NATALY FIGUEIREDO FERREIRA

CHARACTERIZATION OF PROPHAGE-LIKE ELEMENTS IN PLANT PATHOGENIC *Xanthomonas* SPECIES

Dissertation submitted to the Agricultural Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Advisor: Poliane Alfenas-Zerbini

Co-advisors: Flávia de Oliveira Souza
Fernanda Prieto Bruckner

VIÇOSA - MINAS GERAIS
2022

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
F383c
2022

Ferreira, Nataly Figueiredo, 1997-
Characterization of prophage-like elements in plant pathogenic *Xanthomonas* species / Nataly Figueiredo Ferreira. – Viçosa, MG, 2022.
1 dissertação eletrônica (98 f.): il. (algumas color.).

Texto em inglês.
Inclui anexos.
Orientador: Poliane Alfenas Zerbini.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Microbiologia, 2022.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2022.389
Modo de acesso: World Wide Web.

1. Bacteriófagos. 2. Genômica. 3. Adaptação fisiológica. I. Zerbini, Poliane Alfenas, 1977-. II. Universidade Federal de Viçosa. Departamento de Microbiologia. Programa de Pós-Graduação em Microbiologia Agrícola. III. Título.

CDD 22. ed. 579.26

Bibliotecário(a) responsável: Alice Regina Pinto CRB6 2523

NATALY FIGUEIREDO FERREIRA

CHARACTERIZATION OF PROPHAGE-LIKE ELEMENTS IN PLANT PATHOGENIC *Xanthomonas* SPECIES

Dissertation submitted to the Agricultural Microbiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: March 04, 2022.

Assent:

Nataly Figueiredo Ferreira
Author

Poliane Alfenas Zerbini
Adviser

To my grandparents Antônio and Magela.
My aguardian angels. I miss you the most.
I dedicate!

AKNOWLEDGEMENTS

Thank you, initially, to the grandparents, and Antônio Magela, even though I am no longer here physically, were my not to give up. Thank my grandmothers, Santana and Emília, for dedicating all the love, affection and attention in the world to me and my brothers and be our base.

To thank my parents, Camilla and Paulo, for doing everything possible and impossible for me and the brothers to have comfort and freedom so that we can choose who we want to be in this life. Thanks to my brothers, Ian and Bernardo, for being the reason I want to be a better human being, for being my incentive to grow in life and always being able to provide for the best conditions. Thanks for being my greatest companions throughout my journey.

I thank my friends from Sete Lagoas (Pabline and Davisson, Daiane and Lion, Jessiquinha and Wendell, Cacau, Debis, Deborah, Luana, Ana Paula, Daiany and Dayane), for all the moments of companionship and distraction since before the beginning of this long journey, and are still present today!

Thanks to the great friends they made in Viçosa: Jeje, Yan, Thamylles, Pampam, João, Nicolly, Bia, Lelé, Weyber and Iago. Gratitude for all the walk, for all the laughs, for the companionship, for many nights at my Rita and Tdies, for all the moments that will always be etched in memory.

We thank our colleagues at the Virus Laboratory: Rafa, Flávia, Pâmela, Luan, Lohan, Patrícia, Bruna, Túlio and Fernandinha, for the coffee, laughter and for all the strength they gave me to build my work.

I would like to thank Professor Poliane for all the learning that I acquired during this time of the Master's, which enabled me to build and finalize my work.

I thank the other professors of the Graduate Program in Agricultural Microbiology, for their support in the classes that were essential for my growth as a professional.

To the Federal University of Viçosa, for the opportunity to complete the postgraduate course.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for granting the scholarship.

To the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), for granting the scholarship.

To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), to granting the scholarship

“Não são as espécies mais fortes que sobrevivem nem as mais inteligentes, e sim as mais suscetíveis a mudanças.”

Charles Robert Darwin

ABSTRACT

FERREIRA, Nataly Figueiredo, M.Sc., Universidade Federal de Viçosa, March, 2022. **Characterization of prophage-like elements in plant pathogenic *Xanthomonas* species.** Adviser: Poliane Alfenas-Zerbini. Co-advisers: Flávia de Oliveira Souza and Fernanda Prieto Bruckner.

