

Article

Genome Sequencing of *Xanthomonas vasicola* Pathovar *vasculorum* Reveals Variation in Plasmids and Genes Encoding Lipopolysaccharide Synthesis, Type-IV Pilus and Type-III Secretion Effectors

Arthur Wasukira ^{1,2}, Max Coulter ¹, Noorah Al-Sowayeh ¹, Richard Thwaites ³,
Konrad Paszkiewicz ¹, Jerome Kubiriba ², Julian Smith ³, Murray Grant ¹
and David J. Studholme ^{1,*}

¹ Biosciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK;
E-Mails: awasukira@gmail.com (A.W.); mc343@exeter.ac.uk (M.C.); nama203@exeter.ac.uk (N.A.-S.);
k.h.paszkiewicz@exeter.ac.uk (K.P.); m.r.grant@exeter.ac.uk (M.G.)

² National Crops Resources Research Institute (NaCRRI), Kampala 7084, Uganda;
E-Mail: jkubiriba@kari.go.ug

³ The Food and Environment Research Agency, Sand Hutton, York YO41 1LZ, UK;
E-Mails: richard.thwaites@fera.gsi.gov.uk (R.T.); julian.smith@fera.gsi.gov.uk (J.S.)

* Author to whom correspondence should be addressed; E-Mail: d.j.studholme@exeter.ac.uk;
Tel.: +44-(0)-1392-72-4678.

Received: 13 December 2013; in revised form: 10 February 2014 / Accepted: 3 March 2014 /

Published: 18 March 2014

Abstract: *Xanthomonas vasicola* pathovar *vasculorum* (*Xvv*) is the bacterial agent causing gumming disease in sugarcane. Here, we compare complete genome sequences for five isolates of *Xvv* originating from sugarcane and one from maize. This identified two distinct types of lipopolysaccharide synthesis gene clusters among *Xvv* isolates: one is similar to that of *Xanthomonas axonopodis* pathovar *citri* (*Xac*) and is probably the ancestral type, while the other is similar to those of the sugarcane-inhabiting species, *Xanthomonas sacchari*. Four of six *Xvv* isolates harboured sequences similar to the *Xac* plasmid, pXAC47, and showed a distinct Type-IV pilus (T4P) sequence type, whereas the T4P locus of the other two isolates resembled that of the closely related banana pathogen, *Xanthomonas campestris* pathovar *musacearum* (*Xcm*). The *Xvv* isolate from maize has lost a gene encoding a homologue of the virulence effector, *xopAF*, which was present in all five of the sugarcane isolates, while *xopL* contained a premature stop codon in four out of

six isolates. These findings shed new light on evolutionary events since the divergence of *Xvv* and *Xcm*, as well as further elucidating the relationships between the two closely related pathogens.

Keywords: gumming disease; sugarcane; banana *Xanthomonas* wilt

1. Introduction

The bacterial genus, *Xanthomonas*, includes many economically important pathogens of crop plants [1]. Genome sequencing of species and pathovars of *Xanthomonas* has led to insights into the evolution and mechanisms of virulence and their ability to overcome host defences [2]. One pathovar whose study at the molecular level has been relatively limited is *Xanthomonas vasicola* pathovar *vasculorum* (*Xvv*).

Strains of *Xvv* are responsible for gumming disease in sugarcane and are also found naturally infecting some other monocotyledonous hosts, including maize and sorghum [3]. In addition to the importance of gumming disease and *Xvv per se*, strains of this pathovar share a recent common ancestor with the causal agent of banana *Xanthomonas* wilt, namely *Xanthomonas campestris* pathovar *musacearum* (*Xcm*), with which they share an identical *gyrB* DNA sequence [4]. Strains of *Xvv* do not cause disease in banana [4], but *Xcm* can infect maize and sugarcane.

The narrow genetic diversity in *Xcm* [5,6] suggests a population bottleneck, and *Xcm* may represent a single clonal lineage of the species, *X. vasicola*, or a closely related species, that has recently emerged to colonise bananas in east Africa. Whole-genome sequence data can provide greater phylogenetic resolution than analysis of a single gene fragment. Although all isolates shared identical *gyrB* amplicon sequences, analyses of genome-wide sequence data revealed that *Xcm* and *Xvv* comprise two distinct clades, although they are closely related to each other [7]. This suggests that *Xcm* did not arise from *Xvv*, but rather from some currently unknown close relative of *Xvv*, possibly another pathovar belonging to the same species (*X. vasicola*).

Given the close phylogenetic relationship between *Xcm* and *Xvv*, molecular comparison between *Xcm* and *Xvv* is of importance for understanding the evolution of *Xcm* and the adaptation to banana. To better understand the range of diversity within *Xvv*, we analysed available genome sequence data for four isolates of *Xvv* (described in previous publications [7,8]) and new sequence data from a further two isolates of *Xvv*.

There has been some confusion in the literature regarding the taxonomy of the *Xvv* strains, whose genome sequences we analyse here. Lewis Ivey and colleagues [9] listed strains NCPPB (National Collection of Plant Pathogenic Bacteria) 702, NCPPB 1326 and NCPPB 206 as *X. axonopodis* pathovar *vasculorum*. In contrast, Qhobela and Claflin [10] classified strain NCPPB 1326 as *X. campestris* pathovar *vasculorum*. However, Vauterin and colleagues [11] proposed that these strains be included in the species, *X. vasicola* (as pathovar *vasculorum*), on the basis of genomic DNA hybridization studies. A subsequent analysis of cellular fatty acid profiles also confirmed a close affinity between these strains and strains of the species, *X. vasicola*, and only rather distant similarity to species *X. axonopodis* and *X. campestris* [3]. Rademaker and colleagues [12] also list strain NCPPB 206 (synonymous with LMG 8284) as belonging to species *X. vasicola* (pathovar *vasculorum*, i.e., *Xvv*).

Overall, the available evidence overwhelmingly supports the inclusion of these isolates in species *X. vasicola*, and therefore, in this manuscript, we refer to them as *Xvv* (rather than as *X. campestris* or *X. axonopodis*).

We previously generated draft genome sequences for one isolate of *Xcm* and one isolate of *Xvv* and identified several differences between them that might contribute to their distinct host ranges [8]; we found differences between *Xcm* NCPPB 4381 and *Xvv* NCPPB 702 in their repertoires of Type-III secretion system (T3SS) effectors, lipopolysaccharide (LPS) synthesis genes and Type-IV pilus (T4P) genes, but these differences were between just a pair of isolates, and so, it was unclear how generalisable these results were to other isolates of *Xcm* and *Xvv*. In a more recent publication, we reported sequencing the genomes of a further 13 isolates of *Xcm* and three isolates of *Xvv* [7]. The main focus of that study [7] was genetic variation among isolates of *Xcm*, revealing two main phylogenetic groups (or sub-lineages) among *Xcm*. That previous study [7] reported a list of genes that were consistently conserved in *Xcm* and absent from the four *Xvv* genome sequences then available, and the study used data from the four *Xvv* genome sequences to generate a phylogenetic tree; however, no further analysis of the *Xvv* genome sequences was reported.

The current study extends the previous work by systematically searching for genetic differences among *Xvv* isolates, rather than focusing on variation between *Xvv* and *Xcm* or variation among isolates of *Xcm*. Furthermore, included in the current study are sequence data from two additional strains of *Xvv*, bringing the total number of sequenced *Xvv* isolates up to six. The sequence analyses presented here revealed large differences in gene-content among isolates of *Xvv*, both among isolates from sugarcane and also between isolates from sugarcane and a single isolate from maize. Some of these differences in gene content are ascribed to the gain and loss of plasmids, though many of the differences are likely to be chromosomally located.

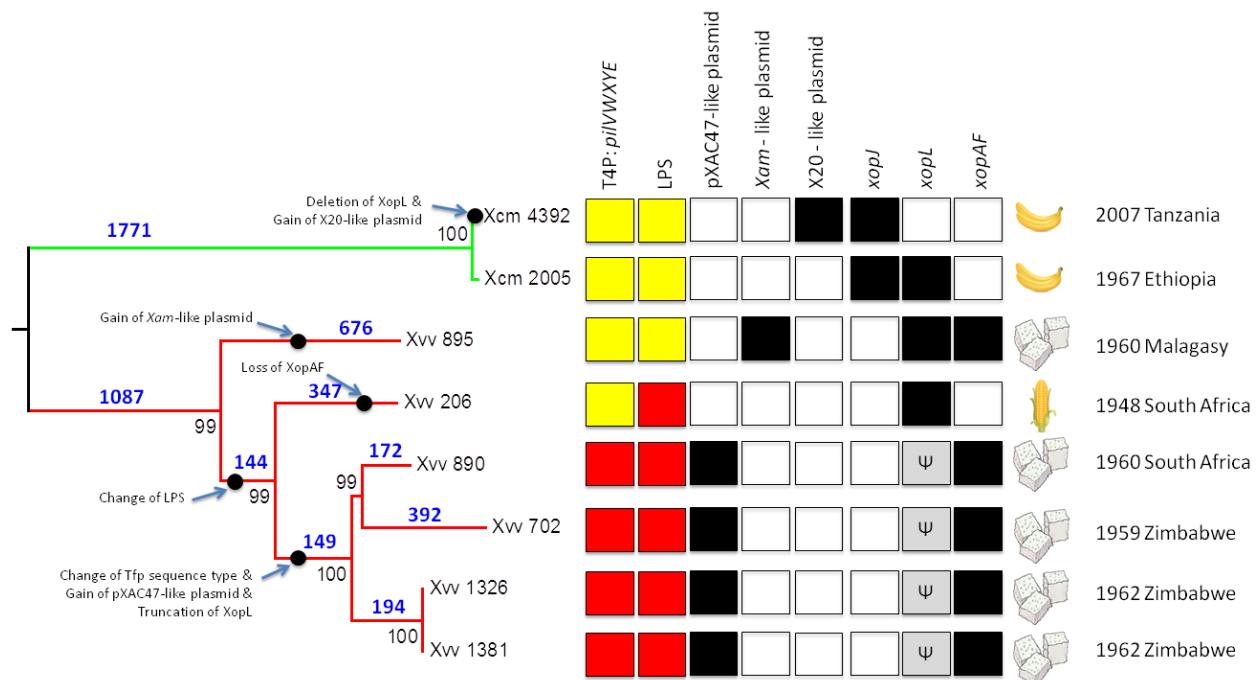
We also report genetic differences implicated in important extracellular structures, such as lipopolysaccharide (LPS) synthesis, the T4P and candidate substrates (*i.e.*, effectors) of the T3SS. Furthermore, further analysis of previously published data revealed some hitherto undetected differences in gene content among *Xcm* isolates, including a homologue of T3SS effector XopL and a plasmid.

2. Results and Discussion

By comparison of genome sequence data, we identified several likely important events in the evolutionary history of *Xvv* and *Xcm*; these are summarized in Figure 1 and include the acquisitions of plasmids and the exchange of genes encoding LPS biosynthesis and T4P, as well as the loss and gain of candidate T3SS effector genes. The evidence supporting the proposal of each of these events is presented in Subsections 2.2–2.9 and Figures 2–6. Further details of these findings are available in the Supplementary Material, which contains an additional 19 figures. The main findings were that: a plasmid similar to pXAC47 is found in a clade of four *Xvv* isolates from sugarcane (Section 2.4); the most ancestrally branching *Xvv* sugarcane isolate (NCPPB 895) may contain a plasmid similar to that from a cassava pathogen (Section 2.5); some (but not all) *Xcm* isolates harbour sequences similar to that from a plasmid in a cotton pathogen (Section 2.6); there is considerable variation in T4P genes among *Xvv* isolates (Section 2.7); there are two distinct sequence types of the LPS biosynthesis gene

cluster among *Xvv* isolates (Section 2.8); and *Xvv* isolates vary with respect to their repertoires of putative T3SS effector genes (Section 2.9).

