Genomics 113 (2021) 2513–2525

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

A comparative genomic analysis of *Xanthomonas arboricola* pv. *juglandis* strains reveal hallmarks of mobile genetic elements in the adaptation and accelerated evolution of virulence

Renata A.B. Assis [a,b], Alessandro M. Varani [c], Cintia H.D. Sagawa [b], José S.L. Patané [d], João Carlos Setubal [e], Guillermo Uceda-Campos [e], Aline Maria da Silva [e], Paulo A. Zaini [b], Nalvo F. Almeida [f], Leandro Marcio Moreira [a,g,*], Abhaya M. Dandekar [b,**]

[a] Center of Research in Biological Science, Federal University of Ouro Preto, Ouro Preto, MG, Brazil
[b] Department of Plant Sciences, University of California, Davis, CA, USA
[c] Faculty of Agricultural and Veterinary Sciences of Jaboticabal (FCAV), Universidade Estadual Paulista (UNESP), Department of Technology, Jaboticabal, SP, Brazil
[d] Cell Cycle Laboratory, Butantan Institute, Sao Paulo, SP, Brazil
[e] Department of Biochemistry, Chemistry Institute, University of Sao Paulo, Sao Paulo, SP, Brazil
[f] School of Computing, Federal University of Mato Grosso do Sul, Mato Grosso do Sul, MS, Brazil
[g] Department of Biological Science, Institute of Exact and Biological Science, Federal University of Ouro Preto, Ouro Preto, MG, Brazil

ARTICLE INFO

Keywords:
T3SS effectors
Copper resistance
Lateral gene transfer
Mobile genetic elements
Replicative transposition
Genome evolution

ABSTRACT

Xanthomonas arboricola pv. *juglandis* (Xaj) is the most significant aboveground walnut bacterial pathogen. Disease management uses copper-based pesticides which induce pathogen resistance. We examined the genetic repertoire associated with adaptation and virulence evolution in Xaj. Comparative genomics of 32 Xaj strains reveal the possible acquisition and propagation of virulence factors via insertion sequences (IS). Fine-scale annotation revealed a Tn*3* transposon (Tn*Xaj417*) encoding copper resistance genes acquired by horizontal gene transfer and associated with adaptation and tolerance to metal-based pesticides commonly used to manage pathogens in orchard ecosystems. Phylogenomic analysis reveals IS involvement in acquisition and diversification of type III effector proteins ranging from two to eight in non-pathogenic strains, 16 to 20 in pathogenic strains, besides six other putative effectors with a reduced identity degree found mostly among pathogenic strains. Yersiniabactin, *xopK*, *xopAI*, and antibiotic resistance genes are also located near ISs or inside genomic islands and structures resembling composite transposons.

1. Introduction

Walnut blight (WB) is a bacterial disease that reduces walnut orchard productivity and negatively influences the quality of the harvested nuts, especially in wet years and on early-leafing varieties [1]. This disease is caused by *Xanthomonas arboricola* pv. *juglandis* (Xaj), a gamma proteobacterium that attacks various tissues of English walnuts, including catkins, pistillate flowers, fruit, green shoots, vegetative buds, and leaves [2]. Diseased fruits are the primary cause of economic loss. The pathogen overwinters in twig lesions, buds, and diseased fruit that remain on the tree [3].

Disease management in both conventional and organic orchards

Abbreviations: ANI, average nucleotide identity; C, C-terminal; GI, genomic island; HGT, horizontal gene transfer; IR, inverted repeats; IS, insertion sequence; LPS, lipopolysaccharides; MGEs, mobile genetic elements; ML, maximum likelihood; MLSA, multilocus sequence analysis; NCBI, National Center for Biotechnology Information; NI, non-identified; NRPS, non-ribosomal peptide synthetases; ORF, open reading frame; PecA, pectate lyase; PKS, polyketide synthases; res, resolution site; RND, resistance-nodulation-division; SWEET, sugars will eventually be exported transporters; T3e, Type III effectors; T3SS, Type III secretion system; T4SS, Type IV secretion system; TALEs, transcription activator-like effectors; TBDR, TonB-dependent receptors; Tn, transposon; TSD, target site duplication; VOC, vertical oozing canker; WB, walnut blight; Xaj, *Xanthomonas arboricola* pv. *juglandis*; Ψ, a pseudogene with premature stop codon or frameshift.

* Correspondence to: L M Moreira, Departamento de Ciências Biológicas (DECBI), Instituto de Ciências Exatas e Biológicas (ICEB), Universidade Federal de Ouro Preto (UFOP), Ouro Preto, MG, Brazil.
** Correspondence to: A M. Dandekar, Department of Plant Sciences, University of California, Davis, CA, USA.
E-mail addresses: lmmorei@gmail.com (L.M. Moreira), amdandekar@ucdavis.edu (A.M. Dandekar).

https://doi.org/10.1016/j.ygeno.2021.06.003
Received 9 November 2020; Received in revised form 1 March 2021; Accepted 1 June 2021
Available online 3 June 2021

requires preventive sprays of copper pesticides to protect green tissue. However, this extensive application has led to soil contamination and the emergence of copper-resistant strains [4,5]. Expanding resistance to copper-based biocides in natural populations of Xaj increases reliance on ethylene bis-dithiocarbamate (EDBC) fungicides as a complement for disease control, negatively affecting the environment [6]. Due to their toxicity and impact on natural ecosystems, the viability of these compounds' continued registration is of immediate and long-term concern for growers. Intensive copper use led to the selection of resistant strains such as *X. arboricola* pv. *juglandis* 417 (Xaj417), isolated from a blighted cv. Chandler walnut fruit in Chico, California [2]. Previous studies determined that copper resistance in Xaj is conferred by *copLAB* and *copABCD* genes present on the chromosome or a plasmid [7,8]. Selection of resistant strains of other *Xanthomonas* species has also been reported under the same management strategy [8,9]. Copper resistance was first observed in xanthomonads in 1994 in Argentina, in *X. citri* pv. *citri* (Xcc), and later on the French islands of Réunion and Martinique [10,11]. These initial reports were followed by characterization of a transmissible plasmid in Xcc carrying a copper resistance operon as passenger genes on mobile genetic elements (MGEs) similar to Tn*3*-like transposons, known to mediate the transfer antibiotic resistance [12–14]. Together, these findings indicate that lateral gene transfer as a key mechanism by which resistance to copper is rapidly acquired within the Xanthomonadaceae family [15].

In addition to copper resistance, comparative genomics has revealed several molecular mechanisms associated with the accelerated evolution and adaptation of genus *Xanthomonas* to diverse agricultural environments and host plants, such as proteins associated with the Type III secretion system (T3SS) [16,17]. T3SS is a highly conserved protein secretion system with structural components encoded by a cluster of hypersensitive response and pathogenicity (*hrp*) genes [18]. T3SS injects unfolded proteins referred to as Type III effectors (T3e) directly inside host cells. Once inside host cells, T3e disrupt host defense signaling pathways and induce disease by interfering with host cellular functions [19–21]. In many plant-pathogenic bacteria, T3e can be recognized by host proteins, activating plant defenses in some interactions and suppressing host defense responses in others, thus limiting or enlarging the pathogen's host range and pathogenic responses [19–23].

Xanthomonas genomes contain many different MGEs, such as insertion sequences (ISs), transposons (Tn), Genomic Islands (GIs) and plasmids [24]. These can be associated with dissemination of T3e and virulence factors [25,26], promoting genetic variation and shaping the genome structure of different strains. For instance, previous studies demonstrated that *Xanthomonas* genomes may carry GIs related to pathogenicity and primary metabolism [27]. GIs have an important role in bacterial fitness and evolution, since they carries accessory genes promoting genome expansion and rearrangements, contributing to microbial adaptation to new environments and emergence of pathogenic strains [28]. Although the complete genomes of other *Xanthomonas* pathovars were comparatively analyzed previously [29–33], these analyses were restricted to just a few *X. arboricola* strains such as pathotypes *X. arboricola* pv. *juglandis* and *X. arboricola* pv. *pruni* [34,35]. Studies comparing and classifying the genetic diversity within *X. arboricola* used a multilocus sequence analysis (MLSA) approach. This type of analysis is limited in its scope as only a few housekeeping genes are used (*fyuA*, *gyrB*, *rpoD*, *atpD*, *dnaK*, *efp* and *glnA*) [7,36,37] instead of the whole genome. Although useful to group closely related strains, this approach lacks the resolution achieved by using complete genome sequences to analyze T3e genes. Here we present a genome-wide comparison of 32 Xaj strains with publicly available sequences, including pathogenic and non-pathogenic strains, to verify the role of horizontal gene transfer on Xaj genomes' evolution. This analysis enabled us to identify genes responsible for the accelerated adaptation to copper-based disease management and evolution of virulence focusing on Xaj417 as the model genome.

2. Results

2.1. Genomic and phylogenomic profile

Thirty-two Xaj genomes were used in this study (Supplementary Table S1). Using bioinformatics and comparative genomics and annotation approaches, we could not detect plasmids in any of the 32 analyzed genomes (Supplementary Table S1). The comparison of all genomes revealed a core genome of 3080 genes (Fig. 1A). A total of 134,871 accessory genes shared between a minimum of two and a maximum of 31 genomes. Unique protein-coding genes ranged from nine to 240 genes per strain, with the least-virulent strains presenting the highest number of unique genes. Functional analysis of these unique genes revealed that genes encoding insertion sequence (ISs) transposases are more common among the more virulent genomes. In the less-virulent genomes, genes encoding proteins without predicted function are more frequent (Fig. 1B). Most of the exclusive genes are classified as pseudogenes due to the presence of a frameshift or a point mutation. Finally, the phylogenetic tree based on the core genome of 3080 orthologous groups (Fig. 1C) showed that most branches have high bootstrap values, excepting branches containing some Chilean strains (CFSAN033078–83, CFSAN033085, CFSAN033087, and CFSAN033088). Also, this phylogenetic tree separates pathogenic from non-pathogenic or low-virulence strains. Global analysis of gene organization, GC% and GC skew content using the copper-resistant strain Xaj417 genome as a reference revealed 23 possible atypical regions that may represent genomic islands (Fig. 2 and Supplementary Table S2). One of these atypical regions illustrates a prophage insertion (GI-3) and another region, GI-19, is characterized as a Tn*3*-like transposable element (Tn*Xaj417*) carrying the copper resistance operon. Interestingly, most of these atypical regions are found exclusively on Xaj417. The prophage insertion is an additional copy of the Type IV secretion system (T4SS) within a genomic island in Xaj417 shared with the two Chilean strains from the same clade (GI-3, Supplementary Table S2).

