BMC Genomics

RESEARCH

Open Access

Unlocking the genetic arsenal of *Xanthomonas arboricola*: new insights into taxonomic classification, pathogenicity and adaptation beyond the effectorome

Renata de A. B. Assis[1,2], Alessandro M. Varani[3], Aidan C. Shands[4], Cintia H. D. Sagawa[2], José S. L. Patané[5], João C. Setubal[6], Paulo A. Zaini[2], Nalvo F. Almeida[7], Robson F. de Souza[8], Camila C. M. Garcia[1,9], Elena G. Orellano[10], James E. Adaskaveg[4], Abhaya M. Dandekar[2] and Leandro M. Moreira[1,9,11*]

Abstract

Background *Xanthomonas arboricola* (*Xar*) is a phytopathogenic bacterial species responsible for economically significant diseases in a wide range of plants, including agricultural, ornamental, and forest species. This study aimed to investigate the genomic basis of host specificity, adaptation, and virulence in *Xar* through comprehensive comparative genomics.

Results A total of 177 genomes from nine *Xar* pathovars were analyzed for evolutionary relationships and effector repertoires. From these, 30 genetically diverse genomes were selected for in-depth comparison. Core, unique, and shared genes were identified and functionally annotated, focusing on their potential roles in adaptation and pathogenicity. Nineteen of the genomes were originally misclassified and did not belong to the *Xar* species. The remaining 158 genomes clustered into three major clades: I (*Xar.* pv. *juglandis*), II (*Xar.* pv. *pruni* + *Xar.* pv. *corylina*), and III (miscellaneous *Xar*). Clades I and II exhibited high effector diversity, ranging from 38 to 54 genes, with *Xar.* pv. *corylina* harboring the most. In contrast, Clade III genomes had significantly fewer effectors, with subclade IIIa containing only 5 and IIIb up to 15. Only one TAL effector was found in nine *Xar.* pv. *corylina* strains (with no conserved RVD patterns) and in both *Xar.* pv. *guizotiae* strains (up to 31 RVDs identified). Phylogenomic and effectorome analyses revealed potential genomic islands acquired via horizontal gene transfer, encoding metal metabolism genes, type II/IV secretion systems, and DNA modification enzymes. Additionally, several gene losses were observed: 19 genomes lacked flagellar assembly genes, 15 lacked nitrate metabolism genes, and 9 lacked cellulose biosynthesis and secretion genes. In contrast, all genomes possessed a lasso peptide biosynthetic cluster, highlighting recurrent genomic rearrangements through insertions and deletions.

Conclusions This study provides a refined understanding of the genetic diversity and adaptive mechanisms in *X. arboricola*, emphasizing gene gain/loss events as central to pathovar-specific metabolic and virulence traits. These

*Correspondence:
Leandro M. Moreira
lmmorei@gmail.com

Full list of author information is available at the end of the article

findings identify novel molecular markers with potential applications in diagnostics and targeted disease control strategies. In particular, the characterization of conserved and lineage-specific effector repertoires provides a framework to inform strategies for breeding resistance through the identification of candidate targets for durable host immunity.

Keywords Horizontal gene transfer, Pathogen-host interaction, Virulence, Adaptation, Effectorome, *Xanthomonas arboricola*, Comparative genomics

Background

Xanthomonas is a genus of Gram-negative bacteria that is of significant agronomic interest. These bacteria are primarily phytopathogenic and cause extensive damage to hundreds of commercially important plants [1]. Since 2002, with the complete sequencing of the genomes of *X. citri* subsp. *citri* strain 306 and *X. campestris* pv. *campestris* strain ATCC33913 [2], thousands of other strains of different species, subspecies and pathovars, pathogenic or not, have been sequenced, thus allowing increasingly robust and detailed studies involving comparative genomics to be carried out [3–6]. From these analyses, a series of functional studies of newly identified genes and biological systems have enabled a better understanding of the molecular interactions in these pathosystems [7–10].

Comparative genomic studies are well-advanced for *Xanthomonas* species that infect citrus (*X. fuscans* and *X. citri* [11–13]), rice (*X. oryzae* [14, 15]), brassicas (*X. campestris* [16, 17]), and Solanaceae (*X. vesicatoria* [9, 18]). However, for other species like *X. arboricola* (*Xar*), commonly associated with tree crops, such studies remain in their early stages. These studies often focus on specific pathovars or the reclassification of some genomes, which limits the exploration of new genes and biological systems related to adaptation and virulence [19–23]. Consequently, our understanding of the genetic and metabolic factors influencing host specificity in *Xar* remains limited. This challenge has been exacerbated by the emergence of new diseases associated with bacteria of this species, particularly in Europe, where the incidence of new cases has steadily increased [24–27].

It is important to highlight that the classification of what we now recognize as an autonomous species is based on the proposition established by Vauterin and collaborators in 1995 [28]. In fact, records regarding *Xar* date back to the early 20th century. The first report was established by Pierce in 1901 when describing the walnut blight disease in English walnut (*Juglans regia*) in the state of California (USA) [29]. Two years later, Smith described a new pathology in Japanese plums (*Prunus salicina*) in the state of Michigan (USA) [30]. In both cases, the causal agents had been classified as belonging to the species *Pseudomonas juglandis* and *Pseudomonas pruni*, later reclassified as independent species *X. juglandis* and *X. pruni*, respectively. Finally, in 1940, Miller and collaborators isolated another pathogen from filbert plants (*Corylus avellana* and *Corylus maxima*), classifying it as *Phytomonas corylina*, later reclassified as *X. corylina* [31]. Over time, as new strains and their associated pathovars were discovered, Vauterin and collaborators proposed a taxonomic reclassification of these microorganisms in 1995 [28]. This proposal, supported by subsequent studies, defined nine pathovars based on host specificity under the umbrella of what is now recognized as the autonomous species *Xar* [28].

This classification remained until 2022, when Zarei and collaborators [23] proposed a new taxonomic reformulation based on a comparative analysis of genomes involving several strains previously considered to belong to the species *Xar*. According to these authors *Xar.* pv. *juglandis* (causal agent of walnut blight - *Juglans regia*), *Xar.* pv. *corylina* (causal agent of bacterial blight of hazelnut - *Corylus avellana*), *Xar.* pv. *pruni* (causal agent of bacterial spot of stone fruits and almonds - *Prunus* ssp.), *Xar.* pv. *celebensis* (pathogenic for banana - *Musa* spp), *Xar.* pv. *fragariae* (causal agent of strawberry bacterial leaf blight - *Fragaria* × *ananassa*), *Xar.* pv. *arracaciae* (causal agent of a bacterial disease identified in arracacha - *Arracacia xanthorrhiza*), and *Xar.* pv. *zantedeschiae* (causal agent of soft rot of arum lily - *Zantedeschia aethiopica*) are pathovars of the species *Xar*, due to the high genomic similarity. In contrast, *Xar.* pv. *guizotiae* (isolated from niger - *Guizotia abyssinica*) and *Xar.* pv. *populi* (causal agent of bark necrosis in poplar - *Populus canadensis* cv. *Regenerata*) were proposed to be members of a new species, since their genomes showed high levels of similarity to each other but were found to be distinct from the genomes of the previous group [23].

Indeed, the lack of comparative genomics studies and the complexity of establishing a precise taxonomic classification hinder efforts to conduct comprehensive genomic comparisons across different pathovars of this species. Consequently, there is a significant need for more in-depth studies to uncover the taxonomic classification, and genetic mechanisms driving the pathogenicity and adaptability of *Xar*.

The objective of this study was to explore the relationship between the effectorome, phylogenomic classification, and host specificity. Additionally, we aimed to identify specific genes responsible for cellular structures, metabolic mechanisms, and virulence induction in interactions with host plants and adaptation to the

environment. To achieve this, we focused on evaluating the genomic content of the pathogen *Xar* through comprehensive comparative genomics analysis. This investigation seeks to enhance our understanding of its pathogenic attributes and address the urgent need for more effective disease control strategies.

Methods

Genome selection for evolutionary analysis

A total of 177 genomes from *Xar*, representing nine distinct pathovars, were selected for comparative genomic analysis. In addition, 72 genomes from 24 other *Xanthomonas* species (three to eight genomes per species) were included to support broader phylogenetic inference (Table S1). All genomes were retrieved from the NCBI RefSeq database on March 7, 2024. Initial phylogenetic screening was conducted using the AMPHORA3 [32] pipeline, followed by a comprehensive pangenome analysis of *Xar* genomes using Panaroo [33] version 1.5.0. Prior to pangenome inference, quality control was performed via Multi-Dimensional Scaling (MDS) to identify and exclude potential outliers or misclassified genomes. Genome-based taxonomic classification was performed using GTDB-Tk (v2.5.2) [34] with default parameters, and assignments were evaluated based on Average Nucleotide Identity (ANI) relative to reference genomes in the GTDB database (Table S2).

A subsequent, more focused analysis was performed using only *Xar* genomes. For this, a phylogenetic tree was constructed based on 164 genomes: 158 previously validated as belonging to *Xar* species, and six additional genomes forming the intermediate clade between *Xar* and *X. gardneri* and originally annotated as *Xar* but potentially belonging to other species. These six genomes were included to assist in robust tree rooting. This phylogenetic reconstruction was based on 263 single-copy core gene families identified across the selected genomes, using a minimum identity threshold of 90%. These core genes were concatenated and aligned, and the final phylogeny was inferred using version 3 IQ-TREE [35]. Tree rooting was performed with the Minimal Ancestor Deviation (MAD) method to ensure accurate evolutionary positioning.

Effectorome analysis

A set of 85 Type III effectors from "The *Xanthomonas* Resource" website, available at www.xanthomonas.org, was analyzed by protein-nucleotide 6-frame translation (tblastn) searches against the set of 166 genomes with e-value ≤1e-50 cutoff. Additionally, given the significance of Transcription Activator-Like Effectors (TALEs) in *Xanthomonas* species, their presence was investigated in other *X. arboricola* genomes. A blastn was done using the list of TALEs available at "The *Xanthomonas* Resource" as a query (Table S3). The search was done initially with maximum similarity and coverage and then decreasing the values to determine not only how many TALEs are present, but also at what level of similarity (Table S4). For genomes where any putative TALE was identified, we conducted an analysis of the genomic region and the composition of Repeat Variable Di-residue (RVD) using the program TALE-NT [36] (Table S5). The alignment of repetitive regions was established using Clustal Omega [37] and a cladogram was generated from the results using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). To complement the BLAST analysis, DeepSecE v0.1.2 [38], a novel tool leveraging the pre-trained protein language model ESM-1b [39], was used to predict putative type I, II, III, IV, and VI secretion system effectors (Table S6).

Subgroup genome selection for comparative genome analysis

From the 164 genomes selected for evolutionary profiling, a subset of 30 genomes was chosen to capture both effectorome diversity and host-range representation. This set also included six genomes potentially misclassified as *Xar*, which were used for phylogenetic tree rooting. Although more genomes became available later, the original 30 genomes still represent the main phylogenetic and functional diversity of the species. This selection strategy ensures that the comparative genomic analysis captures a wide range of genetic and functional diversity within the species. *X. gardneri* strain SM40611 was used as an outgroup.

Prediction of Xar protein families

The 30 genomes previously selected were annotated using the PGAP pipeline version 2024-04–27.build7426. *X. gardneri* strain SM40611 was used as an outgroup in this analysis. All 31 genomes were incorporated into a specific repository (http://jau.facom.ufms.br/arboricola/), and from there, comparative analysis based on the identification of families of orthologous proteins was performed using the OrthologSorter tool [40] (https://git.facom.ufms.br/bioinfo/orthologsorter), which employs blastp [41] and OrthoMCL [42] tools with default parameters to determine orthology. This pipeline generates three main outputs: (a) protein families unique to each genome (singletons), (b) protein families shared among at least two genomes (accessory or flexible genome), and (c) protein families shared across all genomes (core genome).