Figure 1. Overview of some key genetic changes during the evolution of *Xvv* and *Xcm*. The phylogenetic tree on the left-hand side (left) was inferred using the maximum parsimony method based on 39,665 single-nucleotide variants with respect to the chromosome of *Xoo* MAFF 311018. Bootstrap values are expressed as percentages of 500 trials in black. The tree is drawn to scale with branch lengths calculated using the average pathway method. For clarity, the numbers of single-nucleotide differences are indicated on the branches in blue boldface. The square boxes indicate presence (black) or absence (white). The grey boxes marked “Ψ” denote that the *xopL* gene is interrupted by a premature stop codon. The two distinct types of the T4P *pilVWXYZE* gene cluster (see Section 2.7 and Figure 4) are indicated, respectively, by yellow and red shading. The two different sequence types of the lipopolysaccharide (LPS) gene cluster (see Section 2.8 and Figure 6) are indicated, respectively, by yellow and red shading. The illustrations on the right-hand side (right) indicate the origin of each bacterial isolate: banana/enset, sugarcane or maize.



2.1. Overview of Sequence Data

Table 1 gives a brief description of the bacterial isolates from which the sequencing data used in this study originate. Table 2 lists summary statistics for the raw sequence data and Table 2 summarises the *de novo* assembly statistics. The sequence data from *Xcm* and for one *Xvv* isolate (*Xvv* 702) were described in previous publications [7,8]. The sequence data for four of the *Xvv* isolates were mentioned in a previous publication [7], but assembly statistics were not given. Therefore, we provide details of the assemblies here (Table 3). The contiguity of the assemblies for the two newly presented isolates (*Xvv* 890 and *Xvv* 895) are much lower than those of the previously presented assemblies (see N_{50} in

Table 3). However, it should be noted that the incompleteness of *de novo* assemblies does not invalidate the results presented in the current study, since our inferences are based upon comparisons of alignments of raw reads rather than comparisons between assemblies; these alignments consist of unassembled read-pairs aligned against various reference genome sequences, and several examples of such alignments are illustrated in the Supplementary Material. The single exception is Figure 2, in which the comparison consists of alignments between *de novo* assemblies; it is possible that some gaps in the alignments in Figure 2 could arise through the incompleteness of the *de novo* assemblies. Figures 4 and 5 are based on alignments between *de novo* assemblies, but the findings were also validated by the inspection of alignments of raw (unassembled) sequence reads (see Supplementary Material).

Five out of the six *Xvv* strains were originally isolated from sugarcane; the exception is 206, which was isolated from maize. Genomes of *Xvv* 890 and 895 were newly sequenced for this study. Genome sequences of *Xvv* 206, 1326 and 1381 were reported in a previous publication [7], but with only very limited analysis, since the focus of that study was on single-nucleotide polymorphism (SNP) in *Xcm*. The genome sequence of *Xvv* 702 was previously reported and compared against that of *Xcm* 4381 [8]. We also included genome sequence data from several isolates of *Xcm* that have previously been published [7,8], because these are the sequenced genomes most closely related to *Xvv* and probably belong to the same species, *X. vasicola* [4].

Table 1. Bacterial isolates and sequence datasets used in this study.

Isolate (NCPBP number)	Source and date of isolation	Source
<i>Xvv</i> 206 ^a	South Africa 1948	<i>Zea mays</i>
<i>Xvv</i> 702 ^b	Zimbabwe 1959	<i>Saccharum officinarum</i>
<i>Xvv</i> 890 ^c	South Africa 1960	<i>S. officinarum</i>
<i>Xvv</i> 895 ^c	Malagasy Republic 1960	<i>S. officinarum</i>
<i>Xvv</i> 1326 ^a	Zimbabwe 1962	<i>S. officinarum</i>
<i>Xvv</i> 1381 ^a	Zimbabwe 1962	<i>S. officinarum</i>
<i>Xcm</i> 2005 ^a	Ethiopia 1967	<i>Ensete ventricosum</i>
<i>Xcm</i> 2251 ^a	Ethiopia 1969	<i>Musa</i> sp.
<i>Xcm</i> 4379 ^a	Uganda 2007	<i>Musa</i> sp.
<i>Xcm</i> 4380 ^a	Uganda 2007	<i>Musa</i> sp.
<i>Xcm</i> 4381 ^b	Uganda 2007	<i>Musa</i> sp.
<i>Xcm</i> 4383 ^a	Uganda 2007	<i>Musa</i> sp.
<i>Xcm</i> 4384 ^a	Uganda 2007	<i>Musa</i> sp.
<i>Xcm</i> 4387 ^a	D. R. Congo 2007	<i>Musa</i> sp.
<i>Xcm</i> 4389 ^a	Rwanda 2007	<i>Musa</i> sp.
<i>Xcm</i> 4392 ^a	Tanzania 2007	<i>Musa</i> sp.
<i>Xcm</i> 4394 ^a	Tanzania 2007	<i>Musa</i> sp.
<i>Xcm</i> 4395 ^a	Tanzania 2007	<i>Musa</i> sp.
<i>Xcm</i> “Kenyan” ^d	Kenya (year not known)	<i>Musa</i> sp.

^a These sequence data were previously reported in [7]. ^b These sequence data were previously reported in [8].

^c These sequences were newly generated for this study. ^d This sequence assembly was submitted to GenBank by the International Institute of Tropical Agriculture in 2011, but no accompanying manuscript has been published to the best of our knowledge. NCPBP, National Collection of Plant Pathogenic Bacteria.

Table 2. Summary of sequence data.

Isolate (NCPBP number)	Number of read pairs	Read length	Coverage	SRA accession
<i>Xvv</i> 206 ^a	2,579,404	76	70 x	SRR494500.3
<i>Xvv</i> 702 ^b	2,913,785	36	35 x	SRR020202.3
<i>Xvv</i> 890 ^c	4,843,028	67	58 x	SRR1045340
<i>Xvv</i> 895 ^c	2,867,513	67	50 x	SRR1045341
<i>Xvv</i> 1326 ^a	2,365,912	76	63 x	SRR494491.5
<i>Xvv</i> 1381 ^a	2,450,234	76	66 x	SRR494499.3
<i>Xcm</i> 2005 ^a	2,536,030	76	72 x	SRR489154.7
<i>Xcm</i> 4379 ^a	3,652,875	76	102 x	SRR494484.2
<i>Xcm</i> 4380 ^a	4,069,509	76	113 x	SRR494485.2
<i>Xcm</i> 4381 ^b	5,052,905	36	56 x	SRR020203.3
<i>Xcm</i> 4384 ^a	1,976,797	76	55 x	SRR494488.2
<i>Xcm</i> 4392 ^a	2,554,161	76	72 x	SRR494498.3
<i>Xcm</i> 4394 ^a	3,295,956	76	92 x	SRR494489.1
<i>Xcm</i> 4395 ^a	4,117,662	76	117 x	SRR494490.2

^a These sequence data were previously reported in [7]. ^b These sequence data were previously reported in [8].

^c These sequences were newly generated for this study.

Table 3. Summary statistics for *de novo* genome sequence assemblies.

Isolate (NCPBP number)	Number of scaffolds	Scaffolds N ₅₀	Total length (b.p.)	GenBank accession
<i>Xvv</i> 206 ^a	177	103,376	4,825,935	AKBM00000000.1
<i>Xvv</i> 702 ^b	97	205,000	5,478,002	ACHS00000000.1
<i>Xvv</i> 890 ^c	2,565	3,627	4,951,053	AKBN00000000.1
<i>Xvv</i> 895 ^c	1,229	8,362	4,803,807	AKBO00000000.1
<i>Xvv</i> 1326 ^a	253	74,455	4,951,570	AKBK00000000.1
<i>Xvv</i> 1381 ^a	137	108,088	4,958,101	AKBL00000000.1
<i>Xcm</i> 2005 ^a	156	61,233	4,692,764	AKBE00000000.1
<i>Xcm</i> 4379 ^a	95	147,554	4,758,198	AKBF00000000.1
<i>Xcm</i> 4380 ^a	87	149,435	4,751,644	AKBG00000000.1
<i>Xcm</i> 4381 ^b	115	143,688	4,793,900	ACHT00000000.1
<i>Xcm</i> 4384 ^a	84	157,780	4,741,777	AKBH00000000.1
<i>Xcm</i> 4392 ^a	162	90,974	4,728,564	AKBI01000000.1
<i>Xcm</i> 4394 ^a	85	151,917	4,792,617	AKBJ00000000.1
<i>Xcm</i> “Kenyan” ^d	510	36,401	4,907,936	AGFQ01000000.1

^a These sequence data were previously reported in [7]. ^b These sequence data were previously reported in [8].

^c These sequences were newly generated for this study. ^d This sequence assembly was submitted to GenBank by the International Institute of Tropical Agriculture in 2011, but no accompanying manuscript has been published to the best of our knowledge.

Figure 2. Global comparison of *Xvv* and *Xcm* genomes with reference to the *Xoo* chromosome sequence. We aligned each of the genome sequence assemblies (the GenBank accession numbers are listed in Table 1) against the *Xoo* MAFF 311018 chromosome sequence using Basic Local Alignment Search Tool Nucleotide tool BLASTN. The innermost ring indicates the genomic position. The next ring is a plot of G + C content. The remaining 14 concentric rings indicate the presence or absence of BLASTN hits at that position, with one ring corresponding to each of 14 *Xcm* and *Xvv* genome assemblies. To aid clarity, each ring is represented in a different colour. Positions covered by BLASTN alignments are indicated with a solid colour; whitespace gaps represent genomic regions not covered by the BLASTN alignments. The accession numbers of the genome assemblies used in this figure are listed in Table 3. The graphical view of the alignments was rendered using BLAST Ring Image Generator (BRIG).

