Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain Xff49: a new isolate obtained from common beans in Southern Brazil

Frederico Schmitt Kremer¹ · Ismail Teodoro de Souza Junior² · Amanda Munari Guimarães¹ · Rafael dos Santos Danelon Woloski¹ · Andrea Bittencourt Moura² · Luciano da Silva Pinto¹

Received: 22 February 2018 / Accepted: 21 January 2019 / Published online: 9 March 2019
© Sociedade Brasileira de Microbiologia 2019

Abstract

The genus *Xanthomonas* comprises Gram-negative bacteria, many of which are phytopathogens. *Xanthomonas fuscans* subsp. *fuscans* is one of the most devastating pathogens affecting the bean plant, resulting in the common bacterial blight of bean (CBB). The disease is mainly foliar and affects a wide variety of bean species, thus acting as the yield-limiting factor for the bean crop. Here, we report the whole-genome sequencing of a new strain of *X. fuscans* subsp. *fuscans*, named Xff49, isolated from the infected and symptomatic beans from Capão do Leão, Southern Brazil. The genetic analysis demonstrated the presence of single-nucleotide variants (SNVs) in this strain, potentially affecting the mobilome, cell mobility, and inorganic ion metabolism. In addition, the analysis resulted in the identification of a new plasmid similar to the pAX22 derived from *Achromobacter denitrificans*, which was named pIX, along with pIA and pIC, previously reported in other strains of *X. fuscans* subsp. *fuscans*. Xff49 represents the first Brazilian genome of *X. fuscans* subsp. *fuscans* and might provide useful information applicable to the studies of phylogenetics, evolution, and pathogenomics, thereby allowing a better understanding of the genomic features present in the Brazilian strains.

Keywords Common bacterial blight · Single-nucleotide variants · Whole-genome sequencing · Phytopathogen

Introduction

The genus *Xanthomonas* comprises at least 30 species of Gram-negative bacteria, most of which are phytopathogens of plant species with high economic value, such as rice (*X. oryzae*), beans (*X. fuscans*), citrus (*X. citri*), soybean (*X. axonopodis* pv. *glycines*), and many others [1]. *Xanthomonas* species are usually sub-classified into pathovars

and subspecies according to the infected host and tissue, the resulting disease, and/or genomic features. The term pathovar refers to a single or a group of bacterial strains with similar features but differs in host plant specificity from other strains of the same species.

Xanthomonas fuscans subsp. *fuscans* (Xff), along with *X. axonopodis* pv. *phaseoli* (Xap), is one of the most devastating pathogens of beans leading to severe losses in crop yield and productivity. It is the causative agent of the common blight of bean (CBB), which affects not only common bean cultures (*Phaseolus vulgaris* L.) but also lima bean (*P. lunatus* L.) along with many other species [2]. The worldwide distribution of CBB is attributed to the efficient mechanism of transmission among seeds employed by bacteria such that studies report a reduction in production of up to 40% [3], and its consequence is both direct and indirect costs to the farmers. The direct or primary loss may consist of loss of crop yield, quality, and extra costs involved in the post-harvesting processes, whereas indirect or secondary effects may include loss to traders, exporters, consumers, farmers, and government due to low crop productivity. In America and Africa, beans represent an important component of the diet and

Responsible Editor: Rodrigo Galhardo.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s42770-019-00050-0>) contains supplementary material, which is available to authorized users.

✉ Luciano da Silva Pinto
ls_pinto@hotmail.com

¹ Laboratório de Bioinformática e Proteômica, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Capão do Leão, Brazil

² Laboratório de Bacteriologia Vegetal, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, Capão do Leão, Brazil

comprise almost 60% of the daily protein intake [4]. Since beans constitute one of the major sources of protein in these continents, an efficient control of causative agents of CBB that lead to extensive yield loss is a constant issue. Several genetic factors and quantitative trait loci (QTLs) have been reported to be responsible for providing resistance to CBB [5–7], especially in bean lineages grown in Mesoamerica and Andes. Host resistance to CBB is considered to be a practical approach toward controlling crop damage. However, the application and successful insertion of these genetic elements into commercial cultivars have been very limited.

The pathogenesis of both *Xff* and *Xap* strains involves the formation of biofilms by bacteria on the plant surface. This surface growth facilitates penetration into the phytosystem through the stomata, leading to the colonization of the mesophyll, vascular tissues, and parenchyma [8] and helping in the dispersal of infection to other plants. The disease is characterized by the formation of spots and necrosis in the leaves, stems, pods, and seeds; however, many infected plants may be asymptomatic [9]. Although the genomic analysis of *X. fuscans* subsp. *fuscans* strain 4834-R, a highly aggressive strain isolated in France [10, 11], has identified several genes shared among the genus, the virulence factors contributing to CBB are still poorly understood. Some of these genes include those that encode for effector secretory proteins that function through type III and type IV secretion systems (T3SS and T4SS), adhesins, degradation enzymes, lipopolysaccharides (LSP) biosynthesis, and formation of biofilms (e.g., xanthan gum formation) [11]. Bacterial appendages, such as flagella, play a significant role in biofilm formation by pathogenic bacteria. However, as 4834-R strain does not contain a functional flagellum, usually found in strains from the species, other strains have been sequenced to provide a better description of the pathogen [12].