Bacteriophages are viruses that infect prokaryotes that are present in diverse environments. Prophages are bacteriophages with genome into the host genome and have been considered part of bacteria mobilome. For instance, bacteriophages may encode proteins that increase bacteria toxicity and virulence, confer genetic variability, and facilitate horizontal gene transfer, improving bacterial fitness. The *Xanthomonas* genus harbors many phytopathogenic bacteria that impair many crops worldwide. However, there is a scarcity of studies that describes bacteriophage distribution and diversity within this genus. Given this, our study aims to identify and characterize the presence of prophage-like elements in three species belonging to the *Xanthomonas* genus: *X. axonopodis, X. campestris* and *X. citri*. We found prophage-like sequences in 98% of the analyzed genomes, whereby more than one prophage-like was detected within the same bacterial genome, event denominated as polilysogen. The prophages-like from *Xanthomonas* spp. belonged to four viral families (*Myoviridae*, *Siphoviridae, Podoviridae, Inoviridae*), three families belonging to the *Caudovirales* order and one belonging to the *Tubulavirales* order (*Inoviridae*). Furthermore, many prophages-like harbored genes of putative bacterial origin, indicating horizontal gene transfer between the virus and host, such as genes encoding virulence factors. Taken together, these results indicated that prophages-like are widespread in *Xanthomonas* spp., influencing the genome diversity and fitness of these bacteria.

Keywords: Bacteriophage. PHASTER. Genome plasticity. Bacteria fitness.

RESUMO

FERREIRA, Nataly Figueiredo, M.Sc., Universidade Federal de Viçosa, março de 2022. **Caracterização de elementos do tipo profagos em espécies de *Xanthomonas* patogênicas.** Orientadora: Poliane Alfenas-Zerbini. Coorientadoras: Flávia de Oliveira Souza e Fernanda Prieto Bruckner.

Bacteriófagos são vírus que infectam procariotos presentes em diversos ambientes, que variam dos mais comuns aos com condições ambientais extremas. Profagos são bacteriófagos que existem na forma integrada ao genoma do seu hospedeiro e vêm sendo estudados como parte do mobiloma bacteriano. Os profagos influenciam o estilo de vida dos seus hospedeiros de diversas formas, como, por exemplo, codificando proteínas que aumentem a toxicidade e virulência, assim como conferindo variabilidade genética, o que pode ser vantajoso no ambiente em que o microrganismo vive. O gênero bacteriano *Xanthomonas* abrange uma variedade de bactérias fitopatogênicas que acomete diversas culturas agrícolas ao redor do mundo. Estudos que apontam a presença de bacteriófagos nesse gênero e a sua correlação com os seus hospedeiros são pouco explorados até o momento. Visto isso, o objetivo do nosso trabalho foi identificar e caracterizar a presença de profagos em três espécies pertencentes ao gênero *Xanthomonas* (*X. axonopodis, X. campestris* e *X. citri*). Foram encontradas sequências de profagos em 98% dos genomas analisados, sendo que existiram casos em que no mesmo genoma foi encontrado mais de uma sequência, caracterizando polilisogenia. Os vírus identificados pertencem a quatro famílias virais (*Myoviridae*, *Siphoviridae, Podoviridae, Inoviridae*), sendo as três primeiras pertencentes à ordem *Caudovirales* e a última pertencente à ordem *Tubulavirales*. Nos genomas dos profagos foram detectados genes de possível origem bacteriana, significando provável transferência horizontal de genes entre os vírus e os seus hospedeiros, assim como genes que codificam fatores de virulência, indiciando que estes elementos potencialmente influenciam no comportamento e *fitness* da bactéria hospedeira. Juntos, os resultados indicam que existe uma ampla diversidade de profagos nos genomas de *Xanthomonas* spp., e dependendo da atividade e do contexto genômico, podem conferir plasticidade fenotípica a estas bactérias que auxilia no seu modo de vida.

Palavras-chave: Bacteriófagos. PHASTER. Plasticidade genômica. Fitness bacteriano.

