




Genetic and Functional Diversity Help Explain Pathogenic, Weakly Pathogenic, and Commensal Lifestyles in the Genus *Xanthomonas*

Michelle M. Pena^{1,8,†}, Rishi Bhandari^{1,†}, Robert M. Bowers², Kylie Weis¹, Eric Newberry¹, Naama Wagner ³, Tal Pupko ³, Jeffrey B. Jones⁴, Tanja Woyke^{2,5}, Boris A. Vinatzer ⁶, Marie-Agnès Jacques⁷, and Neha Potnis^{1,*}

¹Department of Entomology and Plant Pathology, Auburn University, Auburn, AL, USA

²DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

³The Shmunis School of Biomedicine and Cancer Research, Tel-Aviv University, Tel Aviv, Israel

⁴Department of Plant Pathology, University of Florida, Gainesville, FL, USA

⁵Life and Environmental Sciences, University of California Merced, Merced, CA, USA

⁶School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA

⁷Institut Agro, INRAE, IRHS, SFR QUASAV, University of Angers, Angers F-49000, France

⁸Present address: Department of Plant Pathology, University of Georgia, Tifton, GA, USA

[†]Michelle M. Pena and Rishi Bhandari contributed equally to this manuscript.

*Corresponding author: E-mail: nzp0024@auburn.edu.

Accepted: March 29, 2024

Abstract

The genus *Xanthomonas* has been primarily studied for pathogenic interactions with plants. However, besides host and tissue-specific pathogenic strains, this genus also comprises nonpathogenic strains isolated from a broad range of hosts, sometimes in association with pathogenic strains, and other environments, including rainwater. Based on their incapacity or limited capacity to cause symptoms on the host of isolation, nonpathogenic xanthomonads can be further characterized as commensal and weakly pathogenic. This study aimed to understand the diversity and evolution of nonpathogenic xanthomonads compared to their pathogenic counterparts based on their cooccurrence and phylogenetic relationship and to identify genomic traits that form the basis of a life history framework that groups xanthomonads by ecological strategies. We sequenced genomes of 83 strains spanning the genus phylogeny and identified eight novel species, indicating unexplored diversity. While some nonpathogenic species have experienced a recent loss of a type III secretion system, specifically the *hrp2* cluster, we observed an apparent lack of association of the *hrp2* cluster with lifestyles of diverse species. We performed association analysis on a large data set of 337 *Xanthomonas* strains to explain how xanthomonads may have established association with the plants across the continuum of lifestyles from commensals to weak pathogens to pathogens. Presence of distinct transcriptional regulators, distinct nutrient utilization and assimilation genes, transcriptional regulators, and chemotaxis genes may explain lifestyle-specific adaptations of xanthomonads.

Key words: nonpathogenic xanthomonads, commensal, association analysis, *hrp2* cluster, cell wall-degrading enzymes, gene flow.

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Significance

Numerous ecological strategies exist in the host-associated microbial genera that span the spectrum from pathogenic to weakly pathogenic to nonpathogenic species. Compared to pathogenic species, nonpathogenic relatives have received less attention, specifically their contribution as members of the microbiota toward microbiota-mediated immunity, their interactions with pathogens, including gene flow, and their evolution alongside pathogenic relatives. In this study, we addressed these questions on life history strategies using a diverse collection of nonpathogenic and pathogenic members of the genus *Xanthomonas*. We find that presence or absence of a type III secretion system is not the sole determinant of these lifestyles, but distinct repertoires of cell wall-degrading enzymes and various stress tolerance traits can explain the differences among pathogens and nonpathogens. Based on gene flow and gain/loss patterns of important pathogenicity traits, we find evidence for regressive evolution of nonpathogenic strains from pathogenic relatives. These findings are broadly applicable to plant-associated bacteria, for which most research has focused on pathogenic bacteria and contribution of nonpathogens toward plant–pathogen–microbiota interactions has been largely ignored.

Introduction

The genus *Xanthomonas*, traditionally considered to group plant pathogenic bacteria, encompasses bacterial strains that, although they maintain close association with plants, do not cause apparent disease symptoms in their host of isolation (Vauterin et al. 1996; Essakhi et al. 2015; Merda et al. 2016, 2017; Garita-Cambronero et al. 2017; Martins et al. 2020; Bansal et al. 2021). Nonpathogenic xanthomonads have a varied lifestyle with the ability to colonize the plant hosts and survive in various environments outside the plants, such as rain and aerosols (Vauterin et al. 1996; Mechan Llonetop et al. 2021). Although referred to as nonpathogenic in the context of their phenotype based on artificial inoculation on the host of isolation, it cannot be ruled out that these strains may cause disease in other hosts. Some of these nonpathogenic *Xanthomonas* strains have been isolated together with pathogenic relatives from a diversity of host plants, at times, from the same lesion in infected plants, from asymptomatic hosts, or from seed lots or transplants (Gitaitis 1987; Vauterin et al. 1996). Vauterin et al. (1996) systematically characterized 70 diverse nonpathogenic xanthomonads based on fatty acid methyl ester (FAME) and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) protein patterns. This pioneering study indicated potential new species and the need to address the diversity and relatedness of these nonpathogenic strains to pathogenic strains from an ecological perspective and the practical aspects of disease diagnostics and management strategies. In the last two decades, several studies have addressed this question of diversity, focusing on individual species, using more advanced methods of multilocus sequence typing and genome sequencing (Gonzalez et al. 2002; Cesbron et al. 2015; Essakhi et al. 2015; Triplett et al. 2015; Bansal et al. 2020, 2021; Li et al. 2020). Whole genome-based phylogeny placed some crop-associated nonpathogenic xanthomonads in the species *arboricola* and *cannabis*, both belonging to Group 2 (Cesbron et al. 2015;

Jacobs et al. 2015; Merda et al. 2016, 2017) and some in the newly described species, such as *Xanthomonas sontii* (Bansal et al. 2021), belonging to the early branching clade, Group 1. Apart from *Xanthomonas arboricola* and *Xanthomonas campestris*, which include both pathogenic and nonpathogenic strains that were simultaneously isolated from symptomatic hosts (Lee et al. 2020; Martins et al. 2020), other nonpathogenic strains are only distantly related to their cocolonizing pathogenic strains (Vauterin et al. 1996).

These nonpathogenic strains are diverse in their phylogenetic placements and vary in their makeup of type III secretion systems (T3SSs) and associated type III secretion effectors (T3SEs). The T3SSs, encoded by the *hrp2* cluster and type III effectors and/or their repertoires, are important determinants of pathogenicity in xanthomonads. Pathogenic xanthomonads belonging to *Xanthomonas maliensis*, *Xanthomonas cannabis*, *Xanthomonas pseudoalbilineans*, and *Xanthomonas sacchari* are an exception in that they lack a *hrp2* T3SS, although some possess regulators of T3SS (Studholme et al. 2011; Jacobs et al. 2015; Pieretti et al. 2015; Triplett et al. 2015). Knowing the importance of T3SS and associated T3SEs in the pathogenicity of xanthomonads, it is unsurprising that most nonpathogenic xanthomonads lack a T3SS. Such nonpathogenic xanthomonads lacking T3SS, specifically *X. arboricola*, have been previously referred to as commensal xanthomonads. However, Merda et al. (2017) further showed that some commensal strains of *X. arboricola* possess a T3SS but contain only three to four T3SEs. Yet other commensal strains lack a *hrp2* cluster but have up to four T3SEs anyway (Cesbron et al. 2015).

Given the heterogenic distribution of T3SSs and variable T3SE repertoires, nonpathogenic strains make a suitable model to study the evolutionary history of the *hrp2* T3SS family, T3SEs, and associated regulators in xanthomonads (Cesbron et al. 2015; Merda et al. 2017). Merda et al. (2017) inferred ancestral acquisitions of the *hrp2* cluster

in *Xanthomonas* and indicated loss or subsequent gain in certain clades in 82 genomes spanning the genus *Xanthomonas*. As we uncover additional diversity in the genus *Xanthomonas*, evolutionary gain and loss of different types of T3SSs, including atypical T3SSs (Pieretti et al. 2015; Pesce et al. 2017), in xanthomonads along with T3SEs and regulators can be a valuable approach to understand the role of T3SSs in allowing intimate association of xanthomonads with plants and to further determining their lifestyle.

Some nonpathogens have been referred to as opportunistic pathogens under favorable conditions (Vauterin et al. 1996) and have amyolytic and/or pectolytic activity, which allows them to cause soft rot on their host (Gitaitis 1987; Zarei et al. 2022). *Xanthomonas* strains were recently found in the endosphere of *Arabidopsis* as part of the At-LSPHERE collection. Evaluation of these strains' pathogenicity on wild-type *Arabidopsis* and immunocompromised plants confirmed their opportunistic or conditional pathogenic nature based on aggressive symptoms on immunocompromised plants lacking plant NADPH oxidase (RBOHD; Pfeilmeier et al. 2021, 2024). These studies further highlighted the role of microbial community members and microbiota-induced plant immunity in reducing the prevalence of opportunistic strains. On the other hand, a closely related endophytic *Xanthomonas* strain, WCS2014-23, was identified as a member of the consortium recruited to the rhizosphere of *Arabidopsis* upon foliar infection with the biotrophic pathogen and played a role in induced systemic resistance against the biotrophic pathogen and enhancing plant growth (Berendsen et al. 2018). These findings raise the question of whether nonpathogenic xanthomonads play an important role as resident and functional members of the phyllosphere microbiome or are transient nonfunctional community members. A related question is if the so far uncharacterized xanthomonads that have been isolated from rainwater are bona fide phyllosphere microbiome members that are only transiently present in the atmosphere or if they present a separate population of nonplant-associated xanthomonads (Failor et al. 2017).

In this study, we addressed the following questions by sequencing a collection of 83 *Xanthomonas* strains from diverse hosts, environments, and geographical locations suspected of being nonpathogenic. How diverse are nonpathogenic xanthomonads and how related are they to their pathogenic counterparts? Given that some have been simultaneously isolated with pathogens, do the genomes show evidence of genetic exchange among pathogens and nonpathogens? And can we identify genomic signatures that can explain different life history strategies (commensal, weak pathogens, or pathogens) in association with plants and their survival outside the plant host? Briefly, phylogenetic analysis of these strains showed that they

spanned the entire genus and included potentially new species. Furthermore, the heterogeneous distribution of T3SS and associated T3SEs across pathogenic, weakly pathogenic, and commensal strains suggested a lack of apparent association of these important pathogenicity factors with the lifestyle of *Xanthomonas* strains. Thus, we used an integrated approach of comparative genomics and association analysis to identify the genomic attributes associated with these lifestyles of strains spanning the entire phylogeny of the genus.

Results

General Features of *Xanthomonas* Strains from This Study, Potential New Species, and Core Genome Phylogeny

Eighty-three *Xanthomonas* strains, found to be nonpathogenic or weakly pathogenic based on pathogenicity assay on the host of isolation, were sequenced in this study (supplementary table S1, Supplementary Material online). Some strains were noted as opportunistic based on the observation of restricted leaf spots only associated with wounds and soft rot-like symptoms when inoculated on fruits (Gitaitis et al. 1987). Most of the strains were isolated from various symptomatic and asymptomatic crops, including tomato (*Solanum lycopersicum*; 14 strains), pepper (*Capsicum annuum*; 8 strains), and common bean (*Phaseolus vulgaris*; 27 strains), along with other plant species such as radish (*Raphanus sativus*), walnut (*Juglans*), orange (*Citrus sinensis*), and sunflower (*Helianthus*). In addition to the strains isolated from plant hosts, 18 were recovered from rainwater (supplementary table S1, Supplementary Material online). The genome sizes among the sequenced strains varied from 3.6 Mb for strain 60 to 5.3 Mb for strain F5. The percent GC content ranged from 64.60% for strain 3075 to 69.31% for strain F10. Furthermore, the number of coding sequences (CDS) varied from 3,223 in strain 60 to 4,535 in strain F5. There was no apparent correlation between genome size, CDS, and %GC with the status as host associated and environmental (supplementary fig. S1, Supplementary Material online). Although the median genome size and median number of genes for each strain were 4.87 Mb and 4,208 genes, respectively, one strain (strain 60) showed a reduced genome size of 3.6 Mb and a reduced number of 3,223 coding genes (supplementary table S1, Supplementary Material online).

To determine the taxonomic placement of the 83 newly sequenced strains, Average Nucleotide Identity (ANI), digital DNA–DNA hybridization (dDDH), and Microbial Species Identifier (MiSI) were used. ANI values between the strains and representative *Xanthomonas* strains including type strains of all *Xanthomonas* spp. available in NCBI (supplementary table S2, Supplementary Material online) varied from 79% to 100%. Nonpathogenic strains

sequenced in this study belonged to both *Xanthomonas* groups (Timilsina et al. 2020). For group 1, nine strains belonged to *Xanthomonas euroxanthea*. For group 2, 34 strains belonged to *X. arboricola*, 16 to *X. cannabis*, 9 to *X. campestris*, and 2 to *Xanthomonas euvesicatoria* (supplementary tables S5 to S7 and figs. S2 and S3, Supplementary Material online). The remaining 13 strains showed ANI values between 85% and 94% compared with type strains or representative strains of known *Xanthomonas* species. These strains were assigned to eight cluster-type cliques or singletons according to the MiSI method, indicating the presence of at least eight potentially novel species (supplementary table S8, Supplementary Material online). Given the findings of potentially novel species adding new diversity to the *Xanthomonas* genus, we established a robust phylogenetic tree based on the single-copy genes of these newly sequenced strains along with the available type or representative strains of the *Xanthomonas* genus (Fig. 1). The OrthoFinder analysis assigned most genes (544,723; 99% of the total) to 11,456 orthogroups. There were 1,005 orthogroups in all species, and 819 of these consisted entirely of single-copy genes. The phylogenetic reconstruction using single-copy genes showed a considerable diversity of strains isolated from both plant hosts and the environment broadly distributed throughout the genus (Fig. 1).

