Article

Complete Genome Sequence Reveals Evolutionary and Comparative Genomic Features of *Xanthomonas albilineans* Causing Sugarcane Leaf Scald

Hui-Li Zhang [1,†], **Mbuya Sylvain Ntambo** [1,†], **Philippe C. Rott** [2], **Gongyou Chen** [3], **Li-Lan Chen** [1], **Mei-Ting Huang** [1] and **San-Ji Gao** [1,*]

[1] National Engineering Research Center for Sugarcane, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China; hlzhang@fafu.edu.cn (H.-L.Z.); ntambos@africau.edu (M.S.N.); Lilan_Chen@163.com (L.-L.C.); hmt159379@163.com (M.-T.H.)

[2] CIRAD, UMR BGPI, F-34398 Montpellier, France, and BGPI, Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, F-34398 Montpellier, France; philippe.rott@cirad.fr

[3] School of Agriculture and Biology/State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai 200240, China; gyouchen@sjtu.edu.cn

* Correspondence: gaosanji@fafu.edu.cn
† These authors contributed equally to this work.

check for updates

Abstract: Leaf scald (caused by *Xanthomonas albilineans*) is an important bacterial disease affecting sugarcane in most sugarcane growing countries, including China. High genetic diversity exists among strains of *X. albilineans* from diverse geographic regions. To highlight the genomic features associated with *X. albilineans* from China, we sequenced the complete genome of a representative strain (Xa-FJ1) of this pathogen using the PacBio and Illumina platforms. The complete genome of strain Xa-FJ1 consists of a circular chromosome of 3,724,581 bp and a plasmid of 31,536 bp. Average nucleotide identity analysis revealed that Xa-FJ1 was closest to five strains from the French West Indies and the USA, particularly to the strain GPE PC73 from Guadeloupe. Comparative genomic analysis between Xa-FJ1 and GPE PC73 revealed prophage integration, homologous recombination, transposable elements, and a clustered regulatory interspaced short palindromic repeats (CRISPR) system that were linked with 16 insertions/deletions (InDels). Ten and 82 specific genes were found in Xa-FJ1 and GPE PC73, respectively, and some of these genes were subjected to phage-related proteins, zona occludens toxin, and DNA methyltransferases. Our findings highlight intra-species genetic variability of the leaf scald pathogen and provide additional genomic resources to investigate its fitness and virulence.

Keywords: bacterial disease; *Xanthomonas albilineans*; average nucleotide identity (ANI); structural variation; specific genes

1. Introduction

Xanthomonas is a genus in the gamma subdivision of the Proteobacteria that contains a large number of plant pathogens. Members of the genus cause disease on at least 124 monocots and 268 dicots and provide excellent case studies for the understanding of molecular plant-microbe interactions [1]. Leaf scald caused by *X. albilineans* is an important disease that can have considerable economic impact on sugarcane industries worldwide [2].

X. albilineans colonizes the vascular system of sugarcane leaves and stalks, but is also capable of infecting the parenchyma cells of sugarcane, a unique characteristic differing from other bacterial pathogens with a reduced genome [3]. This bacterial pathogen induces various leaf and stalk symptoms during disease progress [2]. In the initial phase of the disease, *X. albilineans* causes the appearance

of white, narrow, sharply defined leaf stripes which is followed by necrosis and wilting of infected leaves, thus resulting in plant death [4,5]. In mature and diseased stalks, side shoots develop along the stalk from the node buds and basal side shoots are always more developed than those higher up [2,4]. *X. albilineans* produces the toxin albicidin, which has phytotoxic and antibiotic properties. Albicidin is a potent DNA gyrase inhibitor that inhibits chloroplast DNA replication and blocks chloroplast differentiation, thus resulting in the white stripe and chlorotic symptoms on infected leaves [6].

A high genetic diversity has been reported among worldwide strains of *X. albilineans* that are currently distributed in three serovars and six lysovars. Serovar I includes isolates from Australia, Mauritius, South Africa, Guadeloupe, India, and the United States. Serovar II consists of isolates from African countries, and serovar III contains isolates from the Caribbean Islands (Martinique, Guadeloupe, Saint-Kitts), Sri Lanka, and Fiji [7,8]. The serological variability of *X. albilineans* has been confirmed using a combination of monoclonal antibodies and DNA fingerprinting with 38 strains of the pathogen from various geographical locations [9]. At least 10 genetic groups (PFGE-A to PFGE-J) have been described using pulsed-field gel electrophoresis and by multi-locus sequence analysis (MLSA) [10,11]. However, our previous studies showed that the genetic diversity of *X. albilineans* from China is very low. MLSA analysis of 14 strains from this country revealed that they belong to the PFGE group B, based on five housekeeping genes [12,13]. With the advent of sequencing technologies, genome sequence data can help to resolve the phylogeny among the strains of *X. albilineans* by considering mutation and recombination at the whole-genome level [14].

The complete genome sequence has been determined for several *Xanthomonas* species, making these bacteria attractive models to study plant-pathogen interactions at the molecular level. Comparative studies have also improved our understanding of the genome features of various bacterial pathogens. For example, the full genome sequence of *X. albilineans* strain GPE PC73 from Guadeloupe lacks the *hrp* genes encoding a type III secretion system (T3SS) present in almost all gram negative plant pathogenic bacteria [11,15]. Moreover, *X. albilineans* experienced a reduced genome evolution in comparison to other sequenced plant pathogenic xanthomonads. This bacterial species is phylogenetically close to *Xylella fastidiosa*, another xylem-invading pathogen with a reduced genome and that is also missing an *hrp* T3SS [15,16]. Phylogenetic analyses with rRNA sequences excluded *X. fastidiosa* from the *Xanthomonas* group, but phylogenetic analysis with *X. albilineans* genomic sequences suggested that *X. fastidiosa* belongs to the *Xanthomonas* group [15].

So far, genome sequencing has been reported for 15 worldwide strains of *X. albilineans*, including 14 strains with scaffolds, using the Illumina technique, and one strain (GPE PC73) using the Sanger technique [11,15]. This genomic information appears to be very useful to explore the biological characteristics of *X. albilineans*, the reductive genome features, and the intra-species genetic diversity of this pathogen. The aforementioned strains did not include any strain from China, where several outbreaks of sugarcane leaf scald were recently reported [12,17,18]. Therefore, to better understand the genome features of *X. albilineans* between China and other counties, we generated the complete genome without gaps of a representative strain from China (Xa-FJ1) [12] using the PacBio RSII and Illumina HiSeq PE150 platforms. The genome of this strain was compared to the genome of the other sequenced worldwide strains of *X. albilineans*, particularly with the complete genome of strain GPE PC73 from Guadeloupe.

2. Materials and Methods

2.1. Isolation of Bacteria and DNA Preparation

X. albilineans strain Xa-FJ1 was isolated from a leaf section originating from a diseased sugarcane plant of clone YG48 collected at Zhangzhou, Fujian Province, China [12]. A pure culture of this strain was grown with constant shaking at 200 rpm and 28 °C for 48 h in XAS liquid medium [19]. Bacterial genomic DNA was extracted from cultures of Xa-FJ1 using the SDS method [20]. Extracted genomic DNA

was subjected to quality control by agarose gel electrophoresis and quantified using the Qubit v.2.0 fluorometer (Life Technologies, Carlsbad, CA, USA).

2.2. Genome Sequencing and Assembly

Whole-genome sequencing was performed using the PacBio RSII platform [21] and Illumina Hiseq platform at Beijing Novogene Bioinformatics Technology Co., Ltd. A 10-Kb DNA library was constructed using the SMRT bell TM Template kit v.1.0, according to the manufacturer's instructions, and sequenced using single-molecule real-time (SMRT) sequencing technology. A 350 bp sequencing library was prepared with NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs (Beijing) LTD, Beijing, China) in accordance with the manufacturer's recommendations, and sequenced on Illumina Hiseq platform with a 2 × 150 bp paired-end sequencing kit. After quality control of the sequencing data, clean reads from the PacBio platform were assembled de novo with the SMRT Link v.5.0.1 software (Pacific Biosciences, Menlo Park, CA, USA) using the hierarchical genome assembly process (HGAP) [22]. The assembly results were further corrected with Illumina data using the bwa software (https://sourceforge.net/projects/bio-bwa/). The complete sequence of the genome of *X. albilineans* strain Xa-FJ1 has been deposited at GenBank under the accession number CP046570-CP046571.

2.3. Genome Component Prediction and Gene Annotation

Putative open reading frames (ORF) were predicted using the GeneMarkS v.4.17 program [23] (http://topaz.gatech.edu/GeneMark/). Transfer RNA (tRNA) genes were predicted with tRNAscan-SE [24]. Ribosome RNA (rRNA) genes were analyzed using rRNAmmer [25]. Small RNAs (sRNA) were predicted by BLAST against the Rfam database [26,27], and confirmed using the cmsearch program (version 1.1rc4) with default parameters. Gene annotations were determined with the BLASTP program (E-value < 1 ×10^{-5}, identity ≥ 40%, coverage ≥ 40%) and six databases including GO (gene ontology) [28], KEGG (Kyoto encyclopedia of genes and genomes) [29,30], COG (clusters of orthologous groups) [31], NR (non-redundant protein databases) [32], TCDB (transporter classification database) [33], and Swiss-Prot [34]. Genome overview was created by Circos to show the annotation information [35].