SUMMARY

GENERAL INTRODUCTION 10
REFERENCES 17
ABSTRACT 23
INTRODUCTION 24
RESULTS 26
Detection of prophage-like elements in complete genomes of *Xanthomonas axonopodis*, *X. campestris* and *X. citri* species 26
Comparative genomic of *Xanthomonas* prophage-like elements 28
Prophage-like elements genes that may interact with the host 30
DISCUSSION 31
METHODS 35
Detection of prophage-like elements in complete genomes of *Xanthomonas axonopodis*, *X. campestris* and *X. citri* genomes 35
Genomic annotation and organization of prophage-like elements 35
Grouping intact prophage-like elements in clusters and phylogenetic analysis 36
Identification of putative genes that may influence bacteria ecology and fitness 37
FIGURE LEGENDS 38
REFERENCES 41
SUPPLEMENTARY TABLES 45
FIGURES 82

GENERAL INTRODUCTION

Bacteriophages, or phages, are viruses that infect prokaryotes, being the organisms most abundant on the planet, with an estimated 10^{31} particles worldwide (SUTTLE, 2005). Since bacteriophages co-exist with bacteria and archaea, they can be found in every explored biome, from soil, animal cells to the global ocean. In some ecosystems, the bacteriophage number can overtake approximately tenfold the number of bacteria (DION et al., 2020). Many studies highlighted the biological importance of prophages as active agents in modulating the environment. For instance, the heterotopic bacteria in the ocean represent a significant proportion of active biomass in this environment. Therefore, their mortality by phage and the consequent liberation of cellular debris serve as an essential source of nutrients, such as carbon, nitrogen, phosphorus and iron, and other elements that will be made available to the microbial food web (WILHELM and SUTTLE, 1999). Another example of the importance of bacteriophages is their relationship with the human gut, where their presence is named phageome. They can influence human health once they modulate the composition and function of the gut microbiome (MANRIQUE et al., 2016).

The size of phage genomes ranged from 2.3 kb to more than 300 kb packaged on various types of capsids. Some particles show a tail as the phages belonging to the *Caudovirales* order (the most common phages). The phages from the *Tubuvirales* order show a filamentous capsid. Other groups have shown polyhedral capsid (*Corticoviridae, Cystoviridae, Leviviridae, Microviridae,* and *Tectiviridae* families). Lastly, a unique group shows a pleomorphic capsid (*Plasmaviridae* family) (HATFULL, 2008, DION et al., 2020).

A bacteriophage can have four main life cycle types: chronic, pseudolysogenic, lysogenic and lytic (WEINBAUER, 2004). The chronic infection occurs when there is no cell lysis after the replication and assembly of viral particles. The virus progeny is released by extrusion or budding, as for some filamentous viruses and phages infecting *Mycobacterium* spp. (BAESS, 1971). The pseudolysogenic cycle is unstable when there is a lack of nutrients, and the bacterial host cannot replicate its genetic material and synthesize its proteins. The phage genome cannot enter the lytic cycle or the lysogenic state (integrated into the host genome) until better conditions are restored. During this period, the phage remains in a non-integrated form as an episome (FEINER et al., 2015).

The lytic cycle is regulated by the expression of viral genes responsible for directing the host's synthesis machinery to viral replication, synthesizing viral proteins and nucleic acids.

When the process is over, cell lysis is promoted, releasing viral particles into the environment that, at this point, are named virions. Holins and hydrolases regulate the cell lysis process, encoded by the viral genome and resulting in the host cell's death (ACKERMANN, 2001; BROCK et al., 2003). When observing the coculture of lytic phage and the bacteria *in vitro*, the lysed cells are indicated by a clear zone, called lysis plaque, in the middle of a layer of growing bacteria in a culture plate (GALLET et al., 2011). From lysis plaque, it is possible to isolate viral strains that can potentially be used as a tool for biological control of diseases caused by their bacteria host in a phage therapy strategy (AHERN et al., 2014).

On the other hand, some phages can integrate into the host genome, known as the lysogenic cycle. The phages with this feature are called temperate phages and they are maintained and integrated into the bacterium chromosome, multiplying with cell multiplication (CANCHAYA et al., 2004). The phage's DNA remains inactive during this cycle, except for some regulatory genes. These regulatory genes are necessary to maintain the phage in a dormant state by repressing the lytic genes. Nevertheless, stressful conditions that cause DNA damage may lead to the activation of the lytic cycle. In this process, the lytic genes are activated, which causes the disintegration of the virus of its host genome and activates its replication, DNA packaging, and phage particle assembly within the cell, which consequently releases virions and promotes cell lysis. When integrated into its host chromosome, the temperate phages are named prophages. (DAVIES et al., 2016).