Xanthomonas sontii, *X. sacchari*, *Xanthomonas albilineans*, *Xanthomonas hyacinthi*, *Xanthomonas translucens*, *Xanthomonas theicola*, and the two recently described species *Xanthomonas bonasiae* and *Xanthomonas youngii* were the only known species belonging to *Xanthomonas* group 1 (Rodriguez-R et al. 2012; Bansal et al. 2021; Mafakheri et al. 2022). The collection of strains sequenced here adds four new *Xanthomonas* species clusters to group 1, species cluster I (strain F5), species cluster II (strain F1), species cluster IV (strain F10), and species cluster VIII (strains 3307, 3498, and F4), all isolated from citrus plants and rainwater. Another new *Xanthomonas* species cluster, species cluster III (strain 60), from this collection clustered with early branching species at the base of the phylogenetic tree, along with *Xanthomonas retroflexus*.

Our collection also added three new species clusters to *Xanthomonas* group 2 (Fig. 1). Most strains sequenced here belong to clade A, specifically to *X. arboricola* (supplementary tables S5 and S6, Supplementary Material online) and *X. euroxanthea* (supplementary table S7, Supplementary Material online). Strains isolated from bean seeds (CFBP 8151 and CFBP 8152) were closely related to *X. arboricola*, and strains isolated from rainwater (3075 and 3058) belong to potentially novel species clusters, species cluster VI and species cluster V, respectively, within clade A (supplementary table S8, Supplementary Material online; Fig. 1).

Two of our sequenced strains belonged to clade B and were identified as *X. euvesicatoria* (CFBP 7921 and CFBP 7922). Strains in clade C isolated from bean seed, tomato, and nightshade plants belong to *X. cannabis* species (16 strains). This clade also harbors one novel *Xanthomonas* species cluster, species cluster VII (strains 3793 and 4461), isolated from rainwater. Crop-associated *X. campestris* strains (9 strains) isolated from radish, bean, and tomato plants belong to clade D, including pathogenic *X. campestris* strains (Fig. 1).

Distribution of T3SS Clusters across the Phylogeny

We screened three types of T3SS clusters known within the genus *Xanthomonas* across the set of genomes: (i) the *hrp2* cluster present in group 2 xanthomonads (Tampakaki et al. 2010), (ii) the SPI-1 type cluster present in *X. albilineans* (Pieretti et al. 2009), and (iii) the noncanonical *Xtra* cluster present in *X. translucens* (Wichmann et al. 2013). In addition, we also included the T3SS cluster from *Stenotrophomonas* sp. to represent the *sct*-type cluster from a closely related genus. Among the 83 newly sequenced *Xanthomonas* genomes, 24% contained a functional T3SS cluster (supplementary fig. S4, Supplementary Material online). Strain *Xanthomonas* sp. 60 showed the presence of a unique *sct*-type T3SS cluster with a gene organization comparable to the one found in *Stenotrophomonas chelatiphaga* DSM 21508 (supplementary fig. S5, Supplementary Material online). Apart from partial T3SS clusters in strains of *Xanthomonas cucurbitae* and *Xanthomonas fragariae*, we also identified a single gene encoding protein V of the *sct*-type cluster in *X. cannabis* CFBP 8595, *X. cannabis* 8600, *X. cannabis* CFBP 8600, *X. arboricola* F12, *X. arboricola* 84A, and *Xanthomonas* sp. 4461 and a single *hrpF* in *X. arboricola* CFBP 7681 (Fig. 2). T3SS clusters were missing in *Xanthomonas pisi*, *Xanthomonas floridensis*, *Xanthomonas melonis*, *X. maliensis*, *X. sontii*, *X. sacchari*, and *X. retroflexus*. *Xanthomonas phaseoli* pv. *phaseoli* CFBP 412 showed the presence of two types of T3SS, the *hrp2* cluster and the SPI-1 type cluster present in *X. albilineans*, similar to previous findings for strain *X. phaseoli* pv. *phaseoli* CFBP 6164 (Alavi et al. 2008; Fig. 2).

T3SS Clusters Were Gained and Lost Multiple Times in the Genus *Xanthomonas*

To better understand the evolution of T3SS clusters in the genus *Xanthomonas*, we next studied the present and absent patterns of T3SS clusters among the analyzed genomes. The phyletic pattern generated above was input to GLOOME, which maps gain and loss events onto the phylogeny. According to the scenario estimated with GLOOME, there were several independent acquisition and loss events of T3SS clusters during the evolution of xanthomonads. Independent acquisition of the *sct*-type T3SS cluster was

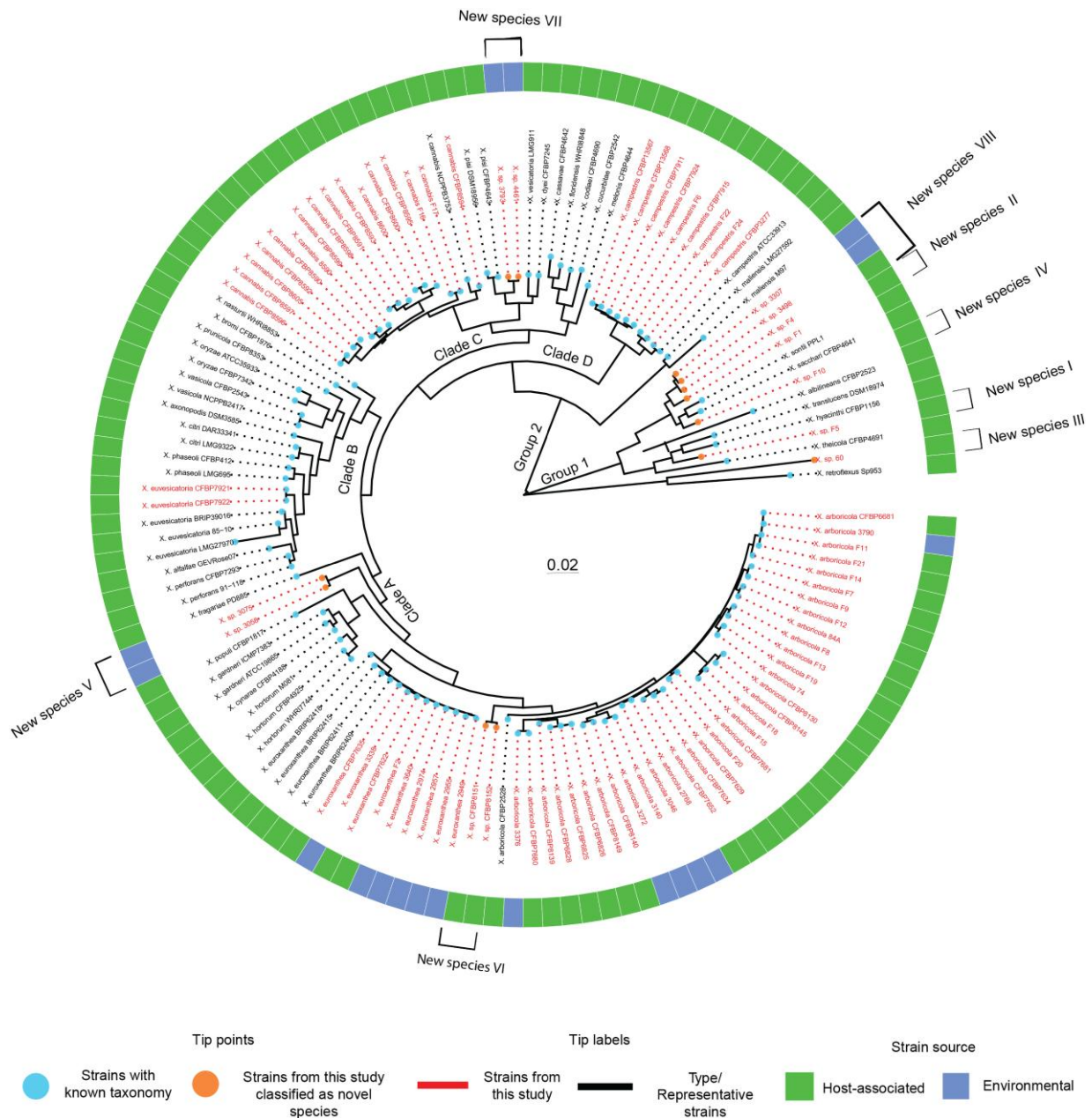


Fig. 1.—Comparative genome analysis demonstrated the presence of eight novel species in the genus *Xanthomonas*. Maximum likelihood phylogeny based on the 819 single-copy genes of 134 strains representing the entire *Xanthomonas* genus. The phylogenetic tree was inferred using OrthoFinder v2.5.2 and drawn with R package *ggtree*. *Xanthomonas* strains from this study are highlighted in red, while the representative/type strains are in black. The tip points are colored orange to show novel species identified from this study, while cyan represents *Xanthomonas* species with known taxonomy. Blue-colored blocks indicate environmental strains, and host-associated strains are indicated by green-colored blocks surrounding the phylogenetic tree.

inferred to have occurred in *Xanthomonas* sp. 60 (Fig. 2). The ancestor of *X. translucens*, *X. hyacinthi*, *Xanthomonas* sp. F5, and *X. theicola* acquired the *Xtra*-type T3SS cluster. This cluster was found to be conserved in the *X. translucens* clade. However, it has been subsequently lost in *Xanthomonas* sp. F5 and partially lost in *X. hyacinthi* and *X. theicola*, as indicated by a partial *Xtr*-type cluster (supplementary fig. S5, Supplementary

Material online; Fig. 2). The SPI-1 type T3SS cluster was independently acquired in *X. albilineans* and *X. phaseoli* CFBP 412.

Next, in addition to the probabilities of gene gain/loss estimated by GLOOME analysis, we noted the genomic context and sequence identities to identify *hrp2* cluster gain/loss events associated with group 2 xanthomonads. Based on the sequence identities of *hrp2* cluster genes, clades

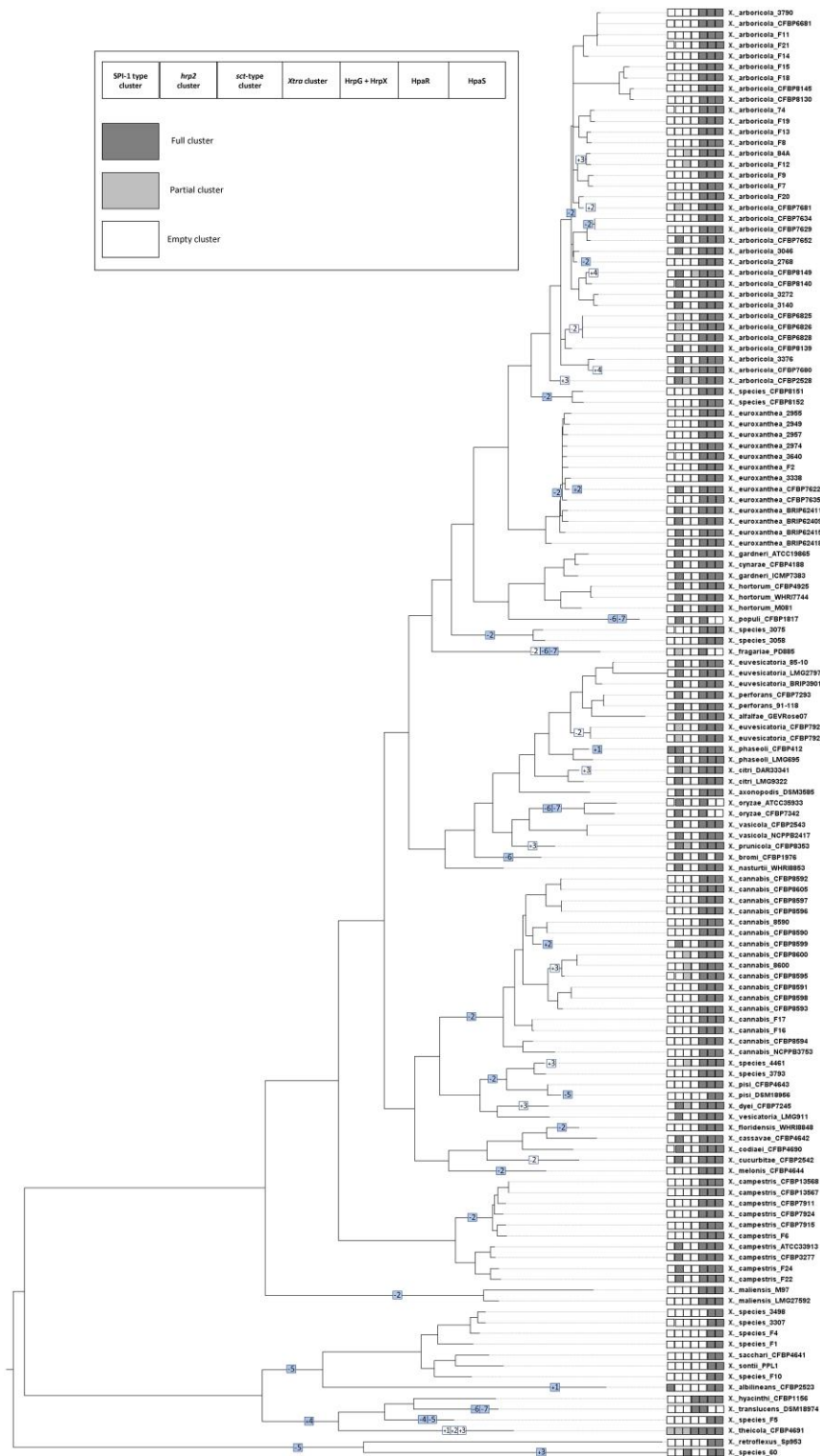


Fig. 2.—Multiple events of gain and loss of T3SS are evident in the genus *Xanthomonas* along with the presence of a novel T3SS cluster. Maximum likelihood phylogeny based on the core proteome and T3SS cluster and regulator gain/loss prediction inferred for the 134 strains representing the entire *Xanthomonas* genus. The presence/absence of the different types of T3SS and regulators are represented by an ordered vector of size 7, such that a dark gray, a light gray, and a white *i*th element in the vector indicates full presence, partial presence, or absence of the *i*th element, respectively. Inferred full and partial T3SS clusters are in color or have a white background, respectively. Acquisition and loss events are represented by plus and minus signs, respectively. For example, a colored +2 indicates a full acquisition of the second T3SS in this figure, i.e. *hrp2*.