Phylogenomic profile

The phylogenomic profile using Orthologsorter was predicted based on the conserved protein families shared by all species analyzed as described above. The families that

had exactly one protein per genome were aligned and concatenated using MUSCLE [43]. After the removal of non-informative alignment columns (positions that may not be conserved, including those with too many gaps or that may be saturated by multiple substitutions) using GBlocks [44], the tree was generated using RAxML [45] with the PROTCAT model, rapid bootstrapping with 1000 replicates, and maximum likelihood search. To compare the phylogenomic results obtained from the core genome, an additional analysis was conducted using MUMi [46]. MUMi quantifies intraspecies variation and conservation at the DNA level to determine similarity and dissimilarity among the samples (genomes). Both phylogenomic topologies were generated and compared.

Metabolic profile

The metabolic profile of *Xar* genomes was analyzed using KEGG [47] to map pathways and understand metabolic functions.

Hypothetical genes reannotation and CAZyme prediction

All hypothetical genes identified and located in exclusive regions of the investigated genomes, as detailed in the analyses presented, were manually re-annotated using Uniprot [48] and KEGG [47]. In cases where a putative new function was attributed based on alignment parameters, such genes were assigned an additional nomenclature (e.g., Hyp/xxxx, where Hyp indicates that the hypothetical protein may have a function of xxxx). Carbohydrate-active enzymes (CAZymes) were predicted for the *Xar* proteomes using dbCAN3 v4.1.4 [49]. Proteins were considered CAZymes if ≥ 2 prediction tools supported the positive hit (Table S7). Final figures were curated and assembled using advanced vector-based graphical composition techniques. This process integrated outputs from specialized bioinformatics tools with high-resolution design protocols to ensure structural alignment, consistent visual identity, and professional-grade clarity of all results.

Results

Phylogenomic classification and host diversity of Xar genomes

A comprehensive phylogenomic analysis was conducted using 177 genomes of *Xanthomonas arboricola* (*Xar*), along with 72 genomes from other *Xanthomonas* species (Fig. 1). This allowed the construction of a robust phylogenetic tree in which 158 *Xar* genomes clustered into a well-defined and isolated clade, while the other 19 genomes grouped unexpectedly outside this cluster (Fig. S1). Among these 19 genomes, nine formed a subcluster between *Xar* and *X. gardneri*, including two strains annotated as *Xar.* pv. *populi* (CFBP3122 and CFBP3123), two as *Xar.* pv. *guizotiae* (CFBP7408 and CFBP7409), and two as *Xar.* pv. *juglandis* (CPBF367 and CPBF426). The remaining ten genomes grouped with other *Xanthomonas* species: CFBP8590 and CFBP8592 with *X. cannabis*; FORF20, FORF21, FORF23, FORF26, PLY3, PLY4, and PLY9 with *X. dyei*; and MEUM1 with *X. hortorum* and *X. gardneri*.

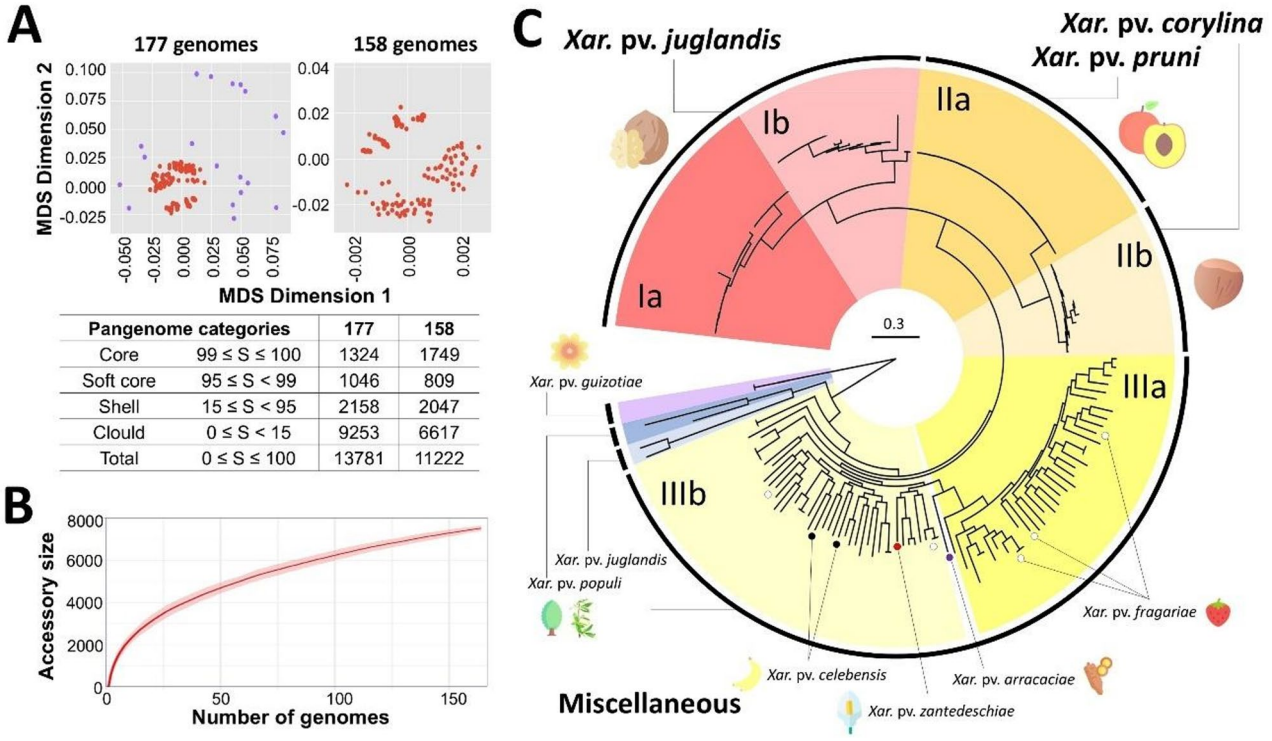
To confirm whether these 19 excluded genomes indeed belong to other species, a refined analysis using Panaroo was performed. Multidimensional scaling (MDS) revealed these genomes to be highly dispersed (Fig. 1A and B), corroborating the initial phylogenomic findings. A subsequent phylogenomic analysis was conducted using the 158 validated *Xar* genomes, along with the six genomes forming the intermediate clade between *Xar.* and *X. gardneri*. This set included two strains each of *Xar.* pv. *guizotiae* (CFBP7408 and CFBP7409), *Xar.* pv. *populi* (CFBP3122 and CFBP3123), and *Xar.* pv. *juglandis* (CPBF367 and CPBF426). Including these genomes enabled tree rooting and expanded host diversity in the analysis. The resulting tree topology (Fig. 1C) revealed three major clades (I, II, and III), each of which could be further divided into subclades (a and b).

Clade I contains only *Xar.* pv. *juglandis* genomes. Subclade Ia contains isolates from Chile (CFSAN and IVIA), the USA (417), and Portugal, while subclade Ib includes genomes from Serbia (OS and OC), China (DW3F3), and New Zealand, highlighting high genomic variability within this pathovar. Clade II consists solely of *Xar.* pv. *pruni* (subclade IIa) and *Xar.* pv. *corylina* (subclade IIb) genomes, which showed less genomic diversity than those of *Xar.* pv. *juglandis*, particularly among *Xar.* pv. *pruni* strains.

Clade III is the largest and most diverse, also subdivided into at least two subclades. Unlike Clades I and II, variability in Clade III does not appear to correlate with host origin. Interestingly, the genome of *Xar.* pv. *arracaciae* marks a potential boundary between subclades IIIa and IIIb but could not be confidently assigned to either group. This may reflect the current under-sampling of this pathovar, as it is represented by a single genome, and additional sequences may resolve it into a distinct subclade (IIIc). Lastly, the six genomes presumed to belong to non-*Xar* species form three well-separated basal subclades, providing tree rooting and further supporting previous phylogenomic results (Fig. S1).

Diverse Type III effectorome patterns across Xar clades

Effectorome analysis revealed substantial variation in the distribution of Type III secretion system effectors (T3Es) across *Xar* strains, with distinct presence/absence patterns among the previously defined clades, including genomes likely misclassified as *Xar* (Fig. 2). Notably, strains showing phylogenomic incongruence exhibited



Pangenome categories		177	158
Core	99 ≤ S ≤ 100	1324	1749
Soft core	95 ≤ S < 99	1046	809
Shell	15 ≤ S < 95	2158	2047
Clould	0 ≤ S < 15	9253	6617
Total	0 ≤ S ≤ 100	13781	11222

Fig. 1 Characterization and phylogeny of the evaluated *Xar* genomes. (**A**) Pangenome analysis of 177 *Xar* genomes obtained from NCBI. Nineteen of these genomes exhibited a certain degree of dissimilarity compared to the others analyzed. The inclusion or exclusion of these genomes significantly impacts genomic comparison parameters. (**B**) Pangenome plot of the 158 genomes validated as belonging to the species *X. arboricola*. (**C**) Phylogenomic analysis of 164 *Xar* genomes. Six of the 19 genomes excluded by Panaroo [33] were used to root the tree. The distribution of *Xar* genomes can be observed in three major clades (I to III), which can be subdivided into well-defined subclades (a and b). The strain Xar pv. *araucariaceae* was classified as independent of clades IIIa and IIIb. This positioning is attributed to its unique phylogenetic branching and its status as the sole representative of this pathovar evaluated in the current dataset

markedly divergent effector repertoires compared to confirmed *Xar* genomes.

The cladogram on the left summarizes the phylogenomic structure shown in Fig. 1. Asterisks (*) in red indicate the 30 reference genomes selected for comparative analyses. The presence/absence matrix displays T3SS effectors across genomes from the three main *Xar* clades, using the same color scheme as in Fig. 1. The bottom row indicates the frequency of each effector: green denotes high frequency, red low frequency. The last six genomes correspond to strains likely belonging to a species other than *Xar* and were used to root the phylogeny shown in Fig. 1.

Of the 85 T3E genes present in *Xanthomonas*, 14 were absent from all genomes analyzed (*hpaT*, *xopAJ*, *xopAK*, *xopAM*, *xopAR*, *xopBG*, *xopC2*, *xopI1*, *xopI2*, *xopO*, *xopS*, *xopU*, *xopW*, and *xopY*). Only three effectors (*avrBS2*, *xopAZ*, and *xopAW*) were conserved across all confirmed *Xar* strains.

Clade I, composed exclusively of *Xar.* pv. *juglandis*, displayed a relatively consistent effector profile, with 38–47 T3Es per genome. Variability stemmed primarily from differential presence of *xopE2*, *xopAH*, *xopAI*, *xopJ1*, and *xopJ5*. Subclade Ia was characterized by *xopE4*, *avrBS1*, *hopH1*, and *hopH2*, while *xopB* and *xopBB* were exclusive to subclade Ib.

Clade II genomes harbored 44–52 T3Es, with clear subclade-specific signatures. *XopAL1*, *xopAL2*, *xopAI*, *xopJ1*, and *xopJ2* were consistently present in subclade IIb, while *xopBA* and *xopAT* were restricted to subclade IIa. These two effectors were exclusive to *Xar.* pv. *pruni* strains, except for *xopBA*, also found in *Xar.* pv. *arracaciae*. Likewise, *xopAF1*, *xopAF2*, *xopE3*, and *xopE5* were shared only between clade II and *Xar.* pv. *arracaciae*. Additional effectors such as *xopBB*, *avrBS1*, *hopH1*, *hopH2*, *avrBS3*, *xopAG1*, *xopAG2*, *xopAO*, *xopD*, and *xopC1* were confined to subclade IIb but displayed heterogeneous distribution.