Prophages express genes that contribute to the maintenance of the host. Among these genes, there are those responsible for the repression of the lytic cycle, for example, the CI protein present in ʎ phage (PTASHNE, 1992). This condition led to the establishment of mutualism in many cases. Prophages develop the ability to cohabit with their hosts, and this coexistence can lead to the development of patterns that modulate their biology, behavior or ecology (HOWARD-VARONA et al., 2017). In some cases, the integrated phage becomes "domesticated" due to the loss of some genes. For example, a phage can lose the genes responsible for viral particle formation. These elements become cryptic or defective phages containing viral elements (BOBAY et al., 2013). Such cryptic phages may benefit bacterial hosts to remain integrated into the chromosomes. For example, the genomes of *Clostridium difficile* contain phage tail-like particles named R-type bacteriocin particles. These particles are produced when the SOS response system of these strains is activated and can inhibit bacterial growth. For that reason, the R-type bacteriocin particles confer a fitness advantage by killing non-lysogenic strains and reducing competition (GEBHART et al., 2012).

The lysogeny can benefit the hosts, therefore establishing a phage-host symbiosis. For example, the ASPE-2 phage infects *Hamiltonella defense*, which infects *Aphidius ervi*, an aphid species. The ASPE-2 phage encodes a toxin that kills developing wasp larvae that attack the aphid, protecting both aphids and bacteria, consequently ensuring the virus's survival (OLIVER et al., 2009). Another example of this type of relationship occurs when the temperate phages enter in lytic cycle, they can modify the surrounding community by lysogenizing other cells. For instance, the *Escherichia coli* phage ʎ, can enter into in lytic cycle and be released into the environment, where it can infect other cells, including susceptible non-lysogenic strains. These non-lysogenic strains are competitors, and at the first moment, the idea is to eliminate these cells. However, when this process occurs with a large amount of this virus, it integrates into that bacterial genome, making that cell immune to further infections by the ʎ virus (GAMA et al., 2013).

Another beneficial symbiotic interaction between phage and host is the lysogenic conversion, in which phages encode proteins that may increase the host's environmental fitness (FEINER et al., 2015). For example, the phage produces toxins that can increase the host virulence, such as ϕRSS1. This bacteriophage was described as a parasite of *Rasltonia solanacearum*, the causative agent of bacterial wilt in many species of plants. *R. solanacearum* strains infected by this phage proved more virulent on host plants. This fact occurred due to some virulence factors being increased in infected cells, such as twitching motility, the synthesis of extracellular polysaccharide (EPS), and an increase in the expression of the global virulence factor *phcA* (ADDY et al., 2012). Furthermore, the phage may lead to the establishment of superinfection exclusion. This process confers “immunity” to the bacterium against other phages and superinfection by the same phage (BONDY-DENOMY and DAVIDSON, 2014). There are different paths to establish the superinfection exclusion, but the most common ways imply modifications on cell envelope components or cellular surfaces (BONDY-DENOMY and DAVIDSON, 2014). To exemplify, a study performed by Newton et al., 2001 revealed a mechanism that involves modification of the O-antigen present on the lipopolysaccharide layer of *Pseudomonas aeruginosa* cells. A bacteriophage called D3 lysogenize the strain PAO1 making changes in the O-antigen 05, changing its serotype. This modification inhibits the superinfection by other phages and the D3 phage itself since the antigen is necessary for binding these elements and, consequently, their infection.