2.2. Expansion of insertion sequences (IS) and genomic islands (GIs) in pathogenic strains

The Xaj genomes are composed, on average, by 0.8% of IS elements. Interestingly, while pathogenic strains retain more IS copies, representing up to 2.6%, in the non-pathogenic strains, the IS copies span up to 0.8% of each genome (Supplementary Table S3). Analysis of IS composition among the Xaj genomes revealed that IS*3*_gr_IS*407* is the only family present in all 32 genomes (Fig. 3). Pathogenic strains showed IS enrichment, mostly from the IS*4*, IS*5* and IS*L3* families. These strains have a core of eight IS families (IS*3*_gr_IS*51*/gr_IS*407*, IS*4*_gr_IS*10*/gr_IS*4*, IS*5*_gr_IS*427*/gr_IS*5*, IS*L3* and IS*110*) that are distributed among 17 different sub-families, compared to three IS families on average in non-pathogenic strains (Fig. 3). Our reference strain (Xaj417) has eight IS families, two of which are shared only with two other Xaj strains each: IS*3*_gr_IS*150* is present in strains CFBP7179 and 7653 while Tn*3* is also present in CFSAN033079 and CFBP1022. These IS and GIs carry many virulence and adaptation factors present within or adjacent to them, contributing significantly to Xaj genome evolution.

2.3. Type III effectome analysis among Xaj genomes

A presence/absence analysis based on tBlastn searches of 63 T3e (Supplementary Table S4) was carried out to understand their evolution in different Xaj lineages and the distribution of T3e among them (Fig. 4A). While the 22 most virulent genomes contain between 16 and 20 T3e, 10 of the non-pathogenic strains (including 424, a less-virulent strain) contained only two to eight effectors. Six of these strains lack genes from the T3SS apparatus and *hpaB*, which is an exit control protein by this pilus structure [38].

The most virulent strains shared sixteen T3e genes: *xopAZ, xopR,*

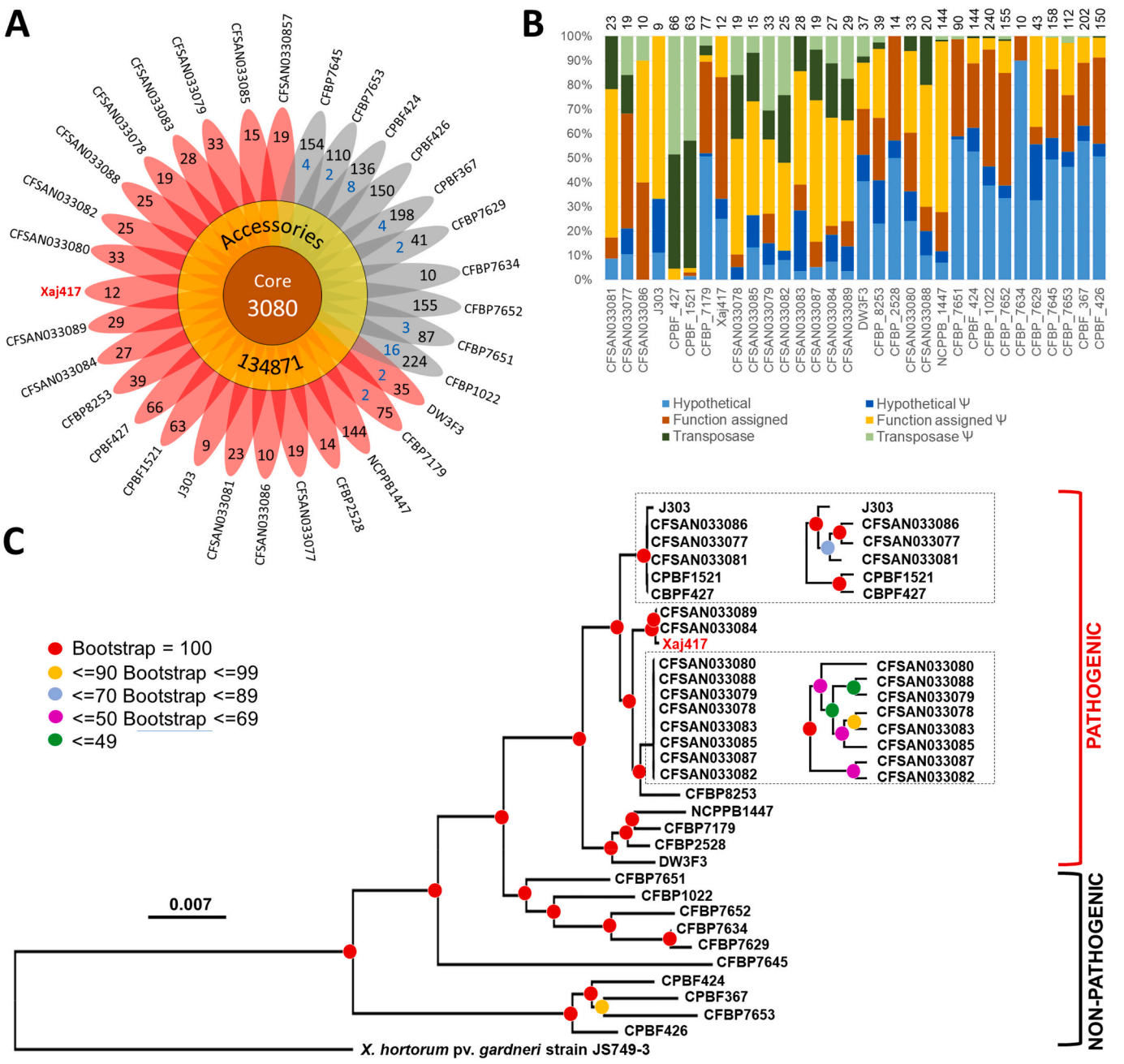

Fig. 1. Comparative analysis of 32 Xaj genomes. (A) Flower plot representing the protein-coding genes shared among Xaj genomes. The orthologous groups shared among all strains (core genome) are indicated in the center. The yellow halo represents the number of accessory genes. The numbers close to the strain codes represent each genome's unique genes (a single copy is shown in black and paralogs in blue). The degree of virulence of each strain is indicated by gray for less and red for more virulent. (B) Stacked bar graph showing the functional annotation percentage of unique genes and pseudogenes (Ψ). Total unique genes in each genome are shown on top. (C) The rooted phylogenetic tree using *X. hortorum* pv. *gardneri* JS749–3 as the outgroup is based on the core genome alignments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

xopAW, avrBs2, hpaA, hrpW, xopF1, xopM, xopZ1, avrXccA2, xopL, xopK, xopN, xopQ, xopX and *xopV* (Fig. 4A). The *xopAI gene* was identified in the American Xaj417, the Chinese DW3F3, the French CFBP7179 and 12 of the 14 Chilean strains (excluding CFSAN033080 and CFSAN033088). The *xopG2*, *xopH1*, and *xopC1* genes were found only in four Chilean strains (J303, CFSAN033086, CFSAN033077, and CFSAN033081) and two Portuguese strains (CPBF1521 and CPBF427), which together form a clade separated from all other strains investigated. Finally, *xopB* was identified only in the French CFBP7179 strain among the most virulent strains, compared to *xopAZ* and *xopR,* which were present in all 32 strains analyzed. Among non-pathogenic strains, the *avrBs2* gene is exclusive to the French strains, which form a single clade (CFBP7651, 1022, 7652, 7634, and 7629). The *hpaA*, *hrpW*, *xopF1* and *xopM* genes were identified only in three French strains (CFBP7651, 1022 and 7652) and the Portuguese strain CPBF424, which in turn is the only pathogenic strain clustered among the non-virulent strains to have the *xopZ1* gene.

The presence/absence analysis of these genes, compared to the distribution of Xaj genomes in the phylogenetic profile, revealed two significant expansions of the effector protein repertoire (Fig. 4A). The first expansion occurred in the pathogenic clade, with the acquisition of genes *xopZ1, avrXccA2, xopL, xopK, xopN, xopQ, xopX, xopV* and *xopAI*, and the second, with the acquisition of *xopG2, xopH1*, and *xopC1* genes

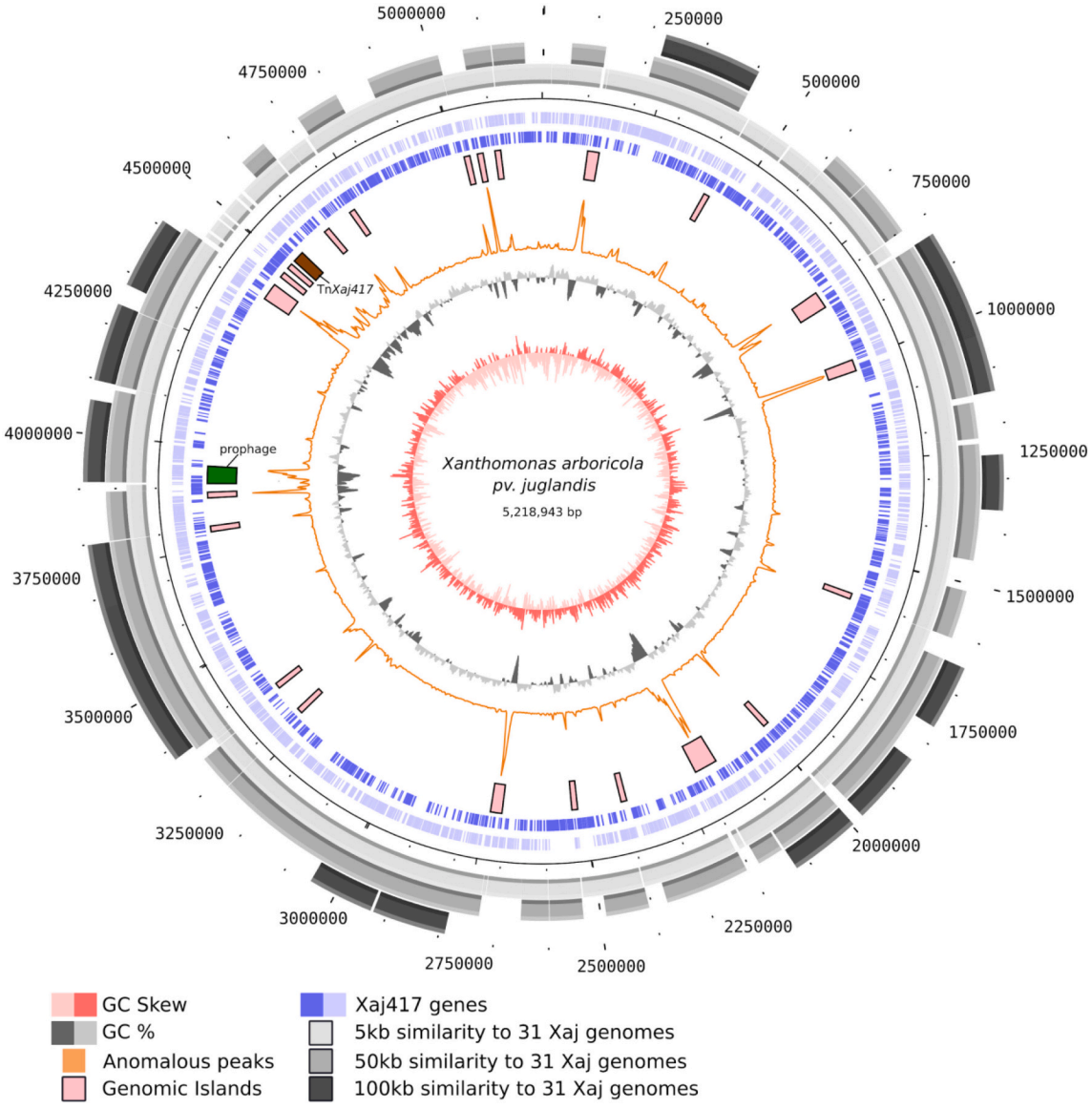

Fig. 2. Circular visualization of predicted genomic islands in the Xaj417 genome (accession number CP012251) based on comparative genome analysis with 31 other *Xanthomonas arboricola* pv. *juglandis* strains. The 23 candidate genomic islands are represented, including one prophage insertion (GI-3) and the Tn*3*-like element Tn*Xaj417* (GI-19). The genome coordinates and details about each GI are described in Supplementary Table S2. Gray segments on the outer tracks represent zones of DNA sequence identity >80% among all strains analyzed.