Clade III exhibited the lowest effector gene content. In subclade IIIa, genomes averaged only four T3Es, including *avrXccA1* and the three core effectors. The most enriched genome in this group, *Xar* F21 (GCF_014199115), carried seven T3Es. Subclade IIIb had a broader effector range (5–15 genes), with *Xar.* pv. *arracaciae* CFBP7407 exhibiting the most diverse effectorome (36 genes), followed by *Xar.* CFBP6827 (22) and

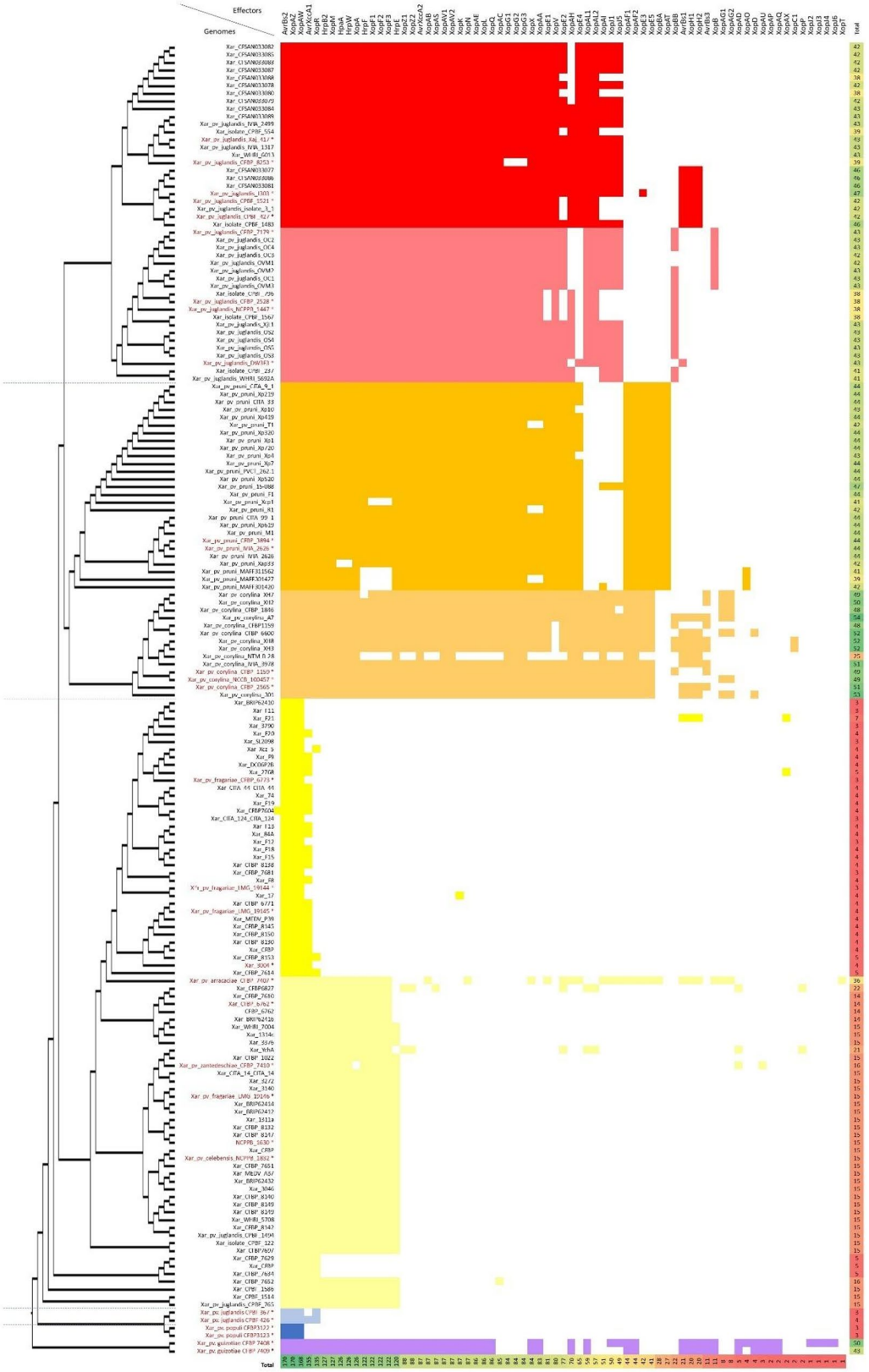


Fig. 2 Occurrence of type III secretion system (T3SS) effectors in 177 *Xar* genomes

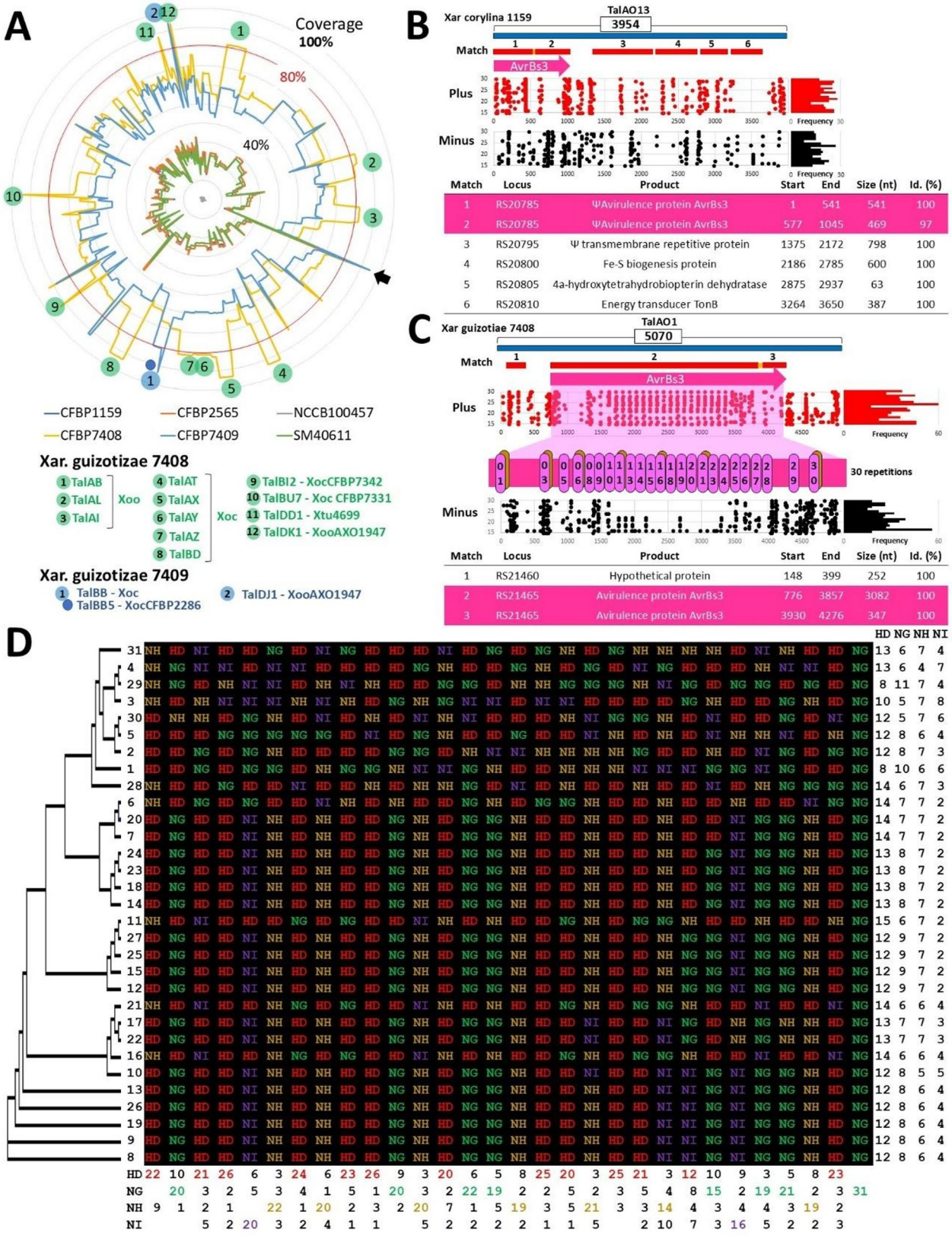


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Prediction of TALEs in the investigated genomes. (**A**) Coverage analysis of the 516 TALEs in the *Xar. corylina*, *Xar. guizotiae* and *X. gardneri* genomes. Numbers within colored circles indicate TALEs with the highest similarity to previously described TALEs from other *Xanthomonas* genomes: Xoo – *X. oryzae* pv. *oryzae*, Xoc – *X. oryzae* pv. *oryzicola*. The black arrow marks TalAO13, the TALE with the highest sequence coverage and identity among the five genomes analyzed. *Xar.* pv. *corylina* NCCB100457 is the only strain without significant similarity or coverage. (**B**) Sequence composition of the genomic region corresponding to TalAO13 in *Xar.* pv. *corylina* 1159 and (**C**) *Xar.* pv. *guizotiae* 7408. Six high-identity matches were found in strain 1159; and three in strain 7408. In both genomes, two matches correspond to genes annotated as *avrBs3*. Match details are provided in the supplementary materials (Table S5). Below the matches, predicted RVD motifs within the TalAO13 region are shown. RVDs on the plus strand are in red; those on the minus strand, in black. Adjacent histograms display the number of tandem repeats (Y-axis: 15 to 30) and their frequency in the analyzed region. In *Xar.* pv. *guizotiae* 7408, 31 repeats were found, each with 30 RVDs. (**D**) RVD repeat composition in the 30 repeats of the *avrBs3* gene from *Xar.* pv. *guizotiae* 7408. The cladogram was constructed from a multiple sequence alignment. Sequence numbers correspond to those in panel C. The last four rows summarize repeat counts and patterns across positions while the last four columns, across sequences

CFBP7630 (21). CFBP7407 uniquely carried multiple effectors, including *xopAB*, *xopN*, *xopX*, *xopE1*, *xopAH*, *xopE4*, *xopAI*, *xopJ1*, *xopJ5*, *xopAF1*, *xopAF2*, *xopE3*, *xopE5*, *xopBA*, *xopB*, *xopAG1*, *xopAG2*, and *xopT*. Conversely, genomes with the lowest effector content in this clade (CFBP7629, GCF_014196045, and CFBP7634) encoded only *avrXccA1* and *xopR*, along with the three core effectors.

Finally, genomes previously identified as non-*Xar* showed distinct effectorome profiles. *Xar.* pv. *juglandis* CPBF367 and CPBF368 contained only three (*avrBS2*, *xopAZ*, and *xopR*) and four (+*avrXccA1*) T3Es, respectively. *Xar.* pv. *populi* strains encoded only *avrBS2*, *xopAZ*, and *xopAW*. Remarkably, *Xar.* pv. *guizotiae* strains CFBP7408 and CFBP7409 exhibited unique repertoires, with *xopAP* and *xopAQ* found exclusively in both genomes. Additionally, *xopJ2*, *xopJ3*, *xopJ4*, and *xopJ6* were unique to CFBP7408, further distinguishing it from all other analyzed strains.

TALE analysis reveals identification and potential functionality in Xar. pv. guizotiae genomes

Given that transcription activator-like effectors (TALEs) are hallmark features of *Xanthomonas* species[48], their presence was assessed across all *Xar* genomes. Surprisingly, most genomes showed no significant matches to any of the 516 TALE sequences described in the literature, with alignment coverages rarely exceeding 5% of query length. Only a few genomes exhibited substantial similarity, particularly those where the *avrBs3* effector was previously identified, including nine of fourteen *Xar.* pv. *corylina* strains and both *Xar.* pv. *guizotiae* strains (CFBP7408 and CFBP7409). In these cases, multiple TALEs showed alignment coverages exceeding 80%, with the TalAO13 sequence reaching up to 95% in the *Xar.* pv. *guizotiae* genomes (Fig. 3A).

Further similarity analysis of the TalAO13 sequence revealed the genomic insertion site of the *avrBs3* gene in these strains. In *Xar.* pv. *corylina*, however, the TALE sequences appeared fragmented. Figure 3B illustrates the alignment of TalAO13 in *Xar.* pv. *corylina* strain 1159, showing correspondence with a truncated *avrBs3*Ψ and four downstream genes. Repeat variable di-residue (RVD) analysis identified the presence of TALE repeat units, but no clear pattern characteristic of functional TALEs, suggesting that these sequences may be non-functional in this pathovar.

In contrast, *Xar.* pv. *guizotiae* strain CFBP7408 carried a complete TalAO13 sequence, showing high identity across the entire length of *avrBs3* (Fig. 3C). RVD analysis confirmed a canonical organization, revealing at least 31 tandem TALE repeats of 30 amino acids each, with the variable di-residues forming combinations of HD, NG, NI, and NH motifs (Fig. 3D), indicative of a potentially functional TALE.

