CRISPR-Cas systems in the plant pathogen *Xanthomonas* spp. and their impact on genome plasticity

Paula Maria Moreira Martins^{a*}; Andre da Silva Xavier^{c*}; Marco Aurelio Takita^a
Poliane Alfemas-Zerbini^b; Alessandra Alves de Souza^{a#}.

*These authors contributed equally to this work

^aCitrus Biotechnology Lab, Centro de Citricultura Sylvio Moreira, Instituto Agronômico de Campinas, Cordeirópolis-SP, Brazil

^bDepartament of Microbiology, Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa-MG, Brazil

^cDepartament of Agronomy/NUDEMAFI, Universidade Federal do Espírito Santo, Brazil.

Key words: Phage, plasmids, *Xanthomonadaceae*, *Xylella*.

Running title: CRISPR-Cas systems in *Xanthomonas* spp.

Abstract

Xanthomonas is one of the most important bacterial genera of plant pathogens causing economic losses in crop production worldwide. Despite its importance, many aspects of basic *Xanthomonas* biology remain unknown or understudied. Here, we present the first genus-wide analysis of CRISPR-Cas in *Xanthomonas* and describe specific aspects of its occurrence. Our results show that *Xanthomonas* genomes harbour subtype I-C and I-F CRISPR-Cas systems and that species belonging to distantly *Xanthomonas*-related genera in *Xanthomonadaceae* exhibit the same configuration of

coexistence of the I-C and I-F CRISPR subtypes. Additionally, phylogenetic analysis using Cas proteins indicated that the CRISPR systems present in *Xanthomonas* spp. are the result of an ancient acquisition. Despite the close phylogeny of these systems, they present significant variation in both the number and targets of spacers. An interesting characteristic observed in this study was that the identified plasmid-targeting spacers were always driven toward plasmids found in other *Xanthomonas* strains, indicating that CRISPR-Cas systems could be very effective in coping with plasmidial infections. Since many effectors are plasmid encoded, CRISPR-Cas might be driving specific characteristics of plant-pathogen interactions.

Introduction

Phytopathogenic bacteria are a global threat to crop production worldwide. *Xanthomonas* spp. is one of the most important genera of phytopathogens since these species can infect at least 120 monocotyledonous and 260 dicotyledonous species of economic importance (1,2). These pathogens are able to live both inside and outside of plant hosts. Regardless of their lifestyle, bacteria are constantly exposed to many different threats, such as the constant pressure in the form of exogenous DNA invasions from both viruses and invading plasmids from other bacteria (3,4). Many basic aspects of how these phytopathogens react and protect themselves from such threats remain understudied.

Bacteriophages (or simply “phages”) are one of the most abundant entities across the biosphere and one of the most potent pathogens of bacteria (5). Many aspects of both bacterial and phage genomes have been shaped by this never-ending war, in which both groups have had to develop defence and attack systems (6,7). In addition to

virus attack, plasmid invasions can also be deleterious to bacteria. The most urgent topic concerning the negative effects of plasmidial invasions can be linked to the so-called “metabolic burden” (4,8), consisting of physiological disturbance due to the presence of exotic genetic material and its associated metabolism that drains important energetic resources of the host cell, negatively impacting its fitness.

For every horizontal genetic transfer that takes place in a prokaryotic cell, specific intra-cellular protection systems may come into play. Despite the fact that genomic rearrangements can lead to positive outcomes, there must be a balance between stability and tolerance of these events (3). Many biological systems have evolved to protect the integrity of the genetic information of prokaryotes. One of the first types of system ever discovered that eradicates exogenous DNA infections at their onset was restriction-modification systems, which recognize self-DNA by its methylation pattern and enzymatically destroy the invader DNA, thereby “restricting” its occurrence (9). Other mechanisms include the extreme abortive infection system, which kills the infected cell, preventing the phage from spreading throughout the bacterial population (10). Curiously, systems that were designed to aid in the maintenance of infective DNAs within cells have been co-opted for other functions. That is the case for the toxin-antitoxin operons (TA), which were originally described as a postsegregational killing system present in plasmids; infected cells that lose these invasive molecules will die, which increases plasmid prevalence among a given bacterial population (11). Few TA systems, such as the *mazEF* (12), *hok/soc* (13) and especially *toxIN* (14,15) systems, have been reported to exclude phages, mainly through the induction of premature cell death after phage invasion.

In the last decade, another bacterial defence system that has been in the spotlight is the CRISPR-Cas system (16). There are three types of CRISPR-Cas systems (I, II and

III), each with many subtypes (17). These systems are basically characterized by the genomic presence of a module of repetitive DNA interpolated by “spacer” sequences consisting of previous invasive DNAs. During the occurrence of another invasion, these spacers are used to positively identify exogenous DNA and oppose the infective molecules (18). With the recent discovery of CRISPR-Cas as a defence mechanism in bacteria, its presence and abundance have been the focus of studies in the genomes of many prokaryotes, especially those of human-associated genera (19–22). However, in-depth analyses are lacking for phytopathogens, even in economically important genera such as the closely related taxa *Xanthomonas* and *Xylella fastidiosa* (23,24). In this work, we performed a genome-wide investigation of CRISPR-Cas in both of these phytopathogens, which cause diseases in different plant species, and showed that these systems may be a driving force for genetic diversity, impacting pathogenicity and host-range distribution.

Materials and Methods

Genome analysis

An in-depth analysis of both prophage and CRISPR arrays (and the identification of putative protospacer targets when CRISPR was present) was carried out in 10 *Xanthomonas* genomes that we previously selected (25). The complete list is shown in Table 1. The *Xylella fastidiosa* strains analysed for CRISPR arrays are also shown in Table 1, and subspecies were selected as phylogenetically representative members of these species (26). We expanded the number of genomes analysed only for the *cas* operon search to strengthen our conclusions about what subtypes of CRISPR-Cas systems are present in the *Xanthomonas* and *Xylella* genera. Therefore, the total

numbers of genomes in this analysis were as follows: 121 *Xanthomonas* strains spanning 27 different species/pathovars (Supplemental File S1), 20 *Xylella* strains of four subspecies (Supplemental File 2S), and 7 other *Xanthomonadaceae* isolates (Supplemental File S3).