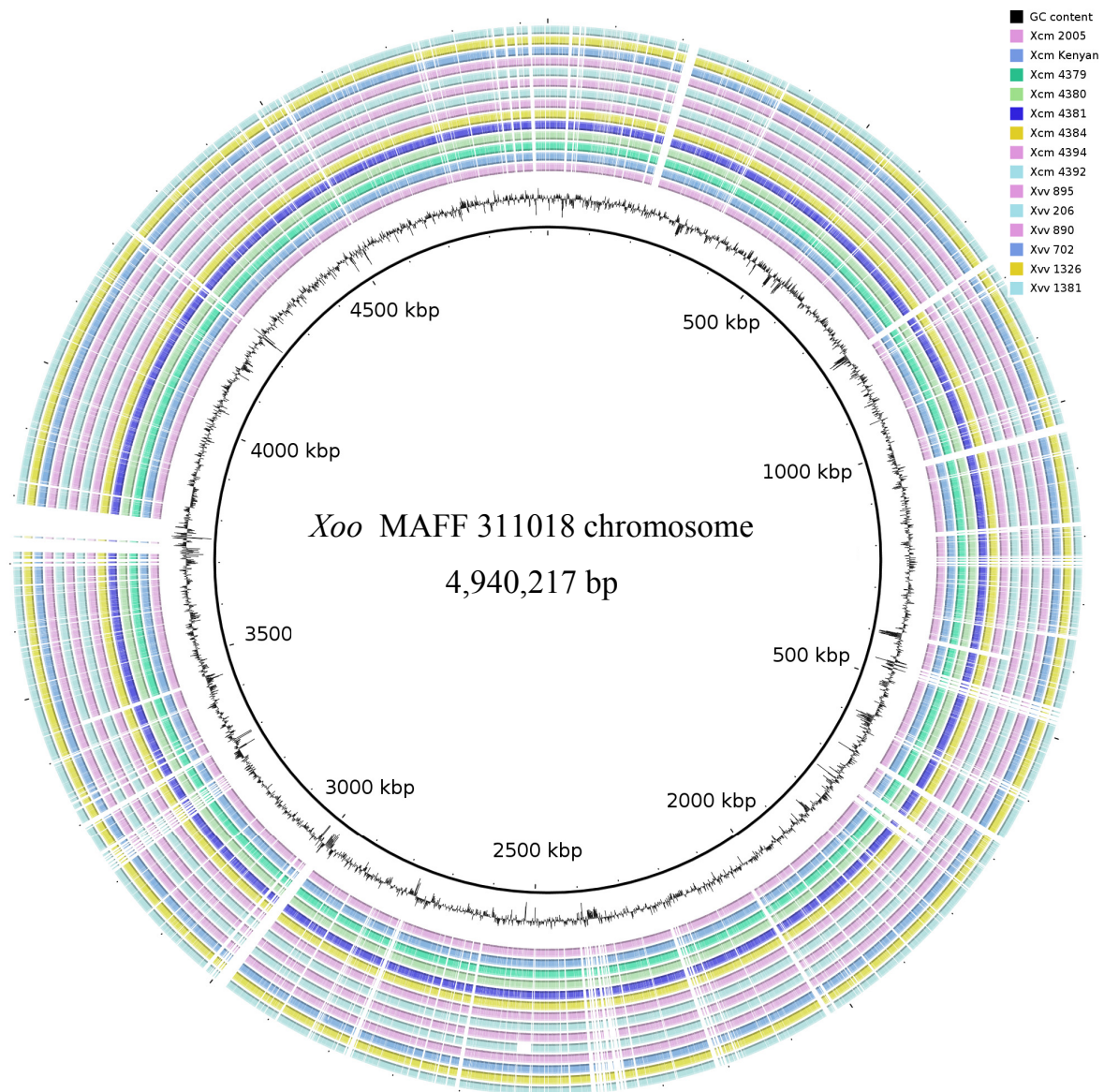


Figure 3. Heatmap depicting *Xvv* and *Xcm* gene-contents with reference to the *Xoo* chromosome sequence. We aligned each of the sets of genomic DNA sequence reads (SRA accession numbers listed in Table 2) against the *Xoo* MAFF 311018 chromosome sequence using BWA. We used the coverageBed tool from BEDtools to calculate the breadth of coverage of each *Xoo* gene by genomic sequence data from each *Xvv* and *Xcm* isolate.

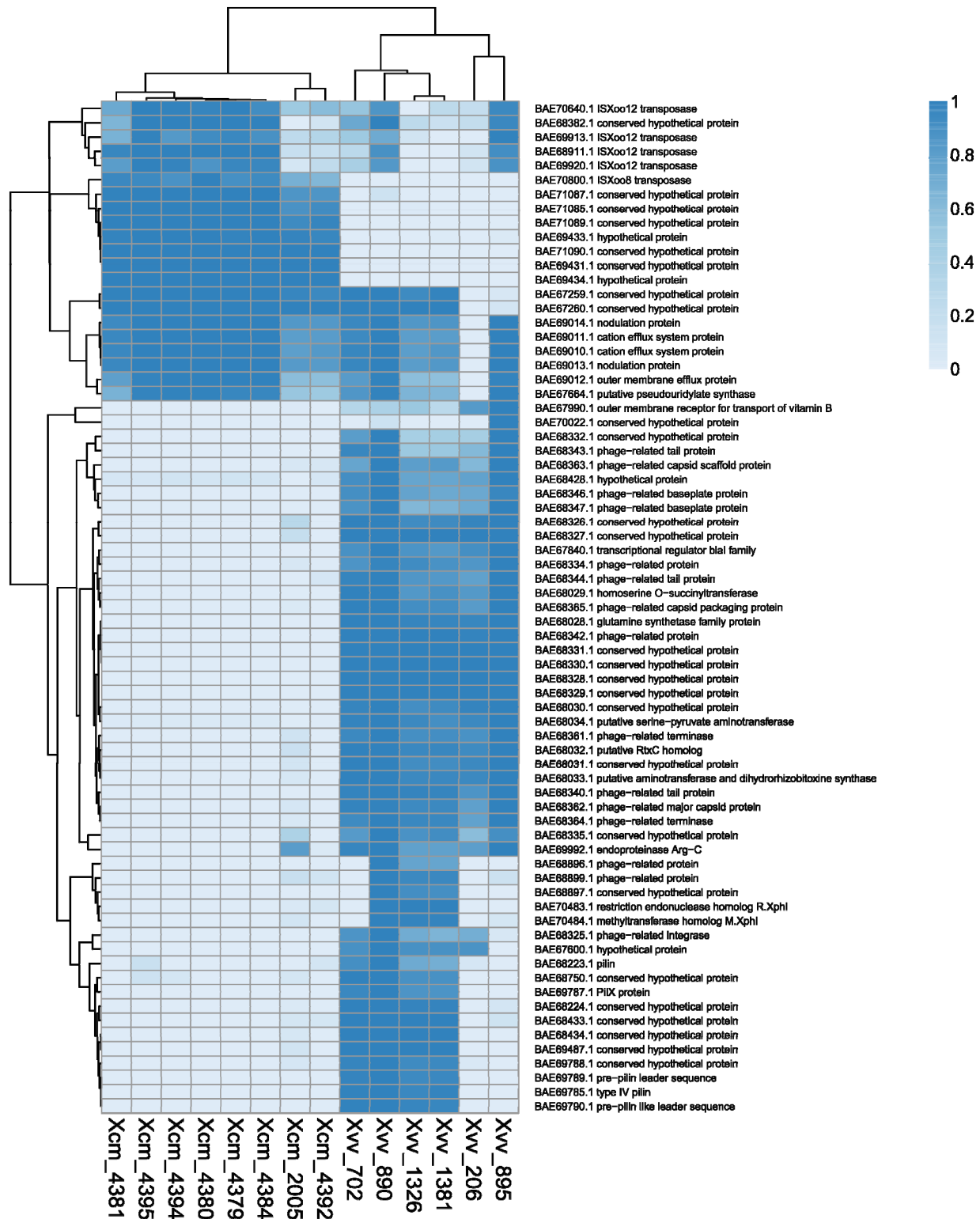


Figure 4. Two sequence types of the *pilVWXYZ* gene cluster in *Xv*: *Xoo*-like and *X. vesicatoria*-like. BLASTN alignments between the genome assemblies are displayed using the Artemis Comparison Tool (ACT). In the chromosome of *Xoo* MAFF 311018, this cluster falls between Positions 3436570 and 3443633 and includes locus tags XOO3030 to XOO3035. It includes the accessions BAE6788, BAE6789, BAE69785, BAE69787 and BAE69790 that appear in the heatmap in Figure 3; accession numbers: AP008229.1 (*Xoo* MAFF 311018) and AEQV01000175.1 (*X. vesicatoria* ATCC 35937). Each genome is represented by a horizontal white bar, labelled with the genomic position, flanked above and below by grey bars. Between each adjacent pair of genomes, BLASTN alignments are indicated by red blocks (for alignments between the positive strands of both sequences) or blue blocks (for alignments between the opposite strands of both sequences). Note, however, that the orientation of the BLASTN alignment (and hence, red or blue colouring) is not biologically significant, but rather just reflects the direction in which the double-stranded genome sequence has been represented in the database. Positions of annotated genes are indicated on the genome sequence of *X. vesicatoria* ATCC 35937. Note that *Xv* 890, *Xv* 702, *Xv* 1326 and *Xv* 1381 share sequence similarity along the whole length of this genomic region, and for most of the region, they also share sequence similarity with *Xoo* MAFF 311018. They differ, over the central part of this genomic region, from *Xv* 206, *Xv* 895 and *Xcm* 2005, all of which show sequence similarity to *X. vesicatoria* ATCC 35937 over this region. See Section 2.7.1.

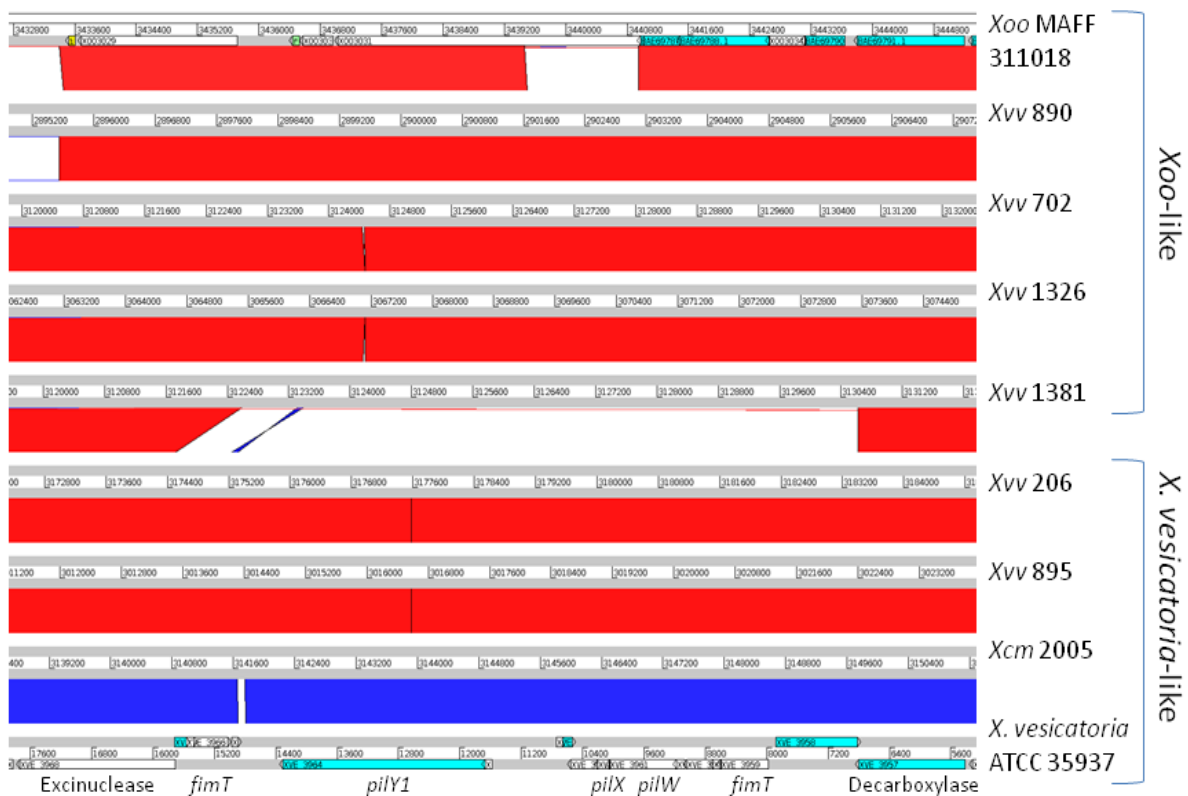
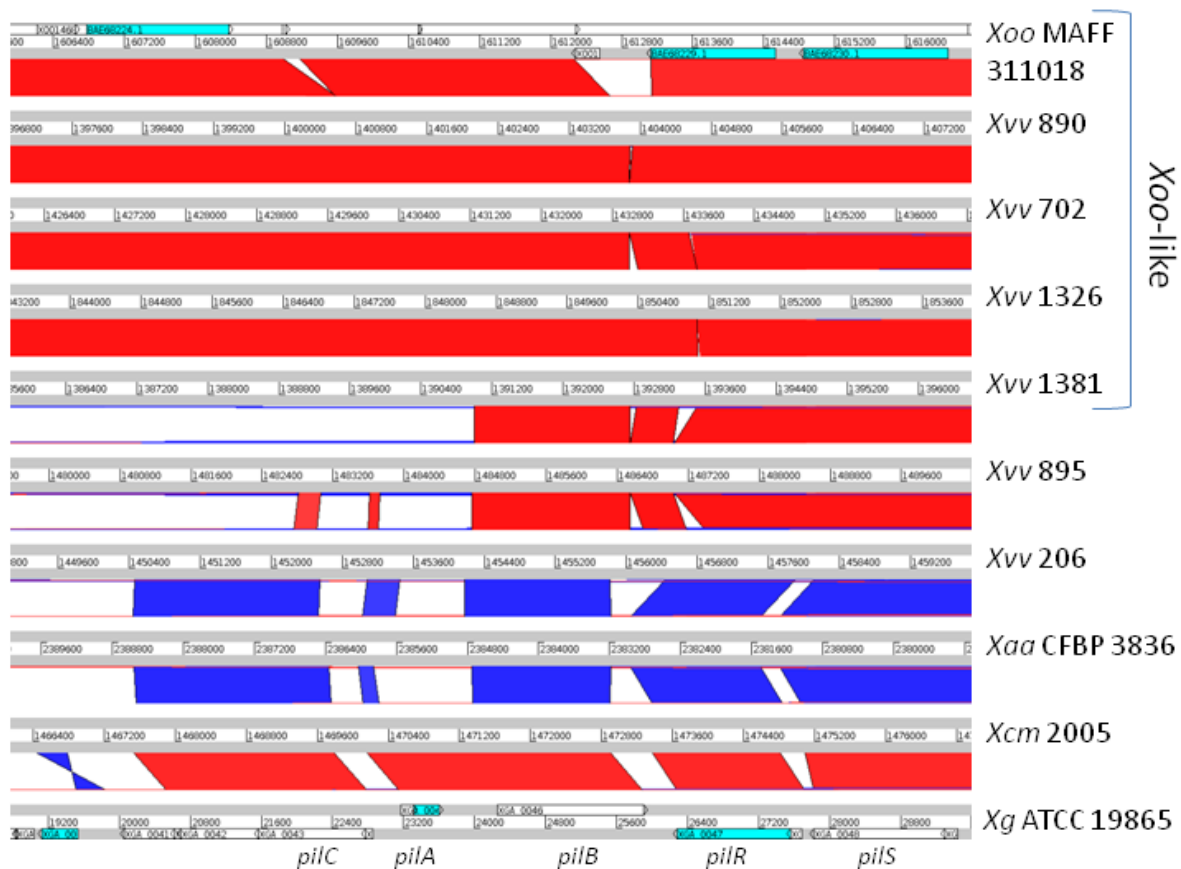