Global comparison involving 30 selected Xar genomes

Our next comparative analysis of *Xar* genomes sought to determine gene content differences that could explain differences in virulence induction and adaptation among the strains considered. Based on the phylogenomic analyses and effectorome composition, a subset of 30 representative genomes was selected for this analysis. The selected genomes are shown in supplementary Table S1.

Within this subgroup of genomes, we identified a core genome composed of 2868 families of orthologous proteins, in addition to 72,610 families of proteins comprising the flexible genome (Fig. 4A). The core genome and the MUMI analysis (Fig. 4B) enabled us to establish a similar phylogenomic profile shown in Fig. S1 and Fig. 1C into well-defined clades with specific pathovars (Fig. 4C and D).

When comparing the number of genes specific to given genomes, a wide divergence was observed, ranging from 185 in *Xar.* pv. *arracaciae* CFBP7407 to only seven in *Xar.* pv. *juglandis* CFBP2528 (Fig. 4A). This variation may be directly correlated with the number of strains for each pathovar investigated. When classified into categories (with assigned function - WAF, hypothetical proteins - Hyp, encoding integrases and transposases - I/T, intact or as pseudogenes - Ψ), additional interesting results emerge (Fig. 4E). For instance, most exclusive genes in *Xar.* pv *juglandis* NCPPB1447 exhibited premature stop codon mutations, suggesting potential inactivity. The number of integrases and transposases is notably high in two strains of *Xar.* pv. *juglandis* isolated from Portugal

(CPBF1521 and 427 strains) and in the two strains of *Xar.* pv. *guizotiae* from Ethiopia, which may correlate with the number of effectors identified in these genomes. Finally, this classification revealed a unique gene repertoire with distinct functions in some genomes, potentially enhancing virulence in *Xar.* pv. *corylina* NCCB100457, *Xar.* pv. *juglandis* CFBP7179, *Xar.* pv. *populi* CFBP3122 and *Xar.* pv *juglandis* CPBF426 genomes.

Among the 182 unique genes identified in *Xar.* pv. *corylina* NCCB100457, 28 clustered in a single genomic region, flanked by integrases that together contribute to the coding of a type 4 secretion/conjugation apparatus (T4SS) (Fig. 5A). With clear signatures of a horizontal transfer, if these genes are functional, they could provide an adaptive advantage to this strain isolated from *Corylus colurna* in the USA. For *Xar.* pv. *juglandis* CFBP7179, 34 unique genes were identified in a large genomic region, with integrases at the upstream end (Fig. 5B). Most of these unique genes are associated with the metabolism of heavy metals and metalloids, including copper, arsenic, mercury, cadmium, and lead. Another set of genes in this region is associated with the synthesis of a type IV secretion/conjugation apparatus (further details will be provided later). In *Xar.* pv. *populi* CFBP3122, a set of 15 unique genes was identified in a region encoding a type II secretion apparatus (T2SS) (Fig. 5C). Finally, in *Xar.* pv *juglandis* CPBF426, six unique genes were identified, four of which are involved in DNA phosphorothioation (DNA PT) in a genomic region flanked by an integrase (Fig. 5D).

Other genes and systems associated with virulence and adaptation

Aiming to expand the analyses of the flexible genome, other genes encoding functional apparatus and biological processes were evaluated. The systems analyzed and their presence and absence in each genome are summarized in Fig. 6A. The global syntenic organization using *Xar.* pv. *juglandis* strain 417 as a reference genome is shown in Fig. 6B. These genes and systems were grouped into six major categories: secretion systems, flagellum, biofilm and quorum sensing, adaptation, secondary metabolites, and cell wall degrading enzymes. For most functions associated with these categories, the genomes are very similar to each other. However, significant discrepancies are observed in some of these systems, as described below.

Secretory systems

Regarding secretion systems, notable discrepancies were observed in the genes encoding T2SS and T4SS, in addition to the extra copy of the T2SS previously mentioned in the *Xar.* pv. *populi* 3122 genome (Fig. 5C). For T3SS, the genes encoding this apparatus were not identified in the genomes of the clade to which *Xar.* pv. *fragariae* (LMG19145, LMG19144, and CFBP6775 strains), *Xar.* pv. *juglandis* 3004, both strains of *Xar.* pv *populi* (CFBP3123 and CFBP3122 strains) and *Xar.* pv. *juglandis* (CPBF367 and CPBF426 strains) belong (Fig. 6A and B. Notably, these genomes also harbor the smallest effectorome repertoires identified in this study (Fig. 2). Analysis of the putative functional profiles of the effectors across the 30 reference strains indicates a marked reduction of these genes and the loss of the T3SS in the aforementioned genomes (Fig. 2). Rather than indicating reduced virulence per se, this pattern suggests that these strains may rely on alternative virulence strategies that are less dependent on T3SS-mediated effector delivery. Such strategies may include the secretion of cell wall–degrading enzymes via the T2SS, production of toxins or secondary metabolites, and other as yet uncharacterized mechanisms. Importantly, these inferences are based on genomic data and require experimental validation. Indeed, other *Xanthomonas* species, such as *X. sacchari* and *X. albilineans*, are known pathogens despite lacking a canonical T3SS, supporting the notion that diverse pathogenicity strategies exist within the genus. The flanking genes at both ends of the genomic region containing the T3SS in the reference genome (*Xar.* pv. *juglandis* strain 417) are found concatenated in the genomes lacking this apparatus (Fig. 6C). This highlights the role of insertion and deletion events in the evolutionary dynamics of *Xanthomonas* species and the genomic erosion observed in the genomes that have lost these genes.

Most genomes lack gene clusters associated with type IV secretion/conjugation apparatus synthesis. Among the genomes that do contain them, there is variation not only in gene composition, but also in potential evolutionary inheritance. For instance, in *Xar.* pv. *corylina* NCCB10047 and *Xar.* pv. *arracaciae* 7407, these genes are inserted in a genomic region exclusive (putative episome) to each species (Fig. 5).

Flagellum assembly

Flagellum synthesis and regulation are complex processes dependent on a series of genes that code for proteins with regulatory and structural functions. In *Xar*, using the *Xar.* pv. *juglandis* 417 as a reference, two large gene clusters were identified. One cluster is located in the central region of this genome, containing genes involved in the chemotaxis process (Che). The other gene cluster is closer to the 10 o'clock region of the genome (Fig. 6A and B). The first cluster is found in all investigated strains, except in *Xar.* pv. *corylina* 2565 which has the second gene cluster identified only in the other two *Xar.* pv. *corylina* and *Xar.* pv. *juglandis* strains, except for those positioned in the most basal clade (426 and 367 strains). The second group is mostly made up of genes that code for the MS, P and L rings that integrate the flagellum into

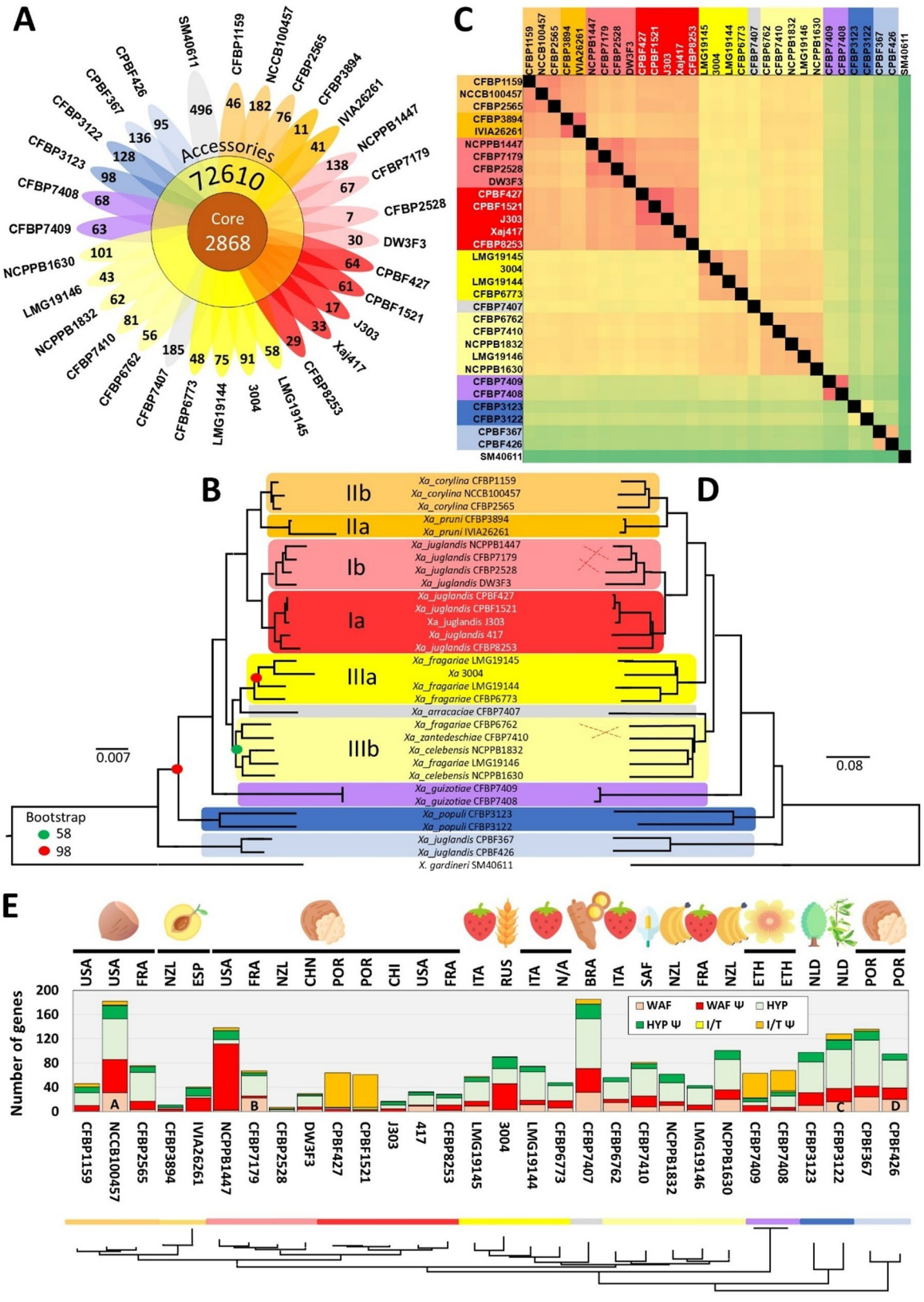


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Comparative genomic analysis involving the 30 previously selected *Xar* genomes. *X. gardneri* strain SM40611 was used as an outgroup in these analyses. (**A**) Floral diagram highlighting a core genome of 2868 orthologous protein families. The colors of the petals were arbitrarily assigned to group or differentiate the genomes based on the phylogenomic profile in Fig. 1. The numbers at the ends of the petals indicate the total number of singleton genes (unique genes) identified in each genome. (**B**) Phylogenomic analysis with the core genome. Circular nodes indicate bootstrap support values below 100%. The absence of circular nodes on the clades indicates maximum bootstrap support (100%). The background colors that classify the genomes are the same as those used in the floral diagram. (**C**) Conservation analysis of DNA sequences using Mummer array [46]. The intersections between lines and columns in red tone variations indicate greater similarity, while in green tones, lower similarity between genomes. The colors assigned to each genome are the same as those shown in the floral diagram. (**D**) Phylogenomic distribution based on Mummer matrix results. Note that there is conservation in the distributions of all clusters defined by the core genome analysis, except for some small internal transitions observed (dotted red lines). (**E**) Assessment, categorization and structural organization of unique genes in each of the genomes evaluated in six categories: WAF (with assigned function), Hyp (Hypothetical) or I/T (encoding integrases or transposases), with the presence or absence of mutations that led them to be classified as pseudogenes (ψ). On top of the panel the hosts and the identification of the country where the strains were isolated are reported (with a three-letter abbreviation). N/A – Not identified or reported. The letters A-D in some of the histogram bars signal the presence of a set of genes that were analyzed regarding their organization and importance in the respective genomes, and which are outlined in detail with the same identification as in Fig. 5

the inner and outer membranes (Fig. 7A). It is interesting to note that the genes flanking this region, at both ends, are also concatenated in genomes that do not have them, signaling the intervention of INDELs in the structure and biological dynamics of *Xar.*

Biofilm production and quorum sensing response

Biofilm formation and quorum sensing responses are critical traits linked to adaptation and virulence in *Xanthomonas*. Among the 30 *Xar* genomes analyzed, notable variation was observed in gene content associated with these processes. Specifically, a genomic region involved in cellulose biosynthesis and secretion was identified, which is absent in certain strains, including *Xar.* pv. *corylina*, *Xar.* pv. *pruni*, two *Xar.* pv. *juglandis* strains from the basal clade (CPBF426 and CPBF367), and *Xar.* pv. *populi* strain CFBP3123 (Fig. 6A and B).