Table 1: Selection of genomes used for CRISPR array searches in both *Xanthomonas* spp. and *Xylella fastidiosa* ssp. genomes. (*) also used for prophage analysis.

<i>Xylella fastidiosa</i> genomes	Accession number
<i>X. f.</i> subsp. <i>fastidiosa</i> Temecula	NC_004556.1
<i>X. f.</i> subsp. <i>pauc</i> a 9a5c	NC_002488.3
<i>X. f.</i> subsp. <i>fastidiosa</i> M23	NC_010577.1
<i>X. f.</i> subsp. <i>multiplex</i> M12	NC_010513.1
<i>X. f.</i> subsp. <i>fastidiosa</i> GB514	NC_017562.1
<i>X. f.</i> subsp. <i>fastidiosa</i> MUL0034	NZ_CP006740.1
<i>X. f.</i> subsp. <i>sandyi</i> Ann-1	AAAM04000275.1
<i>Xanthomonas</i> genomes	
* <i>X. axonopodis</i> pv. <i>citri</i> 306	NC_003919.1
* <i>X. axonopodis</i> Xac29-1	NC_020800.1
* <i>X. citri</i> subsp. <i>citri</i> Aw12879	NC_020815.1
* <i>X. campestris</i> pv. <i>vesicatoria</i> 85-10	AM039952.1
* <i>X. campestris</i> pv. <i>raphani</i> 756C	NC_017271.1
* <i>X. campestris</i> pv. <i>campestris</i> ATCC33913	NC_003902.1
* <i>X. campestris</i> pv. <i>campestris</i> 8004	NC_007086.1
* <i>X. albilineans</i> GPE PC73	NC_013722.1
* <i>X. oryzae</i> pv. <i>oryzicola</i> BLS256	NC_017267.2
* <i>X. oryzae</i> pv. <i>oryzae</i> PXO99A	NC_010717.2

CRISPR arrays and putative protospacer target identification

Based on the strain selection previously performed by Martins et al. (2016), who analysed the TA profiles of 10 *Xanthomonas* genomes that spanned the phylogenetic tree of this genus, we decided to use the same subset of *Xanthomonas* genomes to thoroughly analyse the origin of spacers. When a CRISPR array was identified, its putative protospacer targets were evaluated. For the *Xylella* genus, we selected 7 genomes spanning the four known subspecies (*fastidiosa*, *pauca*, *multiplex* and *sandyi*) regardless of their host range (Table 1). These genomes were submitted to CRISPR Finder (27) (<http://crispr.i2bc.paris-saclay.fr/Server/>), and the output of the CRISPR array when present was subsequently submitted to CRISPR Target (28) (http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html) to identify possible matches to each of the spacers retrieved. The spacer content of every possible CRISPR array was analysed against the phage and plasmid databases provided by CRISPR Target, and to assess possible endogenous targets, we uploaded the bacterial genomes and performed the search again. In the case of a possible positive endogenous match, the sequences retrieved were further localized in the genome to identify the ORF. In these cases, the score threshold assumed for a positive ID was 5 mismatches (29). The full data for the targets identified are provided in Supplemental File S4. The results shown in Supplemental File S4 were classified into four colour-coded categories: unknown (pink), phage (green), plasmid (blue) and endogenous (yellow). To improve the analysis of the quantitative contribution of each of these targets to the composition of the CRISPR array, the numeric values were submitted to the online CIRCOS Table viewer (30) (http://circos.ca/intro/tabular_visualization/).

Prophage search

For the prophage search within the *Xanthomonas* genus, the same selection of 10 genomes used for the CRISPR array search was employed (Table 1) (25). The full chromosomal and plasmidial DNA sequences were submitted to the PHAST (31) and PHASTER (32) tools. Since our main objective was to assess viral infection entry (regardless of whether a full or incomplete prophage was involved), we used the PHAST output. The full PHAST output with the number of prophages found in each genome is shown in Supplemental File S5.

***cas* operon search**

Two different approaches were adopted to assess the presence of the *cas* operon in *Xanthomonas* and *Xylella*. The genomes already added to the CRISPI database (33) (<http://crispi.genouest.org/>) were analysed using this tool. However, the vast majority of the selected genomes are deposited as large contigs in the databases; in such cases, each CRISPR-Cas island was inspected and confirmed using CLC Genomics Workbench version 9.5.3 (QIAGEN) for the purpose of verifying the conservation of the *Cas* operon architecture.

CRISPR-Cas systems with acceptable CRISPR arrays and a *Cas* operon in the vicinity of these CRISPR units were considered valid(18). We considered CRISPR repeats embedded within ORFs to be false positives.

Cas protein phylogeny

The Cas1 amino acid sequences of *Xanthomonadaceae* taxa from two CRISPR subtypes (I-C and I-F) were used for phylogenetic reconstruction along with Cas5d, Cas7 and Cas8c (CRISPR subtype I-C) downloaded from GenBank. The alignments were checked manually, and the evolutionary history was inferred using the maximum

likelihood method based on the Jones-Taylor-Thornton (JTT) evolutionary model. Evolutionary analyses were conducted in MEGA7 (34). Additional phylogenetic trees using the Cas5d, Cas7 and Cas8c proteins (CRISPR subtype I-C) were generated to include the *Xylella taiwanensis* PLS229 taxon in this reconstruction since the operon in this isolate is eroded and does not contain the Cas1 protein, which is generally used in classical analyses.