Further interaction between prophage and host is termed active lysogeny. This process involves reintegrating the prophage genome, which integrates within bacterial

functional genes or regulatory regions. The expression of these genes is controlled by integration and excision of the viral element and is divided into two categories: the reversible and the non-reversible active lysogeny. After the phage excision, it can be reintegrated into the bacterial chromosome or be permanently lost (FEINER et al., 2015). An example of reversible active lysogeny involves the φ10403S-prophage, a phage that maintains itself integrated within the *Listeria monocytogenes* genome, a human pathogen. This phage is integrated within the *comK* gene, the master activator gene of the DNA uptake competence (Com) system. Usually, *comK* is interrupted by this prophage. However, when the bacterium infects the human macrophage cells, the φ10403S-prophage is excised, promoting its expression, which supports the bacterium's evasion from macrophage cells to the cytoplasm to finish its growth and infection. This event is considered reversible active lysogeny because the prophage reintegrates into *L. monocytogenes* after excision and replication chromosome (PASECHNEK et al., 2020). The non-reversible active lysogeny can be exemplified by the *Bacillus subtilis* sporulation, mediated by *skin* phage. This phage integrates into the *sigK* gene from *B. subtilis*. Under such conditions, this gene is not expressed. In adverse conditions that cause damage to the cell, mainly to the cellular DNA, the bacterium enters a sporulation process. In this process, the phage is excised, and the *sigK* gene is activated, which is necessary for the final phase of cell differentiation on sporulation (EICHENBERGER et al., 2004). After this, the *skin* element is lost along with the death of the mother cell, whereas a new *sigK* gene is interrupted by a new *skin* phage within the endospore formed (ERRINGTON, 1993). Therefore, bacteriophages may affect the host cell in many ways, influencing their behavior, ecology and evolution.

Among the prokaryotes that harbor bacteriophages, there are some studied phytopathogenic bacteria. These cells are known for causing a wide range of diseases in different plant species, which causes adverse economic. Symptoms such as blights, spots, tissue rots and cankers are observed that lead to factors that hinder the growth and development of the plant (KANNAN et al., 2015). Every year, it is estimated that the damage caused by these cells to the food production, collectively, ranges about $1 billion worldwide (MANSFIELD et al., 2012; KANNAN et al., 2015).

Phytopathogenic bacteria are also very infected by bacteriophages, and, similar to the examples cited in this study, they can influence their behavior, ecology and evolution. Various viral families have been described as infecting plant pathogenic bacteria, such as *Myoviridae*, *Siphoviridae*, *Podoviridae*, dsDNA viruses, and *Inoviridae*, comprising ssDNA filamentous viruses (VARANI et al., 2013). The phages mentioned above impact the host

genome and biology in multiple ways. For example, prophage-like elements detected in *Rasltonia* spp. contain genes possibly related to antibiotic resistance, bacterial virulence and niche adaptation that may contribute to the host's environmental fitness (GONÇALVES and SOUZA et al., 2021).

Xanthomonas is a genus of Gram-negative obligate aerobic bacteria belonging to the class *Gammaproteobacteria*. It is generally rod-shaped with a unique polar flagellum, and its optimal growth temperature ranges from 25 to 30 ºC (AN et al., 2019). This genus comprises more than 35 species, primarily pathogenic, causing several diseases in more than 400 plant species (RYAN et al., 2011). A feature of this group is the production of xanthomonadin, a halogenated aryl-polyene, a water-insoluble yellow pigment bound to the cellular membrane (POPLOWSKY et al., 1993). Studies indicate that xanthomonadin has a role in preventing photo damage, which enhances the *Xanthomonas* survival on plant leaf surfaces in the epiphytic stage, as described below (HE et al., 2020). It is worth noting that not all *Xanthomonas* species can produce this pigment, such as *X. campestris* pv. *viticola* and *X. axonopodis* pv. *manihotis* (TIMILSINA et al., 2020). The bacteria belonging to the *Xanthomonadales* order are one of the most critical members of soil environments and rhizosphere, reaching two to seven percent of the bacteria microbiome community (AN et al., 2019). This phenomenon occurs because these microorganisms live part of their life cycle outside their host, accessible on the soil or in lesions of fallen leaves (ZHAO et al., 2002). The process of plant infection is divided into the epiphytic and endophytic stages. In the epiphytic phase, the bacteria remain on the leaf surface. The endophytic stage occurs when the bacterium is inside the plant tissues and colonizes the entire plant. After the colonization and multiplication inside the host, bacterium re-emerge on the leaf surface to disperse to a new host, primarily by rain or wind (MOREIRA et al., 2015). During the epiphytic infection, *Xanthomonas* is exposed to the outer part of the plant leaf, being vulnerable to biotic and abiotic stresses. For this reason, *Xanthomonas* spp. have developed protection mechanisms to increase their survival, such as biofilm formation. This structure is responsible for adhesion to the host surface and is composed of extracellular nucleic acids, proteins and exopolysaccharides (EPS) (CASTIBLANCO and SUNDIN, 2016). In *Xanthomonas* spp., this extracellular matrix is also composed of xanthan gum, a non-toxic heteropolysaccharide commonly used in pharmaceutical, cosmetics and food industries, due to its water solubility, and stability and high viscosity (CROSSMAN and DOW, 2004).