Downloaded from https://academic.oup.com/gbe/article/16/4/evae074/7655398 by guest on 05 June 2024

A, C, and D were observed to possess the *Xcc*-type *hrp2* cluster, while clade B, except for *Xanthomonas nasturtii*, possessed the *Xeu*-type *hrp2* cluster. It is possible that replacement of the *Xcc*-type *hrp2* cluster by the *Xeu*-type *hrp2* cluster occurred within clade B strains, except in *X. nasturtii*, through rearrangements. According to GLOOME analysis, regulators of T3SS, HrpX, and HrpG were acquired by a common ancestor of *Xanthomonas* before the split into group 1 and group 2. A single loss event of these regulators occurred in group 2, where these genes were lost on the branch leading to *X. pisi* DSM18956. In group 1, these genes were lost in several independent events: (i) on the branch leading to the common ancestor of *X. albilineans*, *X. sontii*, *X. sacchari*, and *Xanthomonas* sp. strains (F10, F1, F4, 3307, and 3498); (ii) on the branch leading to *Xanthomonas* sp. F5; and (iii) on the branch leading to the common ancestor of *X. retroflexus* and *Xanthomonas* sp. 60. An alternative less parsimonious explanation is that these regulators were only acquired by group 2 *Xanthomonas* and were independently acquired by the cluster of *X. hyacinthi*, *X. translucens*, *X. theicola*, and *Xanthomonas* sp. F5, followed by loss of these genes together with loss of *Xtra*-type T3SS genes in *Xanthomonas* sp. F5. Similar to HrpX and HrpG, the regulators HpaR and HpaS were also acquired by the common ancestor of *Xanthomonas* before the split of group 1 and group 2. These regulators were lost independently on numerous occasions: (i) on the branch leading to *X. translucens* in group 1, (ii) on the branch leading to *Xanthomonas populi* in group 2, and (iii) on the branch leading to *Xanthomonas oryzae* in group 2.

T3SE Repertoires Range from 0 to 41 in Crop-Associated and Environmental Strains

Previous studies involving genome screening for T3SEs of nonpathogenic *X. arboricola* strains indicated low effector gene loads, with a reduced core effector set ranging from zero to four T3SEs, namely, XopR, HpaA, XopF1, and AvrBs2 (Merda et al. 2017). Given the diversity of commensal xanthomonads spanning the entire phylogeny of the genus *Xanthomonas* and the fact that this study also included environmental strains, we hypothesized that low effector loads might be widespread among nonpathogenic xanthomonads that are either plant-associated or environmental strains, owing to their presumed global broad host range (we caution that host range tests have not been conducted for each strain sequenced in this study; thus, we refer to a global broad host range based on previous studies that either recovered nonpathogenic isolates from diverse plant hosts or tested their host range on diverse plants). Surprisingly, our analysis revealed that effector repertoires vary greatly from 0 to 41 (supplementary fig. S6, Supplementary Material online).

Xanthomonas arboricola strains sequenced in this study, although belonging to a monophyletic group, showed considerable variation in the presence/absence of T3SS and the size of effector repertoires, ranging from 1 to 12 known T3SEs. Strains lacking T3SS but containing T3SEs, XopAW, and XopAX included crop-associated and environmental strains. Some crop-associated strains lacking a T3SS possessed an additional T3SE, AvrBs2. Two *X. arboricola* strains (F21 and CFBP 6681) isolated from tomato lacked a T3SS but possessed two T3SEs, AvrBs1 and XopH, in addition to XopAW and XopAX. AvrBs1 and XopH have been identified as plasmid-borne T3SEs in *X. euvesicatoria*, a tomato pathogen (Potnis et al. 2012). Interestingly, a set of strains isolated from rainwater and from diverse crops such as walnut, pepper, and bean seed shared the same effector repertoire (XopR, XopF1, XopF2, AvrBs2, and XopAW), in addition to a functional T3SS. Strains of *X. arboricola* (CFBP 6825, CFBP 6826, and CFBP 6828) isolated from pepper possessed unusually large effector repertoires comprising 10 T3SEs (XopZ2, XopR, XopP, XopF1, XopF2, XopAW, XopAR, XopAL1, XopAD, and AvrBs2) comparable to those found in *X. arboricola* strains pathogenic on walnut (supplementary fig. S6, Supplementary Material online).

Along with the variable presence of T3SSs, crop-associated nonpathogenic strains also varied in effector repertoire size, which ranged from 1 to 10 T3SEs in *X. cannabis* and reached 27 effectors in *X. campestris* (supplementary fig. S6, Supplementary Material online). The strains with a higher number of T3SEs (>7) may suggest their possible pathogenic status, although their host range needs further exploration. Rain-derived *X. euroxanthea* strains lacked T3SSs and possessed a single effector, XopR. Crop-associated *X. euroxanthea* strains isolated from tomato and bean contained a T3SS and T3SEs, XopF1, XopF2, XopZ2, XopAK, and XopR. Rain-derived novel *Xanthomonas* sp. strains (3058, 3075, 3793, and 4461), although lacking canonical T3SSs, possessed orthologs of the HrpG/X master regulators. Strains 3058 and 3075 lacked any known T3SEs, while strains 3793 and 4461 contained homologs of AvrXccA1 and AvrXccA2. XopAW and AvrXccA1 were observed in the crop-associated strains 3793 and 4461 that lacked T3SSs but contained sequences homologous to HrpG/X.

Xanthomonas Strains Can Be Assigned to Three Lifestyle Categories: Commensal, Weakly Pathogenic, and Pathogenic

To determine the role of nonpathogenic strains in the evolution of the genus *Xanthomonas* and to determine the genetic basis of their nonpathogenic lifestyle, strains sequenced in this study and strains with publicly available genome sequences (a total of 337 strains) were assigned

to three lifestyle categories: commensal, weakly pathogenic, and pathogenic. The classification was based on the following: (i) data available in NCBI BioSample, CFBP, NCPPB, and LMG culture collection records as well as peer-reviewed articles (for all strains with publicly available genome sequences); (ii) pathogenicity tests performed on hosts of isolation (for plant-associated strains sequenced in this study); and (iii) inference from the presence of T3SS and T3SEs for environmental strains and those lacking pathogenicity test data (strains lacking T3SS were classified as commensals; strains possessing T3SS and fewer than 12 T3SEs were classified as weakly pathogenic; strains possessing T3SS and more than 12 T3SEs were classified as pathogenic). Whenever possible, classification was based on more than one of the above criteria.

Of the 337 strains, 151 were classified as commensal, 140 as weakly pathogenic, and 46 as pathogenic (supplementary table S3, Supplementary Material online). The plant-associated strains sequenced in this study were largely nonpathogenic when tested on their host of isolation (54 strains), except for 7 strains identified as weakly pathogenic. These results aligned well with the genomic analysis: strains lacking T3SS displayed either no symptoms or atypical symptoms on the host of isolation, and strains with a functional T3SS and a reduced T3SE repertoire were either nonpathogenic or weakly pathogenic. While we cannot exclude that some of these weakly pathogenic strains cause disease in other hosts, we think this is unlikely based on their reduced T3SE repertoire.

Genetic Exchange between Commensal, Weakly Pathogenic, and Pathogenic *Xanthomonas* Strains

To determine whether the strains classified as commensal or weakly pathogenic strains exchange genetic material with crop-associated pathogenic strains, we examined phylogenetic networks inferred from concatenated sequences of 12 housekeeping genes using SplitsTree (Fig. 3). This analysis revealed reticulated events between commensal and pathogenic *Xanthomonas* strains, suggesting recombination. Two evident reticulation events were identified at the intersections of the network, one between pathogenic and commensal or weakly pathogenic strains belonging to *X. arboricola* and another one between species encompassing clades B, C, and D of group 2 and species belonging to group 1, indicating the flow of genetic information between them (highlighted light blue in Fig. 3). Three intrinsic events can be observed by closely analyzing the parallelograms between group 1 and group 2. The first event is localized in the central part of the entire network. This reticulated event links the two main branches involving all species belonging to clades B, C, and D of group 2 and species belonging to group 1 (highlighted purple in Fig. 3). The second event involves clade D

(*X. campestris*), some species belonging to clade C, and the entire group 1 (highlighted orange in Fig. 3). Finally, the third event is exclusively shared between species from clade B and some species belonging to clade C (highlighted green in Fig. 3).

According to the neighbor-net tree, at least two strains isolated from rainwater appear to have resulted from the abovementioned putative recombination events. Within the clade A parallelogram, *X. arboricola* strains (3790, 2768, 3140, 3272, 3046, and 3376) isolated from rainwater showed genetic recombination with type strain *X. arboricola* pv. *juglandis* CFBP 2528 and other crop-associated commensals/weakly pathogenic *X. arboricola* strains. A similar observation was found for *Xanthomonas* sp. strains (3793, 4461, 3307, and 3498) as these environmental strains exchanged genomic content with the ancestors of crop-associated pathogenic and commensal species belonging to clade C and of species belonging to group 1, respectively.

Many exchange events were observed between the ancestors of crop-associated commensal, weakly pathogenic, and pathogenic strains isolated from the same crop hosts. Within the *X. arboricola* clade, two commensal strains (CFBP 7629 and CFBP 7634) and one weakly pathogenic strain (CFBP 7652) isolated from walnut and the pv. *juglandis* pathotype strain (CFBP 2528) showed a reticulated network. *Xanthomonas campestris* ATCC33913, pathogenic on crucifers, also appears to result from the genomic exchange with the ancestor of the commensal strains *X. campestris* (CFBP 13567 and CFBP 13568), both isolated from radish plants.

By screening for mobile genetic elements (MGEs) and associated genes using a computational approach, we further tested the hypothesis that commensal xanthomonads may act as reservoirs carrying fitness and virulence factors that can potentially be transferred to other strains through MGEs. MGEfinder identified at least one predicted mobile genetic island for each phylogenetic cluster and over 300 unique MGEs. Predicted mobile islands containing genes of possible interest and all islands identified as having a mobility gene, such as transposases or integrases, are listed in supplementary table S9A, Supplementary Material online. While many predicted MGEs that we found contain housekeeping genes, some contained genes that may play a role in increasing fitness or virulence. The predicted mobile islands contained genes associated with antimicrobial resistance and genes for bacterial and fungal competition, such as multidrug efflux pumps, type IV secretion system genes, chitinase, and virulence factors. Most MGEs also contained integrase and phage-related genes. T3SE XopAD, flanked on both ends by IS5 transposases, was observed in *X. campestris* strains F24 and F22. This island was also found in *X. campestris* pv. *raphani* 756C (the insertion location is shown in supplementary fig. S7, Supplementary