The secretory proteins were predicted with SignalP (version 4.1, http://www.cbs.dtu.dk/services/SignalP-4.1/) [36] and TMHMM (Version 2.0c, http://www.mybiosoftware.com/tmhmm-2-0c-prediction-transmembrane-helices-proteins.html). Since *X. albilineans* is a bacterial pathogen, pathogenicity and drug resistance data were also retrieved from the pathogen–host interactions database (PHI-base) [37], the virulence factors of pathogenic bacteria database (VFDB) [38], and the antibiotic resistance genes database (ARDB) [39]. Carbohydrate-active enzymes were predicted using the carbohydrate-active enzymes database [40]. CRISPRs were identified using CRISPRdigger (version 1.0, https://github.com/greyspring/CRISPRdigger) [41].

2.4. Average Nucleotide Identity and Phylogenetic Analysis

The genome sequences of 15 worldwide strains of *X. albilineans* and one strain of *Xanthomonas pseudalbilineans* (used as outgroup) were retrieved from NCBI (Table S1). The average nucleotide identities of these 16 strains and strain Xa-FJ1 sequenced in this study were calculated by pairwise genome comparison based on BLAST+ with JSpeciesWS (http://jspecies.ribohost.com/jspeciesws/) mboxciteB42-microorganisms-693790,B43-microorganisms-693790. Core genes and specific genes of the 16 strains of *X. albilineans* were analyzed using the CD-HIT rapid clustering of similar proteins software with a threshold of 50% pairwise identity, and 0.7 length difference cut off of amino acids [32,44,45]. Gene sequences were aligned pairwise and sequence redundancies were removed using the BLAST option of the solar software [46]. The gene family clustering was based on the alignment results using Hcluster-sg software (https://github.com/douglasgscofield/hcluster). A phylogenetic tree, based on 2341

core genes conserved across the 16 strains of *X. albilineans*, was constructed using the neighbor-joining method and 1000 bootstrap replications with TreeBeST (http://treesoft.sourceforge.net/treebest.shtml).

2.5. Comparative Genomic Analysis

Chromosome alignment between strains Xa-FJ1 and GPE PC73 of *X. albilineans* was performed using MUMmer version 3.22 [47] and LastZ version 1.02.00 tools [48,49]. Scattered comparison results were linked into longer forms by the chainNet package. Structural variations (SV) like translocation, inversion and trans + inverse relationships in the comparison blocks were identified according to arrange relationships and relative orientations. The results were visualized using Circos (http://circos.ca/) [35].

3. Results

3.1. General Genomic Features of X. albilineans Strain Xa-FJ1

A total of 775 Mb PacBio clean data (N50 read length of 15,957 bp, quality of 0.84; average read length of 10,889 bp) was generated with an estimated 206× average depth of sequencing coverage. Preliminary assembly was conducted with SMRT Link v.5.0.1 and corrected by the variant Caller module. The assembly result was further corrected with 2141 Mb Illumina clean data (89.99% of bases with quality score >30) using bwa. This genome was assembled into one circular chromosome of 3,724,581 bp (Figure S1A) and one single plasmid of 31,536 bp (Figure S1B). This chromosome contained 3176 predicted genes with an average gene length of 1016 bp. The genome of strain GPE PC73—the only other *X. albilineans* strain with a complete genome sequence in NCBI—contained 3115 putative genes with an average length of 1059 bp (Table 1). Besides the circular chromosome, one plasmid was present in strain Xa-FJ1 but strain GPE PC73 had three plasmids. The nucleotide sequence of the plasmid of strain Xa-FJ1 shared 99.91% identity with PlasmII of strain GPE PC73 (Table S2).

Table 1. General chromosomal features of *X. albilineans* strains Xa-FJ1 from China and GPE PC73 from Guadeloupe.

Element and Characteristics	Xa-FJ1	GPE PC73
Sequencing platform	PacBio RSII, Illumina PE150	Sanger
Coverage	206×, 570×	17×
Size (bp)	3,724,581	3,768,695
G + C content (%)	63	63
No. protein-coding sequences (CDSs)	3176	3115
Coding density (%)	86.66	84
Average length in bp of all CDSs	1016	1059

3.2. Functional Annotation of the Predicted Genes of X. albilineans Strain Xa-FJ1

Gene annotation was performed with 11 different databases (Table S3). Genes involved in metabolism pathways were significantly enriched based on GO, KEGG, and COG databases. Among the 24 subcategories of biological processes of the GO database, the largest category of Xa-FJ1 was assigned to metabolic process (1186 genes) (Figure S2). Using KEGG annotation, 1491 of 2987 annotated genes were involved in metabolism, especially the metabolic pathways belonging to global and overview maps (529 genes) (Figure S3). Amino acid transport and metabolism was ranked the third largest category among the 25 classes of functional categories of the COG database (Figure S4). A total of 279 putative virulence-associated factors were identified based on the VFDB database, including 53 flagella-related genes, 27 genes related to type IV pili, and 35 genes related to different secretion systems (Table S4).

Genes potentially involved in pathogenicity of Xa-FJ1 were also identified by blasting the pathogen–host interactions database (PHI-base) (Figure 1 and Table S3). Among the 179 genes with homologous sequences in the PHI-base, 23 shared homologs with infection-related genes in pathogenic

fungi such as *Magnaporthe oryzae*. Eighty-one and 75 genes were homologous to virulence factors characterized in bacteria pathogenic to animals and plants, respectively. Fifty three of the 75 bacterial virulence factors with hit in Xa-FJ1 were from *Xanthomonas* spp. Six Xa-FJ1 genes, homologous to sequences essential for full virulence in plant pathogenic bacteria, were XaFJ1_GM001161 (*rsmA* in *X. oryzae*), XaFJ1_GM001582 (*galU* in *X. campestris*), XaFJ1_GM001983 (*pstB* in *X. citri*), XaFJ1_GM002139 (*rpoN* in *Erwinia amylovora*), XaFJ1_GM002197 (*vrpA* in *X. citri*), and XaFJ1_GM002596 (*hrpM* in *X. citri*).
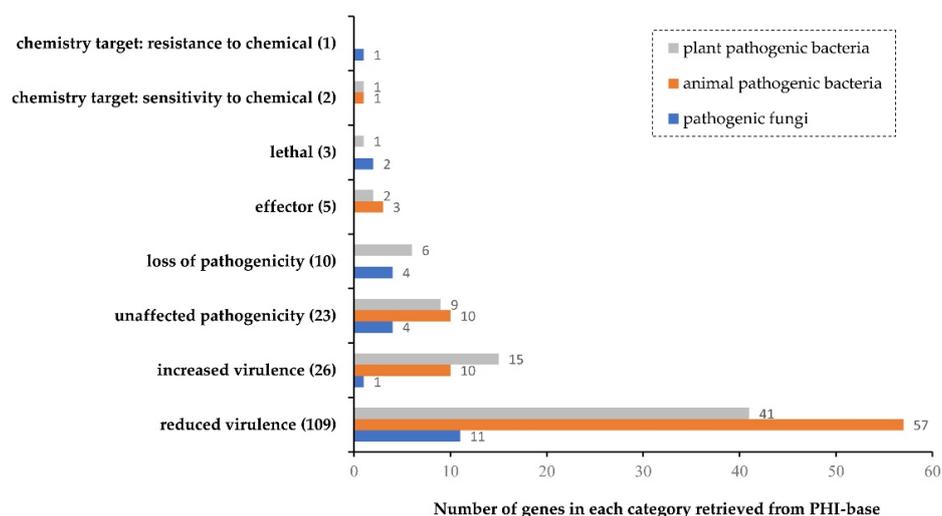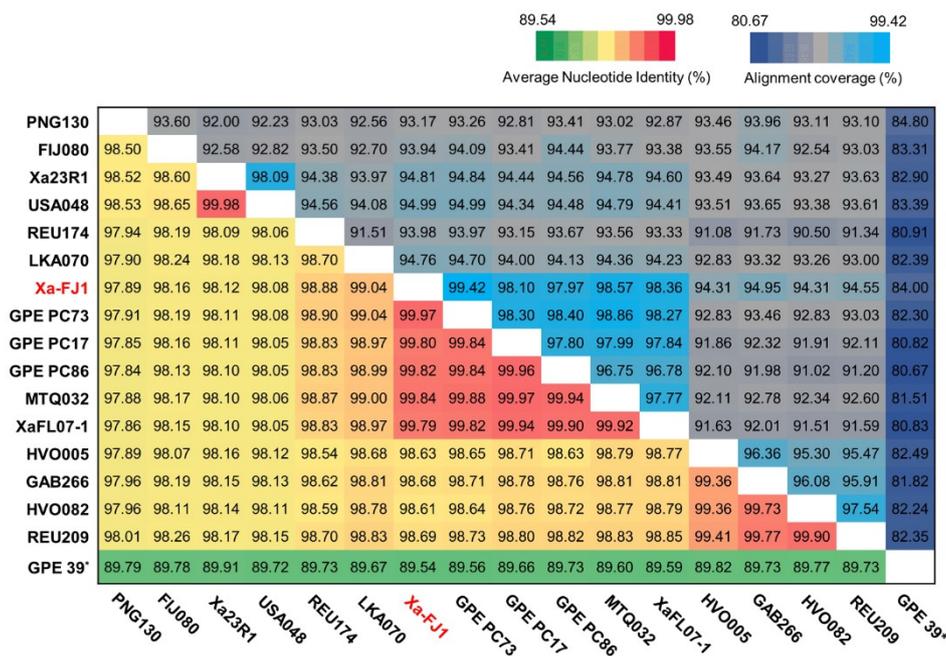

Figure 1. Phenotype classification of putative pathogenicity genes of *X. albilineans* retrieved from the pathogen–host interactions database (PHI-base).