Results

CRISPR repeat assessment

Some of the repeats reported by CRISPR Finder were considered false positives in our analyses. This was the case for one CRISPR locus from each the following 6 genomes: *X. citri* subsp. *citri* Aw12879, *X. citri* subsp. *citri* 306 and *X. axonopodis* Xac29-1, *X. campestris*: *X. campestris* pv. *campestris* ATCC 33913; *X. campestris* pv. *campestris* 8004 and *X. campestris* pv. *raphani* 756C. In all these cases, there was no associated *cas* operon in the vicinity of these repeats, and they were therefore considered false positives. These false-positive sequences and their repeats are shown in Supplemental File S6. No other CRISPR repeat region was dismissed, and they were all considered reliable. A summary of the final count of the number of CRISPR-Cas systems is shown in Supplemental File 8.

The majority of *Xanthomonas* spp. present at least one *cas* operon

Among the twenty-seven different species/pathovars evaluated, 60% presented a putative functional CRISPR-Cas system (Supplemental File S7). For each given

species/pathovar that showed at least one encoded *cas* operon, we observed that all of its strains presented the same type of *cas*. The only exception that we found was *X. oryzae* pv. *oryzicola*, in which one strain presented the subtype I-C system (str. YM15), while no CRISPR-Cas system was present in the other (str. BLS256) (Supplemental File S7). Among the CRISPR-Cas systems described to date (35), we only found Type I systems in *Xanthomonas*, of subtypes I-C and I-F (Supplemental File S6, Figure 1). The less prevalent subtype, I-F, was found exclusively in *X. fragariae*, *X. campestris* pv. *raphani* 756C, *X. albilineans* and *X. hyacinthi* (Figure 2A), while the most prevalent and widespread *Cas* operon found was I-C (Figure 2B), which was present in 15 of the 17 *Xanthomonas* species/pathovars with at least one *Cas* operon (Supplemental File S7). Curiously, *X. albilineans* and *X. hyacinthi* were the only species to present two *Cas* operons, each of which belonged to different subtypes: I-F and I-C. Interestingly, the same situation occurred in only two other *Xanthomonadaceae* species analysed as an outgroup (*Luteimonas huabeiensis* and *Dokdonella koreensis*) (Supplemental File S3, Figure 1); therefore, the presence of more than one *Cas* operon is a rare characteristic in this family. Curiously, none of the *Xylella fastidiosa* genomes analysed presented any CRISPR-Cas system, with the exception of *X. taiwanensis* PLS229, which presented a unique vestigial eroded subtype I-C-like *cas* operon, in which the genes encoding the Cas1, Cas2, Cas3 and Cas4 proteins were absent (Figure 2C).

The Cas1 phylogeny shows ancestral acquisition of the Cas operon among *Xanthomonas* species

For both subtypes I-C and I-F, we observed a common phenomenon concerning the ancestrality of the *Cas* operon among *Xanthomonas* spp. strains (Figure 3A and 3B). Despite the wide range of horizontal gene transfer events in bacteria (36), it is

reasonable to assume that Cas1 was acquired in a unique event of acquisition due to the high identity of this protein between different pathotypes of the same species. For instance, Figure 3A shows that *X. citri* 306 and *X. citri* Aw12879 exhibit identical Cas1 proteins despite their known differential host specificity (37). The same phenomenon was found in other species/pathovars showing different host preferences but presenting Cas1 clustering with high identity. The less prevalent subtype I-F showed exactly the same Cas1 profile, clustering the *Xanthomonas* strains of the same species/pathovar together (Figure 3B). Since no Cas operon or CRISPR was present in the *Xylella fastidiosa* genomes with the exception of *X. taiwanensis* PLS229, the phylogenetic reconstruction of the *Xanthomonadaceae* incorporating *X. taiwanensis* PLS229 was based on the sequences of the other proteins that are still present in this strain (Cas5d, Cas7/Csd1, Cas8c/Csd2) in an attempt to compare the phylogenetic signal of these unusual markers (Figure 4). The same taxon clusters detected in the phylogeny using Cas1 were observed in the trees generated using the Cas5d, Cas7/Csd1, and Cas8c/Csd2 proteins, in which the *Xanthomonas* strains of the same species/pathovars were clustered together, reinforcing the ancestral acquisition of the Cas operon among *Xanthomonas* species. For the *Xanthomonadaceae* analysed here, the phylogenetic signal of Cas1 (Figure 3) can be compared with those of other Cas proteins of subtypes I-C (Figure 4).

The analysis of spacers shows a wide variety of targets

Each CRISPR locus of the strains was thoroughly analysed to assess the targets of each spacer. Although 40% of the strains showed no CRISPR-Cas systems whatsoever, those that harboured at least one such system showed variation in the number of spacers and their targets. The greatest number of these systems was found in

X. campestris pv. *raphani* 756C (99 spacers), followed by *X. oryzae* pv. *oryzae* PXO99A (75 spacers). Additionally, although most of the targets were unknown, those that were identified showed matches with phage, plasmid and endogenous genome sequences (Figure 5).

Our study showed that *X. oryzae* pv. *oryzae* PXO99A presented the greatest number of spacers targeting phages (frequently OP2, OP1 and Xop144) (Figure 5, green squares). In addition, spacer sequences targeting plasmids (Figure 5, blue squares) were found in *X. oryzae* pv. *oryzae* PXO99A and *X. campestris* pv. *raphani* 756C. Both *X. citri* subsp. *citri* 306 and *X. axonopodis* Xac29-1 presented one CRISPR-Cas system. However, the target could not be identified for any of the spacers encoded by their genomes. Likewise, *X. citri* subsp. *citri* Aw12869 presented one CRISPR-Cas system whose targets were not identified; however, a second CRISPR array was also detected in this strain. Despite the lack of an association with a *cas* operon in its vicinity, one of the spacer targets was positively identified in a phage sequence (Supplemental File S4). We therefore considered this second CRISPR array to be a putatively functional CRISPR-Cas system that may operate with the Cas proteins produced *in trans* by the first CRISPR-Cas system. Curiously, spacers targeting endogenous genome sequences were found at the two CRISPR loci only in *Xanthomonas albilineans* GPE PC73 (Figure 5, yellow squares). The percent contribution of each type of target in each CRISPR array is presented in Figure 6. The vast majority of unidentified matches and how the abundance of each category of spacers varies among the *Xanthomonas* strains are notable.