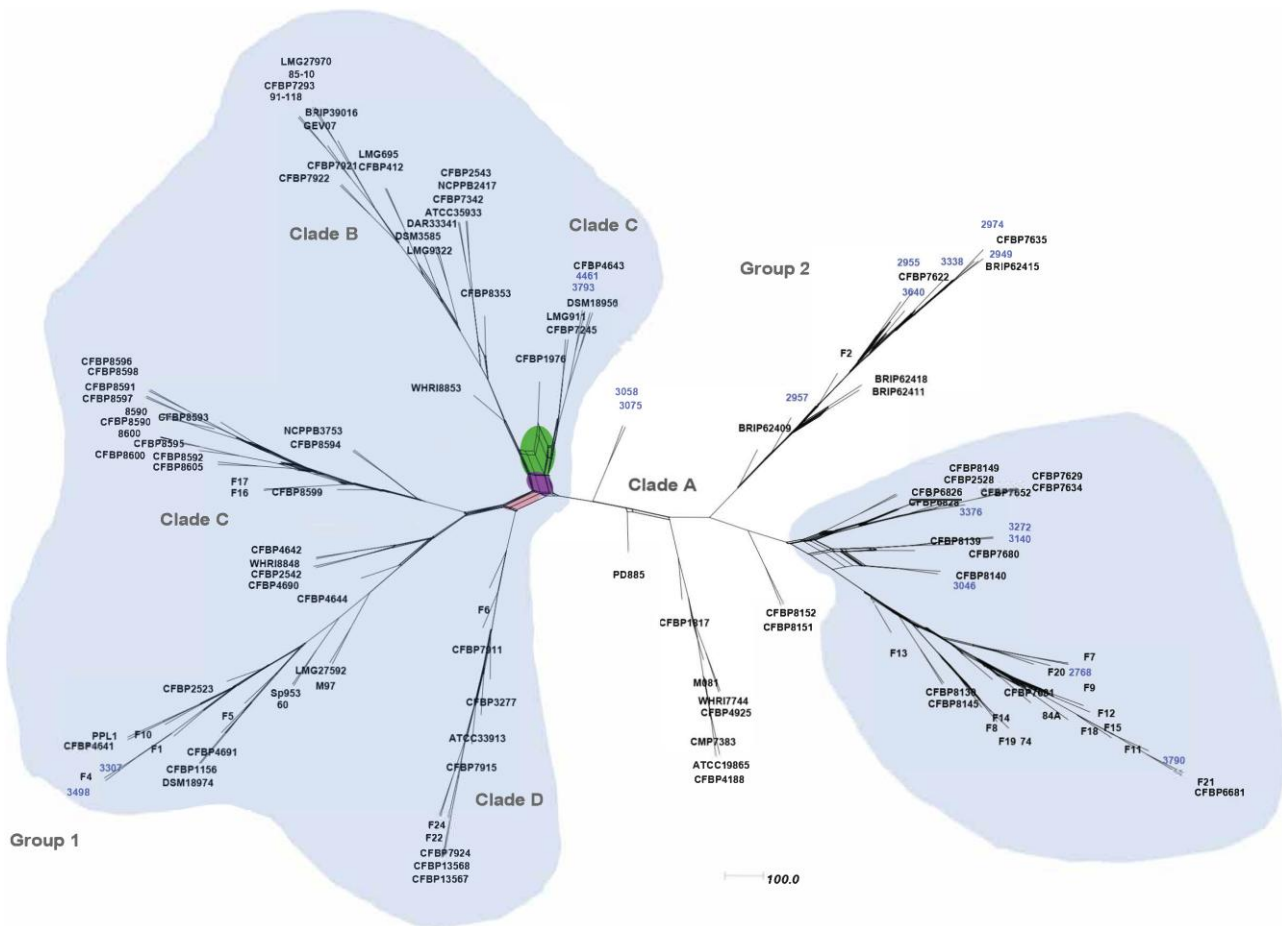


FIG. 3.—Phylogenetic network between pathogens, weak pathogens, and commensals suggests the possibility of several recombination events during their evolutionary history. Neighbor-net tree constructed using SplitsTree software based on 12 concatenated housekeeping gene sequences from the genomes of the 134 *Xanthomonas* strains used in this study, indicating diversity and recombination events. The light blue highlighted areas represent two major clusters containing phylogenetic networks that can accommodate conflicting signals as indicated by reticulation events. Each color (green, purple, and orange) indicates a reticulated event suggesting genetic exchange among different clades within *Xanthomonas* genus phylogeny. *Xanthomonas* genus phylogeny is represented as two groups, early branching species belonging to group 1 and group 2. Group 2 can be further divided into clades A, B, C, and D. Environmental strains (rain derived) are labeled blue. One example of genetic exchange between ancestor of pathogenic and commensal/weakly pathogenic relatives can be the reticulation at the base of branches connecting pathogenic *X. arboricola* strain CFBP2528 and other weakly pathogenic strains CFBP7629, CFBP6826, or CFBP7634.

Material online). The type III effector XopAA, flanked by a putative transposase, was found in *X. euvesicatoria* strains CFBP 7922 and CFBP 7921. An endopolygalacturonase, known for degrading pectin in plant cell walls (Federici et al. 2001), was identified in *Xanthomonas* sp. F1. A peptidoglycan *O*-acetylase, which may alter bacterial cell walls to avoid lysis by an innate immune response (Sychantha et al. 2018), was detected in *X. arboricola* 3272 (supplementary table S9B, Supplementary Material online).

Lifestyle Had a Significant Effect in Determining the Repertoires of Cell Wall–Degrading Enzymes in *Xanthomonas*

We hypothesized that commensal or weakly pathogenic *Xanthomonas* strains possess distinct repertoires of cell

wall–degrading enzymes (CWDEs) compared to those from pathogens. Such repertoires might confer them the ability to utilize a wide range of carbohydrate substrates and colonize diverse host plants. Within the genus *Xanthomonas*, different species generally had similar types of CAZymes but with large variations in the absolute numbers of genes within each category in the CAZy profiles (Fig. 4; supplementary table S10 and fig. S8, Supplementary Material online). Commensals and weak pathogens formed a distinct cluster compared to pathogens, indicative of differences in repertoires of CAZymes in commensals and weak pathogens compared to pathogens. Three commensal strains that clustered together with pathogens were *X. maliensis* LMG27592, *X. maliensis* M97, and *Xanthomonas* sp. 60.

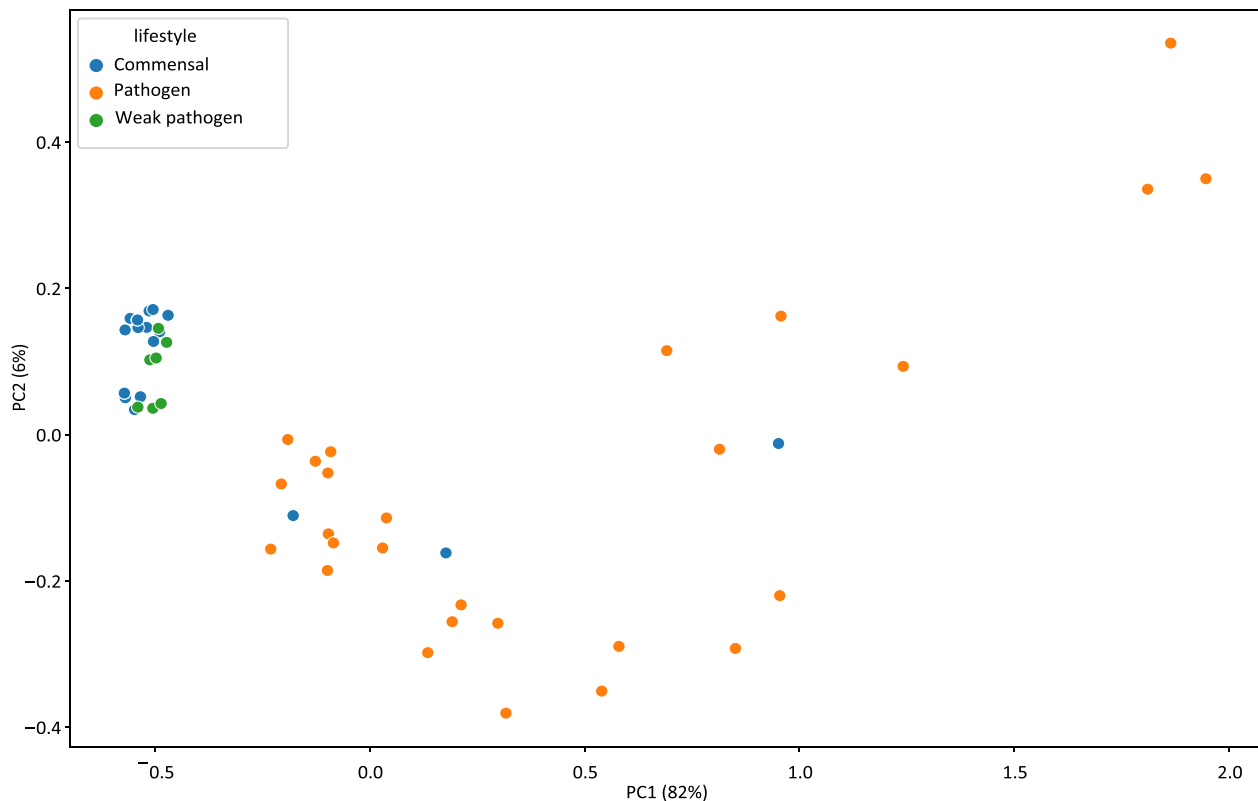


FIG. 4.—*Xanthomonas* lifestyle can be explained by an altered CAZymes landscape. PCA plot showing the contribution of the different bacterial lifestyles in the distribution of gene repertoire for CAZymes.

We assessed the effect of bacterial lifestyle on repertoires of genes coding for CAZymes. Distance-based redundancy analyses (dbRDA) of Jaccard distances and permutational multivariate analysis of variance (PERMANOVA) calculated on the genomic compositions of CAZyme family revealed a significant contribution of lifestyle to the distribution of gene repertoires for CAZyme ($R^2 = 0.06$, $P < 0.05$; Fig. 4). Furthermore, a pairwise comparison of genomes of different lifestyles revealed CAZyme gene repertoire composition across commensals, weak pathogens, and pathogens to be significantly different ($P < 0.05$), suggesting that lifestyle plays an important role in determining the distribution of CAZymes. Principal component analysis (PCA) on a matrix containing CAZymes using Jaccard distances showed increased separation of pathogenic from weakly pathogenic strains and commensals (Fig. 4).

Association Analysis to Identify Genes and Features That Define the Lifestyle of *Xanthomonas* Species

Next, we were interested in identifying genes either involved in the adaptation of *Xanthomonas* species to a commensal, weakly pathogenic, or pathogenic lifestyle. Association analysis was performed on the 26,812 orthogroups from a set of 337 *Xanthomonas* genomes

containing representative strains from different clades that comprised both pathogens, commensals, and weak pathogens (supplementary table S4 and fig. S9, Supplementary Material online). We hypothesized that commensals or weak pathogens and pathogens possess genes or gene families that define their adaptation to the respective lifestyles and have evolved these genes in a phylogeny-independent manner. Presence/absence and copy number matrix of orthologs were used as input to conduct association analysis using Scoary, PhyloGLM, and hyperglm. Candidate genes present/absent or enriched/depleted in commensal, weakly pathogenic, and pathogenic strains were identified (supplementary fig. S10, Supplementary Material online). Overall, each lifestyle category had unique genomic attributes. Interestingly, weak pathogens, as defined here by the presence of a T3SS and 7 to 12 T3SEs, contained overlapping genes with both pathogens and commensals (supplementary fig. S10, Supplementary Material online).

Among the genes identified as enriched in commensals compared to pathogens and weak pathogens, we found those that belonged to the following functional categories: intracellular trafficking and secretion (type VI secretion system proteins and putative effectors), carbohydrate metabolism and transport, phages and transposons,

posttranslational modification, protein turnover and chaperones, replication, and repair and defense mechanisms (specifically those encoding multidrug efflux proteins; [supplementary table S11, Supplementary Material](#) online, Fig. 5).

Enrichment of genes involved in carbohydrate metabolism in commensals may indicate their ability to utilize broader energy sources than pathogens. A distinct set of DNA-binding transcriptional regulators in commensals may allow them to easily switch between different energy sources depending on their availability on a diverse group of hosts. Other genes that may also impart stress tolerance related to a much broader range of conditions associated with plants or environments outside plants include multidrug efflux proteins, chaperones, outer membrane transporters, and type I restriction–modification system genes. During epiphytic colonization of a broad range of hosts, enrichment of genes involved in type VI secretion systems and effectors may allow the commensal xanthomonads to mediate interactions with the resident epiphytic community or pathogenic species (Drebes Dörr and Blokesch 2018).

Next, we identified the genomic attributes shared by commensals and weak pathogens but absent from pathogens that could explain strategies evolved by nonpathogens during their association with a broad range of hosts ([supplementary table S12A and B, Supplementary Material](#) online; Fig. 5). Glyoxalase/bleomycin resistance protein dioxygenase was identified to be enriched in both the commensal and weakly pathogenic xanthomonads. This protein might be involved in providing tolerance to toxins produced by other phyllosphere colonizers, as seen recently with the clone expressing a putative glyoxalase/bleomycin resistance dioxygenase showing neutralizing activity against toxoflavin, produced by *Burkholderia gladioli* (Choi et al. 2018). SGNH hydrolase-like domain containing acetyltransferase protein, AlgX, is involved in alginate biosynthesis in *Pseudomonas aeruginosa* and virulence during cystic fibrosis. Although alginate is known to be involved in epiphytic fitness in *Pseudomonas syringae* (Yu et al. 1999), its contribution toward stress tolerance in xanthomonads remains to be investigated. Another gene belonging to PD-(D/E)XK nuclease superfamily, also characterized in contact-dependent DNA-degrading effectors (Sirias et al. 2020), was identified in both commensals and weak pathogens. Such nuclease effectors may impart a competitive advantage to the commensals and weak pathogens in a mixed species community as they establish their niche. At least three TonB-dependent receptors were identified to be associated with commensals and weak pathogens. Overall, TonB-dependent receptors are overrepresented in xanthomonads and are thought to impart the adaptive ability to xanthomonads to utilize a wide variety of carbohydrates (Blanvillain et al. 2007). Another gene associated with commensal and weak pathogens belonged to the NRAMP

(natural resistance-associated macrophage protein) family, conserved from bacteria to eukaryotes. These proteins function as proton-dependent manganese transporters in bacteria and have evolved to adapt to oxidative environments (Kehres et al. 2000; Cellier et al. 2001). Enrichment of these proteins in commensals and weak pathogens may explain why they can be simultaneously isolated along with pathogens from infected tissue and why they can withstand host defenses, including reactive oxygen species.

We also screened the genomes for *fliC*, flagellin-encoding gene, and, specifically, the canonical flg22 epitope sequence to see the extent of variation within xanthomonads. It is not known whether commensal and weakly pathogenic xanthomonads can evade MAMP-triggered immunity (MTI) induced by flagellin. Polymorphism in the 43rd amino acid residue in the *fliC* sequence, a part of the flg22 epitope, was previously observed in pathogenic xanthomonads (Malvino et al. 2022). More specifically, a change from the canonical residue Val⁴³ to Asp⁴³ allowed pathogenic xanthomonads to escape detection by FLS2 in *Arabidopsis* and tomato (Sun et al. 2006; Malvino et al. 2022). We observed that all commensals and weak pathogens carried the canonical Val⁴³ in the flg22 epitope, except for *X. maliensis* LMG27592. In addition, some pathogenic xanthomonads also had the immunogenic version of flagellin ([supplementary table S3, Supplementary Material](#) online). Some commensals and weak pathogens possessed multiple copies (up to three) of *fliC* in their genome. Based on their similarity to the canonical flg22 epitope residues, they can be expected to possess immunogenic properties. However, variations in other parts of the gene may have a role in modulating immunity in various hosts.