in the genomes that form the clade represented by strains J303, CFSAN033086, CFSAN033077, CFSAN033081, CPBF1521, and CPBF427. We also confirmed that all Xaj genomes lack transcription activator-like effectors (TALEs), as previously noted for the reference strain Xaj417 [39].

In addition to the effectors described above with E-value ≤1e-50, manual curation of the protein sequences identified five other putative effectors: *avrXccA1*, *xopAH*, *xopAL1*, *xopAY* and *xopG1* (Fig. 4A), mostly present in pathogenic strains. A detailed analysis of the alignment, conservation degree of the sequence and region of insertion into the genome was performed (Fig. 4B). A small portion of the AvrXccA1 carboxy-terminal region showed 87% identity with an avirulence protein (AKJ12_19725) annotated as a pseudogene in all strains except Xaj417, CFBP1022, and CFBP7645. In contrast, the amino-terminal region showed low identity with hypothetical protein AKJ12_RS08330. XopAH showed a 29% identity with a protein annotated as AvrB3 (WP_016902298.1), while XopAL1 showed 39% identity with the carboxy-terminal region of a protein annotated as T3e (AKJ12_RS23955) and flanked by IS*1403* and IS*5* family transposases. The XopAY carboxy-terminal region showed an identity of 45 and 52%, respectively, with two hypothetical proteins (WP_026064526.1 and WP_126750867.1), whose genes are arranged in tandem next to XopQ (AKJ12_05475). This gene is located upstream of a hypothetical protein, AKJ12_RS22130, that showed 37% identity with XopG1, which in turn is flanked by IS*1403*.

These five effectors identified by manual curation reveal similarities with known T3e, but with low amino acid sequence identity (27 to 52%), indicating that they may be new effectors in Xaj (Fig. 4B). Among the non-pathogenic strains, only three have *hrp/hrc* genes encoding the T3SS apparatus (CFBP1022, 7651 and 7652) and eight conserved T3e (*xopAZ*, *xopR*, *xopAW*, *avrBs2*, *hpaA*, *hrpW*, *xopF1* and *xopM*) [34]. The other six non-pathogenic strains lack the T3SS apparatus and have a maximum of five T3e, validating other studies [40–54]. We propose here a representative model of the T3e from pathogenic and non-pathogenic Xaj strains (Fig. 5).

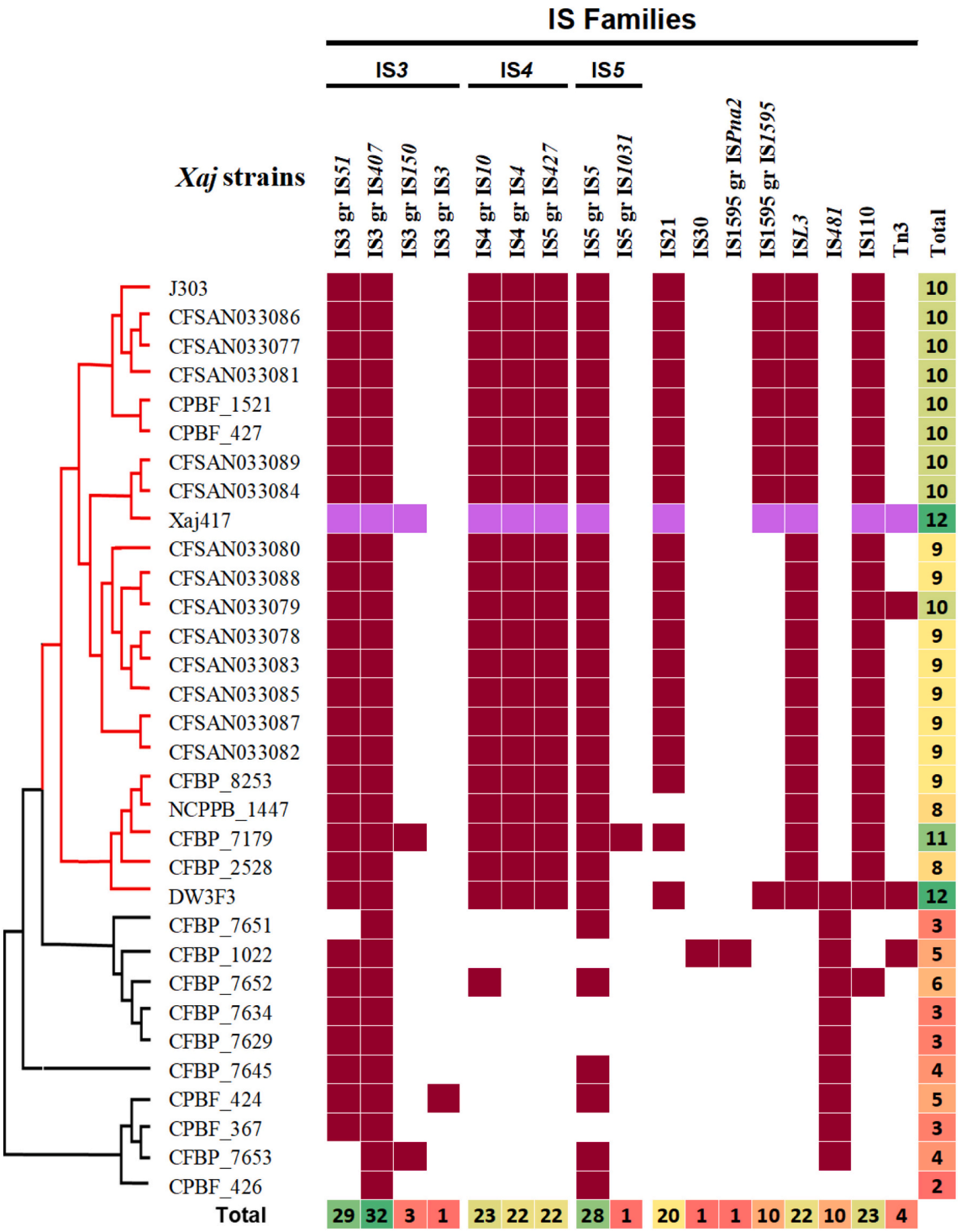

Fig. 3. Analysis of insertion sequence composition among Xaj genomes using the ISsaga tool. The phylogenomic tree is represented on the left side. IS expansion followed T3e expansion in pathogenic strains.

2.4. Copper resistance locus

Genes related to copper homeostasis, including *copL*, *copA*, *copB*, *copM*, *copC*, *copD* and *copF,* formed an operon structure identified in pathogenic strains Xaj417 and CFSAN033079 (Fig. 6). As previously verified in other *Xanthomonas* species [55,56], this copper resistance operon is integrated as passenger genes within a Tn*3* family transposon. This new element carries a copy of tRNA-Ile and is inserted, disrupting a transmembrane protein. It also generates a target site duplication (TSD) of zero and five nucleotides in Xaj417 and CFSAN033079, respectively. Named as Tn*Xaj417* and Tn*XajCFSAN033079*, both elements exhibit typical Tn*3* family features, including the GnGGG motif at the tips of the inverted repeats (IRs), a transposase (TnpA), and two recombination genes: a tyrosine recombinase (TnpT) and a helper protein (TnpS) expressed divergently, with the potential resolution site located downstream of *tnpT*. Interestingly, Tn*Xaj417* exhibits a toxin-antitoxin module that might regulate transposition of the element, while Tn*XajCFSAN033079* lacks one copy of the *copL* gene and has additional