In the present study, we report the whole-genome sequencing of a new strain of *X. fuscans* subsp. *fuscans* named *Xff49*, isolated from infected and symptomatic beans from Capão do Leão, Southern Brazil. To our knowledge, *Xff49* is the first Brazilian strain of *Xff* to be sequenced and made publicly available.

Material and methods

Whole-genome shotgun sequencing (WGS) was performed at the Neoprosperta facility using the Illumina MiSeq platform and a 250-bp paired-end library. Raw reads obtained were analyzed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and pre-processed using the Fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Trimmomatic [13] was utilized for quality filtering and read trimming.

De novo genome assemblies were generated using A5 [14], Velvet [15], CLC Genome Workbench (<https://www.qiagenbioinformatics.com/>), SGA [16], and Ray [17]. The assemblies were merged using CISA [18] to generate a consensus assembly. Contigs and scaffolding generated by CISA were aligned using BLASTn [19] from the NCBI-BLAST+ [20] package to a database of *X. fuscans* genome obtained from GenBank (Supplementary data 1) to separate chromosome contigs from plasmid contigs (“profiling”).

Unmapped contigs were compared to the National Center for Biotechnology Information (NCBI) nucleotide database. Multiple reference-guided scaffolding was performed for the contigs mapped to the chromosome of the bacteria using MeDuSa [21] and the same dataset of *X. fuscans* genomes used for the contig profiling. The scaffolds produced by this method were re-scaffolded using CAR [22] based on the genome of *X. fuscans* subsp. *fuscans* strain 4834-R (GenBank NC_022541.1). This genome was identified by BLAST as the closest available complete genome. The contigs and scaffolds mapped to plasmids were ordered using CAR based on the sequence of the *X. fuscans* plasmids, namely pIA (GenBank FO681495.1) and pIC (GenBank FO681497.1) according to their respective best hits. Then, gap closing was performed using IMAGE [23]. As CAR and IMAGE do not estimate the gap distance during the scaffolding or gap-closing process, all contigs were joined using an arbitrary gap span of 300 base-pairs (bps). The final assembly was evaluated using QUAST [24].

Genome annotation was performed using Genix [25] and InterProScan, and manually curated using Artemis [26]. The protein-coding genes present in the chromosome of *Xff49* and 4834-R were compared using OrthoVeen [27] to identify those genes present in *Xff49* but absent in 4834-R. The genomes of *Xff49* and 4834-R were also analyzed using AntiSMASH [28] to identify gene clusters and metabolic pathways associated with secondary metabolites biosynthesis.

A dataset of 20 genomes of *X. fuscans* and 3 from outgroups (Supplementary data 2), including both draft and finished records, was obtained from GenBank and aligned to the genome of *Xff49* using Mugsy [29]. The process led to the formation of a MAFF file containing the alignments of different syntenic regions identified in the input genomes. The sub-alignments were processed using trimAl [30] and then merged into a single alignment with functions implemented in the BioPython package [31]. The phylogenetic tree was generated using the implementation of the unweighted pair-group method with arithmetic mean (UPGMA) algorithm available in the MEGA X package using 10,000 cycles of bootstrapping [32].

A variant calling analysis was performed based on the genome of *X. fuscans* subsp. *fuscans* strain 4834-R to determine the nucleotide differences at a given locus. Paired-end reads were mapped using SMALT (<https://www.sanger.ac.uk/science/tools/smalt-0>), and the alignments were processed

using SAMTools, BCFTools, and VCFUtils.pl [33]. The variants identified by VCFUtils.pl were filtered using the PyVCF library (<https://github.com/jamescasbon/PyVCF>) to select only those single-nucleotide polymorphisms (SNPs) with $Q \geq 30$. The functional impact of each variant was predicted using SnpEff [34] based on the reference annotation, and the affected genes were annotated using RPS-BLAST+, from the NCBI-BLAST+ package, using the conserved domains database (CDD) [35] to assign molecular functions based on the Clusters of Orthologous Group (COG) database nomenclature [36].

Finally, to identify the presence of genes from the flagellar genes cluster, we have followed the same procedure adopted originally in the characterization of strain 4834-R, which was based on flagellar genes identified in the genome of *X. campestris* pv. *vesicatoria* (GenBank AM039952.1) [11]. In the analysis, we compared the strains Xff49, 4834-R, and *X. fuscans* pv. *aurantifolli* str. ICPB 10535 (*Xac*), which was already reported with a functional polar flagellum and motility when cultured in semi-solid media [37].