We also identified genes that were enriched in pathogens compared to commensals. As expected, the majority of such genes were genes involved in T3SSs and T3SEs ([supplementary table S13 and fig. S10, Supplementary Material](#) online). Other features enriched in pathogens included glycosyl hydrolases, phage proteins, and chemotaxis proteins. Certain phage-related proteins and transposases were also identified as exclusively present in pathogens but absent in commensals, indicating their role in mobilizing pathogen fitness and virulence genes. This finding suggests that it might be possible to devise phage-based control strategies using phages selective toward pathogens ([supplementary table S13, Supplementary Material](#) online). Pathogens also exclusively contained certain transcriptional regulators, genes involved in the c-di-GMP signaling pathway and chemotaxis, lipid metabolism, and carbohydrate and amino acid metabolism compared to commensals ([supplementary table S14, Supplementary Material](#) online).

Genes that were enriched in pathogens compared to nonpathogens (commensals and weak pathogens) included a TonB-dependent receptor involved in Fe transport,

a chemotaxis protein, specific CWDEs (CAZymes), and an anti-sigma factor and AraC transcriptional regulator (Fig. 5; [supplementary table S15A to C, Supplementary Material](#) online). Among the genes depleted in commensals and weak pathogens was an adenylosuccinate synthetase (S-AMPS) involved in de novo purine biosynthesis in the cytoplasm. This pathway is connected to the TCA cycle and, thus, to the central metabolism. This enzyme links GTP hydrolysis to inosine monophosphate (IMP) condensation with L-aspartate to produce adenylosuccinate (S-AMP). It can be hypothesized that pathogens may experience limitations of these compounds in their niche, for example, apoplast or xylem, and, thus, have evolved the ability to synthesize S-AMP. This enzyme was found in the periplasm-enriched fraction of *Xanthomonas citri* but not in *X. fuscans* subsp. *aurantifolii* type B (Zandonadi et al. 2020). Another gene involved in lipid metabolism, annotated as phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase or related enzyme, was associated exclusively with pathogenic *Xanthomonas*. *Xanthomonas campestris* was characterized for its unique lipid metabolism pathways with a bifunctional enzyme for PE synthesis that functions in a serine-dependent or an ethanolamine-dependent pathway, depending upon the availability of substrates in planta (Aktas and Narberhaus 2015).

Discussion

Since nonpathogenic *Xanthomonas* strains were first coisolated with pathogenic strains or in close association with plants or plant debris, the interest in exploring the role of these strains in the evolution of the genus, their pathogenic potential, and their contribution toward the microbiota-mediated extended plant immunity have increased (Vauterin et al. 1996; Gonzalez et al. 2002; Cesbron et al. 2015; Essakhi et al. 2015; Triplett et al. 2015; Bansal et al. 2020, 2021; Li et al. 2020; Martins et al. 2020; Pfeilmeier et al. 2021, 2023; Entila et al. 2024). In this study, we attempted to address these aspects to understand how the diversity spanning the commensal to pathogenic lifestyles across the genus *Xanthomonas* has evolved. We harnessed the unexplored diversity that the collection sequenced in this study brought, particularly the previously poorly explored group 1 *Xanthomonas* species (Studholme et al. 2011). These new genomes were combined with a collection of 1,834 high-quality publicly available *Xanthomonas* genomes to obtain a phylogenetically representative set of 337 genomes spanning the different lifestyles present in the genus for comparative analysis.

We observed variable, but not random, presence/absence of T3SS clusters similar to the observations from previous studies (Cesbron et al. 2015; Fang et al. 2015; Triplett et al. 2015; Merda et al. 2017). This study extends these

findings to include some atypical T3SS clusters not limited to group 1 xanthomonads. This finding of gain of atypical T3SSs in addition to the *hrp2* cluster in some group 2 strains raises important questions about its functional significance and ecological role. The variable presence of T3SSs and effector repertoire sizes in xanthomonads ranging from 0 to 41 T3SEs indicate a plethora of diversity in the lifestyles of xanthomonads associated with plants along a continuum from being commensal endophytes to opportunistic pathogens or weak pathogens to full-blown pathogens.

The diversity of commensals and weak pathogens across the genus phylogeny presents opportunities to study evolution of xanthomonads specifically in the context of important virulence factors such as T3SS, T3SEs, and CWDEs that explain their plant-associated lifestyles. While CWDEs have been proposed to be acquired by the ancestor of xanthomonads (Lu et al. 2008), they seemed to have diversified along the course of evolution as xanthomonads established associations with diverse hosts. The differential repertoire of CWDEs accompanied by unique TonB-dependent receptors among pathogens and commensals may explain the differential niche colonization. While we did not study patterns of gain/loss of specific CWDEs, we observed that commensal and weak pathogens have distinct sets of CWDEs compared to pathogens, similar to the observation with pathogenic and nonpathogenic *X. arboricola* (Cesbron et al. 2015). Different methyl-accepting chemotaxis proteins in commensals and pathogens may also suggest how the perception of the environment may be lifestyle-dependent. Similar to the observations by Merda et al. (2016, 2017), we found evidence for an ancestral gain of the *hrp2* cluster before the split of group 2 xanthomonads. Group 1 xanthomonads have independently acquired different T3SS clusters. Thus, commensals belonging to group 1 xanthomonads may possess ancestral traits that allowed them to closely associate with plants, such as diverse regulators, carbohydrate metabolism, and defense/repair-related genes. Subsequently, upon diversification of group 2 xanthomonads into subsequent species, each displaying a high degree of host specificity, the effector repertoire reshuffling was observed (Hajri et al. 2009). Examining the genomic context of *hrp2* cluster and identities of core *hrp2* genes led us to hypothesize that replacement of Xcc-type *hrp2* cluster by Xeu-type *hrp2* cluster occurred within clade B strains, except in *X. nasturtii*, through rearrangements. Merda et al. (2016, 2017) also observed gene flow among T3SS and T3SEs of group 2 xanthomonads. The subsequent loss of *hrp2* cluster in certain lineages representing commensal xanthomonads was observed (Fig. 2). T3SS and associated T3SE loss may suggest that T3SS may impose a fitness cost. Thus, under this model, commensals belonging to group 2 may have been derived by regressive evolution from pathogenic strains. We also assessed patterns of gain/loss of *hrp2* regulators, *hrpG/X*,

analyzed in this study to identify clues to their involvement in the regulation of T3SS, associated limited repertoire in weak pathogens, or their existence in commensals in the absence of T3SS. Pfeilmeier et al. (2023) found that these master regulators do not regulate T2SS in nonpathogenic *Xanthomonas* strains lacking T3SS. This suggests a distinct regulatory network in play in commensal and pathogenic xanthomonads upon colonization of the host. Our analysis also indicated that many commensal and pathogenic strains engage in gene exchange and possess shared MGEs and carrier genes. Although experiments demonstrating the transfer of T3SSs from pathogenic to nonpathogenic xanthomonads were insufficient to impart pathogenicity phenotype to nonpathogenic strains (Meline et al. 2019), gene flow among certain strains may explain the reacquisition of T3SS. Such gain/loss events of T3SS and associated T3SEs may explain the heterogeneous distribution of the *hrp2* cluster and diversity in effector repertoires across strains. Merda et al. (2016) showed that genetic exchange involving life history traits between pathogenic and nonpathogenic *Xanthomonas* strains occurred likely through horizontal gene transfer and suggested a possibility of nonpathogenic strains acting as reservoirs of traits that allow the emergence of novel pathogenic strains (Meline et al. 2019). In this study, we analyzed genomes for the presence of MGEs and their cargo genes shared among pathogens and nonpathogens. Genes encoding fitness traits including T3SEs, antimicrobial resistance genes, and multidrug efflux pumps were noted in both commensals and closely related pathogens. This finding further highlights the role of commensals as a reservoir of traits that may contribute to the adaptation of pathogens resulting in new outbreaks, as demonstrated by Lee et al. (2020). Also, HGT between pathogen and commensal strains has been demonstrated in some bacteria, converting nonpathogenic strains into pathogenic ones (Brouwer et al. 2013). Such adaptive traits may have been subject to gene transfer among commensal and pathogenic strains.

Importantly, this study also demonstrates that T3SS and T3SEs are not the sole lifestyle determinants in xanthomonads. Commensals and weak pathogens have evolved strategies for tolerance to stresses with distinct sets of chemotaxis proteins, type VI secretion systems, TonB-dependent receptors, chaperones, and transcriptional regulators. We previously examined gain/loss patterns of type VI secretion system clusters across the *Xanthomonas* phylogeny and found evidence for nonrandom acquisition of T6SS and gene flow in core genes and effectors (Liyanapathirana et al. 2022). Further examination of gain/loss patterns of traits enriched in nonpathogens may help assess support for the regressive evolution model to explain the origin of commensal strains.

As for the strains isolated from rainfall, some *X. arboricola* strains cluster phylogenetically with pathogenic strains of

the same species isolated from plants and are indistinguishable in regard to the presence of a T3SS and the size of their effector repertoires. This suggests that they are pathogens that were simply caught while migrating through the atmosphere between host plants. The rainfall isolates identified as *X. euroxanthea* and as members of the novel species cluster VIII are very similar to strains without T3SS and a very small number of T3SEs previously isolated from plants. They can thus also be expected to generally live in association with plants, although as nonpathogens, and they were also isolated from the atmosphere while dispersed from one plant to the other. The situation may be similar for the four nonpathogenic strains isolated from rainfall that belong to the novel species clusters V and VII. However, since no closely related strains from plants are yet known, a lifestyle independent of plants cannot be excluded.

Finally, systematic evaluation of genomic traits associated with lifestyles might guide us as we develop diagnostic strategies to differentiate pathogenic from nonpathogenic strains associated with seed samples or infected field samples. A recent machine learning approach developed to predict the phenotype of plant-associated xanthomonads indicated specific domains associated with pathogenic and nonpathogenic lifestyles (te Molder et al. 2021), many of which overlap with the candidates identified in this study. Our study further confirms that diagnostic methods cannot rely on T3SS gene markers alone to identify pathogenic xanthomonads.

As a genus for which research has been highly focused on pathogenic potential, commensal or nonpathogenic xanthomonads represent a largely ignored component (Vauterin et al. 1996). Unlike pseudomonads, this side of the continuum that spans from commensal to opportunistic to weakly pathogenic lifestyles has not been well studied in xanthomonads. We lack understanding as to what extent nonpathogenic xanthomonads play a role in being evolutionary partners of plants with adaptive value contributing to overall plant health or whether they are just the members that contribute to niche filling, a process influenced by plant traits but are of minor adaptive importance in terms of fitness or growth benefit to the plant host. Hacquard et al. (2017) proposed a system involving multiple layers of barriers for establishing homeostasis with plants. Here, we assess how xanthomonads may have established association with the plants across the continuum of lifestyles from commensals to pathogens using this framework. The first two protective layers are (i) microbiota exhibiting nutritional and niche competition and (ii) plant physical barriers. Based on the association analysis conducted in this study, we identified several traits in commensals, such as enrichment in type VI secretion system genes, transcriptional regulators, and carbohydrate metabolism genes, which may indicate their ability to overcome nutritional and niche competition with other microflora members associated

with a wide range of plant hosts. Distinct CWDE repertoire in commensals, weak pathogens, and pathogens may also indicate their differential ability to overcome epidermal cell barriers. The next layer in maintaining homeostasis with plants is MTI. We found that commensal, weakly pathogenic, and few pathogenic xanthomonads possess canonical flg22 immunogenic epitopes, indicating that plants recognize and mount an innate immune response against them. Whether they can suppress this defense response may depend on the presence of T3SSs and T3SEs. It is hypothesized that a minimal repertoire of T3SEs present in some non-pathogenic *X. arboricola* may help them suppress the MTI (Merda et al. 2017). Some commensal strains from our collection had no known T3SEs, indicating that MTI may explain their low abundance and lower in planta population. Their simultaneous isolation with pathogenic xanthomonads may also suggest that association with pathogens allows them to take advantage of innate immune response suppression performed by the pathogens. However, the ability of commensal and weakly pathogenic xanthomonads to activate MTI may also suggest their contribution toward extending plant immunity against other pathogenic bacteria or fungi, as seen in *Sphingomonas* (Innerebner et al. 2011). Apart from this epitope conservation, we observed variation in the rest of the flagellin sequence of commensals and weak pathogens (supplementary table S3, Supplementary Material online). The importance of such sequence diversification in commensals and multiple copies of flagellin in evading MTI needs to be further explored. MTI suppression by T3SEs has been demonstrated in many pathogenic xanthomonads. Analysis in this study showed that weak pathogens could also cross the MTI barrier due to the presence of a larger set of T3SEs (7 to 12) and intact T3SS. Whether a minimized ancestral T3SE repertoire of commensal xanthomonads is sufficient to overcome the MTI barrier needs to be explored further as commensals lack T3SS, and thus, secretion and translocation of these T3SEs might be of question. However, Merda et al. (2017) indicated the possibility of the secretion of T3SEs, specifically *xopR* and *avrBs2*, mediated by the flagellar apparatus (Journet et al. 2005). These T3SEs may also have functions independent of T3SS. Simultaneous colonization of commensals and pathogenic xanthomonads may also mean that commensal xanthomonads coordinate the action of these T3SEs and share these T3SEs as a public good with the pathogenic members, similar to the phenomenon demonstrated with studies using effectorless *Pseudomonas* strains (Ruiz-Bedoya et al. 2023). Experiments evaluating the effect of coinfection on their collective virulence may further our understanding of the importance of reduced effector repertoires in commensal xanthomonads (Sadhukhan et al. 2024). These T3SEs in commensal strains may also have a role outside the host, similar to that shown for AvrBs1,

IS476, and the associated plasmid offering enhanced overwintering potential (O'Garro et al. 1997). Further functional assessment of MTI-inducing and MTI-suppressing abilities of the commensals and weak pathogens from diverse clades across the phylogeny may clarify whether they can actively or passively cross the MTI barrier and how they may contribute toward microbiota–host homeostasis and ultimately toward plant growth–defense tradeoff and plant fitness (Ma et al. 2021). Although the opportunistic nature of nonpathogens has been documented (Gitaitis et al. 1987), the contribution of T3SS and an associated minimized effector repertoire or distinct CWDE repertoire toward such conditional pathogenicity has not been experimentally validated. Alternatively, conditional pathogenicity may result from altered host–microbiota homeostasis and a compromised immune response rather than the involvement of T3SS or CWDEs alone. This important question of the contribution of nonpathogenic xanthomonads as a member of the microbiota has been investigated by two independent studies that emphasized the role of T2SS and CWDEs in mediating the shift in microbiota and conditional pathogenicity (Entila et al. 2024; Pfeilmeier et al. 2024). Further, Entila et al. (2024) showed that conditional pathogenicity of a nonpathogenic xanthomonad strain, lacking *hrp2* cluster, is kept in check by suppression of CWDEs by plant NADPH oxidase respiratory burst oxidase homolog D (RBOHD) through a negative feedback loop between DAMP-triggered immunity-led reactive oxygen species production by *Arabidopsis* and T2SS/CWDEs.