Spacers targeting *Xanthomonas* plasmids and prophage analyses

In addition to phages, plasmids targeted by the CRISPR-Cas systems were found in some *Xanthomonas* CRISPRs, and it was noteworthy that all of them presented the best matches to common *Xanthomonas* plasmids with high identity. Plasmids from *X. axonopodis* and *X. fuscans* subsp. *fuscans* are targets of the CRISPR-Cas systems of *X. campestris* pv. *raphani* (Figure 7 A, B and C). The other three spacers found in *X. oryzae* exhibited matches with 100% identity to plasmid targets of *X. citri* subsp. *citri* (Figure 7 D, E and F). We observed that *Xanthomonas* strains with more spacers presented fewer plasmids (Table 4). On the other hand, we noted that the *Xanthomonas* strains with many spacers targeting phages were those with more prophages integrated in their genome (Table 2).

Table 2. Presence of CRISPR-Cas versus mobile genetic elements. For the complete data, please see Supplemental Files S5 and S8

Genomes	CRISPR-Cas systems	Plasmids	Prophages / genome
<i>Xanthomonas citri</i> subsp. <i>citri</i> 306	1	<u>Yes</u> (2)	4
<i>Xanthomonas axonopodis</i> XAC29-1	1	<u>Yes</u> (3)	3
<i>Xanthomonas citri</i> subsp. <i>citri</i> Aw12879	2	<u>Yes</u> (2)	6
<i>Xanthomonas campestris</i> pv. <i>raphani</i> 756C	<u>1</u>	No	3
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99A	<u>1</u>	No	15
<i>Xanthomonas albilineans</i> GPE PC73	<u>2</u>	<u>Yes</u> (3)	3
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> 85-10	0	<u>Yes</u> (4)	7
<i>Xanthomonas campestris</i> pv. <i>campestris</i> ATCC 33913	0	No	4
<i>Xanthomonas campestris</i>	0	No	4

pv. campestris 8004			
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BLS256	0	No	6

Discussion

Recently, the CRISPR-Cas system was identified as a defence mechanism in bacteria; however, very little is known about its occurrence, abundance and targets in phytopathogens. Therefore, in this study, we performed a broad genome analysis of CRISPR-Cas in two closely related economically important genera, *Xanthomonas* and *Xylella*. Interestingly, no CRISPR-Cas system was found in *X. fastidiosa*. However, an eroded *Cas* operon was found in *X. taiwanensis*, a distant relative from Taiwan (26), raising the possibility that CRISPR-Cas systems may have been acquired but did not remain over time. It has been reported that entire bacterial lineages may lack CRISPR-Cas systems, as is the case for the *Chlamydiae* phylum among others, which is probably due to the potential deleterious autoimmunity risk that carrying a CRISPR-Cas system may pose (38). In addition, this absence may be a characteristic that is restricted to *Xylella* spp. and is not widespread among the *Xanthomonadaceae* since we also analysed 4 other genera within this family (*Thermomonas*, *Stenotrophomonas*, *Pseudoxanthomonas* and *Luteimonas*), and all of them showed at least one *Cas* operon. It is important to consider that phage-related regions of *Xylella fastidiosa* genomes can account for as much as 15% of the genome (39), which might be a direct result of the absence of CRISPR-Cas systems.

In contrast to the absence of CRISPR-Cas systems in *Xylella*, 60% of *Xanthomonas* spp. showed at least one *Cas* operon. Among the three types of CRISPR-Cas systems and their multiple subtypes (28), only Type I was identified in *Xanthomonas* (subtypes I-C and I-F). Usually, only one subtype was present per

genome, except in *X. albilineans* and *X. hyacinthi*, which presented one copy of each I-C and I-F. We observed that the subtypes present in *Xanthomonas* (I-C, I-F or both together) were consistent among the strains of a particular species, which is in accordance with the observation that strains belonging to the same species usually harbour the same CRISPR-Cas system (40). We also highlight that species belonging to distantly *Xanthomonas*-related genera in *Xanthomonadaceae* presented the same configuration of coexistence of the same I-C and I-F CRISPR subtypes. In addition, our phylogenetic analysis indicated that the CRISPR systems present in *Xanthomonas* spp. are the result of an ancient acquisition.

Despite the similarities of the CRISPR-Cas subtypes in *Xanthomonas* spp. genomes, they presented significant variation in both the number and targets of spacers. The greatest number of spacers targeting sequences was observed in *X. oryzae* pv. *oryzae* PXO99A, which was in agreement with other studies that have emphasized the abundance of spacers in the CRISPR arrays of *X. oryzae* pv. *oryzae* (41,42). Regarding targets, self-targeting endogenous spacers were found only in *X. albilineans*, despite their presence in many bacterial genomes (43). The presence of self nucleic acids in CRISPR arrays indicates a form of autoimmunity that could explain the abundance of degraded CRISPR systems across prokaryotes (37). In addition, endogenous CRISPR spacers have been associated with regulatory mechanisms for repressing phage replication (44)). However, an important characteristic observed in this study was that the identified plasmid-targeting spacers were always driven toward plasmids found in other *Xanthomonas* strains, with *X. oryzae* pv. *oryzae* harbouring many of these spacers and being devoid of any extrachromosomal DNA. The same was true for *X. campestris* pv. *raphani*, which raises the possibility that CRISPR-Cas systems could be very effective in coping with plasmidial infections. Indeed *X. campestris* pv. *vesicatoria* and

X. fastidiosa present more than one plasmid and no functional CRISPR-Cas system. On the other hand, the strain harbouring the greatest number of spacers targeting phages, *X. oryzae* pv. *oryzae* PXO99A, exhibited the greatest number of prophages in the genome, which may be a result of a very challenging environment concerning phage diversity but may also indicate that this system may not be functioning at the same rate at which viruses evolve to evade it (45). Therefore, CRISPR-Cas systems in *Xanthomonas* seem to be very effective in controlling plasmid infections, but they do not show the same success regarding phages. Since many effectors are plasmid encoded, CRISPR-Cas might be driving the specific characteristics of plant-pathogen interactions.