Results and discussion

The results of the genome assembly analysis are displayed in Table 1. After de novo assembly, profiling, and scaffolding, four replicons were reconstructed, including the chromosome and three plasmids. The chromosome of *X. fuscans* subsp. *fuscans* str. Xff49 has a length of approximately 5.2 megabases (Mb), and its scaffold contains 69 gaps, with a few genome translocations and inversions when compared to the reference sequence of *X. fuscans* subsp. *fuscans* strain 4834-R (Fig. 1). We also identified plasmids similar to pLA and pLC from 4834-R (GenBank **FO681495.1** and **FO681497.1**, respectively), although no contigs with genes homologous to pLB was identified (GenBank **FO681496.1**). However, another plasmid, similar to the pAX22 from *Achromobacter xylosoxidans* subsp. *denitrificans* (GenBank **NC_022242.1**), was identified and named “pIX.” A circular map of pIX generated using DNAPlotter, a synteny comparison between pIX and pAX22 generated using Artemis Comparison Tool (ACT), and a scheme representing the differences in gene organization identified between pIX and pAX22 are presented in Fig. 2a, b, c, respectively. When comparing to pAX22, we have identified a

set of 17 genes that are missing in pIX, which include transposases, proteins for the toxin/antitoxin system, metallo-beta-lactamase, and proteins associated with aminoglycoside metabolism. All these genes are located in a single non-syntenic region in the middle of the pAX22; where in pIX, there are only 2 protein-coding genes, one that encodes a cytosine-specific methyltransferase and other that encodes a restriction endonuclease NotI (CKU38_04582). The complete list of genes differing between pAX22 and pIX in this locus is presented in Supplementary Data 3. pIX also included a copy of the of the genes *traA* (CKU38_04589), *traC* (CKU38_04590), *traD* (CKU38_04591), and *traG* (CKU38_04592), indicating that it has conjugative functions, although the limitations in the plasmid assembly does not allow the identification of all genes included in the replicon, and the verification of the other genes from the *tra* region.

An overview of the results from the genome annotation is depicted in Table 2, and the pathogenesis-related genes that were identified on the chromosome and plasmids are summarized in Supplementary data 4. Based on the annotation of the strain 4834-R, we could identify many genes associated with the pathogenesis process, including the xanthan gum biosynthesis pathway, e.g., *gumBCDEFGHIJKLM*, ABC transporters, and a wide variety of effector proteins, including a transcription activator-like effector (TALE) protein, encoded by the gene CKU38_04498. The genome-wide analysis of virulence genes based on the genome of strain 4834-R allowed the identification of more than 20 genes involved in the pathogenesis, including the Type III effectors (T3Es) *AvrB2*, *XopAD*, *XopAE*, *XopAK*, *XopAM*, *XopC2*, *XopE1*, *XopF1*, *XopF2*, *XopI*, *XopJ5*, *XopN*, *XopK*, *XopL*, *XopP2*, *XopP1*, *XopQ*, *XopR*, *XopT*, *XopV*, *XopX*, *XopZ*, and *XfuTAL2*. However, some effectors previously reported in 4834-R, such as *XopAF*, *XopG*, *XfuTAL1*, *XopC1*, *XopE3*, and *XfuTAL1* were not identified in Xff49, indicating that the repertory of T3Es may be slightly different between these two strains.

The annotation of the Xff49 genome allowed the identification of a wide variety of non-coding elements (Supplementary Data 5), such as those from the *Xanthomonas* small RNA families (sX and asX), some of which have already been demonstrated as regulators of the pathogenesis process in other species of the *Xanthomonas* genus [38], such as *X. campestris* and *X. oryzae*, but not in *X. fuscans*. SX13 (CKU38_04192), for example, was already

Table 1 Overview of the metrics calculated by QUAST for the genome sequence of *X. fuscans* subsp. *fuscans* strain Xff49

Replicon	Gaps	CG%	Size (bp)	No. of contigs	N50 (contigs)	L50 (contigs)
Chromosome	69	64.68%	5,182,257	70	112,211	15
Plasmid pLA	3	59.49%	41,822	4	26,697	1
Plasmid pLC	4	57.52%	34,064	5	5176	2
Plasmid pIX	1	63.04%	18,734	2	15,216	1

Xff49

Fig. 1 Structural comparison of the scaffolded chromosome of *X. fuscans* subsp. *fuscans* str. Xff49 and the reference genome of *X. fuscans* subsp. *fuscans* str. 4834-R

pointed as a major regulator of the virulence in *X. campestris* pv. *vesicatoria* due to its effect on the expression of many genes associated with the *Hrp*-regulon [39].

In the annotation of the plasmid pIA, we have identified a non-coding element from the Rfam [40] family RF01695, which potentially transcribes to a C4 antisense-like RNA used by many phages that infect bacteria. This ncRNAs are present in purified phages and were also identified in many bacteria genome as well. Although this feature was absent in the original annotation of the pIA plasmid of 4834-R, a manual analysis using the cmsearch program for the INFERNAL package [41] revealed that it is also present and overlaps a conserved hypothetical protein with no domains identified using InterProScan [42], Pfam [43], or SMART [44].

