter trimming was performed automatically by MiSeq control software (MCS), and additional adapter contamination identified by the NCBI server was removed by manual trimming. Raw reads were assembled *de novo* using CLC Genomics Workbench v7.5 (CLC bio, Aarhus, Denmark) with default settings. The quality of the genomes was accessed using CheckM v1.0.13 (36). Annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

Phylogenomics and taxonogenomics. PhyloPhlAn v0.99 was used to construct the phylogenomic tree based on more than 400 conserved genes from whole-genome proteome data (37). USEARCH v5.2.32 (38), MUSCLE v3.8.31 (39), and FastTree v2.1 (40) were used for ortholog searching, multiple-sequence alignment, and phylogenomic construction, respectively. Taxonogenomic analysis of the strains was performed using the OrthoANlu (41) standalone tool, which employs USEARCH v5.2.32 over BLAST for OrthoANI calculation, and dDDH values were calculated using a genome-to-genome distance calculator (<http://ggdc.dsmz.de/distcalc2.php>).

Pan genome analysis. Pan genome analysis of the strains was performed using Roary v3.12.0 (42). Here, gff files generated by Prokka v1.13.3 (43) were used as input for Roary. Since we were analyzing genomes from different species, the identity cutoff used was 60%. Functional annotation of the core genes identified was performed using EggNOG mapper v1 (44).

Virulence related gene(s) and gene cluster(s) analysis. Protein sequences of genes or gene clusters (15) were used as the query, and tBLASTn searches were performed on genomes. Genomic coordinates corresponding to the gene clusters were identified from tBLASTn, and corresponding gene clusters from the genomes were extracted using SAMtools v1.9 faidx module (45). Here, tBLASTn searches were performed using standalone BLAST+ v2.9.0 (46), and cutoffs used for identity and coverage were 40% and 50%, respectively. Heat maps for the blast searches were generated using GENE-E v3.03215 (<https://software.broadinstitute.org/GENE-E/>).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TABLE S1, XLSX file, 0.1 MB.

TABLE S2, XLSX file, 0.1 MB.

TABLE S3, XLSX file, 0.1 MB.

TABLE S4, XLSX file, 0.1 MB.

TABLE S5, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This work is supported by a project entitled Megagenomics and metagenomics insights into adaptation and evolution of fruit microbiome (GAP0187) and Megagenomic insights into coevolution of rice and its Microbiome (MICRA/MLP0020) to P.B.P.

K.B. and S.K. performed genome sequencing. S.K. performed phylogenomic and comparative analyses. K.B. performed taxonogenomic and virulence gene cluster analyses. K.B. drafted the manuscript with input from S.K. and P.B.P. P.B.P. conceived the study and participated in its design with K.B. All authors have read and approved the manuscript.

We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

- Leysns F, De Cleene M, Swings J-G, De Ley J. 1984. The host range of the genus *Xanthomonas*. Bot Rev 50:308–356. <https://doi.org/10.1007/BF02862635>.
- Hayward A. 1993. The hosts of *Xanthomonas*, p 1–119. In Swings JG, Civerolo EL (eds), *Xanthomonas*. Springer, Dordrecht, Netherlands.
- Chan JW, Goodwin PH. 1999. The molecular genetics of virulence of *Xanthomonas campestris*. Biotechnol Adv 17:489–508. [https://doi.org/10.1016/S0734-9750\(99\)00025-7](https://doi.org/10.1016/S0734-9750(99)00025-7).
- Gottwald TR, Hughes G, Graham JH, Sun X, Riley T. 2001. The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. Phytopathology 91:30–34. <https://doi.org/10.1094/PHYTO.2001.91.1.30>.
- Gottwald TR, Sun X, Riley T, Graham JH, Ferrandino F, Taylor EL. 2002. Geo-referenced spatiotemporal analysis of the urban citrus canker epidemic in Florida. Phytopathology 92:361–377. <https://doi.org/10.1094/PHYTO.2002.92.4.361>.
- Graham JH, Gottwald TR, Cubero J, Achor DS. 2004. *Xanthomonas axonopodis* pv. citri: factors affecting successful eradication of citrus canker. Mol Plant Pathol 5:1–15. <https://doi.org/10.1046/j.1364-3703.2004.00197.x>.
- Brunings AM, Gabriel DW. 2003. *Xanthomonas citri*: breaking the surface. Mol Plant Pathol 4:141–157. <https://doi.org/10.1046/j.1364-3703.2003.00163.x>.
- Bansal K, Midha S, Kumar S, Patil PB. 2017. Ecological and evolutionary insights into *Xanthomonas citri* pathovar diversity. Appl Environ Microbiol 83:e02993-16. <https://doi.org/10.1128/AEM.02993-16>.
- Bansal K, Kaur A, Midha S, Kumar S, Korpole S, Patil PB. 2019. *Xanthomonas sontii* sp. nov., a non-pathogenic bacterium isolated from healthy basmati rice (*Oryza sativa*) seeds from India. bioRxiv <https://www.biorxiv.org/content/10.1101/738047v1>.
- Triplet LR, Verdier V, Campillo T, Van Malderghem C, Cleenwerck I, Maes M, Deblais L, Corral R, Koita O, Cottyn B, Leach JE. 2015. Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. Antonie Van Leeuwenhoek 107:869–881. <https://doi.org/10.1007/s10482-015-0379-5>.