In particular, the quorum sensing (QS) signal molecule Ax21 of *X. oryzae* pv. *oryzae* was retrieved from the PHI-base. Thirteen genes belonging to four operons (*raxSTAB*, *raxPQ*, *raxRH* and *phoPQ*) that are supposed to be required for the well-known effector Ax21 activity in *X. oryzae* pv. *oryzae* were also identified in *X. albilineans* strain Xa-FJ1 (Table S3). Four copies of *raxB* (XaFJ1_GM001779, XaFJ1_GM001890, XaFJ1_GM001891, XaFJ1_GM002569) were included, but *raxA* and *raxST* were not retrieved from the PHI-base. XaFJ1_GM000725 was homologous to gene *raxC* outside the *raxSTAB* locus in *X. oryzae* pv. *oryzae*, and may encode the outer membrane component of the type I secretion system. The retrieved genes of Xa-FJ1 also contained three copies of *raxR* (XaFJ1_GM000629, XaFJ1_GM000871, XaFJ1_GM002789) and one *raxH* ortholog (XaFJ1_GM000872), which form two-component regulatory systems. XaFJ1_GM000222 and XaFJ1_GM000223 were related to the operon *phoP/phoQ* of *X. oryzae*, which is involved in reduction and increase of virulence. XaFJ1_GM002301 and XaFJ1_GM002300 were present in an operon that corresponds to the operon *raxP/raxQ* of *X. oryzae*, which is involved in Ax21 tyrosine sulfation.

3.3. Average Nucleotide Identity and Phylogenetic Analysis among Strains of X. albilineans

Average nucleotide identity (ANI) analysis was conducted with the full-length genome sequence of Xa-FJ1 and 15 other strains of *X. albilineans* and one strain of *X. pseudalbilineans* (GPE 39) (Figure 2 and Table S1). ANI varied from 97.84–99.98%% among the 16 worldwide strains of *X. albilineans*. These strains shared 89.54–89.91% ANI with strain GPE 39 of *X. pseudalbilineans*. Strain Xa-FJ1 from China had 97.89–99.97% ANI with the 15 other strains of *X. albilineans* in the NCBI database. Strain Xa-FJ1 had the highest identity (≥99.79%) with four strains from the French West Indies (GPE PC73, GPE PC17, GPE PC86 from Guadeloupe and MTQ032 from Martinique), and one strain from the USA (XaFL07-1).

89.54	99.98		80.67	99.42
Average Nucleotide Identity (%)			Alignment coverage (%)	

|	PNG130	FIJ080	Xa23R1	USA048	REU174	LKA070	Xa-FJ1	GPE PC73	GPE PC17	GPE PC86	MTQ032	XaFL07-1	HVO005	GAB266	HVO082	REU209	GPE 39*
PNG130		93.60	92.00	92.23	93.03	92.56	93.17	93.26	92.81	93.41	93.02	92.87	93.46	93.96	93.11	93.10	84.80
FIJ080	98.50		92.58	92.82	93.50	92.70	93.94	94.09	93.41	94.44	93.77	93.38	93.55	94.17	92.54	93.03	83.31
Xa23R1	98.52	98.60		98.09	94.38	93.97	94.81	94.84	94.44	94.56	94.78	94.60	93.49	93.64	93.27	93.63	82.90
USA048	98.53	98.65	99.98		94.56	94.08	94.99	94.99	94.34	94.48	94.79	94.41	93.51	93.65	93.38	93.61	83.39
REU174	97.94	98.19	98.09	98.06		91.51	93.98	93.97	93.15	93.67	93.56	93.33	91.08	91.73	90.50	91.34	80.91
LKA070	97.90	98.24	98.18	98.13	98.70		94.76	94.70	94.00	94.13	94.36	94.23	92.83	93.32	93.26	93.00	82.39
Xa-FJ1	97.89	98.16	98.12	98.08	98.88	99.04		99.42	98.10	97.97	98.57	98.36	94.31	94.95	94.31	94.55	84.00
GPE PC73	97.91	98.19	98.11	98.08	98.90	99.04	99.97		98.30	98.40	98.86	98.27	92.83	93.46	92.83	93.03	82.30
GPE PC17	97.85	98.16	98.11	98.05	98.83	98.97	99.80	99.84		97.80	97.99	97.84	91.86	92.32	91.91	92.11	80.82
GPE PC86	97.84	98.13	98.10	98.05	98.83	98.99	99.82	99.84	99.96		96.75	96.78	92.10	91.98	91.02	91.20	80.67
MTQ032	97.88	98.17	98.10	98.06	98.87	99.00	99.84	99.88	99.97	99.94		97.77	92.11	92.78	92.34	92.60	81.51
XaFL07-1	97.86	98.15	98.10	98.05	98.83	98.97	99.79	99.82	99.94	99.90	99.92		91.63	92.01	91.51	91.59	80.83
HVO005	97.89	98.07	98.16	98.12	98.54	98.68	98.63	98.65	98.71	98.63	98.79	98.77		96.36	95.30	95.47	82.49
GAB266	97.96	98.19	98.15	98.13	98.62	98.68	98.71	98.78	98.76	98.81	98.81	98.81	99.36		96.08	95.91	81.82
HVO082	97.96	98.11	98.14	98.11	98.59	98.78	98.61	98.64	98.76	98.72	98.77	98.79	99.36	99.73		97.54	82.24
REU209	98.01	98.26	98.17	98.15	98.70	98.83	98.69	98.73	98.80	98.82	98.83	98.85	99.41	99.77	99.90		82.35
GPE 39*	89.79	89.78	89.91	89.72	89.73	89.67	89.54	89.56	89.66	89.73	89.60	89.59	89.82	89.73	89.77	89.73	

Figure 2. Heat map of average nucleotide identity (ANI) and alignment coverage based on the entire genome sequence of 16 strains of *X. albilineans* and one strain of *X. pseudalbilineans*. Strain Xa-FJ1 sequenced in this study is written in red. ANI values (%) and alignment coverage (%) of each two-genome sequence comparison are shown in the lower triangle and in the upper triangle of the matrix, respectively. * The last bacterial strain (GPE 39) does not belong to *X. albilineans* but to *X. pseudalbilineans* and was used as an outgroup.

The 16 strains of *X. albilineans* were distributed into three distinct clades of a phylogenetic tree based on the core-genome (2341 genes) of these strains. Strain Xa-FJ1 from China was assigned to a specific branch or sub-group of clade III that also included four strains from the French West Indies (GPE PC73, GPE PC17 and GPE PC86 from Guadeloupe, and MTQ032 from Martinique) and one strain from the USA (XaFL07-1) (Figure 3). Strain REU174 from Reunion Island, and strain LKA070 from Sri Lanka, were also assigned to clade III but were located in two other sub-groups. Clade II was formed by three strains from Africa (GAB266, HVO082, and HVO005) and one strain, REU209, from Reunion Island, an island off the coast of East Africa. Clade I contained four strains: One from Fiji (FJI080), one from Papua New Guinea (PNG130), and two from the USA (USA048 and Xa23R1).