Figure 5. Multiple sequence types of the *pilCABRS* gene cluster in *Xv*. BLASTN alignments between the genome assemblies are displayed using the Artemis Comparison Tool (ACT). In *Xoo* MAFF 311018, this cluster is located between Positions 1613866 and 1616597 and encompasses Loci XOO1467 to XOO1475; accession numbers: AP008229.1 (*Xoo* MAFF 311018); AUWN01000008.1 (*X. alfalfae* subsp. *alfalfae* (*Xaa*) CFBP 3836); and AEQX01000002.1 (*X. gardneri* (*Xg*) ATCC 19865). Each genome is represented by a horizontal white bar, labelled with the genomic position, flanked above and below by grey bars. Between each adjacent pair of genomes, BLASTN alignments are indicated by red blocks (for alignments between the positive strands of both sequences) or blue blocks (for alignments between the opposite strands of both sequences). Note, however, that the orientation of the BLASTN alignment (and hence, blue or red colouring) is not biologically significant, but rather just reflects the direction in which the double-stranded genome sequence has been represented in the database. Positions of annotated genes are indicated on the genome sequence of *X. vesicatoria* ATCC 35937. Note that *Xv* 890, *Xv* 702, *Xv* 1326 and *Xv* 1381 share sequence similarity along the whole length of this genomic region, and for most of the region, they also share sequence similarity with *Xoo* MAFF 311018. However, the *pilC* and *pilA* genes are highly divergent among *Xv* 895, *Xv* 206 and *Xcm* 2005, with *Xcm* 2005 having *X. gardneri*-like *pilC* and *pilA* genes. *Xv* 895 and *Xv* 206 have *Xaa*-like *pilC* and *pilA* genes. See Section 2.7.2.



2.2. Phylogenetic Relationships among *Xvv* Strains

We generated a phylogenetic reconstruction of the sequenced *Xvv* and *Xcm* strains based on single-nucleotide polymorphisms called against the reference sequence of the *X. oryzae* pathovar *oryzae* (*Xoo*) MAFF 311018 [13]. The maximum parsimony phylogenetic tree is shown in Figure 1; the maximum likelihood method produced identical topology, and the topology is consistent with the tree that we previously presented [7]. The *Xvv* strains form a monophyletic clade closely related to, but distinct from, *Xcm* and the genetic distances among these sequenced *Xvv* isolates are considerably larger than those among *Xcm* isolates.

Within the sequenced *Xvv* isolates, the single isolate from maize falls within the diversity of sugarcane isolates; in other words, there are not separate monophyletic groups for isolates from the two different hosts. Two of the sugarcane-derived isolates (*Xvv* 1326 and 1381) are indistinguishable on the basis of the SNPs used to generate the tree in Figure 1. They had both been collected from sugarcane in Zimbabwe in 1962 and may be essentially two isolates of the same bacterial population. However, these two isolates are genetically distinct from the isolate collected from sugarcane in Zimbabwe three years earlier (*i.e.*, *Xvv* 702).

2.3. Global Genomic Comparison of Genomes of *Xvv* and *Xcm* Isolates

Alignment of our genome assemblies against a closely related reference chromosome sequence (*Xoo* MAFF 311018) suggested numerous differences in gene content both between *Xvv* and *Xcm* and also among *Xvv* isolates (see Figure 2). Particularly noticeable in Figure 2 is a region of the *Xoo* genome (centred at position 2,508 Mb) that is absent from *Xvv* 206, the isolate from maize. This consists of an 18-kb region of the genome, including the eight loci XOO2253–XOO2263 (GenBank accession numbers: BAE69008.1–BAE69018.1) that includes several predicted efflux proteins of unknown function. The absence of this region in *Xvv* 206 was confirmed by the alignment of sequence reads against the *Xoo* reference genome, independently of any *de novo* assembly.

We identified *Xoo* genes that are differentially present or missing in each of the *Xvv* and *Xcm* isolates (Figure 3). To do this, we aligned the Illumina sequence reads against the *Xoo* chromosome sequence using BWA [14]. We then calculated the breadth of coverage of each *Xoo* gene in each *Xvv* and *Xcm* isolate using the coverageBed tools from the BEDtools suite [15]. Note that this analysis involved the alignment of raw sequence reads against the reference; it was not dependent on any *de novo* assembly of our sequence data. We confirmed differential presence/absence by PCR amplification from genomic DNA (Figure 4). Breadths of coverage for each gene are indicated by the heatmap in Figure 3, which reveals numerous genes, whose presence distinguishes *Xcm* from *Xvv* and also several that distinguish among *Xvv* isolates. However, this approach is limited to the analysis of genes that are in the chromosome of *Xoo*. It excludes plasmids, as well as chromosomal genes in *Xvv* or *Xcm* that are not conserved in *Xoo*. Therefore, we also performed similar analyses based on using genomic assemblies of *Xvv* and *Xcm* genomes instead of the *Xoo* chromosome sequence. The results of these analyses are presented as heatmaps in the Supplementary Material. Some of these differences in gene content are discussed in more detail below.

2.4. *Xvv* Isolates NCPPB 890, 702, 1326 and 1381 Contain Sequences Similar to Plasmid pXAC47

We searched for evidence of plasmids in the *Xvv* and *Xcm* genomes by aligning the Illumina reads against all bacterial plasmid sequences in the RefSeq database as of the October 31, 2013. The *Xvv* Isolates 890, 702, 1326 and 1381 all yielded Illumina sequence reads that cover about 70% of the length of plasmid pXAC47 from *Xanthomonas axonopodis* pathovar *citri* (*Xac*) 29-1 (the RefSeq accession number is NC_020798.1). The genomes of *Xvv* Isolates 890, 702, 1326 and 1381 all contain sequences with extensive similarity to pXAC47, while the other isolates of *Xvv* (and *Xcm*) do not.

2.5. *Xvv* Isolate NCPPB 895 Contains Sequences Similar to a Plasmid from *X. axonopodis* pv. *manihotis* (*Xam*)

We found extensive sequence similarity between *Xvv* 895 and sequences from several recently published draft genome sequences [16] of *Xam*, a pathogen of cassava. These sequences are not found in the other sequenced *Xvv* isolates nor in the sequenced *Xcm* isolates. Specifically, a 34.7-kb *Xvv* 890 contig (GenBank: AKBO01000002.1) shared 99% nucleotide sequence identity with contigs from *Xam* strains NG1, UA306, ORST17 and ORST X27 originally isolated from Nigeria, Columbia, Congo and Togo (the GenBank accession numbers for these *Xam* sequences are, respectively: AKEG01000263.1, AKEN01000114.1, AKEH01000160.1 and AKEJ01000091.1).

Some other strains of *Xam* also contained some slightly less-similar sequences, sharing up to 89% nucleotide sequence identity, including strains IBSBF725, IBSBF436, IBSBF285, IBSBF2820 and UA324. Most of the 41.4-kb sequence from *Xam* ORST X27 is conserved in *Xvv* 895, but absent from the other sequenced *Xvv* and *Xcm* isolates. This sequence is annotated as a (partial) plasmid [16], presumably because it contains several conjugative transfer genes. Overall, these observations indicate that *Xvv* 895 has acquired a plasmid that is closely related to plasmids that are circulating in *Xam* strains, suggesting the inter-species exchange of plasmids over a wide geographic range.

2.6. *Xcm* Isolates NCPPB 4379, 4380, 4383, 4384, 4392 and 4395 Contain Sequences Similar to a Plasmid from *X. citri* pv. *malvacearum* Strain X20

Previously, we found no evidence for the presence of plasmids in *Xcm* 4381 [8]. However, examination of data from *Xcm* Isolates 4379, 4380, 4383, 4384, 4392 and 4395 revealed extensive sequence similarity to a recently reported plasmid sequence from *X. citri* pv. *malvacearum* strain X20, a highly virulent pathogen of cotton isolated in Burkina Faso [17]. This 39-kb plasmid sequence (GenBank: CM002030.1) was not present in any of the sequenced *Xvv* isolates nor in any of the *Xcm* isolates belonging to Sub-lineage I (as defined by [7]); it appears to be restricted to *Xcm* Sub-lineage II among the isolates of which it is present in all, except *Xcm* 4381. The most parsimonious explanation for this pattern of distribution is that this plasmid was acquired by the common ancestor of *Xcm* Sub-lineage II and subsequently lost in *Xcm* 4381. The GenBank accession number for the corresponding plasmid sequence in the *Xcm* 4384 *de novo* assembly is AKBH01000036.1. The average nucleotide sequence identity between the *X. citri* pv. *Malvacearum* plasmid and *Xcm* was 92%, somewhat lower than the 99% identity between *Xvv* 895 and *Xam* plasmid sequences.

2.7. Genetic Variation in the Type-IV Pilus (T4P) Among Isolates of *Xv*

The global analyses of gene content (see Figure 3) revealed differences among the *Xv* isolates with respect to their T4P genes (Figures 4 and 5). For example, Figure 3 shows that *Xv* Isolates 702, 890, 1326 and 1381 are distinguished from the other sequenced isolates by the presence of several T4P-related genes (including GenBank accessions: BAE68223, BAE6788, BAE6789, BAE69785, BAE69787 and BAE69790).

2.7.1. The *pilVWXYZE* Gene Cluster

In *Xanthomonas* species, the T4P apparatus is encoded by clusters of genes scattered over several genomic locations, including two large clusters containing *pilVWXYZE* and *pilCABRS*, respectively. The first of these two clusters falls between a gene encoding an excinuclease ABC subunit B and a gene encoding a decarboxylase-family protein. We found two distinct sequence types at this locus among *Xv* isolates. The corresponding gene cluster in *Xv* 702 encodes homologues of FimT, PilV, PilW, PilX, PilY and PilE and is highly conserved (at least 99% identical nucleotide sequence) in *Xv* Isolates 890, 1326 and 1381 and *Xoo* (92% identity). However, the corresponding gene cluster is quite different in the other two sequenced *Xv* isolates, namely 206 and 895. These two isolates share 99% nucleotide sequence identity with *Xcm* NCPPB 4381 and the other sequenced isolates of *Xcm* and 90% identity with *X. vesicatoria* ATCC 35937 [18].