The comparative analysis of protein-coding genes shared and differentially distributed among Xff49 and 4834-R revealed that there are at least 3632 ortholog groups conserved between these two strains (Fig. 3), although these number might be higher if considering these genes that might be absent in the Xff49 genome assembly as a consequence of misassemblies (e.g., transposase genes). Additionally, 7 groups were identified as exclusive in 4834-R, and 29 in Xff49, and the protein products of these groups are displayed in Table 3.

From the ortholog groups identified as absent in 4834-R based on the OrthoVenn analysis, we have identified one, which presents two copies in the genome (CKU38_01232 and CKU38_00615) that encodes an AbiF-like from the abortive infection system, which is associated with resistance to phage infection. In the context of bacterial plant disease, phages have already been pointed as potential candidates for biocontrol of phytopathogens in many crop species [45]. Therefore, the adaptation of the resistance to phage infection in phytopathogens may become an issue in the future in crop protection. Another ortholog group identified as exclusive in Xff49 encodes a long-chain-specific 3-oxoacyl-acyl carrier protein

reductase (CKU38_02114 and CKU38_02115), which has already been demonstrated as acting in the diffusible signal factor (DSF) quorum sensing regulation, a process that is important in the pathogenesis of many phytopathogens, including *X. campestris* [46]. Additionally, we also have identified two genes (CKU38_04417 and CKU38_04433) that encode outer-membrane efflux pump proteins from the resistance-nodulation-cell division superfamily (RND). The RND family of transporters are widely distributed in Gram-negative bacteria, and usually associated with the active transporting of antibiotics [47]. Finally, two beta-lactamase genes with no correspondence in 4834-R were identified in Xff49 (CKU38_00306 and CKU38_01422). Beta-lactamases confer resistance to beta-lactam antibiotics, which are synthesized by many filamentous fungi such as those from the genera *Aspergillus* and *Penicillium* [48]. The identification of these genes in phytopathogens is particularly important not only because they may provide resistance to molecules secreted by fungi present in the environment and surface of the host plant but also due to the potential risks of horizontal transferring to animal and human pathogens [49]. The complete output from the OrthoVenn analysis is available in Supplementary Data 6.

Based on the AntiSMASH analysis, we were able to identify in Xff49 and 4834-R complete gene clusters responsible for the biosynthesis of xanthoferrin; a siderophore molecule, secreted by bacteria from the *Xanthomonas* genus that is used in the uptake of iron from the environment [50], was identified in both strains. Gene clusters for xanthan and xanthomonadin biosynthetic pathways were also identified in the genome of these strains, although incomplete. Xanthan gum plays an important role in the formation of biofilm and helps the adhesion of the bacteria to the surface on the host plant, thus, been important in its pathogenesis. The xanthan biosynthesis gene cluster was identified with 61% of coverage both in Xff49 and 4834-R [51]. Xanthomonadin, a yellow pigment produced by

Fig. 2 Structure and organization of the pIX plasmid from *X. fuscans* subsp. *fuscans* str. Xff49. **a** Circular map of the pIX plasmid constructed using DNAPlotter. Coding DNA sequences (CDS) are identified as light-blue arrows, and other features (“misc_features”) in the annotation are indicated in each sequence as green arrows; syntenic regions identified between the two plasmids are connected by red blocks when in the same strand or blue when in opposite strands. **b** Synteny comparison between pIX and the pAX22 plasmid from *A. xylosoxidans*

subsp. *denitrificans* constructed using Artemis Comparison Tool. **c** Scheme representing the genes shared between pIX and pAX22 and their organization in each plasmid. CDS are indicated as blue arrows, and their respective “locus_tags” are indicated in boxes connect by gray dashed lines. Green boxes indicate “shared genes,” in which white boxes indicate genes with not homolog in the other plasmid. Shared genes identified between pIX and pAX22 are linked by blue dashed lines

Table 2 Overview of the features identified in the genome sequence of *X. fuscans* subsp. *fuscans* strain Xff49

Replicon	CDS	tRNA	rRNAs	tmRNAs	ncRNAs and regulatory*
Chromosome	4218	61	7	1	48
Plasmid pIA	49	0	0	0	1
Plasmid pIC	37	0	0	0	0
Plasmid pIX	20	0	0	0	0

*Other families of ncRNAs that do not belong to the tRNA, rRNA, or tmRNA families, and regulatory elements (e.g., riboswitches)

Fig. 3 Venn diagram illustrating the genes shared by genomes of Xff49 and 4834-R and those exclusive to each strain as identified by the OrthoVenn program

many bacteria in the *Xanthomonas* genus, is important in the protection of the DNA from ultraviolet light, improving the ability to survive in the environment [52]. The gene cluster for the biosynthesis of this molecule was identified with a coverage of 64% in Xff49 and 71% in 4834-R.