In summary, this study highlights the diversity of lifestyles across the genus phylogeny along the continuum of commensal, weakly pathogenic, and pathogenic strains. We find that T3SS and T3SEs are not the only factors that define these lifestyles. Several niche adaptative factors were identified to be associated with each lifestyle. Commensals establish themselves on different hosts in the presence of various host defenses and competing microflora, as well as derive complex nutrients from a wide range of hosts while sustaining their populations on a broad range of hosts. We also observed distinct CWDE repertoires that distinguish pathogenic versus commensal or weakly pathogenic lifestyles. Conversely, pathogens rely on T3SS and associated T3SEs to subvert defense responses. In the absence of T3SS, commensal bacteria typically carry genes that might enable them to endure environmental stresses rather than actively evade them.

Materials and Methods

Bacterial Strain Collection and Genome Sequencing

Nonpathogenic *Xanthomonas* strains collected from different plant hosts and environmental samples (supplementary table S1, Supplementary Material online) were used for

genomic DNA extraction using the CTAB–NaCl method (William et al. 2012). Degradation and contamination of the genomic DNA were monitored on 0.5% agarose gels. DNA concentration was measured using a Qubit DNA Assay Kit on a Qubit 2.0 Fluorometer (Life Technologies, CA, USA) and submitted to the Joint Genome Institute (JGI) for library preparation and sequencing. Paired-end reads were generated by multiplexing 12 libraries in a single lane on the Illumina NovaSeq (PE150) platform. The raw reads were then trimmed for their quality per JGI standard operating practice (SOP) protocol using BBTtools (v38.86; <http://bbtools.jgi.doe.gov>). Filtered reads were assembled into contigs using SPAdes (v3.13.0; Prjibelski et al. 2020) with 25, 55, and 95 k-mers. The quality of the final assembly was determined by using tRNAscan-SE (Chan and Lowe 2019) to count tRNAs and checkM (Parks et al. 2015) to determine completeness and contamination. Prodigal (v2.6.3; Hyatt et al. 2010) was used to predict CDS on each scaffold. Raw reads, annotation data, and final assembly are in the JGI data portal (<http://genome.jgi.doe.gov>). The information for 83 newly sequenced genomes from this study is in [supplementary table S1, Supplementary Material](#) online. The majority of these strains were further tested for their pathogenicity on the host of isolation according to Koch's postulates and the observations describing symptoms are noted in [supplementary table S1, Supplementary Material](#) online.

Genome-Based Identification of *Xanthomonas* Strains

Genome-based identification was performed among 134 *Xanthomonas* strains, including 83 strains from this study and 51 representative *Xanthomonas* strains from NCBI ([supplementary table S2, Supplementary Material](#) online), including at least one genome sequence of each type strain for all current *Xanthomonas* species, and representatives of some species were compared with the genomes of the strains sequenced in this study ([supplementary table S2, Supplementary Material](#) online). ANI was estimated using all-versus-all strategies using FastANI (v1.1; Jain et al. 2018) and pyani (v0.2.12; Pritchard et al. 2016). We also used the ANI values from the web server LINbase (Tian et al. 2020) and MiSI (Varghese et al. 2015) as additional tools in species circumscription. The MiSI method addresses inconsistencies based on ANI alone and includes alignment fractions and genome-wide ANI values. Additionally, 20 strains representing the novel species diversity were subject to the Type (Strain) Genome Server (<https://tygs.dsmz.de/>) to calculate dDDH values based on the Genome BLAST Distance Phylogeny (Meier-Kolthoff et al. 2013). A combination of ANI, dDDH, and MiSI was used to designate the "new species" status to a given strain only when values were below the accepted threshold ($\leq 95\%$ for ANI and $\leq 70\%$ for dDDH; Kim et al. 2014).

The whole proteome of the 134 *Xanthomonas* strains was compared by OrthoFinder (v2.5.2; Emms and Kelly 2019) to identify orthogroups using the original algorithm (Emms and Kelly 2015). The identified orthogroups were used to infer unrooted gene trees using the BLAST-based hierarchical clustering algorithm DendroBLAST (Kelly and Main 2013). Using this set of unrooted gene trees, the STAG algorithm identified the closest pair of genes from those species to infer an unrooted species tree (Emms and Kelly 2018). The unrooted species tree inferred from STAG was then rooted using the STRIDE algorithm (Emms and Kelly 2017) by identifying well-supported gene duplication events. The resulting cladogram was visualized with R package *ggtree* (Yu et al. 2017).

To identify and visualize possible conflicting signals that would suggest recombination events and evolutionary relationships within the *Xanthomonas* sequence data, multilocus sequence analysis (MLSA) was carried out for 12 housekeeping gene fragments (*gyrB*, *gapA*, *lacF*, *gltA*, *fusA*, *lepA*, *atpD*, *rpoD*, *glnA*, *efp*, *dnaK*, and *fyuA*) using autoMLSA2 (v0.7.1; <https://github.com/davised/automlsa2>). The 12 housekeeping genes have previously been used for MLST schemes by Almeida et al. (2010) and Fischer-Le Saux et al. (2015). The query sequences for these 12 genes were retrieved from Plant Associated and Environmental Microbes Database (PAMDB website: http://genome.ppws.vt.edu/cgi-bin/MLST/search_alleles.pl). The resultant splits.nex file was used in SplitsTree4 (v4.17.0; Huson et al. 2008) to develop a phylogenetic network. Neighbor-net method was used to perform split decomposition analysis. The possibility of recombination events was identified by the branches that form parallelograms (Joseph and Forsythe 2012).

Analysis of the Gain and Loss Dynamics of the T3SS Clusters

Protein sequences of T3SEs representing all effector families, putative T3SEs ([supplementary table S4, Supplementary Material](#) online), and their diversity were also identified in genome sequences using tBLASTn searches. The T3SS-coding genes from five different *Xanthomonas* species (*X. campestris* pv. *vesicatoria* 85-10, *X. campestris* pv. *campestris* ATCC33913, *X. translucens* pv. *translucens* DSM18974, *X. albilineans* CFBP 2523, and *Xanthomonas* sp. 60) were used as query to perform BLASTn searches on *Xanthomonas* strain genomes using autoMLSA2 (v0.7.1) with the cutoffs set to 40% identity and 30% coverage. Heatmaps for the blast searches were generated using the R package *Pheatmap* (v1.0.12).

Branch-specific gain and loss probabilities of *Xanthomonas* T3SS genes during the evolution were inferred with the species tree and presence/absence in the 134 genomes using GLOOME (Cohen et al. 2010).

GLOOME analyzes presence and absence profiles (phyletic patterns) and accurately infers branch-specific and site-specific gain and loss events. We first inferred the gain and loss dynamics of all genes encoding components of the T3SS. Specifically, we searched the genes of four T3SS clusters: (i) *hrp2* cluster derived from *X. campestris* pv. *campestris* ATCC33913 (*hrp2* cluster, 22 genes); (ii) *X. translucens* pv. *translucens* DSM18974 (*Xtra* cluster, 18 genes); (iii) *X. albilineans* CFBP 2523 (SPI-1 type cluster, 11 genes); and (iv) *Xanthomonas* sp. 60 (*sct*-type cluster from this study, 11 genes). We additionally searched for the presence and absence of genes encoding transcription factors involved in T3SS-related pathogenicity: HrpG, HrpX, HpaR, and HpaS (four genes). To identify which T3SS cluster is encoded in each of the 134 genomes, a sequence similarity search was conducted with each of these 66 genes as a query using BLASTp. A hit was considered if the identity percentage was at least 50%, the *E*-value was lower than 10^{-10} , and the coverage was at least 30%. When a hit was detected to two or more genes from different T3SS clusters, the one with the highest bit score was retained. As the phylogenetic tree for the analysis, we used the species tree generated by OrthoFinder. The final GLOOME analysis was performed with default parameter values. Graphical visualization of the tree was done using FigTree (v1.4.4; <http://tree.bio.ed.ac.uk/software/figtree/>).

Prediction of MGEs

To study the genomic differences driven by MGEs in the genus *Xanthomonas*, we used Mobile Genetic Element Finder (MGEfinder, v1.0.6; Durrant et al. 2020). MGEfinder assembles the short reads and aligns them to a reference genome to find insertions (Durrant et al. 2020). *Xanthomonas* reads from this study and representative short reads were downloaded from NCBI's SRA database and trimmed using Trim Galore (v0.6.6; <https://github.com/FelixKrueger/TrimGalore>). Reference genomes were indexed, and the cleaned reads were aligned with BWA-MEM (v0.7.17; Li and Durbin 2009). Target strains were assigned to pathogenic reference genomes according to their phylogenetic placement, generating eight clusters (supplementary table S2, Supplementary Material online). Each cluster contains a representative/type strain and the strains from this study. The predicted mobile genetic islands were annotated using a consensus from BLASTx results in NCBI, JGI, and UniProt. Islands containing carrier genes of potential interest were further analyzed using JGI BLASTp and gene neighborhood viewer to locate transposases or phage-related genes associated with or flanking the island. Additionally, EasyFig (v2.2.2; Sullivan et al. 2011) was used to visualize the insertion location of MGEs within genomes.

Comparison of Secreted Carbohydrate-Active Enzymes

We screened the genomes of commensal, weakly pathogenic, and pathogenic *Xanthomonas* strains for the presence of various genes involved in breakdowns (CEs, PLs, and GHs) and assembly (GTs) of carbohydrates, lignin degradation (AAs), and the carbohydrate-binding module (CBM; Lairson et al. 2008; Kaoutari et al. 2013). Carbohydrate-active enzymes (CAZymes) were assigned to Prokka (v1.14.5; Seeman 2014) protein output files (.faa files) using run_dbcan command (https://github.com/linnabrown/run_dbcan) against the HMMER, DIAMOND, and eCAMI databases with default settings. Final CAZyme domain annotations were the best hits based on the outputs of at least two databases to investigate the genomic potential of various species for carbohydrate utilization. To assess the impacts of different lifestyles on the secreted CAZyme count while considering phylogenetic signals, pairwise phylogenetic distances were created between the genomes using the function tree.distance() from the package biopython phylo (Cock et al. 2009), which was then used to build a PCA. The CAZyme count for each genome from the run_dbcan step was then converted into a distance matrix with the function vegdist (method="jaccard") from the Vegan (v2.6-4) R package (Dixon 2003; Oksanen et al. 2009). PERMANOVA was performed to determine the effect of phylogeny and microbial lifestyle on the distribution of CAZymes with the function adonis2 from the Vegan R package. Pairwise comparisons between the lifestyles were carried out using the function pairwise.perm.manova (from R package RVAidememoire) to understand the difference between lifestyles in terms of their genome content as described in Miyauchi et al. (2020).

Identification of Lifestyle-Associated Genes

We retrieved 1,834 *Xanthomonas* genomes from the NCBI GenBank RefSeq database to ensure a high-quality and minimally biased set of genomes. These genomes were then dereplicated using dRep (v3.2.2; Olm et al. 2017) to dereplicate the complete data set using a 95% minimum genome completeness cutoff and 5% maximum contamination. The ANI threshold to form primary clusters (-pa) was set at 0.95 (species level) and 0.99 (strain level) for the secondary cluster. During secondary comparisons, a minimum level of overlap between genomes was set to 80% coverage. The dereplicated genomes were manually curated to include the 83 new genomes generated in this work and representative genomes from diverse *Xanthomonas* species belonging to groups 1 and 2. A total of 337 genomes were selected for downstream analysis (supplementary table S3, Supplementary Material online). OrthoFinder was used as a clustering approach to compare the whole proteome of the 337 *Xanthomonas* strains with the default settings as previously described. To determine significantly enriched or

depleted protein clusters in different *Xanthomonas* lifestyles, we used the hypergeometric test, PhyloGLM, and Scoary as described in Levy et al. (2018). Among the three methods, the hypergeometric test looks for the overall enrichment of genes without considering the data set's phylogenetic structure. PhyloGLM is a phylogenetic-aware method that eliminates enrichments related to shared ancestry (Ives and Garland 2010), while Scoary combines the phylogeny-aware test, Fisher's exact test, and empirical label switching permutation analysis (Brynildsrud et al. 2016). All these approaches were used on gene presence/absence and gene copy number data and used for PhyloGLM test. A gene was considered significant (i) if it had a $q < 0.05$ for Fisher's exact test and an empirical $P < 0.05$ for Scoary; (ii) if it had a corrected P -value with False Discovery Rate with $q < 0.01$ for hypergeometric test; and (iii) a $P < 0.01$ along with an estimate of < -1.5 or > 1.5 in copy number analysis for PhyloGLM. We used eggNOG-mapper (v2.1.7; Cantalapiedra et al. 2021) to address COG categories to each significantly enriched and depleted protein cluster from the combination of two or more methods (hypergeometric, PhyloGLM, and Scoary) in the different *Xanthomonas* lifestyles. In addition, each ortholog ID was queried across the IMG database to obtain annotation based on COG, KO, TIGRFAM, and Pfam. Heatmaps were generated using the R package ggplot2.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

The work (proposal: 10.46936/10.25585/60001156) conducted by the US Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy operated under Contract No. DE-AC02-05CH11231. We thank CIRM-CFBP (Beaucouzé, INRAE, France, http://www6.inra.fr/cirm_eng/CFBP -Plant-Associated-Bacteria) for strain preservation and supply. We also thank Abi S. A. Marques and Marisa A. S. V. Ferreira for participating in the strain sampling from bean. This work was made possible in part by a grant of high-performance computing resources and technical support from the Alabama Supercomputer Authority.