This region comprises a six-gene cluster organized in tandem, including two regulatory genes (*bcsQ* and *bcsR*) and four structural genes directly involved in polymerization and secretion (*bcsA*, *bcsB*, *bcsZ*, and *bcsC*) (Fig. 7B). In genomes lacking this cluster, the flanking genes at both ends appear directly concatenated, suggesting a complete deletion of the operon.

Genes of interest within a genomic island in the Xar genome associated with adaptation

In the same extensive putative episome region where a copy of T4SS is located in *Xar.* pv. *juglandis* strain 417, several other genes of interest were also observed (Fig. 7C). This includes a group of genes involved in copper resistance, previously well characterized by Assis et al. 2021 [19]. Notably, these metal resistance genes differ from those identified for *Xar.* pv. *juglandis* 7179 (Fig. 5B), suggesting they have the same function but were acquired through distinct horizontal gene transfer events. Additionally, a cluster of genes associated with nitrate metabolism is present (Fig. 7C). Upstream of the metal resistance genes, genes related to nitrate metabolism, including those for internalization and siroheme synthesis, are arranged in tandem. Siroheme is an essential cofactor for the enzyme that converts nitrate into nitrite. This cluster was identified only in *Xar.* pv. *guizotiae*, *Xar.* pv. *populi* 3122, *Xar.* pv. *fragariae*, *Xar.* pv. *celebensis*, *Xar.* pv. *zantedeschiae*, and in strains 426, 367 and 417 of *Xar.* pv. *juglandis* (Fig. 6A and B). To date, there has been no experimental analysis of the importance of these genes in *Xar* metabolism.

Secondary metabolites

Lasso peptide An interesting gene cluster present in all 30 reference strains investigated and functionally reported only in *X. gardneri* is associated with lasso peptide synthesis. The synthesis of this molecule depends on a small gene (LP), which codes for a protein containing two domains: an N-terminal corresponding to the leader peptide and a C-terminal corresponding to the lasso peptide itself. Three other genes (Lp-B2, Lp-Cyc and Lp-Iso) participate in the processing of the active molecule, separating these two peptides and contributing to the folding of the final structure (Fig. 7D). These genes are always arranged in tandem. In all 30 genomes investigated, the genes involved in the processing of the lasso peptide were identified with a high degree of coverage and identity. As for the gene coding for the lasso itself, two tandem copies (LPI and LPII) were identified. Sequence pairing analysis of the unprocessed lasso protein was established using LPI and LPII sequences previously described for *X. gardneri* as Refs. [50]. A high degree of conservation was evident between the LP copies, with specific differences between *Xar* and *X. gardneri* and between the LPI and LPII sequences, specifically in the N-terminal portion (leader peptide). When comparing the fraction of processed peptides, a clear difference is observed between LPI and LPII lasso peptides. However, among the species evaluated the peptides are identical (Fig. 7D).

Other secondary metabolites In addition to the presence of genes encoding the lasso peptide as a signature of the *Xar* species, gene clusters associated with the synthesis of secondary metabolites were identified in practically

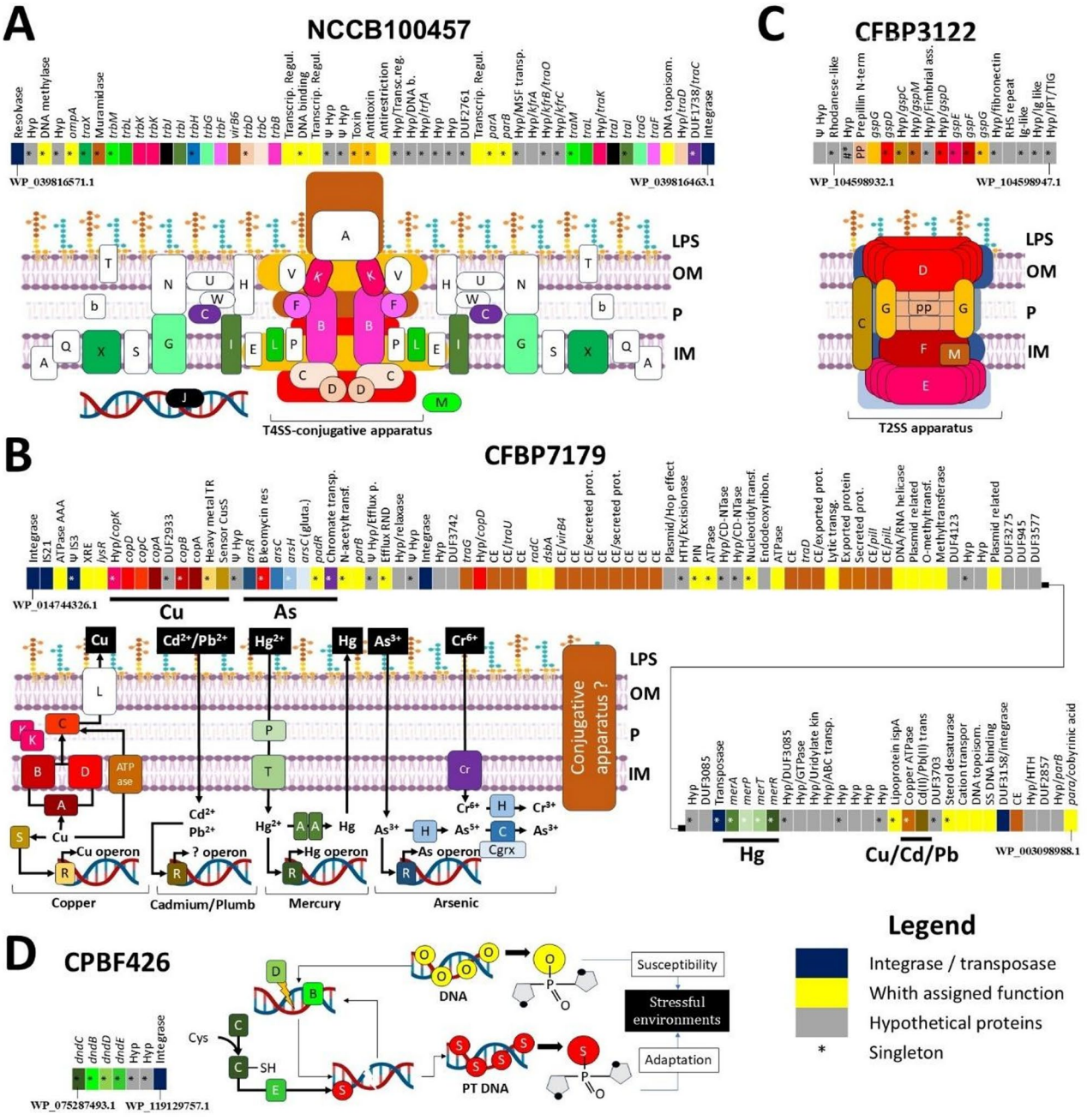


Fig. 5 Analysis of unique genes in four of the investigated genomes. (**A**) Coding region of the T4SS apparatus with putative conjugative function in the genome of *Xar.* pv. *corylina* strain NCCB1000457. Genes encoding proteins exclusive to this genome are marked (*). The presence of integrase and resolvase at the ends of the region accompanied by the high number of unique genes are clear signatures of insertion of this region by a HGT event. The colors assigned to genes are the same as those assigned to proteins in the structural model. (**B**) Extensive genomic region identified in *Xar.* pv. *juglandis* strain CFBP7179 with a high number of unique genes (*) encoding proteins linked to the internalization and metabolism of metals and metalloids (copper - Cu, arsenic - As, cadmium - Cd, lead - Pb, chromium - Cr and mercury - Hg). Additionally, it is possible to identify at least 19 genes associated with the synthesis of a putative T4SS conjugation apparatus. The colors assigned to the genes are the same as those assigned to the proteins in the functional metabolic model. An integrase in one of the flanking regions supports acquisition by HGT. (**C**) Exclusive region of *Xar.* pv. *populi* strain CFBP3122 genome with genes encoding the T2SS apparatus (Gsp). The colors assigned to genes are the same as those assigned to proteins in the structural model. (**D**) Region with genes associated with DNA phosphorothioation (PT), flanked downstream by an integrase, in the genome of *Xar.* pv. *juglandis* strain CPBF426. If functional, these genes confer DNA resistance to degradation events, by replacing the oxygen with sulfur in the phosphate group that makes up the phosphodiester bond. The colors assigned to genes are the same as those assigned to proteins in the structural model