Our variant calling analysis identified 299 SNPs with $QUAL \geq 40$ Q and 19 INDELs with coverage $\geq 30\times$. However, most of these are located in intergenic regions or do not have a functional impact on protein level (synonymous mutation), indicating that the 4834-R and Xff49 strains share a high similarity at the functional level in shared genes. The distribution of the short variants identified in the variant calling analysis in different groups of biological processes indexed in the COG database is displayed in Fig. 4. As indicated, only a few processes, such as those associated with the mobilome (e.g., transposons), inorganic ion metabolism, and signal

Table 3 Ortholog groups identified by OrthoVenn as not conserved between the genomes of the strains Xff49 and 4834-R of *X. fuscans* subsp. *fuscans* and their respective annotations based on the UniProt database

Cluster ID	No. of proteins	Strain	Locus tags	UniProt annotation
Cluster12	4	4834-R	XFF4834R_chr17480 XFF4834R_chr17480 XFF4834R_chr24690 XFF4834R_chr16840	Uncharacterized protein
Cluster152	2	4834-R	XFF4834R_chr25110 XFF4834R_chr13450	Uncharacterized protein
Cluster153	2	4834-R	XFF4834R_chr15000 XFF4834R_chr25050	ISxcd1 transposase
Cluster154	2	4834-R	XFF4834R_chr27360 XFF4834R_chr41000	Xylose isomerase
Cluster155	2	4834-R	XFF4834R_chr23680 XFF4834R_chr23670	Peptidase M23
Cluster156	2	4834-R	XFF4834R_chr32180 XFF4834R_chr10940	Uncharacterized protein
Cluster157	2	4834-R	XFF4834R_chr10950 XFF4834R_chr32170	Uncharacterized protein
Cluster4	10	Xff49	CKU38_01937 CKU38_00009 CKU38_00153 CKU38_01544 CKU38_02299 CKU38_02733 CKU38_03103 CKU38_03107 CKU38_03212 CKU38_03376	ISXax1 transposase
Cluster6	6	Xff49	CKU38_01075 CKU38_00622 CKU38_00790 CKU38_00922 CKU38_00924 CKU38_01250	ISXoo14 transposase
Cluster7	6	Xff49	CKU38_03560 CKU38_00311 CKU38_00368 CKU38_00619 CKU38_03229 CKU38_04226	IS1479 transposase
Cluster9	5	Xff49	CKU38_01236 CKU38_01237 CKU38_01230 CKU38_02128 CKU38_02301	Uncharacterized protein
Cluster16	4	Xff49	CKU38_03206 CKU38_03205 CKU38_03207	Uncharacterized protein

Table 3 (continued)

Cluster ID	No. of proteins	Strain	Locus tags	UniProt annotation
Cluster17	3	Xff49	CKU38_03208 CKU38_00791 CKU38_00623 CKU38_00923 CKU38_01251	Transposase
Cluster27	3	Xff49	CKU38_01540 CKU38_01149 CKU38_03216	Uncharacterized protein
Cluster36	3	Xff49	CKU38_03215 CKU38_01150 CKU38_01541	Uncharacterized protein
Cluster50	3	Xff49	CKU38_01148 CKU38_01539 CKU38_03217	Uncharacterized protein
Cluster54	3	Xff49	CKU38_01542 CKU38_01151 CKU38_03214	Transcriptional regulator
Cluster158	2	Xff49	CKU38_01426 CKU38_00310	Uncharacterized protein
Cluster159	2	Xff49	CKU38_01232 CKU38_00615	Xfu1 transposase
Cluster160	2	Xff49	CKU38_03218 CKU38_01538	Abortive infection bacteriophage resistance protein
Cluster161	2	Xff49	CKU38_01231 CKU38_01985	IS1479 transposase
Cluster162	2	Xff49	CKU38_03527 CKU38_02582	Transposase family protein
Cluster163	2	Xff49	CKU38_00309 CKU38_01425	Glycine/betaine ABC transporter substrate-binding protein
Cluster164	2	Xff49	CKU38_04481 CKU38_00015	Carbohydrate diacid regulator
Cluster165	2	Xff49	CKU38_00307 CKU38_01423	Uncharacterized protein
Cluster166	2	Xff49	CKU38_01155 CKU38_01819	Uncharacterized protein
Cluster167	2	Xff49	CKU38_02115 CKU38_02114	3-oxoacyl-ACP reductase
Cluster168	2	Xff49	CKU38_03941 CKU38_03898	Isocitrate dehydrogenase kinase/phosphatase
Cluster169	2	Xff49	CKU38_00308 CKU38_01424	Uncharacterized protein
Cluster170	2	Xff49	CKU38_01543 CKU38_03213	Uncharacterized protein
Cluster171	2	Xff49	CKU38_00080 CKU38_00100	Transposase
Cluster172	2	Xff49	CKU38_04417 CKU38_04433	Outer-membrane efflux protein
Cluster173	2	Xff49	CKU38_02134 CKU38_02671	Uncharacterized protein
Cluster174	2	Xff49	CKU38_00306 CKU38_01422	Uncharacterized protein
Cluster175	2	Xff49	CKU38_00193 CKU38_00120	Transposase
Cluster176	2	Xff49	CKU38_01406 CKU38_02713	Transposase

transduction, were affected by missense mutations, while most of the other processes might be only affected by intergenic mutations or silent mutations (whose effects are much more difficult to predict considering only genomic data).