Author Contributions

The study was conceptualized and designed by N.P., M.-A.J., B.A.V., and J.B.J. Strains were collected from various sources by N.P., B.A.V., J.B.J., and M.-A.J. R.B. and E.N. processed the samples and conducted greenhouse experiments. T.W. assisted with sequencing and data availability

from JGI. The data were analyzed by R.B., M.M.P., R.M.B., K.W., N.W., T.P., and N.P. Specifically, M.M.P. conducted OrthoFinder analysis, taxonomic placement, pathogenicity assay, T3SS, and T3SE analysis and wrote the initial draft of the manuscript. N.W. and T.P. conducted GLOOME analysis. R.B. analyzed CWDEs. K.W. performed MGEfinder analysis. N.P. carried out association analysis. R.M.B. assisted with constructing heatmaps for association analysis. R.B. formatted figures and tables. The final manuscript was written by R.B. and N.P., with input from all the other coauthors.

Conflict of Interest

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

Data Availability

Sequence data generated from this work, i.e. assembled genomes and their associated annotations, have been deposited in the IMG-JGI and GenBank database under the JGI taxon ID and BioProject accession number as given in [supplementary table S1, Supplementary Material](#) online. This project belongs to the Community Science Program of Joint Genome Institute. Thus, all data generated including raw reads, assembled genomes, and annotations have been archived on the IMG-JGI website as well as on NCBI-SRA and NCBI-RefSeq. Essential metadata including host of isolation, pathogenicity data, year of isolation, and geographical location for isolation have been included in [supplementary table S1, Supplementary Material](#) online, and available on IMG-GOLD database. The strains used in this study are available to the scientific community from the respective author's collections as mentioned in [supplementary table S1, Supplementary Material](#) online. The data that support the findings of this study are available in the [Supplementary Material](#) of this article. In addition, the scripts, input files, intermediate files, and output files used for interpretation of the analyses conducted in this work are made available on <https://github.com/Potnislav/commensal> as well as on DOI: [10.5281/zenodo.8000197](https://doi.org/10.5281/zenodo.8000197) (Pena et al. 2023).

Literature Cited

- Aktas M, Narberhaus F. Unconventional membrane lipid biosynthesis in *Xanthomonas campestris*. *Environ Microbiol*. 2015;17(9): 3116–3124. <https://doi.org/10.1111/1462-2920.12956>.
- Alavi SM, Sanjari S, Durand F, Brin C, Manceau C, Poussier S. Assessment of the genetic diversity of *Xanthomonas axonopodis* pv. *phaseoli* and *Xanthomonas fuscans* subsp. *fuscans* as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive

- hybridizations. *Appl Environ Microbiol.* 2008;74(10):3295–3301. <https://doi.org/10.1128/AEM.02507-07>
- Almeida NF, Yan S, Cai R, Clarke CR, Morris CE, Schaad NW, Schuenzel EL, Lacy GH, Sun X, Jones JB, et al. PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. *Phytopathology* 2010;100(3):208–215. <https://doi.org/10.1094/PHYTO-100-3-0208>.
- Bansal K, Kaur A, Midha S, Kumar S, Korpole S, Patil PB. *Xanthomonas sontoii* sp. nov., a non-pathogenic bacterium isolated from healthy basmati rice (*Oryza sativa*) seeds from India. *Antonie van Leeuwenhoek.* 2021;114(11):1935–1947. <https://doi.org/10.1007/s10482-021-01652-1>.
- Bansal K, Kumar S, Patil PB. Phylogenomic insights into diversity and evolution of nonpathogenic *Xanthomonas* strains associated with citrus. *mSphere* 2020;5(2):e00087-20. <https://doi.org/10.1128/mSphere.00087-20>.
- Berendsen RL, Vismans G, Yu K, Song Y, de Jonge R, Burgman WP, Burmølle M, Herschend J, Bakker PAHM, Pieterse CMJ. Disease-induced assemblage of a plant-beneficial bacterial consortium. *ISME J.* 2018;12(6):1496–1507. <https://doi.org/10.1038/s41396-018-0093-1>.
- Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, Vasse J, Lauber E, Arlat M. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS One* 2007;2(2):e224. <https://doi.org/10.1371/journal.pone.0000224>.
- Brouwer MSM, Roberts AP, Hussain H, Williams RJ, Allan E, Mullany P. Horizontal gene transfer converts non-toxicogenic *Clostridium difficile* strains into toxin producers. *Nat Commun.* 2013;4(1):2601. <https://doi.org/10.1038/ncomms3601>.
- Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 2016;17(1):238. <https://doi.org/10.1186/s13059-015-0866-z>.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38(12):5825–5829. <https://doi.org/10.1093/molbev/msab293>.
- Cellier MF, Bergevin I, Boyer E, Richer E. Polyphyletic origins of bacterial Nramp transporters. *Trends Genet.* 2001;17(7):365–370. [https://doi.org/10.1016/S0168-9525\(01\)02364-2](https://doi.org/10.1016/S0168-9525(01)02364-2).
- Cesbron S, Briand M, Essakhi S, Gironde S, Boureau T, Manceau C, Fischer-Le Saux M, Jacques M-A. Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front Plant Sci.* 2015;6:1126. <https://doi.org/10.3389/fpls.2015.01126>.
- Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol.* 2019;1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.
- Choi J-E, Nguyen CM, Lee B, Park JH, Oh JY, Choi JS, Kim J-C, Song JK. Isolation and characterization of a novel metagenomic enzyme capable of degrading bacterial phytotoxin toxoflavin. *PLoS One* 2018;13(1):e0183893. <https://doi.org/10.1371/journal.pone.0183893>.
- Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England).* 2009;25(11):1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
- Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. GLOOME: gain loss mapping engine. *Bioinformatics (Oxford, England).* 2010;26(22):2914–2915. <https://doi.org/10.1093/bioinformatics/btq549>.
- Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14(6):927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.
- Drebes Dörr NC, Blokesch M. Bacterial type VI secretion system facilitates niche domination. *Proc Natl Acad Sci U S A.* 2018;115(36):8855–8857. <https://doi.org/10.1073/pnas.1812776115>.
- Durrant MG, Li MM, Siranosian BA, Montgomer SB, Bhat AS. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe.* 2020;27(1):140–153.e9. <https://doi.org/10.1016/j.chom.2019.10.022>.
- Emms DM, Kelly S. STAG: species tree inference from all genes. *BioRxiv* 267914. <https://doi.org/10.1101/267914>, 19 February 2018, preprint: not peer reviewed.
- Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 2015;16(1):1–14. <https://doi.org/10.1186/s13059-015-0721-2>.
- Emms DM, Kelly S. STRIDE: species tree root inference from gene duplication events. *Mol Biol Evol.* 2017;34(12):3267–3278. <https://doi.org/10.1093/molbev/msx259>.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):1–14. <https://doi.org/10.1186/s13059-019-1832-y>.
- Entila F, Han X, Mine A, Schulze-Lefert P, Tsuda K. Commensal lifestyle regulated by a negative feedback loop between Arabidopsis ROS and the bacterial T2SS. *Nat Commun.* 2024;15(1):456. doi: 10.1038/s41467-024-44724-2.
- Essakhi S, Cesbron S, Fischer-Le Saux M, Bonneau S, Jacques M-A, Manceau C. Phylogenetic and variable-number tandem-repeat analyses identify nonpathogenic *Xanthomonas arboricola* lineages lacking the canonical type III secretion system. *Appl Environ Microbiol.* 2015;81(16):5395–5410. <https://doi.org/10.1128/AEM.00835-15>.
- Failor KC, Schmale DG, Vinatzer BA, Monteil CL. Ice nucleation active bacteria in precipitation are genetically diverse and nucleate ice by employing different mechanisms. *ISME J.* 2017;11(12):2740–2753. <https://doi.org/10.1038/ismej.2017.124>.
- Fang Y, Lin H, Wu L, Ren D, Ye W, Dong G, Zhu L, Guo L. Genome sequence of *Xanthomonas sacchari* R1, a biocontrol bacterium isolated from the rice seed. *J Biotechnol.* 2015;206:77–78. <https://doi.org/10.1016/j.jbiotec.2015.04.014>.
- Federici L, Caprari C, Mattei B, Savino C, Di Matteo A, De Lorenzo G, Cervone F, Tsernoglou D. Structural requirements of endopolygalacturonase for the interaction with PGIP (polygalacturonase-inhibiting protein). *Proc Natl Acad Sci U S A.* 2001;98(23):13425–13430. <https://doi.org/10.1073/pnas.231473698>.
- Fischer-Le Saux M, Bonneau S, Essakhi S, Manceau C, Jacques M-A. Aggressive emerging pathovars of *Xanthomonas arboricola* represent widespread epidemic clones distinct from poorly pathogenic strains, as revealed by multilocus sequence typing. *Appl Environ Microbiol.* 2015;81(14):4651–4668. <https://doi.org/10.1128/AEM.00050-15>.
- Garita-Cambrero J, Palacio-Bielsa A, López MM, Cubero J. Pan-genomic analysis permits differentiation of virulent and non-virulent strains of *Xanthomonas arboricola* that cohabit *Prunus* spp. and elucidate bacterial virulence factors. *Front Microbiol.* 2017;8:573. <https://doi.org/10.3389/fmicb.2017.00573>.
- Gitaitis RD. Pectolytic xanthomonads in mixed infections with *Pseudomonas syringae* pv. *syringae*, *P. syringae* pv. *tomato*, and *Xanthomonas campestris* pv. *vesicatoria* in tomato and pepper