3.4. Chromosomal Structural Variation Between Strains Xa-FJ1 and GPE PC73 of X. albilineans

GPE PC73 was one of the closest strains to Xa-FJ1 and, as mentioned above, was the only *X. albilineans* strain with a complete genome sequence in the NCBI. Comparative analysis of the chromosomes between Xa-FJ1 and GPE PC73 identified 16 structural variations that were linked to various insertion/deletions (InDels, R1~R16) (Figure 4 and Table 2). In comparison to strain Xa-FJ1, the chromosome of strain GPE PC73 contained an insertion from position 219,691 bp to 271,791 bp (52,101 bp in R1 region). This region included 72 specific genes in the genome of strain GPE PC73 that were absent in strain Xa-FJ1 (Table 2 and Table S5). A plasticity zone (about 7 Kb long) of the chromosome of strains Xa-FJ1 and GPE PC73 had low similarity (40.26%) between the two strains and consisted of four InDels (R5, R6, R7, and R8).
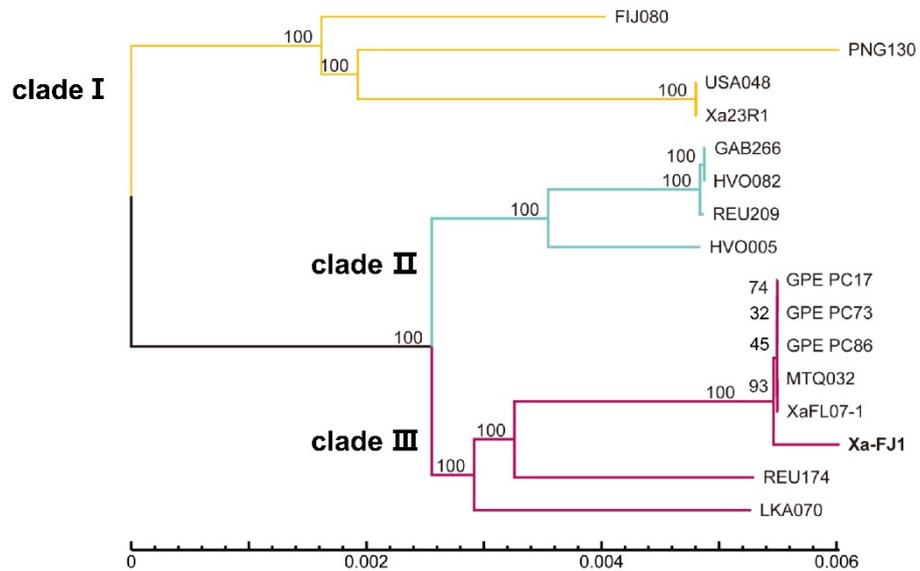
Figure 3. Neighbor-joining phylogenetic tree of *X. albilineans* constructed with the core-genome sequence of 16 strains of *X. albilineans* using TreeBesT software. Bootstrap values based on 1000 replications are indicated at branches. Scale bar units are in number of substitutions per nucleotide.
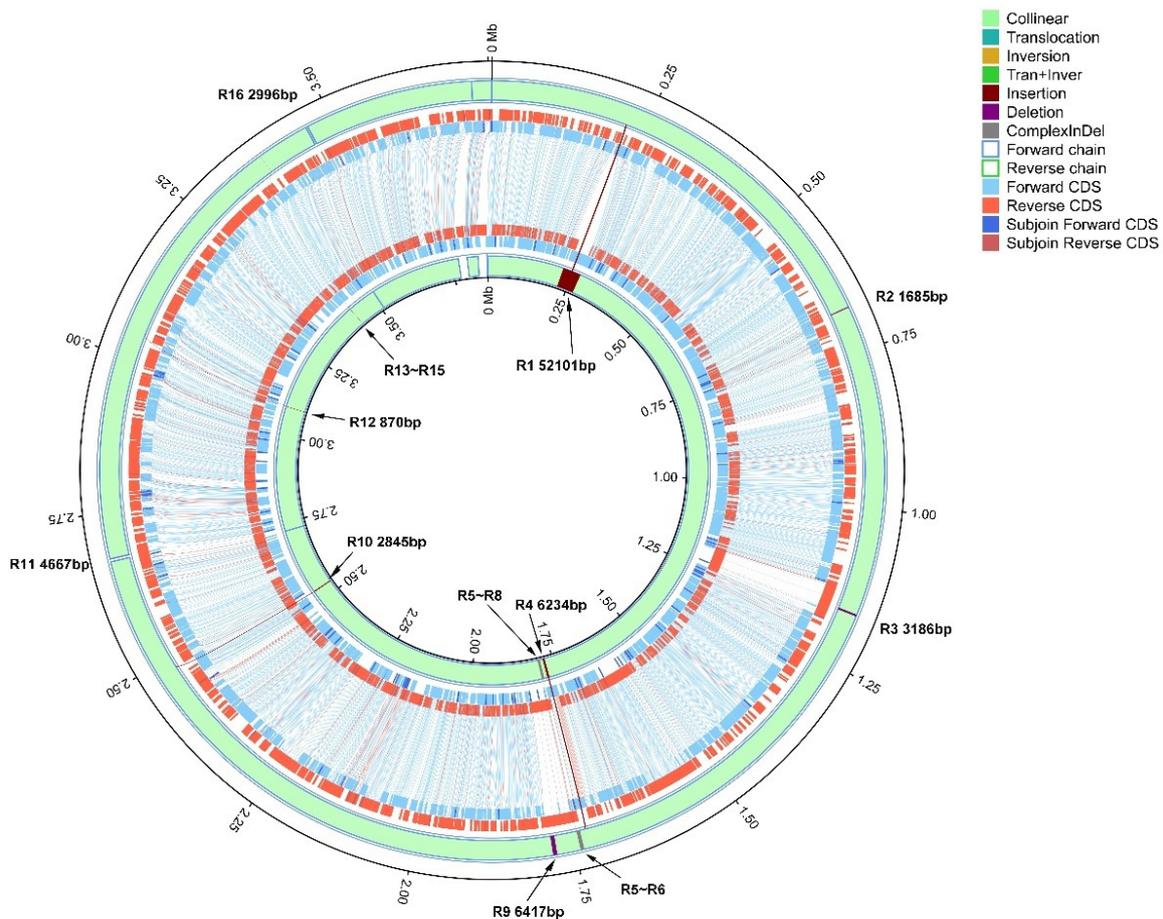
Figure 4. Chromosomal structural variations between strains Xa-FJ1 and GPE PC73 of *X. albilineans*. The chromosomal sequence of strain Xa-FJ1 was designated as the reference. Moving inwards, the first three circles show protein coding sequences (CDSs) conserved in Xa-FJ1, and the inner three circles show CDSs conserved in GPE PC73. The grey lines between the predicted CDSs indicate consistent alignment.

Table 2. Chromosomal structural variation between *X. albilineans* strains Xa-FJ1 from China and GPE PC73 from Guadeloupe (See also Figure 4).

Region	Xa-FJ1		GPE PC73		Variation Type (Reference = Xa-FJ1)	Fragment Length (nt)		Predicted Cause of Variation	Affected Gene(s)
	Begins	Ends	Begins	Ends		Xa-FJ1	GPE PC73		
R1	219,400	219,400	219,691	271,791	Insertion	0	52,101	Prophage integration	XALc_0171-XALc_0242 (72 specific genes in GPE PC73)
R2	679,518	681,202	731,933	731,933	Deletion	1685	0	Recombination/assembly defect	XaFJ1_GM000644; XaFJ1_GM000645; XaFJ1_GM000646
R3	1,165,743	1,168,928	1,216,465	1,216,465	Deletion	3186	0	Recombination/assembly defect	XaFJ1_GM001035; XALc_1056
R4	1,727,177	1,727,177	1,774,582	1,780,815	Insertion	0	6234	Recombination/assembly defect	XaFJ1_GM001510; XALc_1529
R5	1,736,246	1,736,755	1,789,885	1,790,986	Complex Indel	510	1102	Prophage integration	XALC_1536-XALC_1545
R6	1,736,873	1,741,929	1,791,101	1,795,842	Complex Indel	5057	4742	Prophage integration	(10 specific genes in GPE PC73);
R7	1,742,065	1,742,065	1,795,979	1,796,122	Insertion	0	144	Prophage integration	XaFJ1_GM001517-XaFJ1_GM001525
R8	1,742,742	1,742,742	1,796,839	1,797,116	Insertion	0	278	Prophage integration	(9 specific genes in Xa-FJ1)
R9	1,778,246	1,784,662	1,832,620	1,832,620	Deletion	6417	0	Recombination/assembly defect	XaFJ1_GM001532; XALc_1551
R10	2,484,556	2,484,556	2,532,506	2,535,350	Insertion	0	2845	Recombination/assembly defect	XaFJ1_GM002152; XALc_2151; XALc_2152
R11	26,75,326	2,679,992	2,726,120	2,726,120	Deletion	4667	0	Transposable elements	XaFJ1_GM002293; XaFJ1_GM002292; XaFJ1_GM002291; XALc_2290
R12	3,041,680	3,041,680	3,087,815	3,088,684	Insertion	0	870	Transposable elements	XALc_2603; XALc_2604
R13	3,377,378	3,377,378	3,424,367	3,424,696	Insertion	0	330	CRISPR-Cas	Intergenic region which contains
R14	3,377,647	3,377,713	3,424,965	3,424,965	Deletion	67	0	CRISPR-Cas	clustered regulatory interspaced short
R15	3,377,912	3,377,912	3,425,163	3,425,227	Insertion	0	65	CRISPR-Cas	palindromic repeats
R16	3,460,518	3,463,513	3,507,852	3,507,852	Deletion	2996	0	Transposable elements	XaFJ1_GM002988; XaFJ1_GM002989; XaFJ1_GM002990; XALc_2969

Five InDel fragments were flanked by repetitive sequences. Three of them (R3, R4 and R9) were expected to be involved in synthesis of the AMP-binding super family domain of a non-ribosomal peptide synthase. The other two InDel fragments were located at position 679,518 bp to 681,202 bp (R2) in strain Xa-FJ1, and 2,532,506 bp to 2,535,350 bp (R10) in strain GPE PC73, and were predicted to encode hypothetical proteins and a putative methyl-accepting chemotaxis protein, respectively.