The *Xanthomonas* spp. infect many crops worldwide, causing significant economic losses. For example, the bacterial blight in rice is caused by *X. oryzae* pv. *oryzae* and *X. oryzae*

pv. *oryzicola*; banana *Xanthomonas* wilt, caused by *X. campestris* pv. *musacearum*; tomato and pepper can be infected by four *Xanthomonas* spp., causing bacterial spot disease; the bacterial leaf blight in *Eucalyptus* caused by *X. axonopodis* pv. *eucalyptorum*, and citrus canker in several citrus plants by *X. citri* subsp. c*itri* (TIMILSINA et al., 2020).

There are described in the literature some *Xanthomonas* spp. prophages that influence their host's life. For example, the phage Cflt, a filamentous phage from *X. citri* pv. *citri* represents the concept of superinfection exclusion very well once the *X. campestris* cells infected by this phage become immune to a new Cflt infection. This phage produces turbid plaques and does not affect the growth of infected cells, besides having a low growth phage rate. The Cflt phage was an important phage discovery since, before this study, there were reports that only double-strand DNA phages could perform the lysogenic cycle (KUO et al., 1987). The phage XacF1 was found in *X. citri* pv. *citri*, an important causative agent of citrus canker disease. This phage belongs to the *Inoviridae* family (filamentous phage) and has a detrimental effect on its host. The infected cell presents physiological modifications, including reduced motility and a decrease in the growth rate. The biofilm polysaccharide production is negatively affected, and an extreme reduction in host virulence was observed. Thus, this phage can be used as a biological control tool for citrus canker, once this element may decrease plant infection by *X. citri* (AHMAD et al., 2014).

In addition to prophages that infect *Xanthomonas* spp.*,* some lytic phages are also described in the literature, with the primary objective of developing techniques that use these elements in biological control as an alternative to the use of agrochemicals. For instance, the XacN1, a jumbo phage belonging to the *Myoviridae* family, was described as infecting *Xanthomonas citri*. It is a phage with a broad host specificity between the *X. citri* strains, which could be an exciting feature of citrus canker biocontrol (YOSHIKAWA et al., 2018). The KΦ1 phage was identified and characterized in *X. euvesicatoria*, the causative agent of bacterial spot of pepper. This lytic phage has a broad host range showing a promising potential to be used as a biocontrol agent (GAŠIĆ et al., 2018). Furthermore, the Xanthomonas virus XC 2 is a high-specific phage that infects *Xanthomonas campestris* pv. *campestris* have a narrow host range that can be explored to control black rot in crucifers (PEREIRA et al., 2019).

The *Xanthomonas* species causes several crop losses worldwide, but little is known about their mobilome in general and how these elements can affect these strains. Phages play an essential role in many bacterial species, affecting them positively or negatively, contributing to their survival and adaptability, or leading to their mortality. Understanding phages ecology,

diversity and behavior are essential to define methods for their application in the control of several *Xanthomonas* spp. For this reason, in this work, we analyzed 200 complete genomes belonging to three species of the *Xanthomonas* genus (*X. axonopodis*, *X. campestris* and *X. citri*), detecting the presence of prophage-like elements. Also, we characterize them to further speculate on the bio/ecological significance of the probable interactions between prophages and bacteria.