This is the first genus-wide analysis of CRISPR-Cas systems in *Xanthomonas*, and we conclude that the presence or absence of functional CRISPR-Cas systems may be an important driving force of genetic diversity in this genus, either allowing the entry and maintenance of DNAs in the cell or not, which may impose important gene flow restrictions in the course of evolution, consequently impacting the pathogenicity and host-range distribution of *Xanthomonas* spp.

Authors statements

The authors declare the absence of any potential conflict of interest.

Authors contribution

PM conceived and wrote the manuscript, and PM and AX executed the bioinformatics analysis. MAT, PAZ and AADS discussed the results and critically reviewed the manuscript.

Acknowledgements

This work was supported by research grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - 2013/10957-0) and INCT-Citrus (CNPq 465440/2014-2 and FAPESP 2014/50880-0). PM is a FAPESP post-doctoral fellow (2016/01273-9).

References

1. Baldi P, La Porta N. Xylella fastidiosa: Host Range and Advance in Molecular Identification Techniques. Front Plant Sci [Internet]. 2017 [cited 2017 Dec 4];8:944. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28642764>
2. Boulanger A, Noël LD. Xanthomonas Whole Genome Sequencing: Phylogenetics, Host Specificity and Beyond. Front Microbiol [Internet]. 2016 [cited 2017 Dec 4];7:1100. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27470197>
3. Darmon E, Leach DRF. Bacterial genome instability. Microbiol Mol Biol Rev [Internet]. 2014 Mar [cited 2017 Dec 4];78(1):1–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24600039>
4. San Millan A, MacLean RC. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. Microbiol Spectr [Internet]. 2017;5(5):1–12. Available from: <http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.MTBP-0016-2017>

5. Clokie MR, Millard AD, Letarov A V, Heaphy S. Phages in nature. Bacteriophage [Internet]. 2011 Jan [cited 2017 Dec 4];1(1):31–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21687533>
6. Bikard D, Marraffini LA. Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. Curr Opin Immunol [Internet]. 2012 Feb [cited 2017 Nov 29];24(1):15–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22079134>
7. Brüssow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev [Internet]. 2004 Sep [cited 2017 Dec 4];68(3):560–602, table of contents. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15353570>
8. Harrison E, Truman J, Wright R, Spiers AJ, Paterson S, Brockhurst MA. Plasmid carriage can limit bacteria-phage coevolution. Biol Lett [Internet]. 2015 Aug 1 [cited 2017 Dec 4];11(8):20150361. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26268992>
9. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev [Internet]. 2013 Mar [cited 2017 Dec 4];77(1):53–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23471617>
10. Dy RL, Richter C, Salmond GPC, Fineran PC. Remarkable Mechanisms in Microbes to Resist Phage Infections. Annu Rev Virol [Internet]. 2014 Nov 3 [cited 2017 Dec 5];1(1):307–31. Available from: <http://www.annualreviews.org/doi/10.1146/annurev-virology-031413-085500>
11. Gerdes K, Bech FW, Jørgensen ST, Løbner-Olesen A, Rasmussen PB, Atlung T,

et al. Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon. EMBO J [Internet]. 1986 Aug [cited 2015 Nov 23];5(8):2023–9.

Available from:

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1167073&tool=pmcentrez&rendertype=abstract>

12. Hazan R, Engelberg-Kulka H. Escherichia coli mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol Genet Genomics [Internet]. 2004 Sep 14 [cited 2017 Dec 5];272(2):227–34. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/15316771>

13. Pecota DC, Wood TK. Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. J Bacteriol [Internet]. 1996 Apr [cited 2017 Dec 5];178(7):2044–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8606182>

14. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci [Internet]. 2009 Jan 20 [cited 2017 Dec 5];106(3):894–9. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/19124776>

15. Blower TR, Fineran PC, Johnson MJ, Toth IK, Humphreys DP, Salmond GPC. Mutagenesis and Functional Characterization of the RNA and Protein Components of the toxIN Abortive Infection and Toxin-Antitoxin Locus of Erwinia. J Bacteriol [Internet]. 2009 Oct 1 [cited 2017 Dec 5];191(19):6029–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19633081>

16. Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: progress,

- 404 implications and challenges. Hum Mol Genet [Internet]. 2014 Sep 15 [cited 2017
405 Dec 4];23(R1):R40–6. Available from: [https://academic.oup.com/hmg/article-
406 lookup/doi/10.1093/hmg/ddu125](https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddu125)
- 407 17. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al.
408 An updated evolutionary classification of CRISPR–Cas systems. Nat Rev
409 Microbiol [Internet]. 2015 Sep 28 [cited 2017 Dec 29];13(11):722–36. Available
410 from: <http://www.ncbi.nlm.nih.gov/pubmed/26411297>
- 411 18. Hille F, Charpentier E. CRISPR-Cas: biology, mechanisms and relevance. Philos
412 Trans R Soc Lond B Biol Sci [Internet]. 2016 Nov 5 [cited 2018 Apr
413 10];371(1707). Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27672148>
- 414 19. Wang P, Zhang B, Duan G, Wang Y, Hong L, Wang L, et al. Bioinformatics
415 analyses of Shigella CRISPR structure and spacer classification. World J
416 Microbiol Biotechnol [Internet]. 2016 Mar 11 [cited 2017 Dec 29];32(3):38.
417 Available from: <http://link.springer.com/10.1007/s11274-015-2002-3>
- 418 20. Hidalgo-Cantabrana C, Crawley AB, Sanchez B, Barrangou R. Characterization
419 and Exploitation of CRISPR Loci in Bifidobacterium longum. Front Microbiol
420 [Internet]. 2017 Sep 26 [cited 2017 Dec 29];8:1851. Available from:
421 <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01851/full>
- 422 21. Koskela KA, Mattinen L, Kalin-Mänttari L, Vergnaud G, Gorgé O, Nikkari S, et
423 al. Generation of a CRISPR database for *Yersinia pseudotuberculosis* complex
424 and role of CRISPR-based immunity in conjugation. Environ Microbiol
425 [Internet]. 2015 Nov [cited 2017 Dec 29];17(11):4306–21. Available from:
426 <http://doi.wiley.com/10.1111/1462-2920.12816>
- 427 22. Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, et al.