InDel fragments at position R11 and R16 contained XaFJ1_GM002292 and XaFJ1_GM002989, respectively, and are expected to encode putative transposases. Homologous sequences occurred on both sides of XaFJ1_GM002292 and XaFJ1_GM002989, resulting in the similarity of adjacent genes. In contrast, no homologous sequence was found on either side of insertion fragment R12 at loci XALc_2603 and XALc_2604 of strain GPE PC73, although these two genes were also associated to transposases. In addition, R12 had the same sequence as loci XALc_1241-XALc_1245 that were associated to replicative transposition. A sequence variation at position 3.4 Mb (R13-R15) between strains Xa-FJ1 and GPE PC73 included clustered regulatory interspaced short palindromic repeats (CRISPR), downstream of CRISPR-associated protein 2 (*cas2* gene).

Like the genome of GPE PC73, the genome of strain Xa-FJ1 contained two CRISPR/cas systems: CRISPR-1 and CRISPR-2. The CRISPR-2 system was 100% identical between the two strains and the CRISPR-1 system shared 76.77% identity between Xa-FJ1 and GPE PC73 (Figure S5). The CRISPR-1 system of GPE PC73 is associated with seven *cas* genes and contains thirty-four 31-base pair repeats and thirty-three 33- to 38-base pair spacers [11]. The nucleotide sequence of the three spacers of the trailer end (33, 32, 31) varied between strains Xa-FJ1 and GPE PC73. Genome variations R13-R15 between the two strains corresponded to seven spacers of CRISPR-1 that were present in only one of the two strains (Figure S5).

3.5. Chromosomal Specific Genes between Strain Xa-FJ1 and Strain GPE PC73 of X. albilineans

Of the 3176 putative protein-coding sequences (CDSs) manually annotated on the chromosome of *X. albilineans* strain Xa-FJ1, 2998 hit in the CDSs of strain GPE PC73, and another 178 hit in the chromosomal sequence of strain GPE PC73, which were not annotated as CDSs due to different prediction methods. Only 10 predicted CDSs were specific to strain Xa-FJ1 from China (Table 3), nine of which (XaFJ1_GM001517~XaFJ1_GM001525) were located in the plasticity zone (R5-R8, Figure 4 and Table 3). Six genes were associated with hypothetical proteins. XaFJ1_GM001517 was predicted to encode a protein involved in synthesis of the zona occludens toxin (Zot). XaFJ1_GM001523 was expected to code for a DNA-binding protein homologous to the one from *Xanthomonas* phage phiLf. The protein encoded by XaFJ1_GM001524 was similar to the replication initiation protein from *X. translucens*. Thus, these genes might be acquired as prophage. The last specific gene (XaFJ1_GM002989) was located far away from the plasticity zone, and was predicted to encode a transposase with an ATP-binging function (R16).

Table 3. Putative function of the 10 genes of the chromosome of *X. albilineans* strain Xa-FJ1 from China that were not present in the genome of strain GPE PC73 from Guadeloupe.

Specific Gene	Annotation	Location
XaFJ1_GM001517	zona occludens toxin	the plasticity zone, R5-R8
XaFJ1_GM001518	hypothetical protein	the plasticity zone, R5-R8
XaFJ1_GM001519	hypothetical protein	the plasticity zone, R5-R8
XaFJ1_GM001520	hypothetical protein	the plasticity zone, R5-R8
XaFJ1_GM001521	hypothetical protein	the plasticity zone, R5-R8
XaFJ1_GM001522	hypothetical protein	the plasticity zone, R5-R8
XaFJ1_GM001523	DNA-binding protein	the plasticity zone, R5-R8
XaFJ1_GM001524	replication initiation protein	the plasticity zone, R5-R8
XaFJ1_GM001525	hypothetical protein	the plasticity zone, R5-R8
XaFJ1_GM002989	transposase	transposable elements, R16

A total of 138 putative CDSs present in the genome of strain GPE PC73 (82 on the chromosome and 56 on two plasmids) were not found in the genome of strain Xa-FJ1. Strain GPE PC73 had an insertion region of 52 kb (R1, Figure 4) that contained 72 genes specific to GPE PC73 (Table S5). Among these, 53 genes had no predicted function and were considered hypothetical proteins (including one hypothetical secreted protein). Fifteen genes were annotated as hypothetical phage-related proteins, including XALc_0206 (hypothetical phage terminase large subunit protein) and XALc_0242 (putative phage integrase protein). The other four genes in the R1 region were predicted to be two putative N6 adenine-specific DNA methyltransferase proteins (XALc_0178 and XALc_0231), one putative DNA (cytosine-5-)-methyltransferase protein (XALc_0203), and one hypothetical DNA methyltransferase protein (XALc_0202). The 10 specific genes in the plasticity zone of GPE PC73 (R5-R8) included genes predicted to code for three hypothetical phage-related proteins (XALc_1539, XALc_1543 and XALc_1544), one putative filamentous phage cf1c related protein (XALc_1537), and one putative F pilin acetylation protein (XALc_1536).

3.6. Genes Present in Plasmids, PlasmI and PlasmIII, of Strain GPE PC73 of X. albilineans from Guadeloupe and Absent in Strain Xa-FJ1 from China

The plasmid common to strain GPE PC73 (PlasmII) and strain Xa-FJ1 (pXaFJ1) shared 99.91% identity between the two strains, and differed by 29 single nucleotide polymorphisms (SNPs). Two plasmids (PlasmI and PlasmIII) present in strain GPE PC73 were not found in Xa-FJ1. These two plasmids were predicted by the NCBI Prokaryotic Genome Annotation Pipeline to encode 24 and 32 genes, respectively (Table S6). As also reported by Pieretti et al. [11], PlasmI and PlasmIII each harbor an incomplete conjugal transfer system, referring to 12 genes and 11 genes, respectively. This system is considered as an important bacterial factor, helping bacterial adaptation to new hosts. Other genes of these two plasmids are predicted to code for proteins involved in replication and maintenance of plasmids such as plasmid partitioning proteins (XALp_3174, XALr_3236, and XALr_3237), plasmid stabilization system proteins (XALp_3179, XALp_3197, and XALr_3241) and the trans-acting replication initiator TrfA protein (XALr_3265). XALp_3176 of PlasmI has been annotated as a probable fic-family protein involved in adenylation of Rho-family GTPases. In PlasmIII, XALr_3238 is predicted to be a zinc metalloproteinase with a sprT domain that is a regulator of the *bolA* gene during the stationary phase. PlasmIII also harbors three genes (XALr_3245, XALr_3246, and XALr_3266) that putatively code for transcription regulator proteins.

4. Discussion

In recent years, leaf scald of sugarcane has been reported with increasing frequency in five sugarcane-growing regions of China by PCR-based detection [12,17,18]. Very low genetic diversity was found among strains of *X. albilineans* from China, based on sequence analysis of an ABC transporter gene (XALc_1791) and MLSA, suggesting recent spread in this country of a single strain (from genetic group PFGE-B) of the leaf scald pathogen [12,13]. Since 2009, 15 genome sequences of *X. albilineans* strains from 10 worldwide geographical locations have been deposited in GenBank with complete or scaffold sequences. Some genomic information is, however, not accessible for most of these strains because of incomplete sequences. No complete genome sequence of *X. albilineans* from China had been reported until now. Third-generation sequencing technology such as the PacBio system can help to resolve most of the assembly problems by providing long reads, low degree of bias and epigenetic classification [50]. Furthermore, this sequencing platform is competent to determine sequences with tandem repeats, high/low G + C values, and interspersed repeated regions, compared to second-generation sequencing platforms [50,51]. In this study, the complete genome of a representative strain of *X. albilineans* from China (Xa-FJ1) was sequenced without gaps using the PacBio RSII technique and was corrected by Illumina data for higher consensus accuracy.