REFERENCES

Addy, H. S., Askora, A., Kawasaki, T., Fujie, M., & Yamada, T. (2012). The filamentous phage φRSS1 enhances virulence of phytopathogenic *Ralstonia solanacearum* on tomato. *Phytopathology*, 102(3), 244-251. https://doi.org/10.1094/PHYTO-10-11-0277

Ackermann, H. W. (2001). Frequency of morphological phage descriptions in the year 2000. *Archives of Virology*, *146*(5), 843-857. https://doi.org/10.1007/s007050170120

Ahern, S. J., Das, M., Bhowmick, T. S., Young, R., & Gonzalez, C. F. (2014). Characterization of novel virulent broad-host-range phages of *Xylella fastidiosa* and *Xanthomonas*. *Journal of bacteriology*, *196*(2), 459-471. https://doi.org/ 10.1128/JB.01080-13

Ahmad, A. A., Askora, A., Kawasaki, T., Fujie, M., & Yamada, T. (2014). The filamentous phage XacF1 causes loss of virulence in *Xanthomonas axonopodis* pv . *citri* , the causative agent of citrus canker disease. *5(July)*, 1–11. https://doi.org/10.3389/fmicb.2014.00321

An, S. Q., Potnis, N., Dow, M., Vorhölter, F. J., He, Y. Q., Becker, A., Teper, D., Li, Y., Wang, N., Bleris, L., & Tang, J. L. (2019). Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiology Reviews*, *44*(1), 1–32. https://doi.org/10.1093/femsre/fuz024

Baess, I. (1971). Report on a pseudolysogenic mycobacterium and a review of the literature concerning pseudolysogeny. *Acta Pathologica Microbiologica Scandinavica Section B Microbiology and Immunology*, *79*(3), 428-434. https://doi.org/10.1111/j.1699-0463.1971.tb00084.x

Bobay, L. M., Rocha, E. P. C., & Touchon, M. (2013). The adaptation of temperate bacteriophages to their host genomes. *Molecular Biology and Evolution*, *30*(4), 737–751. https://doi.org/10.1093/molbev/mss279

Bondy-Denomy, J., & Davidson, A. R. (2014). When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. *Journal of Microbiology*, *52*(3), 235–242. https://doi.org/10.1007/s12275-014-4083-3

Brock, T. Dale, Madigan, M. T, Martinko, J. M, & Parker, J. (2003). Brock biology of microorganisms. 10th ed. Upper Saddle River (N.J.): Prentice-Hall.

Canchaya, C., Fournous, G., & Brüssow, H. (2004). The impact of prophages on bacterial chromosomes. *Molecular Microbiology*, *53*(1), 9–18. https://doi.org/10.1111/j.1365-2958.2004.04113.x

Castiblanco, L. F., & Sundin, G. W. (2016). New insights on molecular regulation of biofilm formation in plant-associated bacteria. *Journal of integrative plant biology*, *58*(4), 362-372. https://doi.org/10.1111/jipb.12428

Crossman, L., & Dow, J. M. (2004). Biofilm formation and dispersal in *Xanthomonas campestris*. *Microbes and infection*, *6*(6), 623-629. https://doi.org/10.1016/j.micinf.2004.01.013

Davies, E. V., Winstanley, C., Fothergill, J. L., & James, C. E. (2016). The role of temperate bacteriophages in bacterial infection. *FEMS Microbiology Letters*, *363*(5), 1–10. https://doi.org/10.1093/femsle/fnw015

Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews Microbiology*, *18*(3), 125–138. https://doi.org/10.1038/s41579-019-0311-5

Eichenberger, P., Fujita, M., Jensen, S. T., Conlon, E. M., Rudner, D. Z., Wang, S. T., Ferguson, C., Haga, K., Sato, T., Liu, J. S., & Losick, R. (2004). The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biology*, *2*(10). https://doi.org/10.1371/journal.pbio.0020328

Errington, J. (1993). *Bacillus subtilis* sporulation: Regulation of gene expression and control of morphogenesis. *Microbiological Reviews*, *57*(1), 1–33. https://doi.org/10.1128/mmbr.57.1.1-33.1993

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., & Herskovits, A. A. (2015). A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. *Nature Reviews Microbiology*, *13*(10), 641–650. https://doi.org/10.1038/nrmicro3527