- 428 Function of the CRISPR-Cas System of the Human Pathogen *Clostridium*
429 *difficile*. MBio [Internet]. 2015 Sep 1 [cited 2017 Dec 29];6(5):e01112-15.
430 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26330515>
- 431 23. Almeida RPP, De La Fuente L, Koebnik R, Lopes JRS, Parnell S, Scherm H.
432 Addressing the New Global Threat of *Xylella fastidiosa*. Phytopathology
433 [Internet]. 2019 Feb [cited 2019 May 27];109(2):172–4. Available from:
434 <http://www.ncbi.nlm.nih.gov/pubmed/30721121>
- 435 24. Brunings AM, Gabriel DW. *Xanthomonas citri*: breaking the surface. Mol Plant
436 Pathol [Internet]. 2003 May;4(3):141–57. Available from:
437 <http://doi.wiley.com/10.1046/j.1364-3703.2003.00163.x>
- 438 25. Martins PMM, Machado MA, Silva N V., Takita MA, de Souza AA. Type II
439 Toxin-Antitoxin Distribution and Adaptive Aspects on *Xanthomonas* Genomes:
440 Focus on *Xanthomonas citri*. Front Microbiol [Internet]. 2016 May 10 [cited
441 2017 May 23];7:652. Available from:
442 <http://www.ncbi.nlm.nih.gov/pubmed/27242687>
- 443 26. Almeida RPP, Nunney L. How Do Plant Diseases Caused by *Xylella fastidiosa*
444 Emerge? Plant Dis [Internet]. 2015 Nov [cited 2019 May 27];99(11):1457–67.
445 Available from: <http://apsjournals.apsnet.org/doi/10.1094/PDIS-02-15-0159-FE>
- 446 27. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered
447 regularly interspaced short palindromic repeats. Nucleic Acids Res [Internet].
448 2007 Jul 8 [cited 2017 May 23];35(Web Server issue):W52-7. Available from:
449 <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkm360>
- 450 28. Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM. CRISPRTarget.
451 RNA Biol [Internet]. 2013 May 14 [cited 2017 May 23];10(5):817–27. Available

from: <http://www.ncbi.nlm.nih.gov/pubmed/23492433>

29. Shariat N, Timme RE, Pettengill JB, Barrangou R, Dudley EG. Characterization and evolution of Salmonella CRISPR-Cas systems. *Microbiology*. 2015;161(May):374–86.

30. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res* [Internet]. 2009 Sep 18 [cited 2017 Dec 6];19(9):1639–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19541911>

31. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res* [Internet]. 2011 Jul 1 [cited 2017 Sep 12];39(suppl):W347–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21672955>

32. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* [Internet]. 2016 Jul 8 [cited 2017 Sep 12];44(W1):W16–21. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw387>

33. Rousseau C, Gonnet M, Le Romancer M, Nicolas J. CRISPI: a CRISPR interactive database. *Bioinformatics* [Internet]. 2009 Dec 15 [cited 2017 May 23];25(24):3317–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19846435>

34. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* [Internet]. 2016 Jul [cited 2018 Jan 3];33(7):1870–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27004904>

- 476 35. Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin E V., van der Oost J.
477 Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas
478 systems. *Science* (80-) [Internet]. 2016 Aug 5 [cited 2018 Feb
479 4];353(6299):aad5147. Available from:
480 <http://www.ncbi.nlm.nih.gov/pubmed/27493190>
- 481 36. Oliveira PH, Touchon M, Cury J, Rocha EPC. The chromosomal organization of
482 horizontal gene transfer in bacteria. *Nat Commun* [Internet]. 2017 Dec 10 [cited
483 2018 Nov 1];8(1):841. Available from:
484 <http://www.ncbi.nlm.nih.gov/pubmed/29018197>
- 485 37. Jalan N, Kumar D, Yu F, Jones JB, Graham JH, Wang N. Complete Genome
486 Sequence of *Xanthomonas citri* subsp. *citri* Strain Aw12879, a Restricted-Host-
487 Range Citrus Canker-Causing Bacterium. *Genome Announc* [Internet]. 2013
488 May 16 [cited 2017 May 29];1(3):e00235-13-e00235-13. Available from:
489 <http://genomea.asm.org/cgi/doi/10.1128/genomeA.00235-13>
- 490 38. Burstein D, Sun CL, Brown CT, Sharon I, Anantharaman K, Probst AJ, et al.
491 Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence
492 systems. *Nat Commun* [Internet]. 2016 Feb 3 [cited 2018 Jan 2];7:10613.
493 Available from: <http://www.nature.com/doi/10.1038/ncomms10613>
- 494 39. de Mello Varani A, Souza RC, Nakaya HI, de Lima WC, Paula de Almeida LG,
495 Kitajima EW, et al. Origins of the *Xylella fastidiosa* prophage-like regions and
496 their impact in genome differentiation. *PLoS One* [Internet]. 2008 [cited 2018 Jan
497 2];3(12):e4059. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19116666>
- 498 40. Louwen R, Staals RHJ, Endtz HP, van Baarlen P, van der Oost J. The role of
499 CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol Mol Biol*