Phylogenetic analysis of the strains based on the multiple genome comparisons revealed a high similarity between the different strains of *Xff* (Fig. 5). It also revealed that the subspecies *fuscans* and *aurantifolii* are relatively phylogenetically distant. It was also observed that the subspecies *aurantifolii* was subdivided into two main branches, indicating that there might be a higher genetic variability inside this subspecies when compared to the subspecies *fuscans*. It is unclear,

however, how much these two pathovars differ at a functional level, although this difference might be directly associated with the differences in the preferred host for each group.⁵³

In the analysis of the flagellar genes, we have identified the ortholog genes of the gene cluster initially characterized in *X. campestris* pv. *vesicatoria* in the genomes of Xff49, 4834-R, and *X. fuscans* pv. *aurantifolii* str. ICPB 10535 (*Xac*) (Supplementary data 7). The organization of the gene cluster in the different strains was analyzed using the Artemis Comparison Tool, and the results are presented in Supplementary Data 8. By comparing the results for each strain, we were able to identify that Xff49 presents a higher number of genes from these gene

Fig. 4 Distribution of the biological processes affected by single-nucleotide variants identified in the genome of *X. fuscans* subsp. *fuscans* str. Xff49 based on the reference genome of *X. fuscans* subsp. *fuscans* strain 4834-R. Biological processes are named based on the COG classification

clusters when comparing to 4834-R, even being an unfinished genomic record (with some gaps), and its “gene profile” similar to *X. fuscans* pv. *aurantifolii* str. ICPB 10535, which was already experimentally validated, was presenting cell motility and a polar flagellum in previous studies.

By applying a combination of different bioinformatics approaches, we were able to characterize the genomic features present in *X. fuscans* subsp. *fuscans* str. Xff49 and its particularities when comparing with 4834-R, a reference strain for this pathovar.

Fig. 5 Phylogenetic tree constructed based on the alignment of syntenic blocks identified among different strains of *X. fuscans* subsp. *fuscans* and *X. fuscans* subsp. *aurantifolii*. Genomes of *Xylella fastidiosa*, *Lysobacter enzymogenes* and *Stenotrophomonas maltophilia* were used as outgroups

Although the phylogenetic- and variant-calling analysis demonstrated that these strains are closely related, the comparative analysis based on orthologs groups identified some genes that differ among these two strains, including genes associated with antibiotic resistance and regulation of quorum-sensing, which may provide different capabilities of surviving in the environment or in the host. The identification of new antibiotic resistance genes in a phytopathogen may also provide useful information for public health researchers interested in the analysis of the transfer of these genes from environmental microorganisms to human and animal pathogens. Finally, we also demonstrated that the genomic deletion of genes from the flagellar cluster observed in 4834-R is absent in Xff49, and the presence of these genes is similar to the strain ICPB 10535, which was already reported with a functional flagellum.

Conclusion

In the paucity of effective methods to control CBB infection in bean crop, the development of genetically resistant cultivars is considered to be an effective management approach. In this regard, the availability of new genomic data of *X. fuscans* subsp. *fuscans* might provide novel and useful information for researchers working on the analysis of its molecular features, pathogenomics, molecular genetics, and phylogenetics of the genus and species. As Xff49 is the first Brazilian strain to be sequenced and made publicly available, it might be used for multiple genome comparisons and genome plasticity by comparing this strain to other strains isolated in South America. These studies, in turn, may pave the way for the identification of cultivars providing resistance to CBB.

The sequence data from this article have been deposited with the DDBJ/EMBL/GenBank Data Libraries under Accession Numbers CP023294.1 (*X. fuscans* subsp. *fuscans* Xff49 complete chromosome), CP023295.1 (*X. fuscans* subsp. *fuscans* Xff49 plasmid pIA), CP023296.1 (*X. fuscans* subsp. *fuscans* Xff49 plasmid pIC), and CP023297.1 (*X. fuscans* subsp. *fuscans* Xff49 plasmid pIX).