11. Cesbron S, Briand M, Essakhi S, Gironde S, Boureau T, Manceau C, Saux F-L, Jacques M-A. 2015. Comparative genomics of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front Plant Sci* 6:1126. <https://doi.org/10.3389/fpls.2015.01126>.
12. Essakhi S, Cesbron S, Fischer-Le Saux M, Bonneau S, Jacques M-A, Manceau C. 2015. Phylogenetic and variable-number tandem-repeat analyses identify nonpathogenic *Xanthomonas arboricola* lineages lacking the canonical type III secretion system. *Appl Environ Microbiol* 81:5395–5410. <https://doi.org/10.1128/AEM.00835-15>.
13. Fischer-Le Saux M, Bonneau S, Essakhi S, Manceau C, Jacques M-A. 2015. Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. *Appl Environ Microbiol* 81:4651–4668. <https://doi.org/10.1128/AEM.00050-15>.
14. Fang Y, Lin H, Wu L, Ren D, Ye W, Dong G, Zhu L, Guo L. 2015. Genome sequence of *Xanthomonas sacchari* R1, a biocontrol bacterium isolated from the rice seed. *J Biotechnol* 206:77–78. <https://doi.org/10.1016/j.jbiotec.2015.04.014>.
15. Bansal K, Midha S, Kumar S, Kaur A, Sonti RV, Patil PB. 2019. Ecological and evolutionary insights into pathogenic and non-pathogenic rice associated *Xanthomonas*. *bioRxiv* <https://www.biorxiv.org/content/10.1101/453373v2>.
16. Vauterin L, Yang P, Alvarez A, Takikawa Y, Roth DA, Vidaver AK, Stall RE, Kersters K, Swings J. 1996. Identification of non-pathogenic *Xanthomonas* strains associated with plants. *Systematic and Applied Microbiology* 19:96–105. [https://doi.org/10.1016/S0723-2020\(96\)80016-6](https://doi.org/10.1016/S0723-2020(96)80016-6).
17. Kumar S, Bansal K, Patil PP, Patil PB. 2019. Phylogenomics insights into order and families of *Lysobacterales*. *Access Microbiol* 1:e000015. <https://doi.org/10.1099/acmi.0.000015>.
18. Jha G, Rajeshwari R, Sonti RV. 2007. Functional interplay between two *Xanthomonas oryzae* pv. *oryzae* secretion systems in modulating virulence on rice. *Mol Plant Microbe Interact* 20:31–40. <https://doi.org/10.1094/MPMI-20-0031>.
19. Lu H, Patil P, Van Sluys MA, White FF, Ryan RP, Dow JM, Rabinowicz P, Salzberg SL, Leach JE, Sonti R, Brendel V, Bogdanove AJ. 2008. Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PLoS One* 3:e3828. <https://doi.org/10.1371/journal.pone.0003828>.
20. Büttner D, Bonas U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol Rev* 34:107–133. <https://doi.org/10.1111/j.1574-6976.2009.00192.x>.
21. Bernal P, Llamas MA, Filloux A. 2018. Type VI secretion systems in plant-associated bacteria. *Environ Microbiol* 20:1–15. <https://doi.org/10.1111/1462-2920.13956>.
22. Dow JM, Feng J-X, Barber CE, Tang J-L, Daniels MJ. 2000. Novel genes involved in the regulation of pathogenicity factor production within the *rpf* gene cluster of *Xanthomonas campestris*. *Microbiology* 146:885–891. <https://doi.org/10.1099/00221287-146-4-885>.
23. Chatterjee S, Sonti RV. 2002. *rpfF* mutants of *Xanthomonas oryzae* pv. *oryzae* are deficient for virulence and growth under low iron conditions. *Mol Plant Microbe Interact* 15:463–471. <https://doi.org/10.1094/MPMI.2002.15.5.463>.
24. Etchegaray A, Silva-Stenico ME, Moon DH, Tsai SM. 2004. *In silico* analysis of nonribosomal peptide synthetases of *Xanthomonas axonopodis* pv. *citri*: identification of putative siderophore and lipopeptide biosynthetic genes. *Microbiol Res* 159:425–437. <https://doi.org/10.1016/j.micres.2004.09.009>.
25. Pandey A, Sonti RV. 2010. Role of the FeoB protein and siderophore in promoting virulence of *Xanthomonas oryzae* pv. *oryzae* on rice. *J Bacteriol* 192:3187–3203. <https://doi.org/10.1128/JB.01558-09>.