Rev [Internet]. 2014 Mar [cited 2018 Jan 2];78(1):74–88. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/24600041>

41. Midha S, Bansal K, Kumar S, Giriya AM, Mishra D, Brahma K, et al. Population genomic insights into variation and evolution of *Xanthomonas oryzae* pv. *oryzae*. Sci Rep [Internet]. 2017 Jan 13 [cited 2017 May 29];7:40694. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/28084432>

42. Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S, et al. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. BMC Genomics [Internet]. 2008 Jan [cited 2016 Apr 8];9:204. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2432079&tool=pmcentrez&rendertype=abstract>

43. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRIPR: gene regulation or autoimmunity? Trends Genet. 2010;26:335–40.

44. Yang C-D, Chen Y-H, Huang H-Y, Huang H-D, Tseng C-P. CRP represses the CRISPR/Cas system in *Escherichia coli*: evidence that endogenous CRISPR spacers impede phage P1 replication. Mol Microbiol [Internet]. 2014 Jun [cited 2017 May 29];92(5):1072–91. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/24720807>

45. Andersson AF, Banfield JF. Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. Science (80-) [Internet]. 2008 May 23 [cited 2018 Jan 3];320(5879):1047–50. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/18497291>

Figure 1. Overview of Cas operons found in *Xanthomonadaceae* species. *Xanthomonas* spp. predominantly carry Cas operons of subtype I-C, and Cas operons of subtype I-F can be found in some species at a lower frequency. Co-existence of the two CRISPR subtypes occurs in *X. albilineans* and *X. hyacinthi*. Taxa of other *Xanthomonas*-related genera also possess the two CRISPR subtypes and present a similar architecture, either occurring alone or co-existing in one isolate. Interestingly, some Cas operons do not possess the usual architecture and contain putative ORFs of completely unknown function in the stages of molecular execution by canonical CRISPR Type I. The region delimited by the dashed line indicates the species in which coexistence of the two CRISPR subtypes occurs. For the CRISPR I-C subtype, the adaptation and interference (cascade complex) modules are brown and blue, respectively, and for the CRISPR I-F subtypes, they are yellow and green, respectively. Unusual ORFs are represented as red arrows.

Figure 2. Cas operons found in *Xanthomonas* spp. A) *Xanthomonas campestris* pv. *raphani* is one of the strains to present only the I-F Cas operon subtype; B) *Xanthomonas citri* strains consistently presented a unique Cas operon of subtype I-C; C) Eroded Cas operon present in *Xylella taiwanensis*. The genes are depicted as arrows, with its putative gene names above. In green, the ORFs found in this genome, and with an “x”, the genes that are absent, but expected to be found in a subtype I-C CRISPR-Cas system.

548

549 **Figure 3.** Cas1 phylogeny of subtypes I-F and I-C of CRISPR-Cas systems in
550 *Xanthomonas* spp. and other *Xanthomonadaceae* species as outgroups. Highlighted
551 coloured rectangles denote the clusters formed by the same *Xanthomonas* A) Subtype I-
552 C, showing that *X. citri* strains cluster together despite having different hosts; B)
553 subtype I-F, where the same phenomenon of species clustering despite different host
554 preferences was observed. Bootstrap values ($\geq 50\%$) are shown beside each node.

555

556 **Figure 4.** Cas5d, Cas7/Csd1 and Cas8c/Csd2 phylogeny of subtype I-C of CRISPR-Cas
557 systems in *Xanthomonas* spp. and other *Xanthomonadaceae* species as outgroups.
558 Highlighted coloured rectangles denote the clusters formed by the same *Xanthomonas*
559 species/pathovars. The only species of *Xylella* that contains an eroded subtype I-C-like
560 CRISPR-Cas system, *X. taiwanensis*, is highlighted in the tree with a grey circle.
561 Bootstrap values ($\geq 50\%$) are shown beside each node.

562

563 **Figure 5:** Schematic representation of CRISPR repeats and their spacers for each
564 CRISPR locus. Grey squares represent repeats; yellow, pink, green and blue squares
565 represent spacer targets of endogenous, unknown, phage and plasmid sequences,
566 respectively. Numbers I and II are used to identify each CRISPR locus when more than
567 one is found in the same genome. The genomic coordinates are indicated with the
568 numbers above and under the squares. A) XAC (*Xanthomonas citri* subsp. *citri* 306); B)
569 XAC29 (*Xanthomonas axonopodis* Xac29-1); C) XCAW (*Xanthomonas citri* subsp.
570 *citri* Aw12869); D) PXO (*Xanthomonas oryzae* pv. *oryzae* PXO99A); E) XCR

(*Xanthomonas campestris* pv. *raphani* 756C); F) XAL (*Xanthomonas albilineans* GPE PC73).

Figure 6. Percent contribution of each spacer target for each *Xanthomonas* spp. Ribbon colours represent the following categories pink: unknown; yellow: self-targets; green: phage-related; blue: plasmids. Numbers I and II are used to identify each CRISPR locus when more than one is found in the same genome. XAC (*Xanthomonas citri* subsp. *citri* 306); XAC29 (*Xanthomonas axonopodis* Xac29-1); XCAW (*Xanthomonas citri* subsp. *citri* Aw12869); PXO (*Xanthomonas oryzae* pv. *oryzae* PXO99A); XCR (*Xanthomonas campestris* pv. *raphani* 756C); XAL (*Xanthomonas albilineans* GPE PC73). Despite being less frequent, the plasmid targets that we identified were all from *Xanthomonas* plasmids.