Funding This work was financially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Ryan RP, Vorhölter F-J, Potnis N, Jones JB, van Sluys MA, Bogdanove AJ, Dow JM (2011) Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nat Rev Microbiol* 9(5):344–355. <https://doi.org/10.1038/nrmicro2558>
- Akhavan A, Bahar M, Askarian H, Lak MR, Nazemi A, Zamani Z (2013) Bean common bacterial blight: pathogen epiphytic life and effect of irrigation practices. *Springerplus*. 2(1):41. <https://doi.org/10.1186/2193-1801-2-41>.
- Morris CE, Monier J-M (2003) The ecological significance of bio-film formation by plant-associated bacteria. *Annu Rev Phytopathol* 41(1):429–453. <https://doi.org/10.1146/annurev.phyto.41.022103.134521>
- Graham PH, Vance CP (2003) Legumes: Importance and constraints to greater use. *Plant Physiol* 131(3):872–877. <https://doi.org/10.1104/pp.017004>.
- Shi C, Navabi A, Yu K (2011) Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. *BMC Plant Biol* 11(1):52. <https://doi.org/10.1186/1471-2229-11-52>
- Durham KM, Xie W, Yu K, Pauls KP, Lee E, Navabi A (2013) Interaction of common bacterial blight quantitative trait loci in a resistant inter-cross population of common bean. *Link W, ed. Plant Breed* 132(6):658–666. <https://doi.org/10.1111/pbr.12103>
- Zhu J, Wu J, Wang L, Blair MW, Zhu Z, Wang S (2016) QTL and candidate genes associated with common bacterial blight resistance in the common bean cultivar Longyundou 5 from China. *Crop J* 4(5):344–352. <https://doi.org/10.1016/j.cj.2016.06.009>
- Jacques M-A, Josi K, Darrasse A, Samson R (2005) *Xanthomonas axonopodis* pv. *phaseoli* var. *fuscans* is aggregated in stable biofilm population sizes in the phyllosphere of field-grown beans. *Appl Environ Microbiol* 71(4):2008–2015. <https://doi.org/10.1128/AEM.71.4.2008-2015.2005>
- Ruh M, Briand M, Bonneau S, Jacques M-A, Chen NWG (2017) *Xanthomonas* adaptation to common bean is associated with horizontal transfers of genes encoding TAL effectors. *BMC Genomics* 18(1):670. <https://doi.org/10.1186/s12864-017-4087-6>.
- Darrasse A, Carrère S, Barbe V, Boureau T, Arrieta-Ortiz ML, Bonneau S, Briand M, Brin C, Cociancich S, Durand K, Fouteau S, Gagnevin L, Guérin F, Guy E, Indiana A, Koebnik R, Lauber E, Munoz A, Noël LD, Pieretti I, Poussier S, Pruvost O, Robène-Soustrade I, Rott P, Royer M, Serres-Giardi L, Szurek B, van Sluys MA, Verdier V, Vernière C, Arlat M, Manceau C, Jacques MA (2013) Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of *xanthomonads*. *BMC Genomics* 14:761. <https://doi.org/10.1186/1471-2164-14-761>
- Indiana A, Briand M, Arlat M, Gagnevin L, Koebnik R, Noel LD, Portier P, Darrasse A, Jacques MA (2014) Draft genome sequence of the flagellated *Xanthomonas fuscans* subsp. *fuscans* strain CFBP 4884. *Genome Announc* 2(5). <https://doi.org/10.1128/genomeA.00966-14>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Tritt A, Eisen JA, Facciotti MT, Darling AE (2012) An integrated pipeline for de novo assembly of microbial genomes. *PLoS One* 7(9):e42304. <https://doi.org/10.1371/journal.pone.0042304>
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829. <https://doi.org/10.1101/gr.074492.107>
- Simpson JT, Durbin R (2012) Efficient de novo assembly of large genomes using compressed data structures. *Genome Res* 22(3):549–556. <https://doi.org/10.1101/gr.126953.111>
- Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* 17(11):1519–1533. <https://doi.org/10.1089/cmb.2009.0238>
- Lin SH, Liao YC (2013) CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One* 8(3):1–7. <https://doi.org/10.1371/journal.pone.0060843>