The annotations performed with the GO, KEGG and COG databases indicated that a wide variety of metabolic related proteins were encoded by *X. albilineans* strain Xa-FJ1. These proteins

included putative pectinesterase (XaFJ1_GM001902), cellulase (XaFJ1_GM000462) and xylose isomerase (XaFJ1_GM000104), and they were expected to be involved in active metabolic pathways for specific adaptation of this pathogen to the nutrient-poor xylem vessels. Among the 279 genes with homologous sequences in the VFDB database (Table S4), 8 genes have been shown to affect pathogenicity of the causal agent of sugarcane leaf scald by transposon mutagenesis [52]. The pathogen–host interactions database (PHI-base) collects gene phenotype and function data obtained by manual curation of the peer-reviewed literature [37]. Among 179 genes retrieved from PHI-base, 39 genes had putative functions previously reported [11,52]. Moreover, 20 of these genes referred to two-component systems. Among the remaining 140 genes, 98 were associated with reduced virulence or loss of pathogenicity and 5 genes were annotated as effectors, including Ax21. As a quorum sensing signal molecule in *X. oryzae*, Ax21 regulates cell density-dependent expression of up to 489 genes [53]. Thirteen genes involving four operons (*raxSTAB*, *raxPQ*, *raxRH* and *phoPQ*) were also identified in PHI-base, and these genes were associated with the activity of Ax21 in *X. oryzae*. Knockout mutants of these genes increased virulence (hypervirulence) to rice, except for *phoP*. Mutants of *raxST*, *raxA*, *raxB* or *raxC* from *X. oryzae* can evade XA21-mediated immunity to different degrees and induce disease lesions on rice expressing the XA21 receptor [54]. Both *raxH*/*raxR* and *phoP*/*phoQ* serve as two-component regulatory systems which may sense Ax21 and activate downstream signaling [53]. XaFJ1_GM002301 and XaFJ1_GM002300 from Xa-FJ1 are two genes corresponding to the operon *raxP*/*raxQ* of *X. oryzae* which is indispensable for the biological activity of Ax21 in the rice bacterial blight pathogen [55]. The functions of Ax21 and Ax21-related *rax* genes in Xa-FJ1 need to be explored to determine their role in the virulence of *X. albilineans*.

Based on complete sequence analysis, strain Xa-FJ1 from China shared high ANI with five strains of *X. albilineans* from the French West Indies (including GPE PC73 from Guadeloupe) and the USA (Florida). These six strains also clustered together in a phylogenetic tree based on the core-genome sequence of 16 strains of the pathogen. These results are congruent with those obtained in a previous study based on an ABC transporter gene (XALc_1791) and five housekeeping genes [13], suggesting a common evolutionary route for these six strains of *X. albilineans*. We hypothesized that the occurrence of *X. albilineans* in mainland China is linked to former frequent introductions of sugarcane germplasm and cultivars from Taiwan and foreign countries such as the USA, India, and Australia. These introductions occurred in order to enhance the local genetic diversity of parental clones in breeding programs during the 1980–1990s. Import of infected material has most likely occurred because *X. albilineans* can be present in cuttings collected from asymptomatic sugarcane, and the quarantine procedures used in China prior to 1990 were not as robust as they are nowadays [12].

Virulence of *X. albilineans* strains (data from the literature detailed in Table S1) was not correlated with distribution of the strains by ANI analysis. Strains with high ANI values showed great variation in virulence, suggesting that small genome differences (such as point mutations or small indels) were involved in variations of the virulence of *X. albilineans*. Additional virulence data are needed to further investigate the pathogenicity of *X. albilineans* and its relationship with adaption and evolution of this pathogen.

Chromosomal structural variations between strains Xa-FJ1 and GPE PC73 suggested that homologous recombination and horizontal gene transfer were tightly associated to genome evolution of *X. albilineans*. Five variable regions (R2, R3, R4, R9 and R10, Figure 4 and Table 2) appear to be caused by homologous recombination because the sequences flanking the InDel sites are identical in both bacterial strains, and homologous recombination may have occurred during chromosome replication. Alternatively, these regions may have resulted from assembly errors of GPE PC73 because the five InDels have no specific meaning as they only create internal duplicated DNA regions [56]. Horizontal gene transfer (HGT) is a common process responsible in prokaryotes for movement of genes from one organism to another, including antibiotic resistance genes and virulence factors [57]. Based on genome analysis of multiple strains of a given bacterial species or different species, prophages are the main cause of bacteria's short-term intraspecies or interspecies diversity [58]. In this study,

two prophage integrations (R1 and R5-R8) resulted in the addition of different specific genes in the genomes of strains Xa-FJ1 and GPE PC73. Regions R1 and R5–8, differing between Xa-FJ1 and GPE PC73, contain four DNA methyltransferases and an F pilin acetylation protein in strain GPE PC73. Various lytic and lysogenic phages have been proven to encode multi- and mono-specific orphan methyltransferases that have the ability to confer protection from restriction endonucleases of their bacterial hosts [59]. Gene XaFJ1_GM001517 of strain Xa-FJ1 (which is absent in GPE PC73) is predicted to encode the Zot protein (PD0928), which plays a role in pathogenicity of *X. fastidiosa*, suggesting a selective advantage for strain Xa-FJ1 from China [60]. These proteins may contribute to variation in virulence of *X. albilineans* or adaptation to certain sugarcane cultivars.

Strains Xa-FJ1 and GPE PC73 of *X. albilineans* have a smaller chromosome size (3.8 Mb) than those of other *Xanthomonas* species sequenced so far (generally about 5 Mb). *X. albilineans* does not possess a Hrp-type III secretion system, indicating that the pathogenicity of this pathogen must rely on other virulence factors or secretory systems [11]. Furthermore, *X. albilineans* possesses a T3SS system of the SPI-1 (for Salmonella Pathogenicity Island-1) injectisome family, which strongly suggests that this bacterial species also interacts with an unknown insect [15,61]. Zot protein acts like a physiological modulator that is used by animal pathogens, such as *Vibrio cholerae* and *Neisseria meningitidis*, to induce a reversible opening of tight junctions between cells and to increase the paracellular permeability in a non-toxic manner [62]. Presence of this gene is further evidence of a possible association of *X. albilineans* with an animal host that remains to be identified.

Transposable elements can also promote genome plasticity in bacterial genomes [63]. Three regions (R11, R12 and R16) of the *X. albilineans* genome contained transposases that were different between Xa-FJ1 and GPE PC73. The two strains also differed by their CRISPR-Cas signatures. The CRISPR-Cas system is a prokaryotic immune system among bacteria and archaea that provides resistance to foreign genetic elements [64]. Although strains Xa-FJ1 and GPE PC73 showed high ANI, these genomic features suggested rapid DNA adaptation of these trains during their evolution in different environments.

Many *Xanthomonas* strains also carry extra-chromosomal circular DNA in the form of plasmids, which can offer a myriad of benefits to the bacterial host, like virulence traits and antibiotic resistance [65]. Although strains Xa-FJ1 from China, and GPE PC73 from Guadeloupe had the closest evolutionary relationship, PlasmI and PlasmIII from GPE PC73 were absent in Xa-FJ1, thus suggesting that these DNA elements are not critical for basic virulence of the leaf scald pathogen, but might be essential for adaptation of *X. albilineans* to certain environments. The Fic (filamentation induced by cyclic AMP) domains are conserved from bacteria to humans. Fic domain proteins catalyze the addition of AMP, or 'adenylylation', of target host proteins [66]. For example, the secreted antigen IbpA from *Histophilus somni* causes collapse of the host-cell actin cytoskeleton by AMPylation to modify the three mammalian Rho family GTPases [67]. It remains to be determined whether XALp_3176 in PlasmI of GPE PC73 also assists *X. albilineans* to target the Rho-family GTPases in sugarcane. Lytic transglycosylases are bacterial enzymes that catalyze the non-hydrolytic cleavage of the peptidoglycan structures of the bacterial cell wall [68]. They are ubiquitous in bacteria which take part in a series of astonishingly diverse biological processes, such as cell wall metabolism, detection of cell-wall-acting antibiotics, insertion of secretion systems and flagellar assemblies into the cell wall and pathogenesis of certain gram-negative bacteria [69]. As a probable lytic transglycosylase protein, XALr_3249 in PlasmIII could be an attractive new target for the study of cell wall and virulence of *X. albilineans* in specific hosts. The importance of PlasmII shared by both strains also remains to be investigated.