In summary, there are two distinct sequence types at this *pilVWXYZE* locus among *Xv* isolates: (i) *Xoo*-like (in *Xv* 890, 702, 1326 and 1381); and (ii) *X. vesicatoria*-like (*Xv* 895, 206 and all *Xcm*); see Figure 4. The most parsimonious explanation for this pattern is horizontal transfer of the locus in a common ancestor of *Xv* Isolates 890, 702, 1326 and 1381 after splitting from the *Xv* 206 lineage (see Figure 1). A less parsimonious explanation would be multiple acquisitions of the same sequence.

2.7.2. The *pilCABRS* Gene Cluster

In addition to the *pilVWXYZE* gene cluster described above, there is also variation at the *pilCABRS* gene cluster, with *pilA* (locus tag XOO1468 and accession number BAE68223 in *Xoo*) being particularly variable (Figure 5). The pattern of variation at this locus is similar to that at *pilVWXYZE*, insofar as *Xv* Isolates 890, 702, 1326 and 1381 have a *Xoo*-like sequence at this locus (94% nucleotide sequence identity between *Xv* and *Xoo*), whereas *Xv* Isolates 206 and 895 have a different sequence type.

The *pilA* genes are of different sequence types in *Xv* 206, *Xv* 895 and *Xcm*. The nucleotide sequence of *Xv* 206 *pilC* and *pilA* is 87% identical to those of *Xcm*. This degree of sequence identity is significantly lower than for the core genome; most orthologous genes share at least 99% nucleotide sequence identity between *Xv* and *Xcm*. Apart from *Xcm*, the next most similar sequence to *Xv* 206 *pilC* and *pilA* comes from *X. alfalfae* subsp. *alfalfae* (*Xaa*) CFBP 3836 (79% identity). In *Xv* 895, *pilB* is 95% identical to *Xv* 206 and 94% identical to *Xcm*. However, at the *pilA* locus, there is further variation between *Xv* 206, *Xv* 895 and *Xcm*. The most similar sequence (as of November, 2013) in the public databases to the *pilA* gene of *Xv* 895 is *X. alfalfae* subsp. *alfalfae* CFBP 3836 (90% identity).

The *pilA* of *Xcm* shares 94% nucleotide sequence identity with *X. gardneri* ATCC 19865 [18] and shows no detectable nucleotide sequence similarity with any other sequence in the public databases.

The pattern of sequence variation in the *pilCABRS* gene cluster indicates multiple superimposed horizontal transfer events, resulting in *Xvv* having three distinct sequence types of *pilA*: (i) the *Xoo*-type in *Xvv* 702, 890, 1326 and 1381; (ii) the *Xvv* 895 type that is 94% identical to *Xaa*; and (iii) the *Xvv* 206 type that is 79% identical to *Xaa*. The *pilA* of *Xcm* belongs to a fourth, *X. gardneri*-like type.

2.7.3. Concluding Remarks about Variation in T4P Genes

The T4P is a key virulence factor for phytopathogenic bacteria [19]. It performs a range of functions, including twitching motility [20–26] and cell-to-cell adhesion [27,28], thereby playing a role in the formation of micro-colonies and biofilms [29–32], and PilA has been implicated in the transmission of the pathogen to seed [31]. It is clear from our results that there have been several horizontal genetic transfers resulting in the replacement of T4P genes with alternative alleles in *X. vasicola*. It is not clear what functional significance, if any, arises from such allele exchanges, and there is no clear-cut correlation between T4P sequence type and host plant species. However, given the key role of the T4P in bacteria—plant interactions and the previously reported observation that some T4P genes are under selection in *Xanthomonas* species [33]—this may warrant further investigation. It is also possible that a phage might exert a selective pressure on T4P; for example, some phages require the T4P to infect *Pseudomonas aeruginosa* [34].

2.8. There Are Two Distinct Sequence Types of LPS Biosynthesis Clusters in *Xvv*

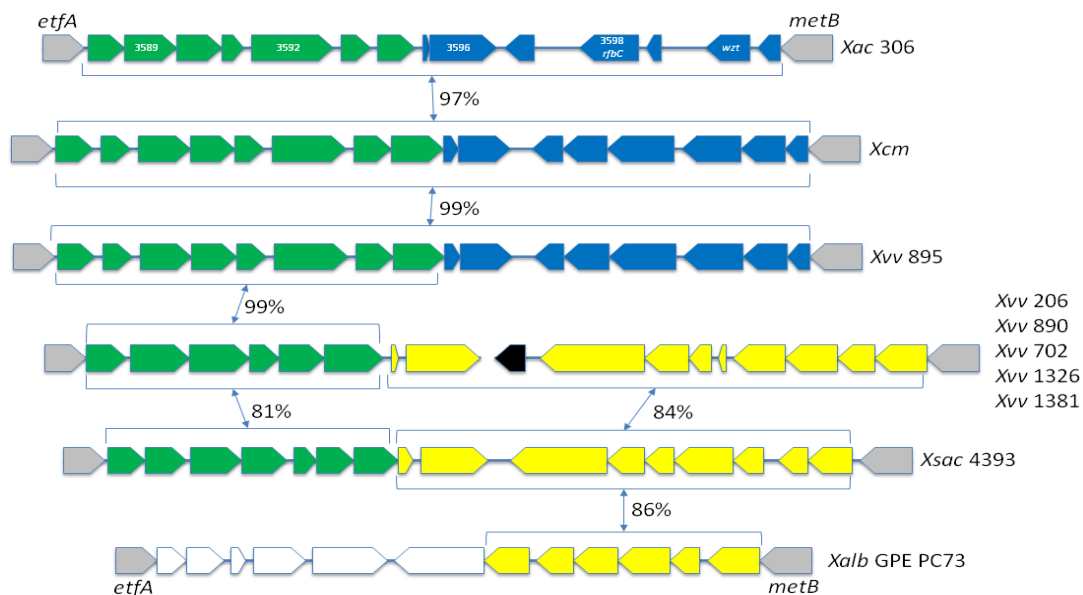
The lipopolysaccharide (LPS) molecules that cover the outer membranes of Gram-negative bacteria [35,36] can be an important virulence factor in plant pathogens [37–46]. Horizontal transfer has led to hyper-variability among different strains within individual *Xanthomonas* species [47–49].

We previously reported [8] that the LPS locus in *Xcm* 4381 most closely matches that of *X. axonopodis* pv. *citri* (*Xac*) strain 306 [50], whereas half of the LPS locus in *Xvv* 702 was not detectably similar to *Xcm* 4381, but rather resembled that of *X. albilineans* strain GPE PC73 [8,51]. We [8] and subsequently others [52] pointed out that this pattern of sequence similarity is incongruent with the close phylogenetic relationship between *Xcm* 4381 and *Xvv* 702 and indicates recent horizontal transfer in one or both strains.

Sequencing of additional isolates indicated that the LPS biosynthesis gene cluster DNA sequence is highly conserved among *Xcm* isolates [7]. However, in the present study, additional genome sequencing revealed variation in this locus among isolates of *Xvv*. As illustrated in Figure 6, *Xvv* 895 shares 99% nucleotide sequence identity with isolates of *Xcm*, which, in turn, share 97% identity with *Xac* 306. However, in *Xvv* 206, 890, 1326 and 1381, the LPS cluster shares at least 99% identity to that of *Xvv* 702. Approximately one half of the LPS cluster in these *Xvv* isolates (adjacent to *etfA*) is 99% identical to that of *Xvv* 895 and *Xcm*. However, the other half (adjacent to *metB*) shares no detectable sequence similarity with *Xvv* 895, but it does share 84% identity with *X. sacchari* [53] and 86% with *X. albilineans* [8,51].

As Pieretti and colleagues noted [52], it is interesting that the *Xv*v 702-type LPS cluster is common to three distinct *Xanthomonas* species that all inhabit the xylem of sugarcane, and thus, there might be opportunities for these species to come into contact with each other and exchange genetic material. However, it is unlikely that this type of LPS is uniquely adapted to evading recognition as a pathogen-associated molecular pattern [54,55] by sugarcane, since *Xv*v 895, isolated from sugarcane, has a *Xac*-type LPS cluster, as does *Xcm*, which can also infect sugarcane [4]. Furthermore, *Xv*v 206 has the *Xv*v 702-type LPS cluster and was originally isolated from maize; this further emphasizes that there is not a clear correlation between LPS cluster type and host plant species. Drivers for variation in the LPS might be interactions with phages [36,37,56,57] or with insect vectors [58].

Figure 6. LPS biosynthesis gene clusters found in sequenced genomes of *Xv*v and related LPS clusters from other xanthomonads. The arrows represent predicted genes and are coloured according to blocks of sequence similarity. In all xanthomonads sequenced to date, the LPS cluster lies between the highly conserved *etfA* and *metB* genes; these are indicated by grey arrows. The black arrow indicates a gene encoding a mobile element protein that is not conserved in *Xv*v NCPPB 895 nor in *Xsac* NCPPB 4393. Pairwise nucleotide sequence identities between blocks of genes are given as percentages. Sequences from each of the *Xcm* sequences share at least 99% identity with *Xcm* NCPPB 2005 over the full length of the LPS cluster. The sequences from *Xv*v NCPPB 206, 890, 1326 and 1381 all share at least 99% identity with that of *Xv*v NCPPB 702. Note that the LPS cluster falls into two distinct scaffolds in the genome assembly for *Xv*v NCPPB 702. Accession numbers: NC_003919.1 (*Xac* 306); AKBE0100022.1 (*Xcm* NCPPB 2005); AKBO01000066.1 (*Xv*v NCPPB 895); ACHS01000119.1 and ACHS01000380.1 (*Xv*v NCPPB 702); AGDB01000011.1 (*Xsac* NCPPB 4393); FP565176.1 (*Xalb* GPE PC73); AKBM01000022.1 and AKBM01000058.1 (*Xv*v 206); AKBN01000985.1, AKBN01000620.1 and AKBN01000538.1 (*Xv*v 890); AKBK01000238.1 and AKBK01000490.1 (*Xv*v 1381); and AKBL01000107.1 and AKBL01000264.1 (*Xv*v 1826).



2.9. *Xvv* Isolates Differ in Their Repertoires of T3SS Effector Genes

The repertoire of T3SS effectors can significantly influence a bacterial phytopathogen's host range [59–62]. Therefore, differences in T3SS repertoires between *Xcm* and *Xvv* are of great interest, as they might partly explain the ability of *Xcm* to cause disease in banana, whilst *Xvv* appears to be non-pathogenic in banana. We previously compared the set of T3SS effectors encoded by *Xcm* 4381 against the set encoded by *Xvv* 702 [8]. In that previous study, we identified only a few differences. Specifically, *Xcm* encodes two homologues of XopJ that are absent from the genome of *Xvv* 702, and *Xvv*702 encodes a homologue of XopAF that is absent from *Xcm* 4381 [8]. However, in the present study, utilizing additional sequence data, we found that XopAF is absent from the genome of *Xvv* 206, though it is present in *Xvv* 895, 890, 1326 and 1381, as well as in *Xvv* 702. It is absent from all sequenced isolates of *Xcm*.