18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
19. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421. <https://doi.org/10.1186/1471-2105-10-421>
20. Bosi E, Donati B, Galardini M, Brunetti S, Sagot MF, Lió P, Crescenzi P, Fani R, Fondi M (2015) MeDuSa: a multi-draft based scaffold. *Bioinformatics* 31(15):2443–2451. <https://doi.org/10.1093/bioinformatics/btv171>
21. Lu CL, Chen KT, Huang SY, Chiu HT (2014) Car: Contig assembly of prokaryotic draft genomes using rearrangements. *BMC Bioinformatics* 15(1):1–10. <https://doi.org/10.1186/s12859-014-0381-3>
22. Tsai IJ, Otto TD, Berriman M (2010) Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol* 11(4):R41. <https://doi.org/10.1186/gb-2010-11-4-r41>
23. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8):1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
24. Kremer FS, Eslabão MR, Dellagostin OA, Pinto L d S (2016) Genix: a new online automated pipeline for bacterial genome annotation. *FEMS Microbiol Lett* 11(16)
25. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16(10):944–945. <https://doi.org/10.1093/bioinformatics/16.10.944>
26. Wang Y, Coleman-Derr D, Chen G, Gu YQ (2015) OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 43(W1):W78–W84. <https://doi.org/10.1093/nar/gkv487>
27. Blin K, Wolf T, Chevrete MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45(W1):W36–W41. <https://doi.org/10.1093/nar/gkx319>
28. Angiuoli SV, Salzberg SL (2011) Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27(3):334–342. <https://doi.org/10.1093/bioinformatics/btq665>
29. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
30. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11):1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
31. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
33. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly (Austin)* 6(2):80–92. <https://doi.org/10.4161/fly.19695>
34. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43(Database issue):D222–D226. <https://doi.org/10.1093/nar/gku1221>
35. Tatusov RL (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28(1):33–36. <https://doi.org/10.1093/nar/28.1.33>
36. Moreira LM, Almeida NF, Potnis N, Digiampietri LA, Adi SS, Bortolossi JC, da Silva AC, da Silva AM, de Moraes FE, de Oliveira JC, de Souza RF, Fancinani AP, Ferraz AL, Ferro MI, Furlan LR, Gimenez DF, Jones JB, Kitajima EW, Laia ML, Leite RP, Nishiyama MY, Rodrigues Neto J, Nociti LA, Norman DJ, Ostroski EH, Pereira HA, Staskawicz BJ, Tezza RI, Ferro JA, Vinatzer BA, Setubal JC (2010) Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifolii*. *BMC Genomics* 11(1):238. <https://doi.org/10.1186/1471-2164-11-238>
37. Abendroth U, Schmidtke C, Bonas U (2014) Small non-coding RNAs in plant-pathogenic *Xanthomonas* spp. *RNA Biol* 11(5):457–463. <https://doi.org/10.4161/ma.28240>
38. Schmidtke C, Abendroth U, Brock J, Serrania J, Becker A, Bonas U, Small RNA (2013) sX13: a multifaceted regulator of virulence in the plant pathogen *Xanthomonas*. *Waldor MK, ed. PLoS Pathog* 9(9):e1003626. <https://doi.org/10.1371/journal.ppat.1003626>
39. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. *Nucleic Acids Res* 31(1):439–441. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=165453&tool=pmcentrez&rendertype=abstract>. Accessed 26 Mar 2015
40. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25(10):1335–1337. <https://doi.org/10.1093/bioinformatics/btp157>
41. Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847–848. <http://www.ncbi.nlm.nih.gov/pubmed/11590104>. Accessed 12 May 2014
42. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Res* 42(Database issue):D222–D230. <https://doi.org/10.1093/nar/gkt1223>
43. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40(Database issue):D302–D305. <https://doi.org/10.1093/nar/gkr931>
44. Buttner C, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A (2017) Bacteriophages and bacterial plant diseases. *Front Microbiol* 8:34. <https://doi.org/10.3389/fmicb.2017.00034>
45. Hu Z, Dong H, Ma J-C, Yu Y, Li KH, Guo QQ, Zhang C, Zhang WB, Cao X, Cronan JE, Wang H (2018) Novel *Xanthomonas campestris* long-chain-specific 3-Oxoacyl-acyl carrier protein reductase involved in diffusible signal factor synthesis. *Lindow SE, ed. MBio* 9(3). <https://doi.org/10.1128/mBio.00596-18>
46. Nikaido H, Takatsuka Y (2009) Mechanisms of RND multidrug efflux pumps. *Biochim Biophys Acta, Proteins Proteomics* 1794(5):769–781. <https://doi.org/10.1016/j.bbapap.2008.10.004>
47. Brakhage AA (1998) Molecular regulation of beta-lactam biosynthesis in filamentous fungi. *Microbiol Mol Biol Rev* 62(3):547–585. <http://www.ncbi.nlm.nih.gov/pubmed/9729600>. Accessed 7 Oct 2018
48. Wright GD (2010) Antibiotic resistance in the environment: a link to the clinic? *Curr Opin Microbiol* 13(5):589–594. <https://doi.org/10.1016/j.mib.2010.08.005>

49. Pandey SS, Patnana PK, Rai R, Chatterjee S (2017) Xanthoferrin, the α -hydroxycarboxylate-type siderophore of *Xanthomonas campestris* pv. *campestris*, is required for optimum virulence and growth inside cabbage. Mol Plant Pathol 18(7):949–962. <https://doi.org/10.1111/mpp.12451>
50. Bianco MI, Toum L, Yaryura PM, Mielnichuk N, Gudesblat GE, Roeschlin R, Marano MR, Ielpi L, Vojnov AA (2016) Xanthan pyruvilation is essential for the virulence of *Xanthomonas campestris* pv. *campestris*. Mol Plant-Microbe Interact 29(9):688–699. <https://doi.org/10.1094/MPMI-06-16-0106-R>
51. Poplawsky AR, Urban SC, Chun W (2000) Biological role of Xanthomonadin pigments in *Xanthomonas campestris* pv. *campestris*. Appl Environ Microbiol 66(12):5123. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC92432/>. Accessed 11 June 2018–5127
52. Huang C-L, Pu P-H, Huang H-J, Sung HM, Liaw HJ, Chen YM, Chen CM, Huang MB, Osada N, Gojobori T, Pai TW, Chen YT, Hwang CC, Chiang TY (2015) Ecological genomics in *Xanthomonas*: the nature of genetic adaptation with homologous recombination and host shifts. BMC Genomics 16(1):188. <https://doi.org/10.1186/s12864-015-1369-8>.