26. Subramoni S, Pandey A, Vishnu Priya M, Patel HK, Sonti RV. 2012. The ColRS system of *Xanthomonas oryzae* pv. *oryzae* is required for virulence and growth in iron-limiting conditions. *Mol Plant Pathol* 13:690–703. <https://doi.org/10.1111/j.1364-3703.2011.00777.x>.
27. Wengelnik K, Rossier O, Bonas U. 1999. Mutations in the regulatory gene *hrpG* of *Xanthomonas campestris* pv. *vesicatoria* result in constitutive expression of all *hrp* genes. *J Bacteriol* 181:6828–6831. <https://doi.org/10.1128/JB.181.21.6828-6831.1999>.
28. Wengelnik K, Bonas U. 1996. *HrpXv*, an AraC-type regulator, activates expression of five of the six loci in the *hrp* cluster of *Xanthomonas campestris* pv. *vesicatoria*. *J Bacteriol* 178:3462–3469. <https://doi.org/10.1128/jb.178.12.3462-3469.1996>.
29. Tsuge S, Nakayama T, Terashima S, Ochiai H, Furutani A, Oku T, Tsuno K, Kubo Y, Kaku H. 2006. Gene involved in transcriptional activation of the *hrp* regulatory gene *hrpG* in *Xanthomonas oryzae* pv. *oryzae*. *J Bacteriol* 188:4158–4162. <https://doi.org/10.1128/JB.00006-06>.
30. Lee S-W, Jeong K-S, Han S-W, Lee S-E, Phee B-K, Hahn T-R, Ronald P. 2008. The *Xanthomonas oryzae* pv. *oryzae* PhoPQ two-component system is required for AvrXA21 activity, *hrpG* expression, and virulence. *J Bacteriol* 190:2183–2197. <https://doi.org/10.1128/JB.01406-07>.
31. Qian W, Han ZJ, Tao J, He C. 2008. Genome-scale mutagenesis and phenotypic characterization of two-component signal transduction systems in *Xanthomonas campestris* pv. *campestris* ATCC 33913. *Mol Plant Microbe Interact* 21:1128–1138. <https://doi.org/10.1094/MPMI-21-8-1128>.
32. Gordon JL, Lefevre P, Escalon A, Barbe V, Cruveiller S, Gagnevin L, Pruvost O. 2015. Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *BMC Genomics* 16:1098. <https://doi.org/10.1186/s12864-015-2310-x>.
33. Midha S, Bansal K, Kumar S, Girija AM, Mishra D, Brahma K, Laha GS, Sundaram RM, Sonti RV, Patil PB. 2017. Population genomic insights into variation and evolution of *Xanthomonas oryzae* pv. *oryzae*. *Sci Rep* 7:40694. <https://doi.org/10.1038/srep40694>.
34. Garita-Cambronero J, Sena-Vélez M, Ferragud E, Sabuquillo P, Redondo C, Cubero J. 2019. *Xanthomonas citri* subsp. *citri* and *Xanthomonas arboricola* pv. *pruni*: comparative analysis of two pathogens producing similar symptoms in different host plants. *PLoS One* 14:e0219797. <https://doi.org/10.1371/journal.pone.0219797>.
35. Midha S, Patil PB. 2014. Genomic flux in *Xanthomonas* group of plant pathogenic bacteria, p 131–153. In Katsy EI (ed), *Plasticity in plant-growth-promoting and phytopathogenic bacteria*. Springer, New York, NY.
36. 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* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
37. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 4:2304. <https://doi.org/10.1038/ncomms3304>.
38. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
39. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
40. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650. <https://doi.org/10.1093/molbev/msp077>.
41. Yoon S-H, Ha S-m, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110:1281–1286. <https://doi.org/10.1007/s10482-017-0844-4>.
42. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>.
43. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
44. Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res* 36:D250–D254. <https://doi.org/10.1093/nar/gkm796>.
45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
46. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.