2.9.1. Gene for XopAF is Absent from *Xvv* 206, Isolated from Maize

We previously reported that *Xvv* 702 encodes a homologue of XopAF (GenBank: ACHS01000051.1, bases 7184–7783; RefSeq: WP_010364039) that is not present in the genome of *Xcm* 4381 [8]. This sequence shares 86% amino acid sequence identity with the XopAF (also known as AvrXv3) from *X. euvesicatoria* that was originally identified as an avirulence factor, inducing the hypersensitive response (HR) in resistant tomato and pepper plants [63]. It is also identical at the amino acid sequence level to proteins encoded by *X. translucens* pv. *translucens* (*Xtt*) DSM 18974 (RefSeq: WP_003475568). In the present study, we found that this *xopAF* gene is present in *Xvv* 895, 890, 1326 and 1381, as well as in *Xvv* 702. It is not present in *Xcm* nor in *Xvv* 206. Thus, XopAF is encoded by *Xvv* isolates from sugarcane, but not by the *Xvv* isolate from maize and not by *Xcm* isolates from banana and enset.

This begs the question of whether XopAF1 contributes to the limitation of host range in *Xvv*; that is, one might hypothesise that XopAF confers avirulence in banana and that the absence of XopAF in *Xcm* enables its pathogenicity in banana. In a recent comparative study of the genomes of different pathotypes of *X. citri* pv. *citri* (*Xca*) [64], the authors noted that XopAF was encoded in the genomes of a narrow-host-range strain, but absent from a closely related broad-host-range strain. Therefore, they hypothesised that XopAF might confer avirulence and contribute to the limitation of the host range. However, their mutational analysis showed that *xopAF* did not affect host range, but it did contribute to the ability of *X. citri* pv. *citri* pathotype A^w (*Xcaw*) to grow in a Mexican lime host plant.

It should be noted that the predicted XopAF protein in *Xcaw* (RefSeq: WP_007652722.1) is much more divergent from the originally described sequence from *X. euvesicatoria* (WP_008577605.1), sharing only 31% amino acid sequence identity, whereas the *Xvv* protein shares 86% identity with *X. euvesicatoria* XopAF. Therefore, the *Xvv* XopAF protein is likely to interact with plants differently. Furthermore, the host plants in question are very different with *X. euvesicatoria* and *Xcaw* infecting dicots and *Xvv* and *Xcm* infecting monocots. However, it is reasonable to suppose that the *Xvv* protein is likely to be a T3SS effector and a potential avirulence factor, given the 86% identity between it and the experimentally characterised *X. euvesicatoria* XopAF [63]. XopAF contains a DNA-binding domain at its C terminus and may allow the pathogen to manipulate its host by affecting the expression of plant genes. It would be interesting to test whether heterologous expression of *xopAF* in *Xcm* would

cause avirulence in banana and whether deletion of *xopAF* in *Xvv* would have any impact on virulence in sugarcane or maize.

In *Xvv* 702, the *xopAF* gene falls between Positions 7181 and 7543, on the reverse strand in GenBank accession ACHS01000051.1. This resides within a genomic region that also encodes several phage-associated proteins, including phage-related lytic enzyme, phage-tail protein, baseplate assembly protein J, phage-tail fibres and phage-tail fibre protein, and may result from the integration of a pro-phage into the *Xvv* genome. In the wheat pathogen, *Xtt* DSM 18974, the *xopAF* gene encoding an identical protein sequence is located in a different genomic context; it resides at Positions 40277 to 40933 in GenBank accession CAPJ01000122.1 (locus tag: BN444_00905). This region of the *Xtt* genome does not contain any obvious phage-related genes, but does contain a predicted transposase for insertion sequence element IS629 (locus tag: BN444_00906), suggesting a mechanism for the mobility of this gene.

2.9.2. Gene Encoding Homologue of XopL Underwent Truncation in Common Ancestor of *Xvv* 890, 702, 1326 and 1381

We also found polymorphism with respect to a *xopL*-like gene in *Xvv*. The genomes of *Xvv* isolates 206 and 895 each encode a protein with 70% amino acid sequence identity to XopL from *X. campestris* pv. *vesicatoria* (*Xcv*) strain 85-10 (RefSeq: YP_364951.1) [65]. The GenBank accession numbers for the *Xvv* 895 and 206 sequences are AKBO01000570.1 (Positions 1804–3768) and AKBM01000211.1 (Positions 29591–31555), where they encode a full-length protein. However, in *Xvv* Isolates 890, 702, 1326 and 1381, there is a single-nucleotide C→A substitution resulting in a TCA codon being transformed to a TAA stop codon. This substitution occurs at Position 14191 in GenBank accession ACHS01000315.1 and is predicted to result in the protein being truncated to 265 amino acids (compared to the full-length 654 amino acids).

This *xopL*-like sequence is completely absent from *Xcm* isolates belonging to Sub-lineage II (*Xcm* 4379, 4380, 4381, 4383, 4384, 4392 and 4395). However, it is present in *Xcm* isolates belonging to Sub-lineage I (*Xcm* 2005, 2251, 4387 and 4389), where each genome encodes a full-length protein. Thus, it appears that this *xopL* homologue may have been lost twice: (i) completely deleted in a common ancestor of *Xcm* Sub-lineage II; and (ii) truncated by a premature stop codon in a common ancestor of *Xvv* 890, 702, 1326 and 1381.

It has recently been demonstrated that XopL possesses E3 ubiquitin ligase activity, induces plant cell death and subverts plant immunity and that the ligase activity is associated with the C-terminal region of the protein [66]. The premature stop codon found in *xopL* of some *Xvv* isolates has split the XopL-encoding open reading frame (ORF) into two.

Interestingly, there is a candidate plant inducible promoter (PIP) box [67–70] upstream of the second ORF, which corresponds to the C-terminal region of XopL, in which the E3 ubiquitin ligase resides. This suggests the hypothesis that this truncated ORF still has the potential to be expressed and be induced *in planta* and that it might still have biochemical activity, though it is unclear whether it would be a substrate of the T3SS. This potential PIP box (sequence TTCCGgccaacatgcagcaaTTCGC) is located at Positions 14107 to 14137 in ACHS01000315.1, which is approximately 160 bp upstream of the C-terminal XopL ORF at 14296 to 15366. There is another PIP box (sequence

TTCGCtacgataaagatgacTTCGC) located at 13300 to 13347, which is approximately 50 bp upstream of the ORF homologous to the XopL N terminus located at 13395 to 14192. The complete set of predicted PIP boxes in *Xvv* 702 and *Xcm* 4381 is tabulated in the Supplementary Material.

2.9.3. Absence of Genes Encoding Homologues of XopJ Distinguishes *Xcm* from *Xvv*

In addition to the homologues of XopAF and XopL, a further two potential T3SS show differential presence/absence among our sequenced strains. We previously reported that *Xcm* 4381 encodes two homologues of XopJ that are absent from *Xvv* 702 [8]. Our subsequent analyses have confirmed that these are conserved in all the sequenced *Xcm* isolates and are absent from all the sequenced *Xvv*. Therefore, these predicted T3SS effectors remain as candidates for contributing to the differences in host range between *Xvv* and *Xcm*.

3. Experimental Section

3.1. Sources of Bacterial Strains

Bacterial strains were obtained from the National Collection of Plant Pathogenic Bacteria (NCPBP) at The Food and Environment Research Agency, UK (Fera). DNA library preparation and genome sequencing using the Illumina GA2x were performed using standard Illumina protocols, as previously described [7,8].

3.2. Preparation of Genomic DNA

For DNA preparation, bacterial strains were grown overnight at 28 °C in 10 mL King Broth shaken at 200 rpm. Bacterial cells were harvested by centrifugation and re-suspended in TE buffer (50 mM Tris-HCl, 40 mM EDTA, pH 8.0) containing 12 µL of 20 mg/mL lysozyme and 10 mg/mL RNase and incubated at 25 °C for 10 min with 17 µL 10% sodium dodecyl sulphate, then incubated on ice for 5 min. Proteins were dissolved with 170 µL of 8 M ammonium acetate, vortexed vigorously for 30 s centrifuged at 4 °C and for 15 min. DNA was precipitated with isopropanol and re-dissolved in 100 µL of 10 mM Tris, pH 8.0, and 1 mM Na₂EDTA.

3.3. Genome Sequencing

We used the Illumina GA2x platform to sequence genomes of *Xvv* strains NCPBP 895 and 890, generating paired sequence reads of length 67 nucleotides, according to the manufacturer's instructions.

3.4. Alignment of Sequence Reads Against Reference Genome Sequences

We used BWA [14] to align GA2x sequence reads against a reference genome sequence and used IGV [71] to visualize the alignments and SAMtools [72] to manipulate the alignments and convert between formats.

3.5. SNP Calling and Phylogenetic Analysis

We used a very conservative approach to infer SNPs from the alignments of Illumina reads against the previously published *Xoo* reference draft genome assembly. To avoid false positives and false negatives, we only used those regions of the *Xoo* genome with a coverage depth of 10 or more for every sequenced *Xcm* and *Xvv* genome and where there was at least 95% consensus among the sequence reads within each isolate. Just over 30% of the length (1,507,606 out of 4,940,217 nt) of the *Xoo* genome fulfilled these two criteria. In other words, for 30% of the *Xoo* chromosome, there was sufficient quantity and consistency in our data to be almost certain of the sequence in all of the eight isolates (six *Xvv* and two *Xcm*; see Figure 1); for the remaining 70% of the genome, there was some degree of ambiguity in the data for one or more of the isolates. The phylogenetic tree was inferred using the Maximum Parsimony method implemented in MEGA5 [73] based on 39,665 single-nucleotide variants with respect to the chromosome of *Xoo* MAFF 311018. Bootstrap values were calculated as percentages of 500 trials.

3.6. Genome Assembly

De novo assembly of Illumina sequence reads was performed using Velvet 1.1.04 [74]. We discarded any sequence reads that contained one or more “N” prior to assembly. It is difficult or impossible to predict the optimal parameter values for Velvet assembly. Therefore, we generated assemblies using a range of combinations of hash length and coverage cut-off and chose the assemblies giving the largest N_{50} values. For the newly presented assemblies (*i.e.*, for *Xvv* 890 and 895), the parameter values were for *Xvv* 895: hash length = 25 and coverage cut-off = 2 and for *Xvv* 890: hash length = 29 and coverage cut-off = 3.

3.7. Identification of Presence and Absence of Genes

We used BEDtools [15] to infer the breadths of coverage for genomic features based on Binary Alignment Map (BAM) files from the Burrows-Wheeler Aligner (BWA) and General Feature Format (GFF) files from Rapid Annotation using Subsystem Technology (RAST) [75]. We used the pheatmap package in R to generate heatmaps [76]. However, it should be noted that incompleteness of *de novo* assemblies do not invalidate these, since our inferences are based upon comparisons of alignments of raw reads rather than comparisons between assemblies; these alignments consist of unassembled read-pairs aligned with BWA against various reference genome sequences, and several examples of such alignments are illustrated in the Supplementary Material. The single exception is Figure 2, in which the comparison consists of alignments between *de novo* assemblies; it is possible that some gaps in the alignments in Figure 2 could arise through the incompleteness of the *de novo* assemblies. Figures 4 and 5 are based on BWA alignments between *de novo* assemblies, but the findings were also validated by the inspection of alignments of raw (unassembled) sequence reads (see the Supplementary Material).

3.8. Visualisation of Genome-Wide Patterns of Sequence Conservation

We used BLASTN [77] to align assembled sequences and visualized the alignments using the Artemis Comparison Tool [78] and BLAST Ring Image Generator (BRIG) [79], which is a wrapper for Circular Genome Viewer (CGView) [80].

3.9. Identification of Potential PIP Boxes

We built a profile hidden Markov model (HMM) based on a multiple sequence alignment of 22 known PIP boxes from *X. vesicatoria* (from Table 3 in [69]) using hmmb from the HMMER 1.8.5 package [81]. The DNA sequence was scanned against this profile-HMM using hmmls from HMMER 1.8.5 with a bit-score cut-off of 10.0.