- transplants. *Phytopathology* 1987;77(4):611–615. <https://doi.org/10.1094/Phyto-77-611>.
- Gitaitis RD, Sasser M, Stall RE. Opportunistic xanthomonads in mixed infections with *Pseudomonas syringae* in tomato and pepper. In *Plant Pathogenic Bacteria: Proceedings of the Sixth International Conference on Plant Pathogenic Bacteria, Maryland, 1985 June 2–7*; (pp. 767–767). Springer, Netherlands; 1987.
- Gonzalez C, Restrepo S, Tohme J, Verdier V. Characterization of pathogenic and nonpathogenic strains of *Xanthomonas axonopodis* pv. *manihotis* by PCR-based DNA fingerprinting techniques. *FEMS Microbiol Lett.* 2002;215(1):23–31. <https://doi.org/10.1111/j.1574-6968.2002.tb11365.x>.
- Hacquard S, Spaepen S, Garrido-Oter R, Schulze-Lefert P. Interplay between innate immunity and the plant microbiota. *Annu Rev Phytopathol.* 2017;55(1):565–589. <https://doi.org/10.1146/annurev-phyto-080516-035623>.
- Hajri A, Brin C, Hunault G, Lardeux F, Lemaire C, Manceau C, Boureau T, Poussier S. A “Repertoire for repertoire” hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One* 2009;4(8):e6632. <https://doi.org/10.1371/journal.pone.0006632>.
- Huson DH, Klopper T, Bryant D. SplitsTree 4.0-computation of phylogenetic trees and networks. *Bioinformatics* 2008;14(1):68–73. <https://doi.org/10.1093/bioinformatics/14.1.68>.
- Hyatt D, Chen G-L, Locascio P, Land M, Larimer F, Hauser L. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- Innerebner G, Knief C, Vorholt JA. Protection of *Arabidopsis thaliana* against leaf-pathogenic *Pseudomonas syringae* by *Sphingomonas* strains in a controlled model system. *Appl Environ Microbiol.* 2011;77(10):3202–3210. <https://doi.org/10.1128/AEM.00133-11>.
- Ives AR, Garland T Jr. Phylogenetic logistic regression for binary dependent variables. *Syst Biol.* 2010;59(1):9–26. <https://doi.org/10.1093/sysbio/syp074>.
- Jacobs JM, Pesce C, Lefevre P, Koebnik R. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front Plant Sci.* 2015;6:431. <https://doi.org/10.3389/fpls.2015.00431>.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018;9(1):5114. <https://doi.org/10.1038/s41467-018-07641-9>.
- Joseph S, Forsythe S. Insights into the emergent bacterial pathogen *Cronobacter* spp., generated by multilocus sequence typing and analysis. *Front Microbiol.* 2012;3:397. <https://doi.org/10.3389/fmicb.2012.00397>.
- Journet L, Hughes KT, Cornelis GR. Type III secretion: a secretory pathway serving both motility and virulence. *Mol Membr Biol.* 2005;22(1-2):41–50. <https://doi.org/10.1080/09687860500041858>.
- Kaoutari AE, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol.* 2013;11(7):497–504. <https://doi.org/10.1038/nrmicro3050>.
- Kehres DG, Zaharik ML, Finlay BB, Maguire ME. The NRAMP proteins of *Salmonella typhimurium* and *Escherichia coli* are selective manganese transporters involved in the response to reactive oxygen. *Mol Microbiol.* 2000;36(5):1085–1100. <https://doi.org/10.1046/j.1365-2958.2000.01922.x>.
- Kelly S, Main PK. DendroBLAST: approximate phylogenetic trees in the absence of multiple sequence alignments. *PLoS One* 2013;8(3):e58537. <https://doi.org/10.1371/journal.pone.0058537>.
- Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol.* 2014;64(Pt_2):346–351. <https://doi.org/10.1099/ijs.0.059774-0>.
- Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* 2008;77(1):521–555. <https://doi.org/10.1146/annurev.biochem.76.061005.092322>.
- Lee YA, Yang PY, Huang SC. Characterization, phylogeny, and genome analyses of nonpathogenic *Xanthomonas campestris* strains isolated from brassica seeds. *Phytopathology.* 2020;110(5):981–988. <https://doi.org/10.1094/PHTYO-08-19-0319-R>.
- Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, et al. Genomic features of bacterial adaptation to plants. *Nat Genet.* 2018;50(1):138–150. <https://doi.org/10.1038/s41588-017-0012-9>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England).* 2009;25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li T, Mann R, Sawbridge T, Kaur J, Auer D, Spangenberg G. Novel *Xanthomonas* species from the perennial ryegrass seed microbiome—assessing the bioprotection activity of non-pathogenic relatives of pathogens. *Front Microbiol.* 2020;11:1991. <https://doi.org/10.3389/fmicb.2020.01991>.
- Liyanapathirana P, Wagner N, Avram O, Pupko T, Potnis N. Phylogenetic distribution and evolution of type VI secretion system in the genus *Xanthomonas*. *Front Microbiol.* 2022;13:840308. <https://doi.org/10.3389/fmicb.2022.840308>.
- Lu H, Patil P, Van Sluys M-A, White FF, Ryan RP, Dow JM, Rabinowicz P, Salzberg SL, Leach JE, Sonti R, et al. Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PLoS One* 2008;3(11):e3828. <https://doi.org/10.1371/journal.pone.0003828>.
- Ma K-W, Niu Y, Jia Y, Ordon J, Copeland C, Emonet A, Geldner N, Guan R, Stolze SC, Nakagami H, et al. Coordination of microbe-host homeostasis by crosstalk with plant innate immunity. *Nat Plants.* 2021;7(6):814–825. <https://doi.org/10.1038/s41477-021-00920-2>.
- Mafakheri H, Taghavi SM, Zarei S, Rahimi T, Hasannezhad MS, Portier P, Fischer-Le Saux M, Dimkić I, Koebnik R, Kuzmanović N, et al. Phenotypic and molecular-phylogenetic analyses reveal distinct features of crown gall-associated *Xanthomonas* strains. *Microbiol Spectr.* 2022;10(1):e0057721. <https://doi.org/10.1128/spectrum.00577-21>.
- Malvino ML, Bott AJ, Green CE, Majumdar T, Hind SR. Influence of flagellin polymorphisms, gene regulation, and responsive memory on the motility of *Xanthomonas* species that cause bacterial spot disease of solanaceous plants. *Mol Plant Microbe Interact.* 2022;35(2):157–169. <https://doi.org/10.1094/MPMI-08-21-0211-R>.
- Martins L, Fernandes C, Blom J, Dia NC, Pothier JF, Tavares F. *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut. *Int J Syst Evol Microbiol.* 2020;70(12):6024–6031. <https://doi.org/10.1099/ijsem.0.004386>.
- Mechan Llontop ME, Tian L, Sharma P, Heflin L, Bernal-Galeano V, Haak DC, Clarke CR, Vinatzer BA. Experimental evidence pointing to rain as a reservoir of tomato phyllosphere microbiota. *Phytobiomes J.* 2021;5(4):382–399. <https://doi.org/10.1094/PBIOMES-04-21-0025-R>.

- Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* 2013;14(1):1–14. <https://doi.org/10.1186/1471-2105-14-60>.
- Meline V, Delage W, Brin C, Li-Marchetti C, Sochard D, Arlat M, Rousseau C, Darrasse A, Briand M, Lebreton G, et al. Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of *Xanthomonas*. *Mol Plant Pathol.* 2019;20(1):33–50. <https://doi.org/10.1111/mpp.12737>.
- Merda D, Bonneau S, Guimbaud J-F, Durand K, Brin C, Boureau T, Lemaire C, Jacques M-A, Fischer-Le Saux M. Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems. *Environ Microbiol Rep.* 2016;8(5):572–581. <https://doi.org/10.1111/1758-2229.12397>.
- Merda D, Briand M, Bosis E, Rousseau C, Portier P, Barret M, Jacques M-A, Fischer-Le Saux M. Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in *Xanthomonas* plant pathogens. *Mol Ecol.* 2017;26(21):5939–5952. <https://doi.org/10.1111/mec.14343>.
- Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G, et al. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat Commun.* 2020;11(1):5125. <https://doi.org/10.1038/s41467-020-18795-w>.
- O'Garro LW, Gibbs H, Newton A. Mutation in the *avrBs1* avirulence gene of *Xanthomonas campestris* pv. *vesicatoria* influences survival of the bacterium in soil and detached leaf tissue. *Phytopathology* 1997;87(9):960–966. <https://doi.org/10.1094/PHYTO.1997.87.9.960>.
- Oksanen J, Kindt R, Legendre P, O'hara B, Simpson GL, Solymos P, Stevens MHH, Wagner H. *vegan*: community ecology package. R package version 1.15-4. Vienna, Austria: R foundation for statistical computing; 2009.
- Olm MR, Brown CT, Brooks B, Banfield JF. Drep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 2017;11(12):2864. <https://doi.org/10.1038/ismej.2017.126>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Pena M, Bhandari R, Bowers R, Weis K, Newberry E, Wagner N, Pupko T, Jones JB, Woyke T, Vinatzer B, et al. Dataset for “Genetic and functional diversity help explain pathogenic, weakly pathogenic, and commensal lifestyles in the genus *Xanthomonas*” Commensal; Zenodo; 2023; [dataset] <https://doi.org/10.5281/zenodo.8000197>.
- Pesce C, Jacobs JM, Berthelot E, Perret M, Vancheva T, Bragard C, Koebnik R. Comparative genomics identifies a novel conserved protein, HpaT, in proteobacterial type III secretion systems that do not possess the putative translocon protein HrpF. *Front Microbiol.* 2017;8:1177. <https://doi.org/10.3389/fmicb.2017.01177>.
- Pfeilmeier S, Petti GC, Bortfeld-Miller M, Daniel B, Field CM, Sunagawa S, Vorholt JA. The plant NADPH oxidase RBOHD is required for microbiota homeostasis in leaves. *Nat Microbiol.* 2021;6(7):852–864. <https://doi.org/10.1038/s41564-021-00929-5>.
- Pfeilmeier S, Werz A, Ote M, Bortfeld-Miller M, Kirner P, Keppler A, Hemmerle L, Gäbelein CG, Pestalozzi CM, Vorholt JA. Leaf microbiome dysbiosis triggered by T2SS-dependent enzyme secretion from opportunistic *Xanthomonas* pathogens. *Nat Microbiol.* 2024;9:136–149.
- Pieretti I, Pestic A, Petras D, Royer M, Süßmuth RD, Cociancich S. What makes *Xanthomonas albilineans* unique amongst xanthomonads? *Front Plant Sci.* 2015;6:289. <https://doi.org/10.3389/fpls.2015.00289>.
- Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques M-A, et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genom.* 2009;10(1):1–15. <https://doi.org/10.1186/1471-2164-10-616>.
- Potnis N, Minsavage G, Smith JK, Hurlbert JC, Norman D, Rodrigues R, Stall RE, Jones JB. Avirulence proteins AvrBs7 from *Xanthomonas gardneri* and AvrBs1.1 from *Xanthomonas euvesicatoria* contribute to a novel gene-for-gene interaction in pepper. *Mol Plant Microbe Interact.* 2012;25(3):307–320. <https://doi.org/10.1094/MPMI-08-11-0205>.
- Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2016;8(1):12–24. <https://doi.org/10.1039/C5AY02550H>.
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics.* 2020;70:e102. <https://doi.org/10.1002/cpb1.102>.
- Rodriguez-R LM, Grajales A, Arrieta-Ortiz ML, Salazar C, Restrepo S, Bernal A. Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiol.* 2012;12(1):1–14. <https://doi.org/10.1186/1471-2180-12-43>.
- Ruiz-Bedoya T, Wang PW, Desveaux D, Guttman DS. Cooperative virulence via the collective action of secreted pathogen effectors. *Nat Microbiol.* 2023;8(4):640–650. <https://doi.org/10.1038/s41564-022-01276-9>.
- Sadhukhan S, Jacques MA, Potnis N. Influence of co-occurring weakly pathogenic bacterial species on bacterial spot disease dynamics on tomato. *Plant Dis.* 2024;108(1):190–199. <https://doi.org/10.1094/PDIS-05-23-0837-RE>.
- Seeman T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30(14):2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Sirias D, Utter DR, Gibbs KA. A family of contact-dependent nuclease effectors contain an exchangeable, species-identifying domain. *bioRxiv* 2020-02. <https://doi.org/10.1101/2020.02.20.956912>, 20 February 2020, preprint: not peer reviewed.
- Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J, Grant M. Draft genome sequences of *Xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel).* 2011;2(4):1050–1065. <https://doi.org/10.3390/genes2041050>.
- Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27(7):1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>.
- Sun W, Dunning FM, Pfund C, Weingarten R, Bent F. Within-species flagellin polymorphism in *Xanthomonas campestris* pv. *campestris* and its impact on elicitation of *Arabidopsis* FLAGELLIN SENSING2-dependent defenses. *Plant Cell.* 2006;18(3):764–779. <https://doi.org/10.1105/tpc.105.037648>.
- Sychantha D, Brott AS, Jones CS, Clarke AJ. Mechanistic pathways for peptidoglycan O-acetylation and de-O-acetylation. *Front Microbiol.* 2018;9:2332. <https://doi.org/10.3389/fmicb.2018.02332>.
- Tampakaki AP, Skandalis N, Gazi AD, Bastaki MN, Panagiotis FS, Charova SN, Kokkinidis M, Panopoulos NJ. 2010. Playing the “harp”: evolution of our understanding of *hrp/hrc* genes. *Annu Rev Phytopathol.* 48(1):347–370. <https://doi.org/10.1146/annurev-phyto-073009-114407>.
- te Molder D, Poncheewin W, Schaap PJ, Koehorst JJ. Machine learning approaches to predict the plant-associated phenotype of

- Xanthomonas* strains. BMC Genom. 2021;22(1):1–14. <https://doi.org/10.1186/s12864-021-08093-0>.
- Tian L, Huang C, Mazloom R, Heath LS, Vinatzer BA. LINbase: a web server for genome-based identification of prokaryotes as members of crowdsourced taxa. Nucleic Acids Res. 2020;48(W1):W529–W537. <https://doi.org/10.1093/nar/gkaa190>.
- Timilsina S, Potnis N, Newberry EA, Liyanapathiranage P, Iruegas-Bocardo F, White FF, Goss EM, Jones JB. *Xanthomonas* diversity, virulence and plant-pathogen interactions. Nat Microbiol Rev. 2020;18(8):415–427. <https://doi.org/10.1038/s41579-020-0361-8>.
- Triplett LR, Verdier V, Campillo T, Van Malderghem C, Cleenwerck I, Maes M, Deblais L, Corral R, Koita O, Cottyn B, et al. Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. Antonie Van Leeuwenhoek. 2015;107(4):869–881. <https://doi.org/10.1007/s10482-015-0379-5>.
- Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kypides NC, Pati A. Microbial species delineation using whole genome sequences. Nucleic Acids Res. 2015;43(14):6761–6771. <https://doi.org/10.1093/nar/gkv657>.
- Vauterin L, Yang P, Alvarez A, Takikawa Y, Roth DA, Vidaver AK, Stall RE, Kersters K, Swings J. Identification of non-pathogenic *Xanthomonas* strains associated with plants. Syst Appl Microbiol. 1996;19(1):96–105. [https://doi.org/10.1016/S0723-2020\(96\)80016-6](https://doi.org/10.1016/S0723-2020(96)80016-6).
- White FF, Potnis N, Jones JB, Koebnik R. The type III effectors of *Xanthomonas*. Mol Plant Pathol. 2009;10(6):749–766. <https://doi.org/10.1111/j.1364-3703.2009.00590.x>.
- Wichmann F, Vorhölter F-J, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. Mol Plant Pathol. 2013;14(6):576–588. <https://doi.org/10.1111/mpp.12030>.
- William S, Feil H, Copeland A. Bacterial genomic DNA isolation using CTAB. Sigma 2012;50:6876.
- Yu J, Peñaloza-Vázquez A, Chakrabarty AM, Bender CL. Involvement of the exopolysaccharide alginate in the virulence and epiphytic fitness of *Pseudomonas syringae* pv. *syringae*. Mol Microbiol. 1999;33(4):712–720. <https://doi.org/10.1046/j.1365-2958.1999.01516.x>.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8(1):28–36. <https://doi.org/10.1111/2041-210X.12628>.
- Zandonadi FS, Ferreira SP, Alexandrino AV, Carnielli CM, Artier J, Barcelos MP, Nicolela NCS, Prieto EL, Goto LS, Belasque J. Periplasm-enriched fractions from *Xanthomonas citri* subsp. *citri* type A and *X. fuscans* subsp. *aurantifolia* type B present distinct proteomic profiles under in vitro pathogenicity induction. PLoS One 2020;15(12):e0243867. <https://doi.org/10.1371/journal.pone.0243867>.
- Zarei S, Taghavi SM, Osdaghi E. Reassessment of the occurrence and distribution of *Xanthomonas arboricola* on fruit trees in Iran. Plant Pathol. 2022;71(9):1859–1869. <https://doi.org/10.1111/ppa.13632>.

Associate editor: Li-Jun Ma