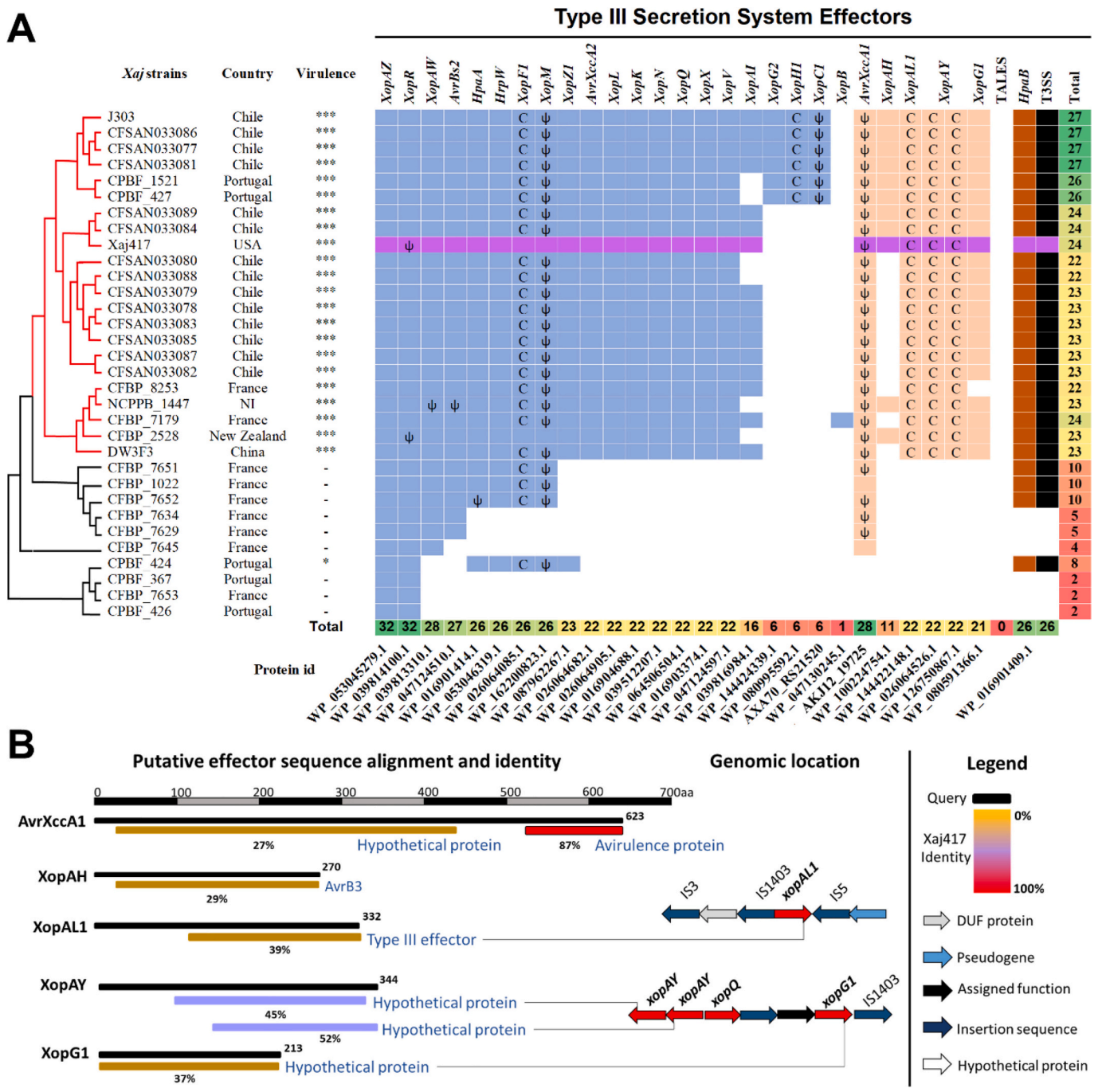

Fig. 4. Effectome of Xaj. (A) Cladogram representing the phylogenetic organization of the Xaj strains presented in Fig. 1C, where red lines denote pathogenic and black lines non-pathogenic strains. For each genome, information about symptom severity in the host (indicated by the number of *) is shown. Colored squares represent the presence of effectors in the matrix: blue squares represent effectors with *E*-value ≤1e-50; beige squares represent effectors identified by manual curation with *E*-value ≤1e-10; brown squares represent the *hpaB* gene that codes for a protein that controls secretion through T3SS; and black squares represent the T3SS apparatus cluster. The total number of effectors by genome (colored columns on the right side) or representativeness in the genomes (colored columns on the bottom) are colour-coded: green represents more abundant and red less abundant. The reference strain Xaj417 is highlighted in purple. (B) T3e with *E*-value ≤1e-10 were manually curated and considered to be putative effectors, according to protein sequence alignment, degree of identity, and genomic location. C: C-terminal; Ψ: a pseudogene with premature stop codon or frameshift; and NI: non-identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

passenger genes related to resistance-nodulation-division (RND) family transporters, which might be involved with the efflux of heavy metals (Fig. 6A).

Both Tn*Xaj417* and Tn*XajCFSAN033079* were registered at the TnCentral database (https://tncentral.proteininformationresource.org/) [83]. Two genes present in these transposons and encoding for copper-binding proteins (*copA* and *copB*) and *copL*, a metal-binding regulatory protein, are duplicated in another genomic region. In contrast with the transposon copies, these other copies are present in all 32 strains (Fig. 6B). The amino acid sequence of the CopA copy within Tn*Xaj417* is very different (only 73% similarity) from the other copy in the Xaj417 chromosome, although the predicted protein structure is conserved (Fig. 6C). For CopB, the amino acid sequence has 64% similarity and 51% identity between the chromosomal and Tn*Xaj417* copies.

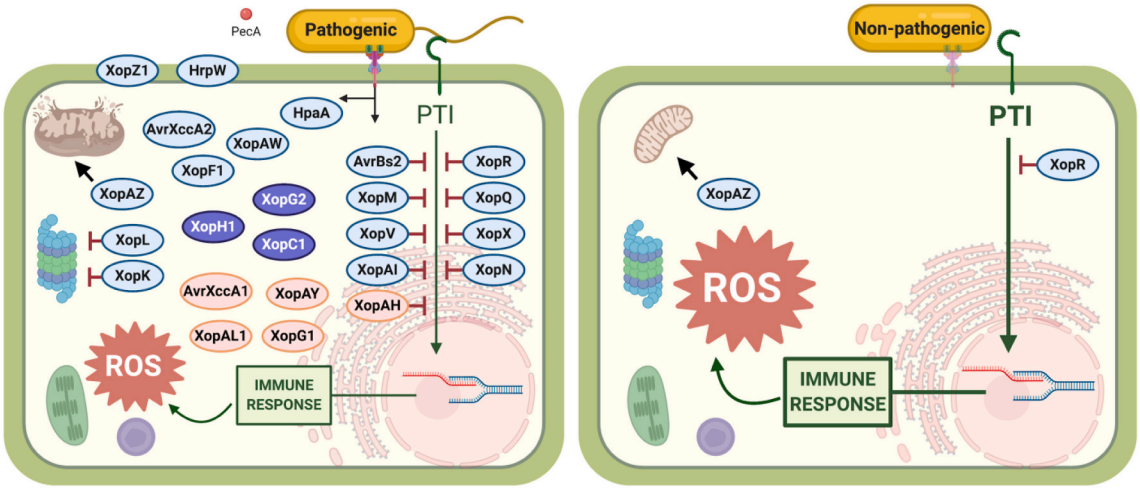

Fig. 5. Comparative distribution of T3e in pathogenic and non-pathogenic Xaj strains. Effectors with *E*-value ≤1e-50 are colored in blue and effectors identified by manual curation with E-value ≤1e-10 are colored in beige. Effectors in purple are only found in the Chilean strains J303, CFSAN033086, CFSAN033077 and CFSAN033081, and in two Portuguese strains, CPBF1521 and CPBF427, which together form a clade separated from all other strains investigated. PecA: pectate lyase. Created with BioRender.com. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We hypothesize that these transposon copies are the result of HGT, as suggested by a synteny analysis (Supplementary Fig. S1).

2.5. Other virulence and adaptation-associated genes

The presence/absence of genes associated with secretory systems and metabolic pathways deemed essential to adaptation and virulence in other *Xanthomonas* spp. were determined for the 32 Xaj strains, using Xaj417 as a reference genome (Fig. 7). All 32 strains have 72 to 82 TonB-dependent receptors (TBDR), Sec and Tat secretion pathways, and the secretory systems T1SS, T2SS, T4SS, but not T6SS. In addition, only strains from the clade containing Xaj417 have, among T1SS genes, an exclusive copy of hemolysin coding gene WP_053045432.1, among the other common copies. All 32 strains have two distinct clusters that code for the T2SS apparatus and genes for flagellum, quorum sensing (*rpf*) and xanthomonadin (*gum*). The non-pathogenic strains lack a large gene cluster associated with flagellum formation. In addition, pectate lyase, cellulose biosynthesis, amino acid transport permease and lipopolysaccharide (LPS) hypervariable region are present in all pathogenic strains and absent in non-pathogenic strains, excepting members of the clade containing CFBP7651, 1022, 7652, 7634 and 7629. Among the non-pathogenic strains that have these genes, only strain 7651 lacks the LPS hypervariable region. Finally, transfer (Tra) conjugative genes are present in all strains from the first two high-virulence clades in the phylogenetic tree, which contain all 14 Chilean strains, CPBF 1521–427, Xaj417, CFBP7179 and DW3F3 (Fig. 7).

2.6. Non-ribosomal peptide synthetases (NRPS)/polyketide synthases (PKS)

Five clusters responsible for the synthesis of xanthomonadin I, xanthoferrin siderophore, and three NRPS/PKS were identified (Supplementary Fig. S2). The first NRPS/PKS (region 4 in Supplementary Fig. S2), identified through antiSMASH [57], is a hybrid NRPS/PKS responsible for yersiniabactin synthesis located in a genomic island. The other two NRPS-predicted clusters show no similarity with any known cluster from the antiSMASH database, suggesting that these NRPS may be found exclusively in the Xaj417 genome (regions 1 and 6 from Supplementary Fig. S2). One of them contains Gramicidin S synthase, identified by IslandViewer [58], in a genomic island that might have been acquired by HGT (region 1 in Supplementary Fig. S2). This cluster is located in region 6 from Alien Hunter, according to Orthologsorter (http://jau.facom.ufms.br/arboricola3/), one of the largest HGT regions of the Xaj417 genome.

IS and transposable elements are directly implicated in the emergence of Xaj417 virulence and pathogenicity. Two of the previously mentioned T3e (XopK and XopAI), a copper resistance locus and the polyketide synthase (PKS) encoding yersiniabactin are located near ISs or inside structures resembling a composite transposon and GIs in the Xaj417 genome (Fig. 8). XopK is located close to a cluster of ISs encompassing many genes related to synthesis of the antibiotic clavulanic acid and fungicide nikkomycin (Fig. 8A). XopAI is surrounded by ISs, forming a composite-like transposon between IS*Xac3* (Fig. 8B). Also, a classic GI structure, carrying IS*Xcd1*/IS*Xca1* composite transposons encoding a tyrosine recombinase and the yersiniabactin gene cluster, is inserted into a tRNA-Arg (Fig. 8C).

2.7. Estimated trees for nine genes associated with IS

Using Xaj417 sequences as a reference, we built individual phylogenetic analyses for both copies of *copA* and *copB* and for the antibiotic-resistance gene cluster from the IS3_gr_IS407 family that encodes for synthesis of the antibiotic clavulanic acid and the fungicide nikkomycin. This cluster is composed of five genes: clavaminate synthase, sulfotransferase, aminotransferase, SanC and ATPase (Fig. 8A). Gene trees had between 85 and 2598 tips (Supplementary Table S5).