4. Conclusions

Here, we analyse draft genome sequences for two isolates of *Xvv* to augment the four previously published [7,8] draft genome sequences of *Xvv*. Comparative analyses of these genome sequences and previously published genome sequences of the closely related pathovar, *Xcm*, have revealed extensive differences in gene content among *Xvv*. This manuscript describes some of these differences in detail, including differences in plasmid content, LPS biosynthesis clusters, T4P and T3SS effectors. The main evolutionary events are summarized graphically in Figure 1. As well as providing some insight into evolutionary events within *Xvv*, these sequence analyses also further refine our understanding of the genomic differences between *Xvv* and the very closely related *Xcm*, which is a recently emerging pathogen in banana and enset; the availability of the sequence from multiple isolates allows us to distinguish between inherent variation within *Xvv* that might confound attempts to identify important genetic differences between the two pathovars and for functional analysis of important virulence factors.

It is clear that *Xcm* is genetically highly monomorphic [5–7]; here, we show that, apart from several phage-related genes and the SNPs in the core genome described previously [7], the few genetic differences among *Xcm* isolates can be explained by the acquisition of a plasmid in *Xcm* Sub-lineage II, which is not present in Sub-lineage I, nor in at least one isolate of Sub-lineage II (*Xcm* 4381). Additionally, the two *Xcm* sub-lineages differ in that members of Sub-lineage II have lost a gene encoding a homologue of XopL; interestingly, this gene has acquired a premature stop codon in four of the six *Xvv* isolates, suggesting that isolates of both *Xcm* and *Xvv* have independently converged on eliminating XopL.

In contrast to the limited genetic diversity within *Xcm*, there is considerable diversity within *Xvv*, both at the level of SNPs in the core genome and at the level of gene content. Some of the differences in gene content are ascribable to the acquisition of two different plasmids (one in *Xvv* 895 and one in *Xvv* 890, 702, 1326 and 1381), but there are also differences in chromosomally located gene clusters, such as those encoding LPS biosynthesis and T4P.

Overall, this work suggests hypotheses for future work towards understanding the molecular basis for the ability of *Xcm* to emerge as an important pathogen of banana and enset. For example, one consistent difference is that all sequenced *Xcm* isolates encode two homologues of XopJ that are absent from all sequenced isolates of *Xvv*. In *X. campestris* pv. *vesicatoria*, this T3SS effector has been shown

to interfere with salicylic acid-dependent defence responses to attenuate the onset of necrosis and to alter host transcription [82]; it will be enlightening to test the contribution of the two XopJ homologues in *Xcm* interaction with banana and enset. Furthermore, understanding the emergence of *Xcm* will require the study of genome sequences of a wider range of strains within the species, *X. vasicola*, to which *Xcm* probably belongs [4]; several isolates are available in strain collections for *X. vasicola* pv. *holcicola* [83], but there is also a need to survey other as yet unknown members of the species that might inhabit the centre of origin, perhaps colonizing other monocot plants.

Acknowledgments

This study was supported in part by the National Agriculture Research Organisation, Uganda, under the MSI/World Bank grant 2009. The authors wish to thank Karen Moore and Alex Moorhouse for their invaluable technical assistance with sequencing and Marta De Torres Zabala for expert guidance to Arthur Wasukira and Max Coulter in the laboratory.

Author Contributions

Conceived the study: David J. Studholme, Murray Grant, Julian Smith, Arthur Wasukira and Jerome Kubiriba. Wrote the manuscript and prepared figures: David J. Studholme, Julian Smith, Murray Grant and Arthur Wasukira. Performed laboratory-based work: Arthur Wasukira, Max Coulter and Richard Thwaites. Supervised DNA sequencing: Konrad Paszkiewicz. Bioinformatics analysis: David J. Studholme and Noorah Al-Sowayeh.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Hayward, A.C. The hosts of *Xanthomonas*. In *Xanthomonas*; Springer: Dordrecht, The Netherlands, 1993; pp. 1–119.
2. Ryan, R.P.; Vorhölter, F.-J.; Potnis, N.; Jones, J.B.; van Sluys, M.-A.; Bogdanove, A.J.; Dow, J.M. Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* **2011**, *9*, 344–355.
3. Dookun, A.; Stead, D.E.; Autrey, L.J. Variation among strains of *Xanthomonas campestris* pv. *vasculorum* from Mauritius and other countries based on fatty acid analysis. *Syst. Appl. Microbiol.* **2000**, *23*, 148–155.
4. Aritua, V.; Parkinson, N.; Thwaites, R.; Heeney, J.V.; Jones, D.R.; Tushemereirwe, W.; Crozier, J.; Reeder, R.; Stead, D.E.; Smith, J. Characterization of the *Xanthomonas* sp causing wilt of enset and banana and its proposed reclassification as a strain of *X. vasicola*. *Plant Pathol.* **2008**, *57*, 170–177.
5. Odipio, J.; Tusiime, G.; Tripathi, L. Genetic homogeneity among Ugandan isolates of *Xanthomonas campestris* pv. *musacearum* revealed by randomly amplified polymorphic DNA analysis. *Afr. J. Biotechnol.* **2009**, doi: 10.4314/ajb.v8i21.66028.

6. Aritua, V.; Nanyonjo, A.; Kumakech, F.; Tushemereirwe, W. Rep-PCR reveals a high genetic homogeneity among Ugandan isolates of *Xanthomonas campestris* pv *musacearum*. *Afr J Biotechnol.* **2007**, *6*, 179–183.
7. Wasukira, A.; Tayebwa, J.; Thwaites, R.; Paszkiewicz, K.; Aritua, V.; Kubiriba, J.; Smith, J.; Grant, M.; Studholme, D.J. Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pathovar *musacearum*. *Genes.* **2012**, *3*, 361–377.
8. Studholme, D.J.; Kemen, E.; MacLean, D.; Schornack, S.; Aritua, V.; Thwaites, R.; Grant, M.; Smith, J.; Jones, J.D.G. Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas* wilt. *FEMS Microbiol. Lett.* **2010**, *310*, 182–192.
9. Ivey, M.L.L.; Tusiime, G.; Miller, S.A. A Polymerase Chain Reaction Assay for the Detection of *Xanthomonas campestris* pv. *musacearum* in Banana. *Plant Dis.* **2010**, *94*, 109–114.
10. Qhobela, M.; Claflin, L.E. Eastern and southern African strains of *Xanthomonas campestris* pv. *vasculorum* are distinguishable by restriction fragment length polymorphism of DNA and polyacrylamide gel electrophoresis of membrane proteins. *Plant Pathol.* **1992**, *41*, 113–121.
11. Vauterin, L.; Hoste, B.; Kersters, K.; Swings, J. Reclassification of *Xanthomonas*. *Int. J. Syst. Bacteriol.* **1995**, 472–489.
12. Rademaker, J.L.W.; Louws, F.J.; Schultz, M.H.; Rossbach, U.; Vauterin, L.; Swings, J.; de Bruijn, F.J. A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology* **2005**, *95*, 1098–1111.
13. Salzberg, S.L.; Sommer, D.D.; Schatz, M.C.; Phillippy, A.M.; Rabinowicz, P.D.; Tsuge, S.; Furutani, A.; Ochiai, H.; Delcher, A.L.; Kelley, D.; *et al.* Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genom.* **2008**, *9*, 204.
14. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760.
15. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842.
16. Bart, R.; Cohn, M.; Kassen, A.; McCallum, E.J.; Shybut, M.; Petriello, A.; Krasileva, K.; Dahlbeck, D.; Medina, C.; Alicai, T.; *et al.* High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. USA.* **2012**, *109*, E1972–E1979.
17. Cunnac, S.; Bolot, S.; Serna, N.F.; Ortiz, E.; Szurek, B.; Noël, L.D.; Arlat, M.; Jacques, M.-A.; Gagnevin, L.; Carrere, S.; *et al.* High-quality draft genome sequences of two *Xanthomonas citri* pv. *malvacearum* strains. *Genome Announc.* **2013**, doi:10.1128/genomeA.00674-13.
18. Potnis, N.; Krasileva, K.; Chow, V.; Almeida, N.F.; Patil, P.B.; Ryan, R.P.; Sharlach, M.; Behlau, F.; Dow, J.M.; Momol, M.T.; *et al.* Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper. *BMC Genom.* **2011**, *12*, 146.
19. Taguchi, F.; Ichinose, Y. Role of type IV pili in virulence of *Pseudomonas syringae* pv. *tabaci* 6605: Correlation of motility, multidrug resistance, and HR-inducing activity on a nonhost plant. *Mol. Plant. Microbe. Interact.* **2011**, *24*, 1001–1011.
20. Liu, H.; Kang, Y.; Genin, S.; Schell, M.A.; Denny, T.P. Twitching motility of *Ralstonia solanacearum* requires a type IV pilus system. *Microbiology* **2001**, *147*, 3215–3229.

21. Mattick, J.S. Type IV pili and twitching motility. *Annu. Rev. Microbiol.* **2002**, *56*, 289–314.
22. De La Fuente, L.; Burr, T.J.; Hoch, H.C. Mutations in type I and type IV pilus biosynthetic genes affect twitching motility rates in *Xylella fastidiosa*. *J. Bacteriol.* **2007**, *189*, 7507–7510.
23. Li, Y.; Hao, G.; Galvani, C.D.; Meng, Y.; De La Fuente, L.; Hoch, H.C.; Burr, T.J. Type I and type IV pili of *Xylella fastidiosa* affect twitching motility, biofilm formation and cell-cell aggregation. *Microbiology* **2007**, *153*, 719–726.
24. Li, Y.-Q.; Wan, D.-S.; Huang, S.-S.; Leng, F.-F.; Yan, L.; Ni, Y.-Q.; Li, H.-Y. Type IV pili of *Acidithiobacillus ferrooxidans* are necessary for sliding, twitching motility, and adherence. *Curr. Microbiol.* **2010**, *60*, 17–24.
25. Pelicic, V. Type IV pili: E pluribus unum? *Mol. Microbiol.* **2008**, *68*, 827–837.
26. Burdman, S.; Bahar, O.; Parker, J.K.; De La Fuente, L. Involvement of type IV Pili in pathogenicity of plant pathogenic bacteria. *Genes*. **2011**, *2*, 706–735.
27. Jenkins, A.T.A.; Buckling, A.; McGhee, M.; French-Constant, R.H. Surface plasmon resonance shows that type IV pili are important in surface attachment by *Pseudomonas aeruginosa*. *J. R. Soc. Interface* **2005**, *2*, 255–259.
28. Heijstra, B.D.; Pichler, F.B.; Liang, Q.; Blaza, R.G.; Turner, S.J. Extracellular DNA and Type IV pili mediate surface attachment by *Acidovorax temperans*. *Antonie Van Leeuwenhoek* **2009**, *95*, 343–349.
29. Roine, E.; Raineri, D.M.; Romantschuk, M.; Wilson, M.; Nunn, D.N. Characterization of type IV pilus genes in *Pseudomonas syringae* pv. *tomato* DC3000. *Mol. Plant. Microbe. Interact.* **1998**, *11*, 1048–1056.
30. Shime-Hattori, A.; Iida, T.; Arita, M.; Park, K.-S.; Kodama, T.; Honda, T. Two type IV pili of *Vibrio parahaemolyticus* play different roles in biofilm formation. *FEMS Microbiol. Lett.* **2006**, *264*, 89–97.
31. Darsonval, A.; Darrasse, A.; Meyer, D.; Demarty, M.; Durand, K.; Bureau, C.; Manceau, C.; Jacques, M. The Type III secretion system of *Xanthomonas fuscans* subsp. *fuscans* is involved in the phyllosphere colonization process and in transmission to seeds of susceptible beans. *Appl. Environ. Microbiol.* **2008**, *74*, 2669–2678.
32. Varga, J.J.; Therit, B.; Melville, S.B. Type IV pili and the CcpA protein are needed for maximal biofilm formation by the gram-positive anaerobic pathogen *Clostridium perfringens*. *Infect. Immun.* **2008**, *76*, 4944–4951.
33. Mhedbi-Hajri, N.; Darrasse, A.; Pigné, S.; Durand, K.; Fouteau, S.; Barbe, V.; Manceau, C.; Lemaire, C.; Jacques, M.-A. Sensing and adhesion are adaptive functions in the plant pathogenic xanthomonads. *BMC Evol. Biol.* **2011**, *11*, 67.
34. Heo, Y.-J.; Chung, I.-Y.; Choi, K.B.; Lau, G.W.; Cho, Y.-H. Genome sequence comparison and superinfection between two related *Pseudomonas aeruginosa* phages, D3112 and MP22. *Microbiology* **2007**, *153*, 2885–2895.
35. Lerouge, I.; Vanderleyden, J. O-antigen structural variation: Mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol. Rev.* **2002**, *26*, 17–47.
36. Raetz, C.R.H.; Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **2002**, *71*, 635–700.
37. Hendrick, C.A.; Sequeira, L. Lipopolysaccharide-DEFECTIVE Mutants of the Wilt Pathogen *Pseudomonas solanacearum*. *Appl. Environ. Microbiol.* **1984**, *48*, 94–101.