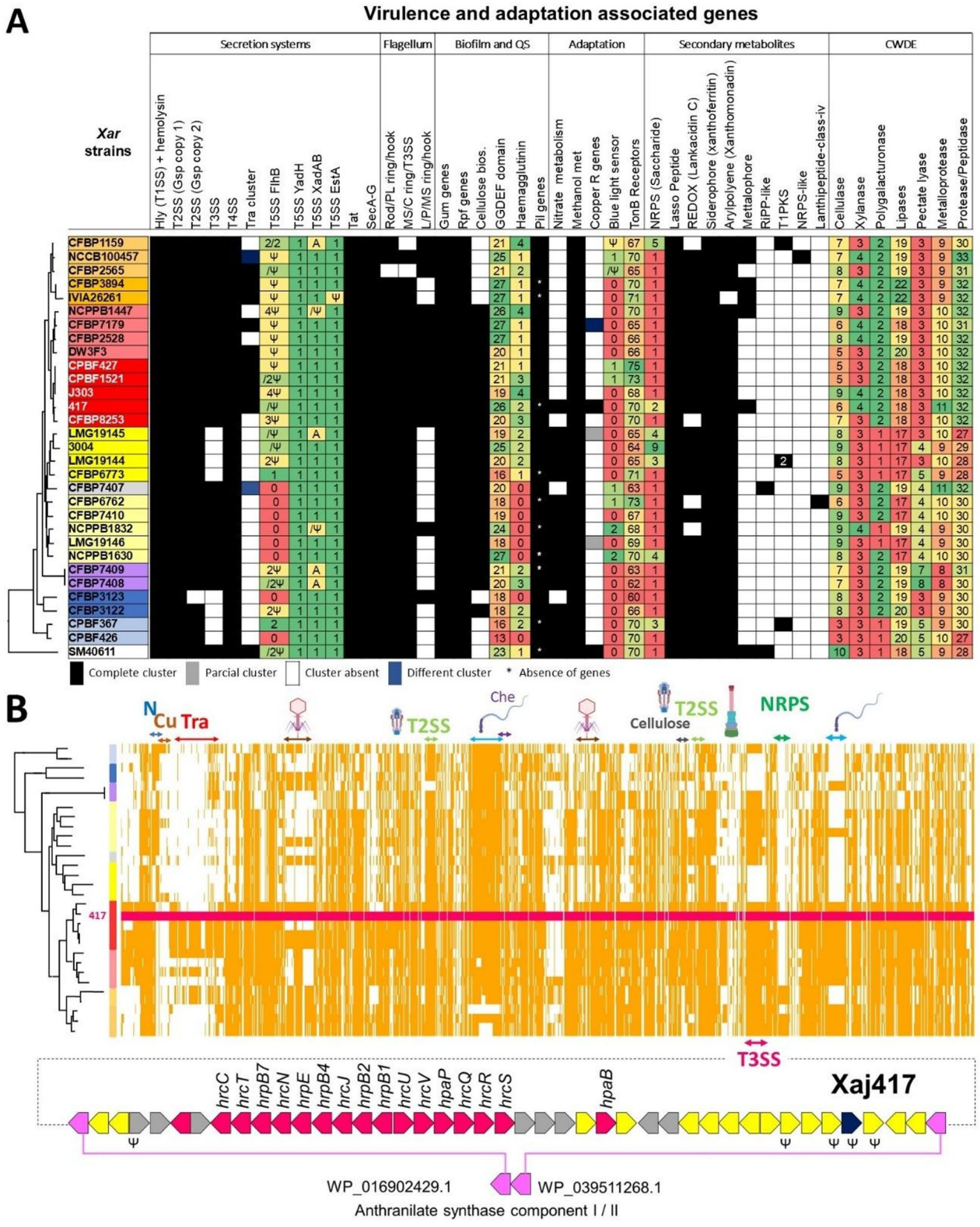


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Comparison of other genes and cellular processes associated with *Xar* virulence and adaptation. (**A**) Analysis of the presence and absence of genes associated with the synthesis of other secretory systems, involved with regulation and flagellum synthesis, biofilm production and quorum sensing, adaptation, biosynthesis of secondary metabolites and degradation of plant cell wall. Presence of the respective system is shown in black; absence is shown in white. An asterisk (*) indicates the absence of specific genes within the set associated with that function. Partial loss is indicated in gray. Blue indicates the presence of a system or gene set that differs from the reference model used for comparison. In systems with quantitative gene variation, the numbers indicate their respective representations, accompanied by a heat map analysis. (**B**) Global synteny of the investigated genomes using the genome of *Xar.* pv. *juglandis* strain 417 as a reference. In orange, the presence of genes is highlighted. Alongside the figure are some of the functional groups presented (secretory systems type II and III - T2SS/T3SS, NRPS, chemotaxis - Che, conjugation - Tra, resistance to metals - Cu, nitrogen metabolism - N, phage insertions, cellulose synthesis - Cel, and flagellum synthesis). At the bottom of the figure, a presence/absence analysis of the genomic region encoding the T3SS apparatus (in red) is shown, using *Xaj417* as the reference genome. Genes with unrelated functions are shown in yellow, and flanking genes in pink. In genomes lacking this region, the upstream and downstream flanking genes (pink) appear concatenated, suggesting a possible INDEL event at this locus

all genomes, such as synthesis of lankacidin C, xanthoferritin, and xanthomonadin (Fig. 6A and B). Others, however, are pathovar-specific or shared with reduced groups of strains. However, the greatest variation observed refers to the number of clusters involved in the synthesis of non-ribosomal peptides (NRPs), mostly present in a single copy, but with some genomes containing up to nine distinct clusters, as is the case of *Xar.* pv. *juglandis* 3004.

Cell wall-degrading enzymes (CWDE)

The composition of the main hydrolytic enzymes secreted by *Xanthomonas* was evaluated in the 30 genomes investigated. Although there are small differences in the number of these enzymes between genomes, none of these differences have functional significance to be shown (Fig. 6A).

Discussion

Bacteria belonging to the *Xanthomonas* genus have significant economic importance due to their ability to infect and damage a wide range of plant hosts [51]. Since 2002, following the complete genome sequencing of *X. citri* subsp. *citri* and *X. campestris* pv. *campestris*, several studies have focused on characterizing the gene repertoires of both pathogenic and non-pathogenic strains [2]. These efforts have significantly improved our understanding of shared genes and species-specific features [4].

A similar scenario is observed in *X. arboricola*. However, studies on this species remain less comprehensive compared to other well-characterized Xanthomonads. Although *Xar* has been recognized as a distinct species since 1995 [28], several reclassification proposals have emerged, based on host specificity, virulence profiles, or genomic similarity [23, 52]. Currently, over one hundred complete genomes across nine pathovars are available in public databases (NCBI: taxon 56448), making *Xar* an excellent candidate for comparative genomic studies. Investigating the genomic specificities of each pathovar can elucidate correlations between gene content and virulence potential in compatible hosts, thereby informing more rational and effective control strategies [53]. This study presents a comparative genomic analysis to identify new sets of genes potentially involved in adaptation and virulence in *Xar*.

Phylogenomic analyses revealed misclassifications of several strains. For example, strains CFBP8590 and CFBP8592 clustered within the *X. cannabis* clade, with no available metadata on host or origin. Strains FORF20, FORF21, FORF23, FORF26, PLY3, PLY4, and PLY9, all isolated from *Arabidopsis thaliana* in France, clustered with *X. dyei*, a species also associated with *A. thaliana.* Strain MEUM1, likewise isolated from *A. thaliana*, clustered with *X. hortorum* or *X. gardneri*. These findings corroborate those of Zarei et al. [23], who demonstrated that these strains are misclassified and should be reassigned to other species.

Phylogenomic reconstruction allowed the classification of confirmed *Xar* strains into three major clades, designated I, II, and III. Clades I and II show pathovar-specific clustering. Clade I contains only *Xar.* pv. *juglandis* genomes and is subdivided into two subclades (Ia and Ib). This division may reflect distinct evolutionary profile and geographical origins, as Chilean strains were distributed across both subclades, suggesting at least two separate introduction events. Notably, the effector repertoires (effectoromes) also differed between subclades: some effectors were exclusive to Ia, while others were restricted to Ib.

Clade II includes *Xar.* pv. *pruni* and *Xar.* pv. *corylina*, each forming a well-defined subclade (IIa and IIb, respectively). *Xar.* pv. *pruni* genomes showed remarkable genomic and effectorome homogeneity, while *Xar.* pv. *corylina* exhibited greater diversity in both aspects. Interestingly, certain effectors, such as xopBA and xopAT, were exclusively and consistently present in *Xar.* pv. *pruni*, potentially indicating a role in host specificity, although experimental validation would be required to confirm this.

Clade III displayed the greatest host diversity and was also subdivided into two subclades (IIIa and IIIb). Although the associated hosts varied, effectorome composition remained relatively conserved within each subclade. Genomes in IIIa harbored only four T3E genes, whereas IIIb genomes encoded around 15. Notable

exceptions include *Xar.* pv. *arracaceae* (strain CPBF426), isolated from *Arracacia xanthorrhiza* in Brazil, which carries 36 T3E genes, and strains from *Capsicum annuum* (CFBP6827, Poland) and *Salvia splendens* (YchA, China), with 22 and 21 T3Es respectively.

The most striking case involved *Xar.* pv. *guizotiae*, represented by two strains (CFBP7408 and CFBP7409) isolated from *Guizotia abyssinica* in Ethiopia. Previous studies proposed reclassifying these strains as *X. guizotiae* sp. nov., and our results confirm their divergence from *Xar* and reveal an extensive effectorome: 50 T3E genes in CFBP7408 and 43 in CFBP7409. Four T3Es were unique to CFBP7408 among all genomes examined.

All these results regarding genomic diversity and effectorome heterogeneity indicate that the variation in effectorome complexity among the investigated *Xar* genomes reflect the challenges in classifying strains within this species. In some cases, effectorome distribution appears to be directly related to the associated host, as observed for most organisms in the *Xar* III clade. However, in the remaining genomes analyzed, such a correlation is not evident. Instead, effectorome composition shows a stronger relationship with genomic similarity, thus corroborating the phylogenomic distribution instead of an association with the pathovars or associated hosts. This supports the notion that genetically similar strains, even when isolated from different hosts and causing different symptoms, may in fact represent the same microorganism. This highlights the value of comparative genomics as a tool for uncovering new avenues of functional investigation [54]. In the context of plant–pathogen molecular interactions, such insights are essential for expanding fundamental knowledge and guiding the development of more effective disease control strategies [55, 56].

The effectorome analysis revealed a wide range in T3E content across genomes. Some strains lacked effectors entirely, while others possessed extensive repertoires capable of manipulating host immunity. To further investigate effector diversity, we examined the presence of transcription activator-like effectors (TALEs) [57]. It is important to highlight that in some *Xanthomonas* genomes it is possible to identify a significant set of TALE-encoding genes with these functions [58, 59], while in others, such genes were not identified, as previously demonstrated for the *Xar.* pv. *juglandis* strain 417 genome [60]. Analysis of the presence and absence of TALEs in the investigated genomes demonstrated that these molecular tools for activating conditioned gene expression responses in plants appear to be a functional attribute found only in *Xar.* pv. *guizotiae* strains [57]. Only in these strains we determined the presence of the complete *avrBs3* gene, with all the complement of RVD repeats, which have been shown to be associated with these gene expression regulatory genes [61]. Based on their broad repertoire of effectors (the largest among the strains investigated), we hypothesize that *Xar* pv. *guizotiae* strains may have elevated virulence potential, mediated by the plant-pathogen interactions of these proteins. However, this remains a genomic-based prediction, and experimental validation (e.g., pathogenicity assays, host-range tests, or expression studies) is required to confirm these effects. Nevertheless, these strains represent an interesting biological model for functional research and a potential concern in the possible emergence of new diseases [62], especially on the African continent, where they were isolated.

In addition to investigating the evolutionary profile and composition of the effectorome, this work also aimed to understand the contribution of other genes and cellular processes that could contribute to adaptation to different environments and hosts. In these terms, a robust comparative genomic analysis involving 30 of the 177 initially selected genomes was performed. The selection of these genomes was based on the diversity of associated hosts and the composition of the effectorome. It was found that the phylogeny based on the core genome of this subset of organisms reflected a distribution very similar to that presented for the larger set of genomes, but with an evident grouping of strains based on the hosts of isolation. This behavior was not observed in the genomes belonging to the *Xar* III clade, described previously, which adds evidence to the hypothesis that some strains contained in this clade may actually represent another *Xanthomonas* species [23].

Analysis of the composition of the unique genes of each strain revealed the presence of intriguing genes that may confer molecular benefits to the strains possessing them [63]. This is the case of *Xar.* pv. *corylina* strain NCCB100457 that carries additional copies of genes involved in the synthesis of the T4SS secretion apparatus, which could confer greater conjugation potential [64] or bactericidal activity in competition events [65]. Another example is the additional copy of T2SS in *Xar.* pv. *populi* strain CFBP3122, which could give it greater efficiency in the distribution of hydrolytic enzymes or effectors [7, 66]. The presence of the genetic cluster involved with DNA phosphorothioate (PT) could confer *Xar.* pv *juglandis* strain CPBF426 greater DNA stability once subjected to stress conditions [67]. All these gains in gene functions clearly outline the importance of horizontal gene transfer events and transposition elements in the genetic diversity observed in *Xanthomonas* [68] as fundamental in adaptation and induction of virulence [19].