5. Conclusions

We reported the first complete genome of a strain of *X. albilineans* (Xa-FJ1) from China using the PacBio RSII and Illumina HiSeq platforms. ANI analysis revealed that strain Xa-FJ1 had the closest evolutionary relationships with five strains of the pathogen from the French West Indies and the USA (Florida), particularly with strain GPE PC73 from Guadeloupe. These strains of *X. albilineans* have been associated with the most recently reported outbreaks and/or aerial spread of leaf scald in

China, Florida, several Caribbean Islands, and Cuba. Genome comparison analysis suggested that horizontal gene transfer and homologous recombination were tightly linked to genome evolution of *X. albilineans*. Our findings contribute additional genomic resources to further investigate the diversity and pathogenicity of the causal agent of sugarcane leaf scald.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2607/8/2/182/s1, Figure S1: Genome organization and gene distribution in strain Xa-FJ1 of *X. albilineans* from China. (A) Chromosome, and (B) Plasmid. From the outside to the inner circles: Coordinates (MB), protein-coding genes, COG annotations, KEGG annotations, GO annotations, ncRNA genes, percent GC content and GC skew. The outer part of the second, third, fourth, and fifth circle illustrates the forward strand, and the inner parts illustrate the reverse strand of the genome. Different colors of gene annotations refer to different categories summarized in the legend with colored squares (from the left to the right column: COG, KEGG, GO, and ncRNA). The red color of the GC content line plot indicates that GC percentage is higher than average while the green color of the same line plot indicates that GC percentage is lower than average. The purple and green inner circle shows the GC skew (G−C)/(G + C) using a 3724 bp window in Figure S1A and a 500 bp window in Figure S1B. Figure S2: GO annotation of the coding sequences of the genome of *X. albilineans* strain Xa-FJ1. Among the 24 subcategories of biological processes of the GO database, the largest category of Xa-FJ1 was assigned to the metabolic process (1186 genes). Figure S3: KEGG annotation of the coding sequences of the genome of *X. albilineans* strain Xa-FJ1. A total of 1491 genes of 2987 annotated genes were involved in metabolism, especially the metabolic pathways belonging to global and overview maps (529 genes). Figure S4: COG functional classification of the coding sequences of the genome of *X. albilineans* strain Xa-FJ1. Amino acid transport and metabolism was ranked the third largest term among the 25 classes of functional categories of the COG database. Figure S5: Comparison of the spacer distribution in the CRISPR-1 and CRISPR-2 systems between *X. albilineans* strains GPE PC73 and Xa-FJ1. Each box represents a CRISPR spacer and the spacer positions were numbered inside each box from the trailer end spacer (number 1) to the leader end spacer (as reported by Pieretti et al. [11]). Meaning of the colored boxes: blue = spacer identical between the two strains; orange and grey = sequence of spacer different between the two strains; yellow, green and purple = presence of additional spacer(s). R13-R15 refer to InDels described in the text. Table S1: Characteristics of 16 strains of *X. albilineans* from worldwide locations and one strain of *X. pseudalbilineans* used in this study. Table S2: General features of the plasmid of strain Xa-FJ1 of *X. albilineans* from China compared to the three plasmids of strain GPE PC73 from Guadeloupe. Table S3: Summary of gene annotation of strain Xa-FJ1. Table S4: Virulence factors retrieved from the VFDB database. Table S5: Putative function of 82 genes present in the chromosome of strain GPE PC73 of *X. albilineans* from Guadeloupe and absent in strain Xa-FJ1 from China. Table S6: Putative function of the genes identified in plasmids PlasmI and PlasmIII of strain GPE PC73 of *X. albilineans* from Guadeloupe.

References

1. Meyer, D.F.; Bogdanove, A.J. Genomics-driven advances in *Xanthomonas* biology. In *Plant Pathogenic Bacteria: Genomics and Molecular Biology*; Jackson, R.W., Ed.; Caister Academic Press: Norfolk, UK, 2009; pp. 147–161.
2. Rott, P.; Davis, M.J. Leaf scald. In *A Guide to Sugarcane Diseases*; Rott, P., Bailey, R.A., Comstock, J.C., Croft, B.J., Saumtally, S., Eds.; CIRAD/ISSCT: Montpellier, France, 2000; pp. 38–44.
3. Mensi, I.; Vernerey, M.S.; Gargani, D.; Nicole, M.; Rott, P. Breaking dogmas: The plant vascular pathogen *Xanthomonas albilineans* is able to invade non-vascular tissues despite its reduced genome. *Open Biol.* **2014**, *4*, 130116. [CrossRef] [PubMed]
4. Rott, P.; Mohamed, I.S.; Klett, P.; Soupa, D.; de Saint-Albin, A.; Feldmann, P.; Letourmy, P. Resistance to leaf scald disease is associated with limited colonization of sugarcane and wild relatives by *Xanthomonas albilineans*. *Phytopathology* **1997**, *87*, 1202–1213. [CrossRef] [PubMed]

5. Rott, P.; Marguerettaz, M.; Fleites, L.; Cociancich, S.; Girard, J.-C.; Pieretti, I.; Gabriel, D.W.; Royer, M. Unravelling pathogenicity of *Xanthomonas albilineans*, the causal agent of sugarcane leaf scald. In Proceedings of the 27th Congress of the International Society of Sugarcane Technologists, Veracruz, Mexico, 7–11 March 2010; pp. 1112–1121.

6. Hashimi, S.M.; Wall, M.K.; Smith, A.B.; Maxwell, A.; Birch, R.G. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob. Agents Chemother.* **2007**, *51*, 181–187. [CrossRef]

7. Rott, P.; Arnaud, M.; Baudin, P. Serological and lysotypical variability of *Xanthomonas albilineans* (Ashby) Dowson, causal agent of sugarcane leaf scald disease. *J. Phytopathol.* **1986**, *116*, 201–211. [CrossRef]

8. Rott, P.; Davis, M.J.; Baudin, P. Serological variability in *Xanthomonas albilineans*, causal agent of leaf scald disease of sugarcane. *Plant Pathol.* **1994**, *43*, 344–349. [CrossRef]

9. Alvarez, A.; Schenck, S.; Benedict, A. Differentiation of *Xanthomonas albilineans* strains with monoclonal antibody reaction patterns and DNA fingerprints. *Plant Pathol.* **1996**, *45*, 358–366. [CrossRef]

10. Davis, M.J.; Rott, P.; Warmuth, C.J.; Chatenet, M.; Baudin, P. Intraspecific genomic variation within *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Phytopathology* **1997**, *87*, 316–324. [CrossRef]

11. Pieretti, I.; Royer, M.; Barbe, V.; Carrere, S.; Koebnik, R.; Couloux, A.; Darrasse, A.; Gouzy, J.; Jacques, M.-A.; Lauber, E.; et al. Genomic insights into strategies used by *Xanthomonas albilineans* with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genom.* **2012**, *13*, 658. [CrossRef]

12. Lin, L.-H.; Ntambo, M.S.; Rott, P.C.; Wang, Q.-N.; Lin, Y.-H.; Fu, H.-Y.; Gao, S.-J. Molecular detection and prevalence of *Xanthomonas albilineans*, the causal agent of sugarcane leaf scald, in China. *Crop Prot.* **2018**, *109*, 17–23. [CrossRef]

13. Ntambo, M.S.; Meng, J.-Y.; Rott, P.C.; Royer, M.; Lin, L.-H.; Zhang, H.-L.; Gao, S.-J. Identification and characterization of *Xanthomonas albilineans* causing sugarcane leaf scald in China using multilocus sequence analysis. *Plant Pathol.* **2019**, *68*, 269–277. [CrossRef]

14. Bansal, K.; Midha, S.; Kumar, S.; Patil, P.B. Ecological and evolutionary insights into *Xanthomonas citri* pathovar diversity. *Appl. Environ. Microbiol.* **2017**, *83*, e02993-16. [CrossRef] [PubMed]

15. Pieretti, I.; Royer, M.; Barbe, V.; Carrere, S.; Koebnik, R.; Cociancich, S.; Couloux, A.; Darrasse, A.; Gouzy, J.; Jacques, M.-A.; et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *Bmc Genom.* **2009**, *10*, 616. [CrossRef] [PubMed]

16. Giampetruzzi, A.; Saponari, M.; Loconsole, G.; Boscia, D.; Savino, V.N.; Almeida, R.P.P.; Zicca, S.; Landa, B.B.; Chacon-Diaz, C.; Saldarelli, P. Genome-wide analysis provides evidence on the genetic relatedness of the emergent *Xylella fastidiosa* genotype in Italy to isolates from central America. *Phytopathology* **2017**, *107*, 816–827. [CrossRef] [PubMed]

17. Zhang, R.Y.; Shan, H.L.; Li, W.F.; Cang, X.Y.; Wang, X.Y.; Yin, J.; Luo, Z.M.; Huang, Y.K. First report of sugarcane leaf scald caused by *Xanthomonas albilineans* (Ashby) Dowson in the province of Guangxi, China. *Plant Dis.* **2017**, *101*. [CrossRef]

18. Zhang, R.-Y.; Wang, X.-Y.; Shan, H.-L.; Li, J.; Li, W.-F.; Cang, X.-Y.; Luo, Z.-M.; Yin, J.; Huang, Y.-K. Identification and phylogenetic analysis of *Xanthomonas albilineans* (Ashby) Dowson based on multiple gene sequences in Yunnan province, China. *Sugar Tech* **2019**, *21*, 794–801. [CrossRef]

19. Davis, M.J.; Rott, P.; Baudin, P.; Dean, J. Evaluation of selective media and immunoassays for detection of *Xanthomonas albilineans*, causal agent of sugarcane leaf scald disease. *Plant Dis.* **1994**, *78*, 78–82. [CrossRef]

20. Lim, H.J.; Lee, E.H.; Yoon, Y.; Chua, B.; Son, A. Portable lysis apparatus for rapid single-step DNA extraction of *Bacillus subtilis*. *J. Appl. Microbiol.* **2016**, *120*, 379–387. [CrossRef]

21. Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* **2012**, *13*, 341. [CrossRef]

22. Chin, C.S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E.; et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563–569. [CrossRef]

23. Besemer, J.; Lomsadze, A.; Borodovsky, M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **2001**, *29*, 2607–2618. [CrossRef]

24. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [CrossRef] [PubMed]

25. Lagesen, K.; Hallin, P.; Rodland, E.A.; Staerfeldt, H.H.; Rognes, T.; Ussery, D.W. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **2007**, *35*, 3100–3108. [CrossRef] [PubMed]

26. Gardner, P.P.; Daub, J.; Tate, J.G.; Nawrocki, E.P.; Kolbe, D.L.; Lindgreen, S.; Wilkinson, A.C.; Finn, R.D.; Griffiths-Jones, S.; Eddy, S.R.; et al. Rfam: Updates to the RNA families database. *Nucleic Acids Res.* **2009**, *37*, D136–D140. [CrossRef] [PubMed]

27. Nawrocki, E.P.; Kolbe, D.L.; Eddy, S.R. Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **2009**, *25*, 1335–1337. [CrossRef]

28. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]

29. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, D277–D280. [CrossRef]

30. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K.F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **2006**, *34*, D354–D357. [CrossRef]

31. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; et al. The COG database: An updated version includes eukaryotes. *BMC Bioinform.* **2003**, *4*, 41. [CrossRef]

32. Li, W.; Jaroszewski, L.; Godzik, A. Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* **2002**, *18*, 77–82. [CrossRef]

33. Saier, M.H., Jr.; Reddy, V.S.; Tsu, B.V.; Ahmed, M.S.; Li, C.; Moreno-Hagelsieb, G. The Transporter Classification Database (TCDB): Recent advances. *Nucleic Acids Res.* **2016**, *44*, D372–D379. [CrossRef]

34. Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28*, 45–48. [CrossRef] [PubMed]

35. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef] [PubMed]

36. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [CrossRef] [PubMed]

37. Urban, M.; Pant, R.; Raghunath, A.; Irvine, A.G.; Pedro, H.; Hammond-Kosack, K.E. The Pathogen-Host Interactions database (PHI-base): Additions and future developments. *Nucleic Acids Res.* **2015**, *43*, D645–D655. [CrossRef]

38. Chen, L.; Xiong, Z.; Sun, L.; Yang, J.; Jin, Q. VFDB 2012 update: Toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res.* **2012**, *40*, D641–D645. [CrossRef] [PubMed]

39. Liu, B.; Pop, M. ARDB–Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **2009**, *37*, D443–D447. [CrossRef] [PubMed]

40. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238. [CrossRef]

41. Ge, R.; Mai, G.; Wang, P.; Zhou, M.; Luo, Y.; Cai, Y.; Zhou, F. CRISPRdigger: Detecting CRISPRs with better direct repeat annotations. *Sci. Rep.* **2016**, *6*, 32942. [CrossRef]

42. Richter, M.; Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19126–19131. [CrossRef]

43. Richter, M.; Rossello-Mora, R.; Oliver Glockner, F.; Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **2016**, *32*, 929–931. [CrossRef]

44. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]

45. Li, W.; Jaroszewski, L.; Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **2001**, *17*, 282–283. [CrossRef] [PubMed]

46. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

47. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* **2004**, *5*, R12. [CrossRef]

48. Chiaromonte, F.; Yap, V.; Miller, W. Scoring pairwise genomic sequence alignments. In *Biocomputing 2002*; World Scientific: Singapore, 2001; pp. 115–126.

49. Harris, R.S. Improved Pairwise Alignment of Genomic DNA. Ph.D. Thesis, The Pennsylvania State University, State College, PA, USA, 2007.

50. Nakano, K.; Shiroma, A.; Shimoji, M.; Tamotsu, H.; Ashimine, N.; Ohki, S.; Shinzato, M.; Minami, M.; Nakanishi, T.; Teruya, K.; et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Hum. Cell* **2017**, *30*, 149–161. [CrossRef]

51. Ferrarini, M.; Moretto, M.; Ward, J.A.; Surbanovski, N.; Stevanovic, V.; Giongo, L.; Viola, R.; Cavalieri, D.; Velasco, R.; Cestaro, A.; et al. An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genom.* **2013**, *14*, 670. [CrossRef]

52. Rott, P.; Fleites, L.; Marlow, G.; Royer, M.; Gabriel, D.W. Identification of new candidate pathogenicity factors in the xylem-invading pathogen *Xanthomonas albilineans* by transposon mutagenesis. *Mol. Plant Microbe Interact.* **2011**, *24*, 594–605. [CrossRef]

53. Wang, L.; Jiang, W.; Zhang, Y.; Wang, S.; Ma, R.; Sun, W. Ax21-triggered immunity plays a significant role in rice defense against *Xanthomonas oryzae* pv. oryzicola. *Phytopathology* **2013**. [CrossRef]

54. Bahar, O.; Pruitt, R.; Luu, D.D.; Schwessinger, B.; Daudi, A.; Liu, F.; Ruan, R.; Fontaine-Bodin, L.; Koebnik, R.; Ronald, P. The *Xanthomonas* Ax21 protein is processed by the general secretory system and is secreted in association with outer membrane vesicles. *PeerJ* **2014**, *2*, e242. [CrossRef]

55. Shen, Y.; Sharma, P.; da Silva, F.G.; Ronald, P. The *Xanthomonas oryzae* pv. *oryzae raxP* and *raxQ* genes encode an ATP sulphurylase and adenosine-5′-phosphosulphate kinase that are required for AvrXa21 avirulence activity. *Mol. Microbiol.* **2002**, *44*, 37–48. [CrossRef]

56. Bobay, L.M.; Ochman, H. The evolution of bacterial genome architecture. *Front. Genet.* **2017**, *8*, 72. [CrossRef] [PubMed]

57. Koonin, E.V.; Makarova, K.S.; Aravind, L. Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu. Rev. Microbiol.* **2001**, *55*, 709–742. [CrossRef] [PubMed]

58. Ochman, H.; Lawrence, J.G.; Groisman, E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **2000**, *405*, 299–304. [CrossRef] [PubMed]

59. Murphy, J.; Mahony, J.; Ainsworth, S.; Nauta, A.; van Sinderen, D. Bacteriophage orphan DNA methyltransferases: Insights from their bacterial origin, function, and occurrence. *Appl. Environ. Microbiol.* **2013**, *79*, 7547–7555. [CrossRef] [PubMed]

60. Zhang, S.; Chakrabarty, P.K.; Fleites, L.A.; Rayside, P.A.; Hopkins, D.L.; Gabriel, D.W. Three new Pierce's disease pathogenicity effectors identified using *Xylella fastidiosa* biocontrol strain EB92-1. *Plos ONE* **2015**, *10*, e0133796. [CrossRef]

61. Pieretti, I.; Pesic, A.; Petras, D.; Royer, M.; Sussmuth, R.D.; Cociancich, S. What makes *Xanthomonas albilineans* unique amongst xanthomonads? *Front. Plant Sci.* **2015**, *6*, 289. [CrossRef]

62. Bullman, S.; Lucid, A.; Corcoran, D.; Sleator, R.D.; Lucey, B. Genomic investigation into strain heterogeneity and pathogenic potential of the emerging gastrointestinal pathogen *Campylobacter ureolyticus*. *Plos ONE* **2013**, *8*, e71515. [CrossRef]

63. Jangam, D.; Feschotte, C.; Betran, E. Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* **2017**, *33*, 817–831. [CrossRef]

64. Horvath, P.; Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **2010**, *327*, 167–170. [CrossRef]

65. Martins, P.M.; Machado, M.A.; Silva, N.V.; Takita, M.A.; de Souza, A.A. Type II toxin-antitoxin distribution and adaptive aspects on *Xanthomonas* genomes: Focus on *Xanthomonas citri*. *Front. Microbiol.* **2016**, *7*, 652. [CrossRef]

66. Xiao, J.; Worby, C.A.; Mattoo, S.; Sankaran, B.; Dixon, J.E. Structural basis of Fic-mediated adenylylation. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1004–1010. [CrossRef] [PubMed]

67. Worby, C.A.; Mattoo, S.; Kruger, R.P.; Corbeil, L.B.; Koller, A.; Mendez, J.C.; Zekarias, B.; Lazar, C.; Dixon, J.E. The fic domain: Regulation of cell signaling by adenylylation. *Mol. Cell* **2009**, *34*, 93–103. [CrossRef] [PubMed]

68. Scheurwater, E.; Reid, C.W.; Clarke, A.J. Lytic transglycosylases: Bacterial space-making autolysins. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 586–591. [CrossRef] [PubMed]

69. Dik, D.A.; Marous, D.R.; Fisher, J.F.; Mobashery, S. Lytic transglycosylases: Concinnity in concision of the bacterial cell wall. *Crit. Rev. Biochem. Mol. Biol.* **2017**, *52*, 503–542. [CrossRef] [PubMed]