3. Discussion

Comparative genomics of closely related bacteria is a powerful approach to gain an insight into the emergence and evolution of pathogenesis [16,30,35]. By examining complete gene sets within genomes from both pathogenic and non-pathogenic isolates, an unbiased phylogenetic tree can be drawn providing insights for specific genetic loci [34,35,40]. We compared the gene repertoires of 32 *Xanthomonas arboricola* pv. *juglandis* genomes to better understand the plant-pathogen interactions that lead to walnut bacterial blight. For over a century, management of xanthomonad diseases have relied on intensive use of copper-based pesticides. While successfully suppressing disease development, these chemicals accumulate in the soil, posing risks to human, plant, and environmental health. There are increasing reports of copper-resistant strains in these pathosystems, exacerbating the need for new management strategies. Intensifying copper use is problematic, given increasing regulatory oversight constraints of residue levels imposed by

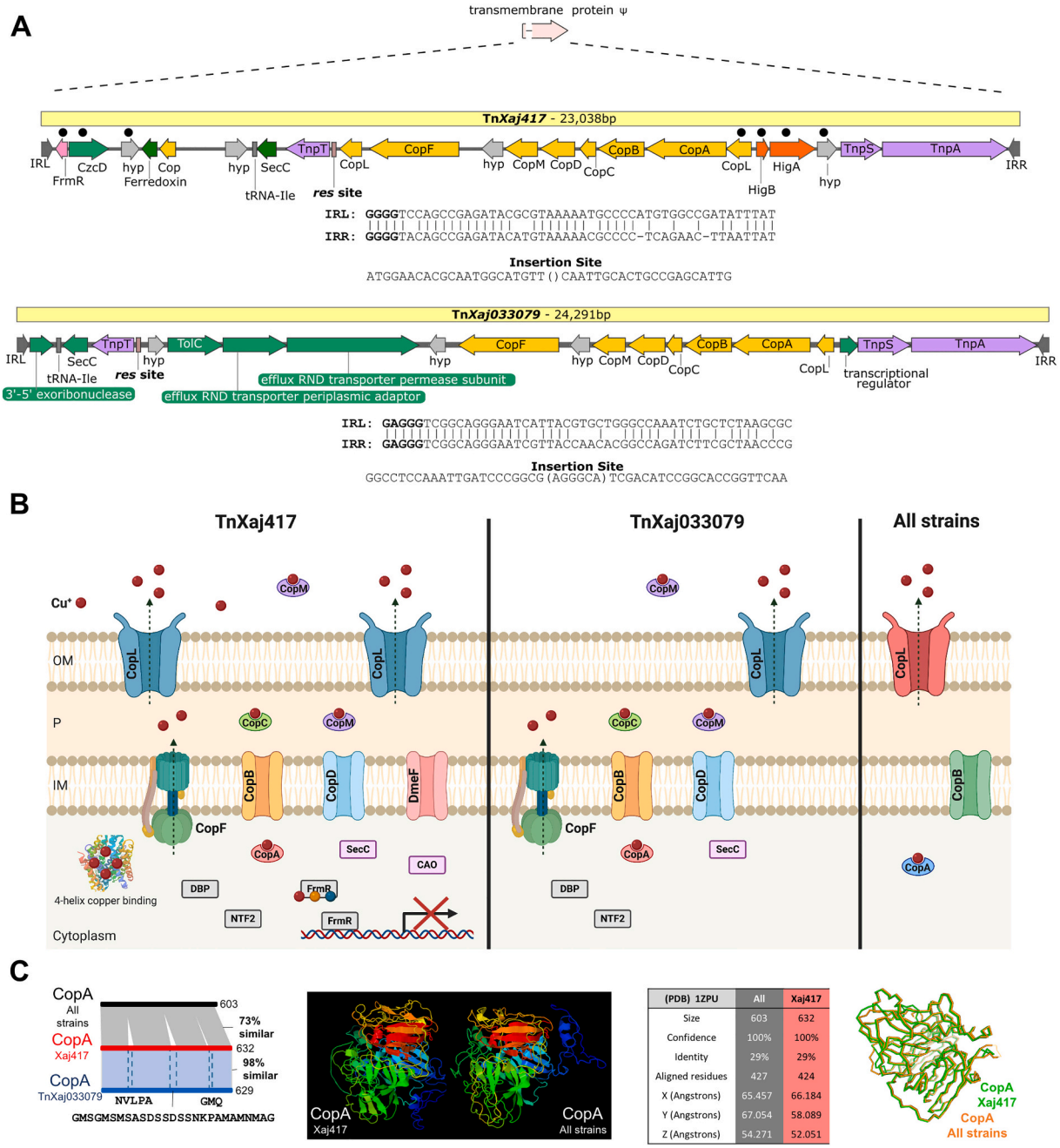

Fig. 6. Copper resistance operon. (A) Copper resistance Tn identified in Xaj417 and in the Chilean strain CFSAN033079. *TnXaj417* exhibits typical Tn*3* family features, including a transposase and two recombination genes, with the potential resolution site located between the recombination genes. Genes in the transposon with a black dot represent singleton genes in Xaj417. (B) Molecular mechanism of copper-resistance in Xaj417 and CFSAN033079, compared to *copLAB* present in all strains. The genes from Tn*Xaj417* and Tn*XajCFSAN033079* are represented on the left and the copies present in all strains on the right side. Xaj417 has a metal/formaldehyde-sensitive transcriptional repressor, four-helix bundle copper-binding proteins, a copper amine oxidase, toxin-antitoxin system HigA and DmeF and a hypothetical protein. CFSAN033079 presents inside the Tn element a set of genes related to the RND family of transporters and a single copy of *copL*. CopA, CopB, CopC and CopM are copper-binding proteins, CopD is a transmembrane transporter protein, CopL is a metal-binding regulatory protein and CopF is a putative copper-transporting p-type ATPase. DmeF: CDF family Co(II)/Ni(II) efflux transporter; SecC: zinc chelation protein; CAO: copper-amine oxidase; DBP: DNA-binding protein; NTF2: nuclear transport factor 2 family; and FrmR: metal/formaldehyde-sensitive transcriptional repressor. Created with BioRender.com. (C) Alignment between *copA* found in all strains and the second copy of *copA* found in strains Xaj417 and CFSAN033079.

the marketplace.

X. arboricola strains comprise pathogens of fruit trees. The pathovars *pruni*, *corylina* and *juglandis* are the most closely related although they present defined groups according to their phytopathogenic specialization [59–61]. Although little is known about the genetic and metabolic basis of host specificities in *X. arboricola* and what led to the emergence of new diseases within this species, the pathogenic strains among different pathovars are closely related to each other despite host specificity similar to what is shown in this study. The bacterial isolates comprising our dataset are representative of Xaj genomes from distinct continents and include both pathogenic and non-pathogenic strains, some of which are copper-resistant. Among pathogenic strains, Xaj417 and CFBP2528 are the causal agent of WB in California and New Zealand respectively, while CFBP7179 and CFBP8253 are responsible for walnut

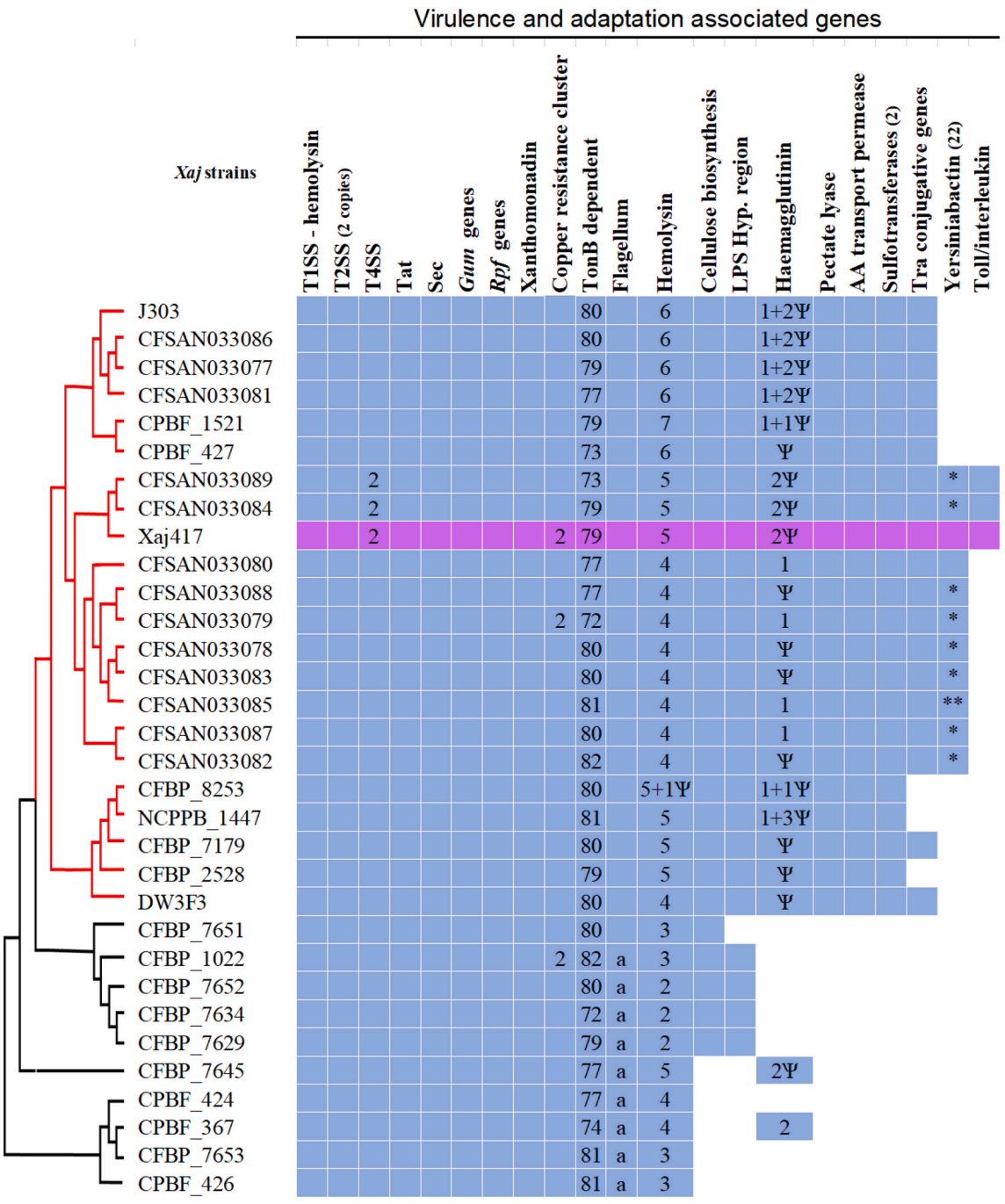

Fig. 7. Presence/absence matrix of genes associated with adaptation and virulence in *Xanthomonas* species. The matrix organization considered the distribution of genomes in the phylogeny of Fig. 1C. a: absence of a large gene cluster associated with flagellum formation; *: absence of a gene in the cluster; and Ψ: gene with premature stop codon or frameshift.

vertical oozing canker (VOC) in France. Symptoms in walnut leaves and fruits are similar to classical bacterial WB, but VOC causes vertical cankers on trunks and branches [34,41]. The other strains responsible for WB disease were isolated from infected walnut trees in southern Chile (14) or Portugal (3), one isolated from blighted immature fruit in China, and NCPPB1447 (missing isolation source). Strain CPBF1521 was isolated from symptomatic leaves of an ornamental walnut in a public site in Portugal [42], together with strains CPBF424 and CPBF367, recently classified as *X. euroxanthea* based on MLSA, average nucleotide identity (ANI) and digital DNA–DNA hybridization [37].