Analysis of the species' flexible genome enabled the identification and classification of a set of genes with significant potential to enhance adaptive capabilities [69]. In addition to T4SS and T2SS already discussed, another interesting set was related to genes encoding

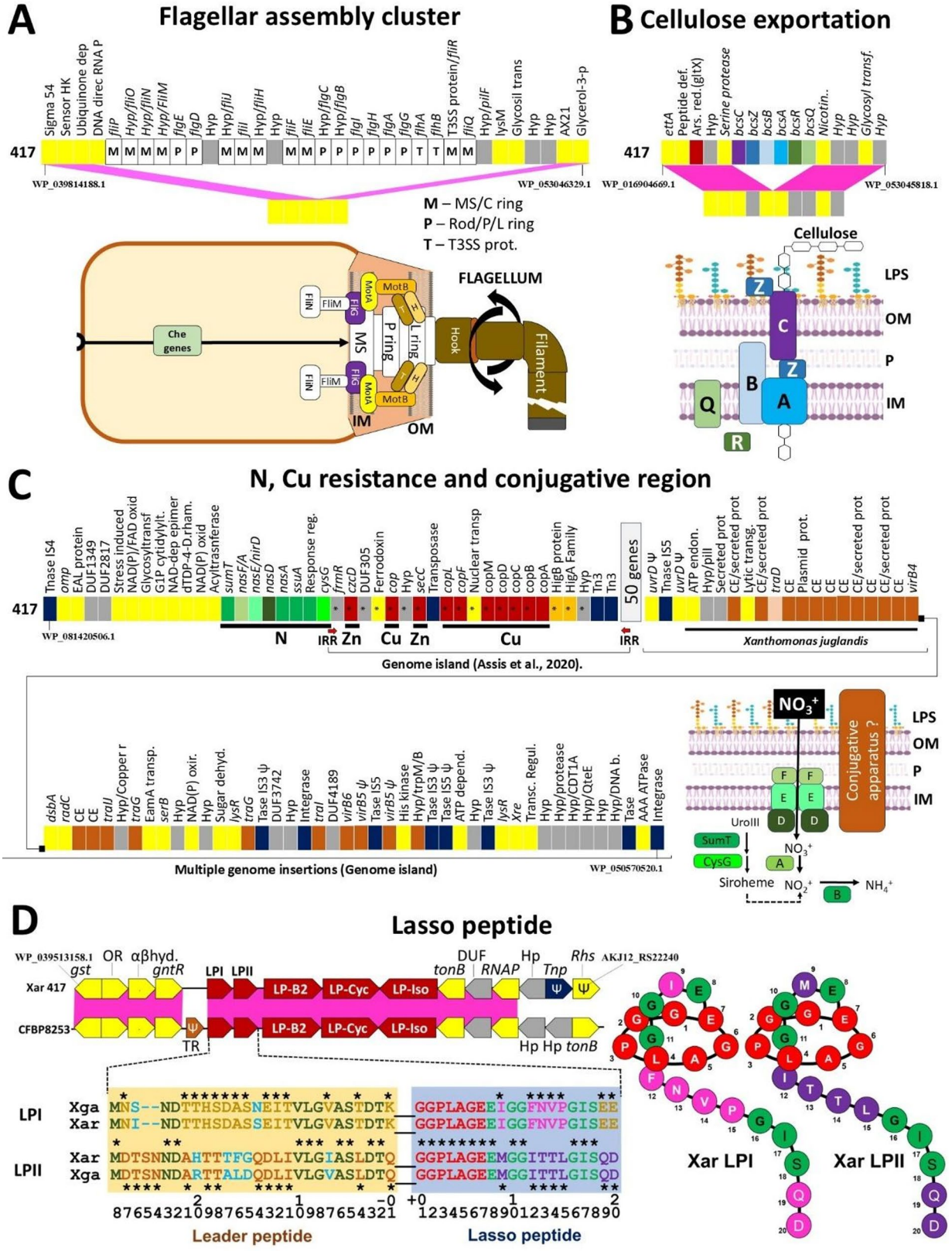


Fig. 7 (See legend on next page.)

(See figure on previous page.)

Fig. 7 Analysis of the composition and syntenic organization of other gene clusters in the *Xar* genomes. (**A**) Genomic region associated with flagellar assembly. The gene organization observed in the representation of *Xar.* pv. *juglandis* strain 417 was only observed in other strains of *Xar.* pv. *juglandis*, in strains of *Xar.* pv. *corylina*, and in strains CFBP3123 of *Xar.* pv. *populi* and NCPPB1832 from *Xar.* pv. *celebensis*. The flanking genes are concatenated in genomes where this region was not identified, indicating a gain or loss mediated by INDEL events. The colors assigned to genes are the same as those assigned to proteins in the structural model. (**B**) Synteny analysis, presence and absence of genes involved in cellulose synthesis in *Xar* genomes, a model that represents the proteins encoded by the respective genes and their functions in the process of synthesis and secretion of this polymer. The flanking genes at both ends are found in tandem in genomes that do not have this cluster. (**C**) Broad genomic region identified in the reference genome *Xar.* pv. *juglandis* strain 417. In addition to a region that confers resistance to copper, previously characterized by Assis and collaborators [19], this region also presents an upstream cluster of genes involved with nitrate metabolism, and downstream genes involved with the synthesis of the T4SS apparatus. A profile of the structural and metabolic function of genes involved in nitrate metabolism is outlined. (**D**) Synteny of genes within the region encoding the lasso peptide. The genes that participate in the structuring of the functional peptide are present in tandem (LP-B2 - biosynthesis B2 protein, LP-Cyc - cyclase and LP-I -isopeptidase): OR – oxidoreductase; Hp – Hypothetical proteins; DUF – domain unknown function. On the bottom is shown the identity and conservation profile between the sequences of lasso peptides I and II in *Xar* using as references the sequences previously described in *X. gardneri* (Xga) [50]. The asterisk (*) indicates conservation between all sequences (central position) or specifically between the structures of each LP (top and bottom). On the right is a model of the possible composition of the two lasso peptide structures identified in *Xar* (LPI and LPII). The colors and numbering are the same as those identified in the alignment profile

the structural proteins of the T3SS apparatus: precisely the genomes that do not have these genes were the same ones that presented the lowest number of effectors [70]. Clade III strains display broad host-range pathogenicity despite a markedly reduced T3SS effector repertoire, suggesting that alternative virulence strategies underlie their success. Considering that the acquisition of this apparatus by horizontal transfer can confer pathogenicity to hitherto non-pathogenic strains [71], and that similarly, the absence of this apparatus has already been described as a condition for the loss of pathogenicity potential in a series of species [53], these observations suggest that strains devoid of this apparatus are predicted to be less reliant on T3SS-mediated virulence compared to those that have it. Together, these observations suggest that specific combinations or expression levels of cell wall–degrading enzymes (CWDEs), non-ribosomal peptides, lasso peptides, or other secondary metabolites may serve as alternative primary virulence factors in Clade III strains, potentially compensating for the reduced T3SS effector repertoire. These inferences remain predictions based on genomic data and require experimental validation. In addition, some of the investigated strains also lack identifiable gene clusters associated with flagellum synthesis and regulation. Given that both structures share genes and biological functions [72], we can assume that loss of pathogenicity potential not only arises from the inability to secrete effector proteins but also from lack of flagellum-dependent motility [73]. For others, however, the loss of genes encoding the flagellar apparatus could be compensated by the activity of genes associated with the T3SS, as previously demonstrated for *X. fuscans* [13].

Regarding the set of genes associated with the flexible genome, three additional clusters were highlighted: the nitrate metabolism gene cluster, genes involved in cellulose synthesis and secretion, and those associated with lasso peptide synthesis. The presence of a nitrate metabolism gene cluster provides a phytopathogen with ecological and metabolic flexibility, enabling adaptation to diverse environments, exploitation of host resources, regulation of virulence, and enhanced survival [74]. Similarly, the presence of cellulose synthesis and secretion genes and lasso peptide synthesis genes plays a crucial role in biofilm formation, host colonization, microbial competition, immune evasion, stress resistance, and potential virulence regulation [75].

Together, these findings suggest that both metabolic pathways and secreted factors contribute to pathogen adaptation and may represent targets for disease control strategies, including approaches aimed at disrupting pathogen metabolism. In addition, the characterization of effector repertoires across *Xar* lineages provides complementary opportunities for disease management. Conserved effectors may represent promising targets for broad-spectrum resistance through the deployment of corresponding resistance (R) genes, while lineage-specific effectors could inform breeding strategies tailored to particular pathovars or host systems. Furthermore, understanding effector diversity and distribution may support the design of more durable resistance by anticipating pathogen adaptation and effector loss or diversification.

Specifically, genes encoding lasso peptides have been identified in a limited number of *Xanthomonas* species, with *X. gardneri* being the most well-characterized, producing Xanthomonins [50]. In the present study, all investigated strains contained this gene cluster, with a composition identical to that described in *X. gardneri*, suggesting that *Xar* species also produce these macrolactam peptides as a defense mechanism [76].

In summary, this genomic comparative analysis identified both unique and shared genes that may confer adaptive advantages to the *Xar* strains under investigation. These findings highlight potential targets for functional studies aimed at gaining deeper insights into the biology of these phytopathogens, thereby paving the way for new control strategies and integrated disease management approaches.

Conclusions

The comparative genomic analyses performed in this work revealed that there is an immense organizational complexity in the genomes belonging to the species *X. arboricola*, and this complexity is directly related to the environment, host and composition of the unique and flexible genome, resulting from insertion and deletion events that may contribute to the virulence and adaptation profiles of the pathovars and strains investigated. This was exemplified by the diversity of the effectorome, and in the set of structural and fundamental genes in biological processes, such as those associated with the synthesis of secretion, conjugation and motility apparatuses. Other interesting genomic features included the composition of genes linked to specialized metabolic functions, such as genes involved in PT DNA, cellulose synthesis and export, nitrogen metabolism, and genes coding for the synthesis of molecules with antimicrobial action, such as the diversity of NRPs and the production of lasso peptide. This extensive molecular repertoire makes *Xar* bacteria highly versatile organisms and informative models for studying molecular interactions with the diverse microenvironments of different host plants.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12916-3.

Supplementary Material 1. Fig. S1. Phylogenomic analysis of 249 Xanthomonas genomes. The dataset includes 177 genomes initially annotated as X. arboricola (Xar) and 72 genomes from 24 other Xanthomonas species. Genomes identified as X. arboricola (158 in total) are highlighted with a salmon background. Notably, nineteen putative Xar genomes cluster within other clades, supporting the hypothesis that they may belong to different species.

Supplementary Material 2. Table S1. List and general characteristics of the 249 genomes used to outline the species' phylogenomic profile, including the 177 genomes preliminarily annotated as Xar.

Supplementary Material 3. Table S2. Genome-based taxonomic classification of analyzed genomes using GTDB-Tk.

Supplementary Material 4. Table S3. Blast results for each of the 516 TALEs evaluated in each of the 30 genomes selected for comparative genomic analyses.

Supplementary Material 5. Table S4. Grouping of TALEs by family and coverage and identity profiles.

Supplementary Material 6. Table S5. Analysis of RVD composition in Xar genomes. Pathovars guizotiae 7408 and corylina 1159.

Supplementary Material 7. Table S6. Prediction of putative type I, II, III, IV, and VI secretion system effectors using DeepSecE v0.1.237.

Supplementary Material 8. Table S7. Distribution and functional classification of Carbohydrate-Active Enzymes (CAZymes) identified in the genomes of Xar.

Authors' contributions

RABA, AMV, NFA, AMD, LMM designed the work and selected the genomes. RABA, AMV, AS, JSLP, JCS, NFA, LMM performed genome comparisons and identified the protein families. AMV, JSLP, NFA performed phylogenetic analysis. RABA, AMV, CHDS, JSLP, JCS, PAZ, RFS, AMD, LMM interpreted the findings. AS, CHDS, JSLP, JCS, PAZ, RFS, CCMG, EGO, NFA, JEA, AMD, LMM contributed additional interpretations and general manuscript comments. RABA, AMV, AS, CHDS, JCS, PAZ, CCMG, EGO, NFA, JEA, AMD, LMM wrote the paper. All authors revised the paper.

Funding

This research was funded by the California Walnut Board. LMM (305552/2025-2), JCS, NFA and AMV have a research fellowship from CNPq. LMM has UFOP grants. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Data availability

Publicly available genomes used in this study were retrieved from the NCBI RefSeq and GenBank databases. Accession numbers for each genome are provided in Additional file 1: Table S1. The NCBI GenBank database is available at https://www.ncbi.nlm.nih.gov/genbank/.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Center of Research in Biological Science, Federal University of Ouro Preto, Ouro Preto, MG, Brazil
[2]Department of Plant Sciences, University of California, Davis, CA, USA
[3]Department of Agricultural and Environmental Biotechnology, College of Agricultural and Veterinary Sciences, São Paulo State University (UNESP), Jaboticabal, SP, Brazil
[4]Microbiology & Plant Pathology Department, University of California, Riverside, CA, USA
[5]Laboratório Especial de Ciclo Celular, Instituto Butantan, São Paulo, SP, Brazil
[6]Department of Biochemistry, Chemistry Institute, University of Sao Paulo, Sao Paulo, SP, Brazil
[7]School of Computing, Federal University of Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, MS, Brazil
[8]Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil
[9]Department of Biological Science, Institute of Exact and Biological Science, Federal University of Ouro Preto, Ouro Preto, MG, Brazil
[10]Instituto de Biología Molecular y Celular de Rosario, Facultad de Ciencias Bioquímicas y Farmacéuticas, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de Rosario, Rosario, Argentina
[11]Departamento de Ciências Biológicas (DECBI), Instituto de Ciências Exatas e Biológicas (ICEB), Universidade Federal de Ouro Preto (UFOP), Ouro Preto, MG 35400-000, Brasil

Received: 4 August 2025 / Accepted: 30 April 2026
Published online: 06 May 2026

References

1. An S-Q, et al. Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen Xanthomonas. FEMS Microbiol Rev. 2020;44:1–32.
2. da Silva, A. C. R. *et al.* Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. *Nature. 2002* 417, 459–463.
3. Moreira, L. M., De Souza, R. F., Digiampietri, L. A., Da Silva, A. C. R. & Setubal, J. C. Comparative analyses of Xanthomonas and Xylella complete genomes. *OMICS. 2005* 9, 43–76.
4. Ryan RP, et al. Pathogenomics of Xanthomonas: understanding bacterium–plant interactions. Nat Rev Microbiol. 2011;9:344–55.