38. Drigues, P.; Demery-Lafforgue, D.; Trigalet, A.; Dupin, P.; Samain, D.; Asselineau, J. Comparative studies of lipopolysaccharide and exopolysaccharide from a virulent strain of *Pseudomonas solanacearum* and from three avirulent mutants. *J. Bacteriol.* **1985**, *162*, 504–509.
39. Jayaswal, R.K.; Bressan, R.A.; Handa, A.K. Effects of a mutation that eliminates UDP glucose-pyrophosphorylase on the pathogenicity of *Erwinia carotovora* subsp. *carotovora*. *J. Bacteriol.* **1985**, *164*, 473–476.
40. Schoonejans, E.; Expert, D.; Toussaint, A. Characterization and virulence properties of *Erwinia chrysanthemi* lipopolysaccharide-defective, phi EC2-resistant mutants. *J. Bacteriol.* **1987**, *169*, 4011–4017.
41. Kingsley, M.T.; Gabriel, D.W.; Marlow, G.C.; Roberts, P.D. The opsX locus of *Xanthomonas campestris* affects host range and biosynthesis of lipopolysaccharide and extracellular polysaccharide. *J. Bacteriol.* **1993**, *175*, 5839–5850.
42. Dow, J.M.; Osbourn, A.E.; Wilson, T.J.; Daniels, M.J. A locus determining pathogenicity of *Xanthomonas campestris* is involved in lipopolysaccharide biosynthesis. *Mol. Plant. Microbe. Interact.* **1995**, *8*, 768–777.
43. Titarenko, E.; López-Solanilla, E.; García-Olmedo, F.; Rodríguez-Palenzuela, P. Mutants of *Ralstonia (Pseudomonas) solanacearum* sensitive to antimicrobial peptides are altered in their lipopolysaccharide structure and are avirulent in tobacco. *J. Bacteriol.* **1997**, *179*, 6699–6704.
44. Li, J.; Wang, N. The gpsX gene encoding a glycosyltransferase is important for polysaccharide production and required for full virulence in *Xanthomonas citri* subsp. *citri*. *BMC Microbiol.* **2012**, *12*, 31.
45. Yan, Q.; Hu, X.; Wang, N. The novel virulence-related gene nlxA in the lipopolysaccharide cluster of *Xanthomonas citri* ssp. *citri* is involved in the production of lipopolysaccharide and extracellular polysaccharide, motility, biofilm formation and stress resistance. *Mol. Plant Pathol.* **2012**, *13*, 923–934.
46. Li, J.; Wang, N. The wxacO gene of *Xanthomonas citri* ssp. *citri* encodes a protein with a role in lipopolysaccharide biosynthesis, biofilm formation, stress tolerance and virulence. *Mol. Plant Pathol.* **2011**, *12*, 381–396.
47. Patil, P.B.; Bogdanove, A.J.; Sonti, R.V. The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in xanthomonads that infect rice, citrus and crucifers. *BMC Evol Biol* **2007**, *7*, 243.
48. Patil, P.; Sonti, R. Variation suggestive of horizontal gene transfer at a lipopolysaccharide (lps) biosynthetic locus in *Xanthomonas oryzae* pv. *oryzae*, the bacterial leaf blight pathogen of rice. *BMC Microbiol.* **2004**, *4*, 40.
49. Lu, H.; Patil, P.; van Sluys, M.-A.; White, F.F.; Ryan, R.P.; Dow, J.M.; Rabinowicz, P.; Salzberg, S.L.; Leach, J.E.; Sonti, R.; *et al.* Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PLoS One* **2008**, *3*, e3828.
50. Da Silva, A.C.R.; Ferro, J.A.; Reinach, F.C.; Farah, C.S.; Furlan, L.R.; Quaggio, R.B.; Monteiro-Vitorello, C.B.; van Sluys, M.A.; Almeida, N.F.; Alves, L.M.C.; *et al.* Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **2002**, *417*, 459–463.

51. Pieretti, I.; Royer, M.; Barbe, V.; Carrere, S.; Koebnik, R.; Cociancich, S.; Couloux, A.; Darrasse, A.; Gouzy, J.; Jacques, M.-A.; 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.
52. 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.
53. Studholme, D.J.; Wasukira, A.; Paszkiewicz, K.; Aritua, V.; Thwaites, R.; Smith, J.; Grant, M. Correction: Studholme *et al.* Draft Genome Sequences of *Xanthomonas sacchari* and Two Banana-Associated Xanthomonads Reveal Insights into the *Xanthomonas* Group 1 clade. *Genes* **2011**, *2*, 1050–1065.
54. Dow, M.; Newman, M.-A.; von Roepenack, E. The induction and modulation of plant defense responses by bacterial lipopolysaccharides. *Annu. Rev. Phytopathol.* **2000**, *38*, 241–261.
55. Meyer, A.; Pühler, A.; Niehaus, K. The lipopolysaccharides of the phytopathogen *Xanthomonas campestris* pv. *campestris* induce an oxidative burst reaction in cell cultures of *Nicotiana tabacum*. *Planta* **2001**, *213*, 214–222.
56. Keshavarzi, M.; Soylu, S.; Brown, I.; Bonas, U.; Nicole, M.; Rossiter, J.; Mansfield, J. Basal defenses induced in pepper by lipopolysaccharides are suppressed by *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Plant. Microbe. Interact.* **2004**, *17*, 805–815.
57. Yang, Y.-C.; Chou, C.-P.; Kuo, T.-T.; Lin, S.-H.; Yang, M.-K. PilR enhances the sensitivity of *Xanthomonas axonopodis* pv. *citri* to the infection of filamentous bacteriophage Cf. *Curr. Microbiol.* **2004**, *48*, 251–261.
58. Pal, S.; Wu, L.P. Pattern recognition receptors in the fly: Lessons we can learn from the *Drosophila melanogaster* immune system. *Fly* **2009**, *3*, 121–129.
59. Sarkar, S.F.; Gordon, J.S.; Martin, G.B.; Guttman, D.S. Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* **2006**, *174*, 1041–1056.
60. Hajri, A.; Brin, C.; Hunault, G.; Lardeux, F.; Lemaire, C.; Manceau, C.; Boureau, T.; Poussier, S. A “repertoire for repertoire” hypothesis: Repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One* **2009**, *4*, e6632.
61. Hajri, A.; Pothier, J.F.; Fischer-Le Saux, M.; Bonneau, S.; Poussier, S.; Boureau, T.; Duffy, B.; Manceau, C. Type three effector genes distribution and sequence analysis provides new insights into pathogenicity of plant pathogenic *Xanthomonas arboricola*. *Appl. Environ. Microbiol.* **2012**, *78*, 371–384.
62. Hajri, A.; Brin, C.; Zhao, S.; David, P.; Feng, J.-X.; Koebnik, R.; Szurek, B.; Verdier, V.; Boureau, T.; Poussier, S. Multilocus sequence analysis and type III effector repertoire mining provide new insights into the evolutionary history and virulence of *Xanthomonas oryzae*. *Mol. Plant Pathol.* **2012**, *13*, 288–302.
63. Astua-Monge, G.; Minsavage, G.V.; Stall, R.E.; Davis, M.J.; Bonas, U.; Jones, J.B. Resistance of tomato and pepper to T3 strains of *Xanthomonas campestris* pv. *vesicatoria* is specified by a plant-inducible avirulence gene. *Mol. Plant. Microbe. Interact.* **2000**, *13*, 911–921.

64. Jalan, N.; Kumar, D.; Andrade, M.O.; Yu, F.; Jones, J.B.; Graham, J.H.; White, F.F.; Setubal, J.C.; Wang, N. Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. *BMC Genom.* **2013**, *14*, 551.
65. Thieme, F.; Koebnik, R.; Bekel, T.; Berger, C.; Boch, J.; Büttner, D.; Caldana, C.; Gaigalat, L.; Goesmann, A.; Kay, S.; *et al.* Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* **2005**, *187*, 7254–7266.
66. Singer, A.U.; Schulze, S.; Skarina, T.; Xu, X.; Cui, H.; Eschen-Lippold, L.; Egler, M.; Srikumar, T.; Raught, B.; Lee, J.; Scheel, D.; Savchenko, A.; Bonas, U. A pathogen type III effector with a novel E3 ubiquitin ligase architecture. *PLoS Pathog.* **2013**, *9*, e1003121.
67. Jiang, W.; Jiang, B.-L.; Xu, R.-Q.; Huang, J.-D.; Wei, H.-Y.; Jiang, G.-F.; Cen, W.-J.; Liu, J.; Ge, Y.-Y.; Li, G.-H.; *et al.* Identification of six type III effector genes with the PIP box in *Xanthomonas campestris* pv. *campestris* and five of them contribute individually to full pathogenicity. *Mol. Plant. Microbe. Interact.* **2009**, *22*, 1401–1411.
68. Büttner, D.; Bonas, U. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* **2010**, *34*, 107–133.
69. Koebnik, R.; Krüger, A.; Thieme, F.; Urban, A.; Bonas, U. Specific binding of the *Xanthomonas campestris* pv. *vesicatoria* AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J. Bacteriol.* **2006**, *188*, 7652–7660.
70. Wengelnik, K.; Bonas, U. HrpXv, an AraC-type regulator, activates expression of five of the six loci in the hrp cluster of *Xanthomonas campestris* pv. *vesicatoria*. *J. Bacteriol.* **1996**, *178*, 3462–3469.
71. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26.
72. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079.
73. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739.
74. Zerbino, D.R.; Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829.
75. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75.
76. R Development Core Team. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* **2013**, *1*, 409.
77. 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.
78. Carver, T.J.; Rutherford, K.M.; Berriman, M.; Rajandream, M.-A.; Barrell, B.G.; Parkhill, J. ACT: The artemis comparison tool. *Bioinformatics* **2005**, *21*, 3422–3423.

79. Alikhan, N.-F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **2011**, *12*, 402.
80. Stothard, P.; Wishart, D.S. Circular genome visualization and exploration using CGView. *Bioinformatics* **2005**, *21*, 537–539.
81. Eddy, S.R. Profile Hidden Markov Models. *Bioinformatics* **1998**, *14*, 755–763.
82. Üstün, S.; Bartetzko, V.; Börnke, F. The *Xanthomonas campestris* Type III Effector XopJ Targets the Host Cell Proteasome to Suppress Salicylic-Acid Mediated Plant Defence. *PLoS Pathog.* **2013**, *9*, e1003427.
83. Qhobela, M.; Leach, J.E. Characterisation of strains of *Xanthomonas campestris* pv. *holcicola* by PAGE of membrane proteins and by REA and RFLP analysis of genomic DNA. *Plant Dis.* **1991**, *75*, 32–36.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).