The phylogenetic tree topology of the strains revealed the emergence of the pathogenicity of Xaj genomes, following the virulence pattern described in the literature (Fig. 4) [34,35]. Interestingly, no TALE sequence was found in any of the Xaj strains. This is consistent with reports that activation of sugar transport is not essential to induce virulence in this pathosystem, as described previously for Xaj417 and we now expand to include other Xaj strains. Moreover, 81 TBDR were identified in the Xaj417 genome, which may be associated with specialization to use plant carbohydrates and other available biomolecules [62]. Although the number of TBDR is similar among all strains, the ability to scavenge plant-derived carbohydrates instead of TALE effectors may play a key role in Xaj during its interaction with host plants. There may also be an assimilation of secondary metabolites through TBDRs, including phenolic compounds, that could impair the

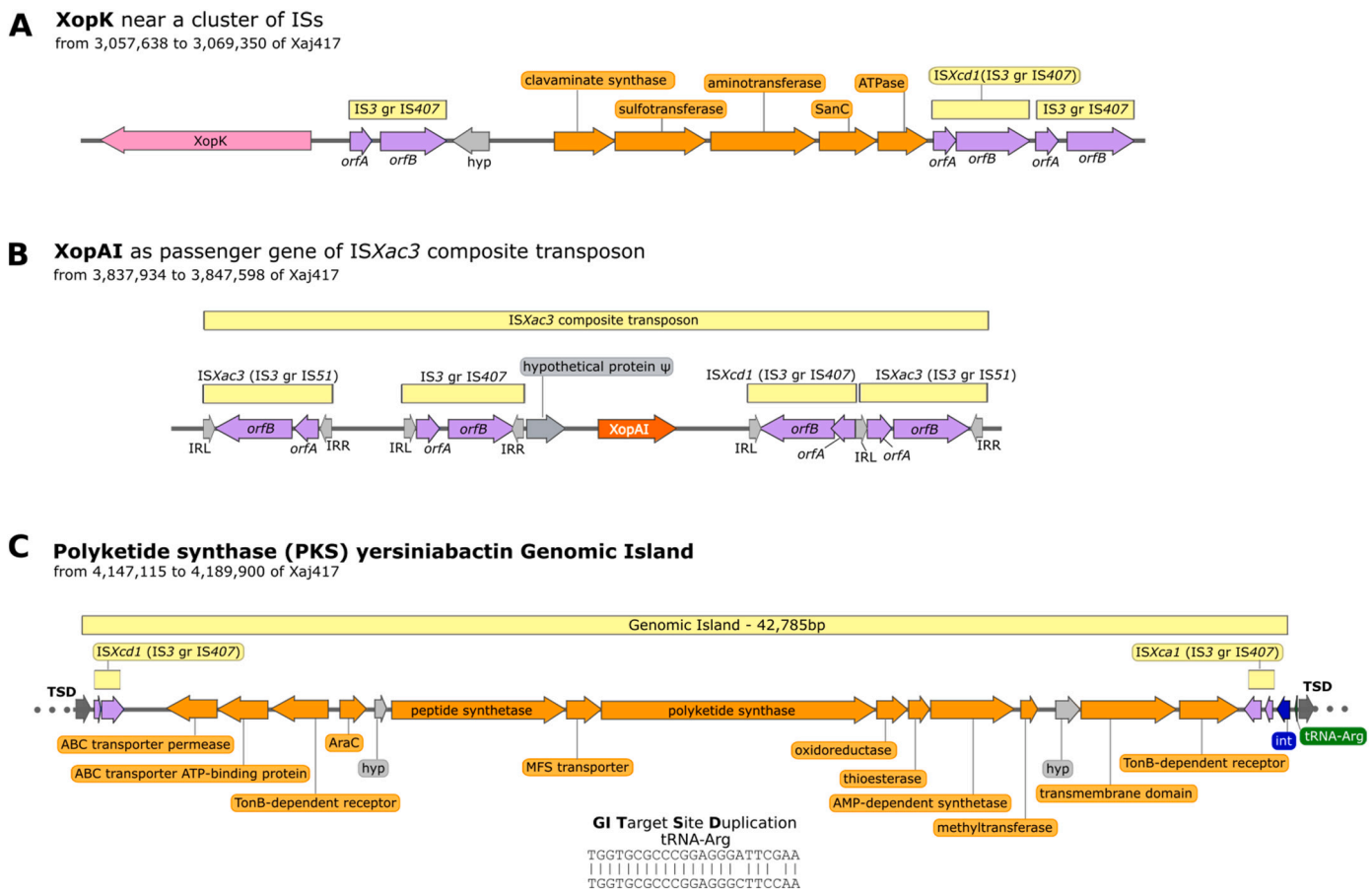

Fig. 8. Genes located near ISs or inside composite-like transposon structures. (A) XopK is located close to a cluster of ISs from the IS*3*_gr_IS*407* family associated with antibiotic and fungicide synthesis genes. (B) XopAI is surrounded by IS*Xac3,* forming a composite transposon between IS*Xac3*. (C) Genomic Island (GI) that codes for a non-ribosomal peptide synthetase named yersiniabactin. The GI carries a putative composite transposon structure formed by IS*Xcd1* and IS*Xca1*, both from the same IS family (IS*3*_gr_IS*407*) and is inserted into a tRNA-Arg, generating a target site duplication (TSD).

defense response and favor disease development.

The pathogenic strains possess a larger repertoire of IS, prophages, and GIs than non-pathogenic strains, corroborating previous results [34] and reinforcing the importance of MGEs in the evolution of pathogenic strains in *X. arboricola* (Fig. 1B). In addition, although the pathogenic strains have less unique genes, they have more transposases what could be explained by spread of adaptive genes through microbial populations via recombination without eliminating genome-wide diversity [63]. Least-virulent strains having a larger number of unique genes is what would be expected under a scenario of hitchhiking effect [64,65], which may have been the case once a strain acquired the copper-resistant genes by HGT. Within this scenario, it is likely that such a genome would rapidly increase its frequency compared to the remaining ones in the same population due to positive selection. This would, in turn, lead to a lower frequency of all other population variants. The end result would be a much lower frequency of genomes with different unique genes in the new genomic pool, than in the original population prior to the HGT event. Furthermore, increase in transposable elements is also associated with lineages experiencing positive selection [63]. Another indicative of the intensity of positive selection acting on the genomes bearing copper-resistant genes is the presence of more pseudogenes in non-pathogenic genomes. This is in agreement with a scenario of recent accelerated genome reduction [63,66], possibly as an arms race for faster replication rates which would also aid in faster spreading [66].

Phylogenomic analyses support the parallel evolution of most virulent strains as they form a paraphyletic group, associated with T3e repertoire expansion and exclusive IS propagation, mostly from the IS*4*, IS*5*, and IS*L3* families (Fig. 3). Interestingly, *xopK* (located in tandem with antibiotic-resistance genes), *xopAI*, a new copper resistant operon, and PKS-related yersiniabactin are located near ISs or inside composite-like transposon structures and GIs. These arrangements exemplify the potential impact of IS and GIs on evolution of the Xaj417 genome (Fig. 8). XopK is located close to a cluster of ISs containing five genes related to antibiotic synthesis (Fig. 8A). Clavulanic acid is an antibiotic effective against bacteria and nikkomycin is a fungicide that acts as a strong competitive inhibitor of chitin synthetases in fungi and insects [67]. In addition, XopAI is in a region under evolutionary pressure, surrounded by an IS forming a composite-like transposon between IS*Xac3*. This composite transposon is formed by two inverted repeats from two separate transposons moving together as one unit and carrying the DNA between them (Fig. 8B).

Fine-scale annotation reveals that the copper resistance operon is integrated as passenger genes of a new Tn*3* family transposon, found exclusively in Xaj417 and the Chilean CFSAN033079 strain (Fig. 6A). Detailed analysis of these lateral transfer islands revealed the presence of important genes that may be associated with adaptation, virulence, and tolerance to metals such as copper. The similarity between the Chilean and Californian strains is not restricted to the Tn3 region, as they form a clade in the phylogenetic tree built with all 3080 genes of the known Xaj core genome (Fig. 1A). Given the geographic isolation of these strains, it is tempting to speculate whether a pathogenic strain was carried from one site to another in infected plant material, as such extensive convergent evolution would be highly unlikely.

Synteny analysis of the chromosomal and transposed *copA* and *copB* genes shows that the evolutionary history of each pair is different suggesting that the transposon copies may be the result of recent HGT. The

chromosomal copies were found in more than 50 other *Xanthomonas* genomes with the same genomic neighborhood. Moreover, genomes other than *Xanthomonas* showed up only when gene sequence similarity was considerably decreased starting with *Xylella* followed by a *Stenotrophomonas* genome. In contrast, the transposon copies were found only in four other *Xanthomonas* genomes, with the fifth being a *Stenotrophomonas* strain, followed by several *Pseudomonas* genomes. In addition, the genomic neighborhood of *copA* and *copB* is different in these other genomes; the only conservation is of the pair itself. In other words, the operon architecture of the transposon copies (Fig. 6A) is unique to Xaj417 and to CFSAN033079 with respect to genomes available at the Integrated Microbial Genomes Resource [68] as of February 2021, adding evidence for the key contribution of this operon to the pathogenicity of the Xaj417 and CFSAN033079 strains. Analysis of the Xaj417 genome also revealed secondary metabolite biosynthesis gene clusters such as the PKS-related siderophore yersiniabactin, and gramicidin S synthase.

Here we performed the first comparative analysis of available sequenced genomes of Xaj and described the secreted effectome and repertoire of genes related to copper tolerance within this genetically diverse group. Sympatric populations may affect genetic dynamics in the emergence of new strains, since pathogenic and non-pathogenic strains are found together on walnut buds [69]. Our analysis also detected important differences among non-virulent strains in which different subsets of the full virulence complement are represented, such as genes for LPS variability and flagellum assembly. Knowledge of the dynamics of virulence determinants within orchard ecosystems can lead to new therapies by developing drugs targeting these pathogen effectors or transgenic plants expressing the host interaction partners, leading to broad-spectrum resistance. These findings also aid development of new molecules to use as antibiotics to potentially reduce dependence on copper-based biocides.