Figure 7. Selected examples of representative alignments between the putative transcribed crRNA and the protospacer. Alignments A, B and C are from spacers found in *X. campestris* pv. *raphani* 756C, and alignments D, E and F are from spacers found in *X. oryzae* pv. *oryzae* PXO99A. The protospacer identities are as follows: A) *X. axonopodis* Xac29-1 plasmid pXAC47; B) *X. fuscans* subsp. *fuscans* 4834- plasmid pla; C) *X. fuscans* subsp. *fuscans* 4834- plasmid pla; D) *X. citri* subsp. *citri* MN12 plasmid pXAC64; E) *X. citri* subsp. *citri* A306 plasmid pXAC64; F) *X. citri* subsp. *citri* NT17 plasmid pXAC64

Subtype I-C/Dvulg

Subtype I-F/Ypest

Two CRISPR Subtypes

		Cascade complex	I-C canonical architecture	I-F canonical architecture	
<i>X. translucens</i> pv. <i>cerealis</i> CFPB 2541			Yes	—	
<i>X. translucens</i> pv. <i>cerealis</i> NCPPB1943			Yes	—	
<i>X. translucens</i> pv. <i>poae</i> CNC2-P4			Yes	—	
<i>X. translucens</i> pv. <i>poae</i> strain ATCC 33804			Yes	—	
<i>X. translucens</i> pv. <i>arrhenatheri</i> UPB455			Yes	—	
<i>X. translucens</i> pv. <i>arrhenatheri</i> LMG 727			Yes	—	
<i>X. vasicola</i> NCPPB 1395			Yes	—	
<i>X. vasicola</i> NCPPB 902			Yes	—	
<i>X. cassavae</i> CFBP 4642			Yes	—	
<i>X. axonopodis</i> pv. <i>vasculorum</i> NCPPB 900			Yes	—	
<i>X. bromi</i> LMG947			Yes	—	
<i>X. oryzae</i> pv. <i>oryzicola</i> YM15			Yes	—	
<i>X. oryzae</i> pv. <i>oryzae</i> KACC 10331			Yes	—	
<i>X. oryzae</i> pv. <i>oryzae</i> PXO99A			Yes	—	
<i>X. axonopodis</i> Xac29-1			Yes	—	
<i>X. citri</i> subsp. <i>citri</i> A306			Yes	—	
<i>X. citri</i> subsp. <i>citri</i> Aw12879			Yes	—	
<i>Rhodanobacter thiooxydans</i> FW104-T7			Yes	—	
<i>Pseudoxanthomonas spadix</i> BD-a59			Yes	—	
<i>Stenotrophomonas ginsengisoli</i> DSM 24757			Yes	—	
<i>Xylella taiwanensis</i> PLS229			No ^a	—	
<hr/>					
<i>X. albilineans</i> XaFL07-1			No ^b		Yes
<i>X. albilineans</i> GPE PC73			No ^b		Yes
<i>X. hyacinthi</i> DSM 19077			Yes		Yes
<i>Luteimonas huabeiensis</i> HB2			Yes		No ^b
<i>Dokdonella koreensis</i> DS-123			Yes		Yes
<hr/>					
<i>X. campestris</i> pv. <i>raphani</i> 756C			—		Yes
<i>X. campestris</i> pv. <i>raphani</i> CFBP 5828R			—		Yes
<i>X. fragariae</i> Fap29			—		Yes
<i>X. fragariae</i> LMG 25863			—		Yes
<i>Rhodanobacter</i> sp. 67-28			—		Yes
<i>Rhodanobacter</i> sp. SCN 66-43			—		No ^b
<i>Stenotrophomonas acidaminiphila</i> JCM 13310			—		Yes
<i>Stenotrophomonas nitrireducens</i> DSM 12575			—		Yes
<i>Thermomonas fusca</i> DSM 15424			—		No ^b

X. campestris pv. *raphani*

A	B	C
sp. 98	sp. 44	sp. 47
A - T	G - C	G - C
A - T	G - C	U - A
G - C	A - G	C - G
G - C	G - C	A - T
G - C	U - A	U - A
U - T	C - G	G - C
G - C	G - C	G - C
A - T	G - C	A - T
C - G	G - C	C - G
G - C	G - C	G - C
A - T	G - C	G - C
A - C	A - T	U - A
G - C	G - C	G - C
A - T	G - C	C - G
C - G	C - G	G - C
A - T	G - C	C - G
A - T	U - A	A - T
C - G	G - C	C - A
U - A	G - C	A - T
A - T	G - C	U - A
C - G	G - C	C - G
G - C	C - G	C - G
A - T	A - T	A - T
U - A	U - A	C - G
C - G	G - C	G - C
A - T	G - C	U - A
G - C	G - C	G - C
A - T	U - A	U - A
A - T	G - C	G - C
G - C	G - C	G - C
C - A	G - C	U - A
U - A	G - C	A - T

X. oryzae pv. *oryzae*

D	E	F
sp. 68	sp. 69	sp. 70
A - T	C - G	A - T
A - T	C - G	C - G
G - C	G - C	G - C
C - G	A - T	U - A
U - A	C - G	A - T
G - C	C - G	G - C
A - T	U - A	A - T
C - G	C - G	A - T
U - A	A - T	C - G
A - T	A - T	G - C
G - C	G - C	A - T
C - G	G - C	G - C
U - A	C - G	C - G
C - G	G - C	G - C
G - C	C - G	C - G
A - T	G - C	U - A
A - T	U - A	U - A
G - C	U - A	G - C
A - T	C - G	U - A
G - C	C - G	U - A
C - G	U - A	A - T
C - G	C - G	U - A
A - T	C - G	A - T
A - T	G - C	G - C
G - C	G - C	C - G
C - G	U - A	G - C
G - C	U - A	C - G
C - G	C - G	U - A
A - T	A - T	C - G
G - C	C - G	G - C
U - A	A - T	C - G
G - C	C - G	C - G
A - T	A - T	C - G
C - G	A - T	A - T
		C - G