5. Bart, R. *et al.* High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. U.S.A. 2012,* 109, E1972-1979.
6. Fonseca, N. P. *et al.* Analyses of Seven New Genomes of Xanthomonas citri pv. aurantifolii Strains, Causative Agents of Citrus Canker B and C, Show a Reduced Repertoire of Pathogenicity-Related Genes. *Front. Microbiol. 2019,* 10.
7. Assis, R. de A. B. *et al.* Identification and analysis of seven effector protein families with different adaptive and evolutionary histories in plant-associated members of the Xanthomonadaceae. *Scientific Reports. 2017,* 7, 16133.
8. Ruh M, Briand M, Bonneau S, Jacques M-A, Chen NWG. Xanthomonas adaptation to common bean is associated with horizontal transfers of genes encoding TAL effectors. BMC Genomics. 2017;18:670.
9. Potnis N, et al. Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. BMC Genomics. 2011;12:146.
10. Li T, et al. Comparative genomic analysis of two *Xanthomonas oryzae* pv. *oryzae* strains isolated from low land and high mountain paddies in Guangxi, China. *Front Microbiol*. 2022;867633.
11. Gordon JL, et al. Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. *BMC Genomics*. 2015;1098.
12. Wu Y, et al. Complete genome sequence of *Xanthomonas citri* subsp. *citri* CQ13, an alternative model strain to study citrus bacterial canker in China. PhytoFrontiers. 2023;484–6.
13. Moreira LM, et al. Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of Xanthomonas fuscans subsp. aurantifolii. BMC Genomics. 2010;11:238.
14. Midha S, et al. Population genomic insights into variation and evolution of *Xanthomonas oryzae* pv. *oryzae*. *Sci Rep*. 2017;40694.
15. Chien C-C, Chou M-Y, Chen C-Y, Shih M-C. Analysis of genetic diversity of *Xanthomonas oryzae* pv. *oryzae* populations in Taiwan. *Sci Rep*. 2019;316.
16. Chen D, et al. Comparative genomic analysis of *Xanthomonas campestris* pv. *campestris* isolates BJSJQ20200612 and GSXT20191014 provides novel insights into their genetic variability and virulence. Front Microbiol. 2022. https://doi.org/10.3389/fmicb.2022.833318.
17. Bellenot C, Carrère S, Gris C, Noël LD, Arlat M. Genome sequences of 17 strains from eight races of *Xanthomonas campestris* pv. *campestris*. Microbiol Resour Announc. 2022;e00279-22.
18. Bianco, M. I. *et al.* Genomic and phenotypic insight into Xanthomonas vesicatoria strains with different aggressiveness on tomato. *Front Microbiol, 2023,* 14, 1185368.
19. Assis, R. A. B. *et al.* A comparative genomic analysis of Xanthomonas arboricola pv. juglandis strains reveal hallmarks of mobile genetic elements in the adaptation and accelerated evolution of virulence. *Genomics, 2021,* 113, 2513–2525.
20. Garita-Cambronero, J., Palacio-Bielsa, A., López, M. M. & Cubero, J. Comparative Genomic and Phenotypic Characterization of Pathogenic and Non-Pathogenic Strains of Xanthomonas arboricola Reveals Insights into the Infection Process of Bacterial Spot Disease of Stone Fruits. *PLOS ONE, 2016,* 11, e0161977.
21. Fernandes, C. *et al.* Comparative Genomics of Xanthomonas euroxanthea and Xanthomonas arboricola pv. juglandis Strains Isolated from a Single Walnut Host Tree. *Microorganisms, 2021,* 9, 624.
22. Gétaz, M., Blom, J., Smits, T. H. M. & Pothier, J. F. Comparative genomics of Xanthomonas fragariae and Xanthomonas arboricola pv. fragariae reveals intra- and interspecies variations. *Phytopathology Research. 2020,* 2, 17.
23. Zarei, S. et al. Taxonomic Refinement of Xanthomonas arboricola. *Phytopathology®* , 2022, 112, 1630–1639.
24. Boudon, S., Manceau, C. & Nottéghem, J.-L. Structure and Origin of Xanthomonas arboricola pv. pruni Populations Causing Bacterial Spot of Stone Fruit Trees in Western Europe. *Phytopathology, 2005,* 95, 1081–1088.
25. 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, 4651–4668.
26. Merda, D. *et al.* Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems. *Environmental Microbiology Reports, 2016,* 8, 572–581.
27. Palacio-Bielsa, A., Roselló, M., Cambra, M. A. & López, M. M. First Report on Almond in Europe of Bacterial Spot Disease of Stone Fruits Caused by Xanthomonas arboricola pv. pruni. *Plant Dis, 2010,* 94, 786.
28. Vauterin L, Hoste B, Kersters K, Swings J. Reclassification of *Xanthomonas*. Int J Syst Evol Microbiol. 1995;45:472–89.
29. Pierce NB. Walnut bacteriosis. Bot Gaz. 1901;31:272–3.
30. Smith, E. F. Observations on a hitherto unreported bacterial disease the cause of which enters the plant through ordinary stomata. *Wellcome Collection, 1903,* 1854–1927.
31. Miller, P. W., Bollen, W. B., Simmons, J. E., Gross, H. N. & Barss, H. P. The pathogen of Filbert bacteriosis compared with Phytomonas juglandis, the cause of Walnut blight. - University of California Davis. 194 AD, 30, 713–733.
32. Wu M, Eisen JA. A simple, fast, and accurate method of phylogenomic inference. Genome Biol. 2008;9:R151.
33. Tonkin-Hill G, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020;21:180.
34. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. https://doi.org/10.1093/bioinformatics/btz848.
35. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
36. Doyle EL, et al. TAL effector-nucleotide targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res. 2012;40:W117–22.
37. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27:135–45.
38. Zhang Y, et al. DeepSecE: a deep-learning-based framework for multiclass prediction of secreted proteins in Gram-negative bacteria. Research. 2023;6:0258.
39. Rives A, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci U S A. 2021;118:e2016239118.
40. Setubal, J. C., Almeida, N. F. & Wattam, A. R. Comparative Genomics for Prokaryotes. in *Comparative Genomics: Methods and Protocols* (eds Setubal, J. C., Stoye, J. & Stadler, P. F.) 55–78 (Springer, New York, NY, 2018). https://doi.org/10.1007/978-1-4939-7463-4_3.
41. Camacho C, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
42. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
44. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52.
45. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
46. Deloger M, El Karoui M, Petit M-A. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. J Bacteriol. 2009;191:91–9.
47. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
48. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31.
49. Zheng J, et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51:W115–21.
50. Hegemann JD, et al. Xanthomonins I-III: a new class of lasso peptides with a seven-residue macrolactam ring. Angew Chem Int Ed Engl. 2014;53:2230–4.
51. Leyns F, Cleene MD, Swings J-G, De Ley J. The host range of the genus *Xanthomonas*. Botanical Review. 1984;50:308–56.
52. Kałużna M, et al. *Xanthomonas arboricola* pv. *juglandis* and pv. *corylina*: Brothers or distant relatives? Genetic clues, epidemiology, and insights for disease management. Mol Plant Pathol. 2021;22:1481–99.
53. Pena MM, et al. Genetic and functional diversity help explain pathogenic, weakly pathogenic, and commensal lifestyles in the genus *Xanthomonas*. Genome Biol Evol. 2024;16:evae074.
54. Bornstein K, Gryan G, Chang ES, Marchler-Bauer A, Schneider VA. The NIH comparative genomics resource: addressing the promises and challenges of comparative genomics on human health. BMC Genomics. 2023;24:575.
55. Medina CA, et al. The role of type III effectors from *Xanthomonas axonopodis* pv. *manihotis* in virulence and suppression of plant immunity. Mol Plant Pathol. 2017;19:593–606.
56. Long, J. *et al.* Non-TAL Effectors From Xanthomonas oryzae pv. oryzae Suppress Peptidoglycan-Triggered MAPK Activation in Rice. *Front. Plant Sci. 2018,* 9.

57. Zhang B, Han X, Yuan W, Zhang H. TALEs as double-edged swords in plant–pathogen interactions: progress, challenges, and perspectives. Plant Commun. 2022;3:100318.
58. Zárate-Chaves CA, et al. TAL Effector Repertoires of Strains of Xanthomonas phaseoli pv. manihotis in Commercial Cassava Crops Reveal High Diversity at the Country Scale. Microorganisms. 2021;9:315.
59. Denancé N, et al. Two ancestral genes shaped the *Xanthomonas campestris* TAL effector gene repertoire. New Phytol. 2018;219:391–407.
60. Jiang S. Genome-wide profiling and phylogenetic analysis of the SWEET sugar transporter gene family in walnut and their lack of responsiveness to *Xanthomonas arboricola* pv. *juglandis* infection. Int J Mol Sci. 2020. https://doi.org/10.3390/ijms21041251.
61. Morbitzer R, Römer P, Boch J, Lahaye T. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A. 2010;107:21617–22.
62. Sileshi GW, Gebeyehu S. Emerging infectious diseases threatening food security and economies in Africa. Glob Food Secur. 2021;28:100479.
63. Luneau JS, et al. Genome-wide identification of fitness determinants in the *Xanthomonas campestris* bacterial pathogen during early stages of plant infection. New Phytol. 2022;236:235–48.
64. Gordils-Valentin L, Ouyang H, Qian L, Hong J, Zhu X. Conjugative type IV secretion systems enable bacterial antagonism that operates independently of plasmid transfer. Commun Biol. 2024;7:1–14.
65. Cenens W et al. Bactericidal type IV secretion system homeostasis in Xanthomonas citri. PLoS Pathog. 2020 16, e1008561.
66. Pfeilmeier S, et al. Leaf microbiome dysbiosis triggered by T2SS-dependent enzyme secretion from opportunistic Xanthomonas pathogens. Nat Microbiol. 2024;9:136–49.
67. Wang L, Jiang S, Deng Z, Dedon PC, Chen S. DNA phosphorothioate modification—a new multi-functional epigenetic system in bacteria. FEMS Microbiol Rev. 2018;43:109–22.
68. Chen NWG, et al. Horizontal gene transfer plays a major role in the pathological convergence of Xanthomonas lineages on common bean. BMC Genomics. 2018;19:606.
69. Greenspan RJ. The flexible genome. Nat Rev Genet. 2001;2:383–7.
70. Merda D, et al. Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in Xanthomonas plant pathogens. Mol Ecol. 2017;26:5939–52.
71. Meline V, et al. Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of Xanthomonas. Mol Plant Pathol. 2018;20:33–50.
72. Diepold, A. & Armitage, J. P. Type III secretion systems: the bacterial flagellum and the injectisome. *Philos Trans R Soc Lond B Biol Sci*, *2015*, 370, 20150020.
73. Liang X, et al. Flagella-Driven Motility Is Critical to the Virulence of Xanthomonas fragariae in Strawberry. Plant Dis. 2023;107:3506–16.
74. Van Alst NE, Picardo KF, Iglewski BH, Haidaris CG. Nitrate sensing and metabolism modulate motility, biofilm formation, and virulence in *Pseudomonas aeruginosa*. Infect Immun. 2007;75:3780–90.
75. Tuleski TR, et al. Cellulose production increases sorghum colonization and the pathogenic potential of *Herbaspirillum rubrisubalbicans* M1. Sci Rep. 2019;9:4041.
76. Solbiati JO, et al. Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic Microcin J25. J Bacteriol. 1999;181:2659–62.

Publisher's Note