4. Conclusion

Plant pathogenic bacteria cause devastating losses to growers worldwide and many can adapt to new hosts or management strategies with outbreaks of hypervirulent strains. The emergence of new strains is favored in the "open laboratory" of the orchard ecosystem, in which new plant material frequently arrives from nurseries and local microbiota constantly exchange genetic material. Here we showed how mobile genetic elements shape *X. arboricola* pv. *juglandis* genomes and contribute to the evolution of this bacterial group. The genome comparisons detailed pathogen-specific multigene signatures near MGEs. Among the many molecular functions hitchhiking on these MGEs are common virulence strategies such as secreted proteases and pectinases and more adaptive systems such as T3e. The gene shuffling and expanding capabilities of this bacterial group also include acquisition of an additional copper-resistance cluster that enables this pathogen to survive current copper-based management used in walnut orchards. Through the action of MGEs, these new features can be incorporated into the bacterial genome and spread quickly across local orchard populations. Understanding the characteristics and dynamics of these drivers of diversity will improve monitoring systems for emergent diseases and inform target-specific therapeutics.

5. Material and methods

5.1. Genome selections

Thirty-two genomes of Xaj available in the NCBI database as of April 2020 were used for comparative genomic analysis (Supplementary Table S1). Xaj417 was used as the reference in all analyses since its complete genome sequence is deposited in NCBI.

5.2. Protein family's prediction

A platform was developed to analyze Xaj genomes from the NCBI database to identify target proteins associated with WB development to inform new strategies to combat plant disease. Using the Orthologsorter tool [70] available at restricted access domain http://jau.facom.ufms.br/arboricola3/, analysis of the presence or absence of each protein family among all 32 genomes allowed us to identify unique, flexible, and conserved protein families that are shared by a subset of bacterial strains.

5.3. Estimating the phylogenomic tree

The phylogenomic profile generated using Orthologsorter was predicted based on the conserved protein families shared by all 32 strains. First OrthoMCL found all shared protein families. Families with exactly one protein from each genome were aligned using MUSCLE [71] and concatenated. After removal of non-informative columns using GBlocks [72] (columns that may not be conserved, including columns with too many gaps, or that may be saturated by multiple substitutions), the tree was generated using RaxML [73] with the PROTCAT model, rapid bootstrapping with 100 replicates and maximum likelihood search.

5.4. Type III effectome prediction

A set of 63 type III effectors from the *Xanthomonas* Resource website (http://www.xanthomonas.org/tools.html) was analyzed by protein-nucleotide six-frame translation (tBlastn) searches [74] with *E*-value $\leq$1e-50 against the set of 32 Xaj genomes. For homologous genes not identified in the automated analysis, an additional search approach was performed. The protein fasta sequences of the 63 type III effectors were analyzed manually in the Blast search tool of Orthologsorter, a local database containing all proteins from the 32 genomes. Sequences with E-value $\leq$1e-10 were considered putative effectors, and the genomic region was examined for signs and signatures that could strengthen the hypothesis of new effector sequences. Additionally, all known TALEs and their sequences listed at The *Xanthomonas* Resource (http://www.jstacs.de/downloads/All_TALE_DNA_sequences.fasta) were downloaded and a nucleotide Blast (Blastn) search was performed using all 516 known TALEs as the query and all 32 Xaj genomes as the database.

5.5. Genome map and genomic islands prediction

The Xaj417 genome map was constructed using Artemis DNAPlotter and BRIG [75,76]. Genomic islands, phage regions and horizontally transferred regions were predicted as described [28]. Identification and analysis of insertion sequences were carried out with the ISsaga tool [77].

5.6. Gene tree estimation of nine genes associated with IS

To estimate which taxon was the probable source of putative HGTs, we conducted individual phylogenetic analyses for nine genes: two *copA* copies, two *copB* copies, clavaminate synthase, sulfotransferase, aminotransferase, SanC and ATPase (Figs. 6 and 8A). An online tBlastn search in GenBank was done for each Xaj417 query's protein sequence, up to the limit of at most 1000 genomes as subjects (allowing more than one hit per genome, hence a hit set >1000 could be retrieved), or while a subject's E-value was lower than 0.05, whichever was reached first. After retrieving all hits, individual gene multiple alignments were performed in Muscle [71], with default parameters. Subsequently, TrimAl [78] was used to discard alignment sites with defined amino acid states in fewer than four sequences (the minimum number of taxa needed for phylogenetic informativeness), meaning all other sequences had gaps in that site position. Phylogenetic searches were done in IQTree v1.6.12 using the GTR + I + G evolutionary model [79,80]. Rooting was done in

MAD [81]. Finally, the topology was checked in FigTree [82] to locate the sister-group to the clade where Xaj417 belonged.

5.7. Synteny analysis

Using as query each of the *copA* copies in the Xaj417 genome, we obtained the genomic neighborhoods of these genes at the Integrated Microbial Genomes Resource [68], using the bidirectional best hit criterion for determining orthologs.

5.8. Metabolic profile

AntiSMASH [57] was used to analyze secondary metabolite biosynthesis gene clusters and IslandViewer [58] was used to predict genomic islands in Xaj417.

Data availability

Data sharing is not applicable to this article as no datasets were generated during the current study.

Author's contributions

RABA, AMD and LMM designed the work and selected the genomes. RABA, JSLP, JCS, GUC, AMDS, AMV, NFA and LMM performed genome comparisons and identified the protein families. JSLP and NFA designed phylogenetic trees. RABA, CHDS, JSLP, JCS, AMV, PAZ, AMD and LMM interpreted findings. RABA, CHDS, JSLP, JCS, GUC, AMDS, AMV, PAZ, AMD and LMM contributed additional interpretations and general manuscript comments. RABA, CHDS, PAZ, AMD and LMM wrote and revised the paper.

Funding

This work was supported by the following agencies: Coordination for the Improvement of Higher-Level Personnel (CAPES – Brazil) BIGA Project, CFP 51/2013, process 3385/2013, and the California Walnut Board. LMM, JCS, NFA and AMV were funded in part by a research fellowship from CNPq. LMM was funded in part by a research fellowship from PROPP-UFOP. NFA was funded in part by FUNDECT-MS (TOs 141/2016, 007/2015).

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgments

The authors thank Lauro Ângelo de Moraes and Izinara Rosse from the Multi-user Bioinformatics Laboratory at the Federal University of Ouro Preto, Brazil, for technical assistance with the TALEs analysis. We thank Mick Chandler (Georgetown University, USA) and Gipsi Lima Mendes (Université Catholique de Louvain, Belgium) for their assistance in annotating Tn3 elements and toxin and antitoxin systems.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.06.003.

References

[1] C.H.D. Sagawa, et al., Proteome analysis of walnut bacterial blight disease, Int. J. Mol. Sci. 21 (2020) 7453.
[2] U.P. Pereira, et al., Complete genome sequence of *Xanthomonas arboricola* pv. *juglandis* 417, a copper-resistant strain isolated from *Juglans regia* L, Genome Announc. 3 (2015).
[3] S.E. Lindow, et al., Epidemiological approaches to the control of walnut blight disease, Annual Walnut Research Reports (2000) 303–327.
[4] L. Gardan, T. Brault, E. Germain, Copper resistance of *Xanthomonas campestris* pv. *juglandis* in French walnut orchards and its association with conjugative plasmids, Acta Horticulturae (Netherlands) (311) (1993) 259–265. ISHS.
[5] P. Radix, F. Seigle-Murandi, G. Charlot, Walnut blight: development of fruit infection in two orchards, Crop Prot. 13 (1994) 629–631.
[6] J.E. Adaskaveg, et al., Epidemiology and management of walnut blight, California Walnut Board Rep. (2011) 1–14.
[7] D. Giovanardi, et al., Morphological and genotypic features of *Xanthomonas arboricola* pv. *juglandis* populations from walnut groves in Romagna region, Italy, Eur. J. Plant Pathol. 145 (2016) 1–16.
[8] Y.A. Lee, M. Hendson, N.J. Panopoulos, M.N. Schroth, Molecular cloning, chromosomal mapping, and sequence analysis of copper resistance genes from *Xanthomonas campestris* pv. *juglandis*: homology with small blue copper proteins and multicopper oxidase, J. Bacteriol. 176 (1994) 173–188.
[9] D. Giovanardi, D. Dallai, E. Stefani, Population features of *Xanthomonas arboricola* pv. *pruni* from *Prunus* spp. orchards in northern Italy, Eur. J. Plant Pathol. 147 (2017) 761–771.
[10] D. Richard, et al., First report of *Xanthomonas citri* pv. *citri* Pathotype a causing Asiatic Citrus canker in Martinique, France, Plant Dis. 100 (2016) 1946. https://apsjournals.apsnet.org/doi/full/10.1094/PDIS-02-16-0170-PDN.
[11] D. Richard, et al., First report of copper-resistant *Xanthomonas citri* pv. *citri* Pathotype a causing Asiatic Citrus canker in Réunion, France, Plant Dis. 101 (503) (2016).
[12] X.-N. Niu, et al., Complete sequence and detailed analysis of the first indigenous plasmid from *Xanthomonas oryzae* pv. *oryzicola*, BMC Microbiol. 15 (2015) 233.
[13] U. Lauf, C. Müller, H. Herrmann, The transposable elements resident on the plasmids of *Pseudomonas putida* strain H, Tn5501 and Tn5502, are cryptic transposons of the Tn3 family, Mol. Gen. Genet. MGG 259 (1998) 674–678.
[14] S.R. Partridge, Analysis of antibiotic resistance regions in gram-negative bacteria, FEMS Microbiol. Rev. 35 (2011) 820–855.
[15] D. Richard, et al., Adaptation of genetically monomorphic bacteria: evolution of copper resistance through multiple horizontal gene transfers of complex and versatile mobile genetic elements, Mol. Ecol. 26 (2017) 2131–2149.
[16] S.-Q. An, et al., Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*, FEMS Microbiol. Rev. 44 (2020) 1–32.
[17] S. Timilsina, et al., *Xanthomonas* diversity, virulence and plant–pathogen interactions, Nat. Rev. Microbiol. 18 (2020) 415–427.
[18] J.R. Alfano, A. Collmer, The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death, J. Bacteriol. 179 (1997) 5655–5662.
[19] J.R. Alfano, A. Collmer, Type III secretion system effector proteins: double agents in bacterial disease and plant defense, Annu. Rev. Phytopathol. 42 (2004) 385–414.
[20] S.R. Grant, E.J. Fisher, J.H. Chang, B.M. Mole, J.L. Dangl, Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria, Annu. Rev. Microbiol. 60 (2006) 425–449.
[21] J.T. Greenberg, B.A. Vinatzer, Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells, Curr. Opin. Microbiol. 6 (2003) 20–28.
[22] D. Büttner, U. Bonas, Regulation and secretion of *Xanthomonas* virulence factors, FEMS Microbiol. Rev. 34 (2010) 107–133.
[23] D. Gürlebeck, F. Thieme, U. Bonas, Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant, J. Plant Physiol. 163 (2006) 233–255.
[24] R.P. Ryan, et al., Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions, Nat. Rev. Microbiol. 9 (2011) 344–355.
[25] R.M. Ferreira, et al., A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other Xanthomonads, mBio 6 (2015).
[26] A.C.P. Oliveira, et al., Transposons and pathogenicity in *Xanthomonas*: acquisition of murein lytic transglycosylases by TnXax1 enhances *Xanthomonas citri* subsp. *citri* 306 virulence and fitness, PeerJ 6 (2018).
[27] W.C. Lima, A.C.M. Paquola, A.M. Varani, M.-A. Van Sluys, C.F.M. Menck, Laterally transferred genomic islands in Xanthomonadales related to pathogenicity and primary metabolism, FEMS Microbiol. Lett. 281 (2008) 87–97.
[28] D.O. Alvarenga, L.M. Moreira, M. Chandler, A.M. Varani, A practical guide for comparative genomics of mobile genetic elements in prokaryotic genomes, Methods Mol. Biol. Clifton NJ 1704 (2018) 213–242.
[29] R. Bart, 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. 109 (2012) E1972–E1979.
[30] J.L. Gordon, 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 16 (2015) 1098.
[31] R.A.B. Assis, et al., Identification and analysis of seven effector protein families with different adaptive and evolutionary histories in plant-associated members of the Xanthomonadaceae, Sci. Rep. 7 (2017) 16133.
[32] J.S.L. Patané, et al., Origin and diversification of *Xanthomonas citri* subsp. *citri* pathotypes revealed by inclusive phylogenomic, dating, and biogeographic analyses, BMC Genomics 20 (2019) 700.
[33] N.P. Fonseca, 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. 10 (2019).

[34] S. Cesbron, et al., Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation, Front. Plant Sci. 6 (2015).
[35] J. Garita-Cambronero, A. Palacio-Bielsa, M.M. López, J. Cubero, 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 11 (2016), e0161977.
[36] J.M. Young, D.-C. Park, H.M. Shearman, E. Fargier, A multilocus sequence analysis of the genus *Xanthomonas*, Syst. Appl. Microbiol. 31 (2008) 366–377.
[37] L. Martins, et al., *Xanthomonas euroxanthea* sp. nov., a new xanthomonad species including pathogenic and non-pathogenic strains of walnut, Int. J. Syst. Evol. Microbiol. 70 (12) (2020) 6024–6031.
[38] D. Büttner, D. Gürlebeck, L.D. Noël, U. Bonas, HpaB from *Xanthomonas campestris* pv. *vesicatoria* acts as an exit control protein in type III-dependent protein secretion, Mol. Microbiol. 54 (2004) 755–768.
[39] S. Jiang, et al., 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. 21 (2020).
[40] S. Essakhi, et al., Phylogenetic and variable-number tandem-repeat analyses identify nonpathogenic *Xanthomonas arboricola* lineages lacking the canonical type III secretion system, Appl. Environ. Microbiol. 81 (2015) 5395–5410.
[41] A. Hajri, et al., Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France, Plant Pathol. 59 (2010) 1014–1022.
[42] C. Fernandes, J. Blom, J.F. Pothier, F. Tavares, High-quality draft genome sequence of *Xanthomonas arboricola* pv. *juglandis* CPBF 1521, isolated from leaves of a symptomatic walnut tree in Portugal without a past of phytosanitary treatment, Microbiol. Resour. Announc. 7 (2018).
[43] B. Kearney, B.J. Staskawicz, Widespread distribution and fitness contribution of *Xanthomonas campestris* avirulence gene avrBs2, Nature 346 (1990) 385–386.
[44] C.-F. Wei, et al., A *Pseudomonas syringae* pv. *tomato* DC3000 mutant lacking the type III effector HopQ1–1 is able to cause disease in the model plant *Nicotiana benthamiana*, Plant J. Cell Mol. Biol. 51 (2007) 32–46.
[45] D. Teper, S. Sunitha, G.B. Martin, G. Sessa, Five *Xanthomonas* type III effectors suppress cell death induced by components of immunity-associated MAP kinase cascades, Plant Signal. Behav. 10 (2015), e1064573.
[46] C. Akimoto-Tomiyama, et al., XopR, a Type III effector secreted by *Xanthomonas oryzae* pv. *oryzae*, suppresses microbe-associated molecular pattern-triggered immunity in *Arabidopsis thaliana*, Mol. Plant-Microbe Interactions® 25 (2011) 505–514.
[47] M. Metz, et al., The conserved *Xanthomonas campestris* pv. *vesicatoria* effector protein XopX is a virulence factor and suppresses host defense in Nicotiana benthamiana, Plant J. Cell Mol. Biol. 41 (2005) 801–814.
[48] J.-G. Kim, et al., *Xanthomonas* T3S effector XopN suppresses PAMP-triggered immunity and interacts with a tomato atypical receptor-like kinase and TFT1, Plant Cell 21 (2009) 1305–1323.
[49] L. Wang, X. Tang, C. He, The bifunctional effector AvrXccC of *Xanthomonas campestris* pv. *campestris* requires plasma membrane-anchoring for host recognition, Mol. Plant Pathol. 8 (2007) 491–501.
[50] F.F. White, N. Potnis, J.B. Jones, R. Koebnik, The type III effectors of *Xanthomonas*, Mol. Plant Pathol. 10 (2009) 749–766.
[51] W. Ma, et al., A *Xanthomonas oryzae* type III effector XopL causes cell death through mediating ferredoxin degradation in *Nicotiana benthamiana*, Phytopathol. Res. 2 (2020) 16.
[52] J. Qin, et al., The *Xanthomonas* effector XopK harbours E3 ubiquitin-ligase activity that is required for virulence, New Phytol. 220 (2018) 219–231.
[53] C. Song, B. Yang, Mutagenesis of 18 type III effectors reveals virulence function of XopZ(PXO99) in *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant-Microbe Interact*, MPMI 23 (2010) 893–902.
[54] J. Vandroemme, et al., *Xanthomonas arboricola* pv. *fragariae*: what's in a name? Plant Pathol. 62 (2013) 1123–1131.
[55] F. Behlau, B.I. Canteros, G.V. Minsavage, J.B. Jones, J.H. Graham, Molecular characterization of copper resistance genes from *Xanthomonas citri* subsp. *citri* and *Xanthomonas alfalfae* subsp. *citrumelonis*, Appl. Environ. Microbiol. 77 (2011) 4089–4096.
[56] F. Behlau, A.M. Gochez, J.B. Jones, Diversity and copper resistance of *Xanthomonas* affecting citrus, Trop. Plant Pathol. 45 (3) (2020) 200–212.
[57] M.H. Medema, et al., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, Nucleic Acids Res. 39 (2011). W339–346.
[58] C. Bertelli, et al., IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets, Nucleic Acids Res. 45 (2017) W30–W35.
[59] M. Scortichini, M.P. Rossi, Genetic diversity of *Xanthomonas arboricola* pv. *fragariae* strains and comparison with some other *X. arboricola* pathovars using repetitive PCR genomic fingerprinting, J. Phytopathol. 151 (2003) 113–119.
[60] L. Vauterin, B. Hoste, K. Kersters, J. Swings, Reclassification of *Xanthomonas*, Int. J. Syst. Evol. Microbiol. 45 (1995) 472–489.
[61] N. Parkinson, C. Cowie, J. Heeney, D. Stead, Phylogenetic structure of *Xanthomonas* determined by comparison of gyrB sequences, Int. J. Syst. Evol. Microbiol. 59 (2009) 264–274.
[62] S. Blanvillain, et al., Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria, PLoS One 2 (2007), e224.
[63] Y. Liu, P.M. Harrison, V. Kunin, M. Gerstein, Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes, Genome Biol. 5 (2004) R64.
[64] N.L. Kaplan, R.R. Hudson, C.H. Langley, The 'hitchhiking effect' revisited, Genetics 123 (1989) 887–899.
[65] J. Masel, Genetic drift, Curr. Biol. CB 21 (2011) R837–R838.
[66] C.-H. Kuo, H. Ochman, The extinction dynamics of bacterial Pseudogenes, PLoS Genet. 6 (2010), e1001050.
[67] H.-P. Fiedler, R. Kurth, J. Langhärig, J. Delzer, H. Zähner, Nikkomycins: microbial inhibitors of chitin synthase, J. Chem. Technol. Biotechnol. 32 (1982) 271–280.
[68] I.-M.A. Chen, et al., IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes, Nucleic Acids Res. 47 (2019) D666–D677.
[69] C. Fernandes, P. Albuquerque, L. Cruz, F. Tavares, Genotyping and epidemiological metadata provides new insights into population structure of *Xanthomonas* isolated from walnut trees, bioRxiv 397703 (2018).
[70] J.C. Setubal, N.F. Almeida, A.R. Wattam, Comparative genomics for prokaryotes, in: J.C. Setubal, J. Stoye, P.F. Stadler (Eds.), Comparative Genomics: Methods and Protocols, Springer, 2018, pp. 55–78.
[71] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[72] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552.
[73] L. Li, C.J. Stoeckert, D.S. Roos, OrthoMCL: identification of Ortholog groups for eukaryotic genomes, Genome Res. 13 (2003) 2178–2189.
[74] C. Camacho, et al., BLAST+: architecture and applications, BMC Bioinformatics 10 (2009) 421.
[75] T. Carver, N. Thomson, A. Bleasby, M. Berriman, J. Parkhill, DNAPlotter: circular and linear interactive genome visualization, Bioinforma. Oxf. Engl. 25 (2009) 119–120.
[76] N.-F. Alikhan, N.K. Petty, N.L. Ben Zakour, S.A. Beatson, BLAST ring image generator (BRIG): simple prokaryote genome comparisons, BMC Genom. 12 (2011).
[77] A.M. Varani, P. Siguier, E. Gourbeyre, V. Charneau, M. Chandler, ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes, Genome Biol. 12 (2011) R30.
[78] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, Bioinforma. Oxf. Engl. 25 (2009) 1972–1973.
[79] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274.
[80] O. Chernomor, A. von Haeseler, B.Q. Minh, Terrace aware data structure for Phylogenomic inference from Supermatrices, Syst. Biol. 65 (2016) 997–1008.
[81] F.D.K. Tria, G. Landan, T. Dagan, Phylogenetic rooting using minimal ancestor deviation, Nat. Ecol. Evol. 1 (2017) 193.
[82] A. Rambaut, A.J. Drummond, D. Xie, G. Baele, M.A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7, Syst. Biol. 67 (2018) 901–904.
[83] Karen Ross, et al., TnCentral: a prokaryotic transposable element database and web portal for transposon analysis, BioRxiv (2021), https://doi.org/10.1101/2021.05.26.445724.