Gallet, R., Kannoly, S., & Wang, N. (2011). Effects of bacteriophage traits on plaque formation. *BMC microbiology*, *11*(1), 1-16. https://doi.org/10.1186/1471-2180-11-181

Gama, J. A., Reis, A. M., Domingues, I., Mendes-Soares, H., Matos, A. M., & Dionisio, F. (2013). Temperate bacterial viruses as double-edged swords in bacterial warfare. *PloS one*, *8*(3), e59043. https://doi.org/10.1371/journal.pone.0059043

Gašić, K., Kuzmanović, N., Ivanović, M., Prokić, A., Šević, M., & Obradović, A. (2018). Complete genome of the *Xanthomonas euvesicatoria* specific bacteriophage KΦ1, its survival and potential in control of pepper bacterial spot. *Frontiers in microbiology*, *9*, 2021. https://doi.org/10.3389/fmicb.2018.02021

Gebhart, D., Williams, S. R., Bishop-Lilly, K. A., Govoni, G. R., Willner, K. M., Butani, A., ... & Scholl, D. (2012). Novel high-molecular-weight, R-type bacteriocins of *Clostridium difficile*. *Journal of bacteriology*, *194*(22), 6240-6247. htpps://doi.org/ 10.1128/JB.01272-12

Gonçalves, O. S., Souza, F. O., Bruckner, F. P., Santana, M. F., & Alfenas-Zerbini, P. (2021). Widespread distribution of prophages signaling the potential for adaptability and pathogenicity evolution of *Ralstonia solanacearum* species complex. *Genomics*. *113*(January), 992–1000. https://doi.org/10.1016/j.ygeno.2021.02.011

Hatfull, G. F. (2008). Bacteriophage genomics. *Current Opinion in Microbiology*, *11*(5), 447-453. https://doi.org/10.1016/j.mib.2008.09.004

He, Y. W., Cao, X. Q., & Poplawsky, A. R. (2020). Chemical Structure, Biological Roles, Biosynthesis and Regulation of the Yellow Xanthomonadin Pigments in the Phytopathogenic Genus *Xanthomonas*. *Molecular Plant-Microbe Interactions*, *33*(5), 705-714. https://doi.org/10.1094/MPMI-11-19-0326-CR

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME Journal*, *11*(7), 1511–1520. https://doi.org/10.1038/ismej.2017.16

Kannan, V. R., Bastas, K. K., & Devi, R. S. (2015). 20 Scientific and Economic Impact of Plant Pathogenic Bacteria. *Sustainable approaches to controlling plant pathogenic bacteria*, 369. https://doi.org/10.1201/b18892

Kuo, T. T., Lin, Y. H., Huang, C. M., Chang, S. F., Dai, H., & Feng, T. Y. (1987). The lysogenic cycle of the filamentous phage Cflt from *Xanthomonas campestris* pv. *citri*. *Virology*, *156*(2), 305-312. https://doi.org/10.1016/0042-6822(87)90410-7

Manrique, P., Bolduc, B., Walk, S. T., van der Oost, J., de Vos, W. M., & Young, M. J. (2016). Healthy human gut phageome. *Proceedings of the National Academy of Sciences*, *113*(37), 10400-10405. https://doi.org/10.1073/pnas.1601060113

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., ... & Foster, G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular plant pathology*, *13*(6), 614-629. https://doi.org/10.1111/j.1364-3703.2012.00804.x

Moreira, L. M., Facincani, A. P., Ferreira, C. B., Ferreira, R. M., Inês, M., Ferro, T., Gozzo, F. C., Cezar, J., Oliveira, F. De, Ferro, J. A., & Soares, M. R. (2015). Chemotactic signal transduction and phosphate metabolism as adaptive strategies during citrus canker induction by *Xanthomonas citri*. *Functional & integrative genomics*, *15*(2), 197–210. https://doi.org/10.1007/s10142-014-0414-z

Newton, G. J., Daniels, C., Burrows, L. L., Kropinski, A. M., Clarke, A. J., & Lam, J. S. (2001). Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. *Molecular Microbiology*, *39*(5), 1237–1247. https://doi.org/10.1046/j.1365-2958.2001.02311.x

Oliver, K. M., Degnan, P. H., Hunter, M. S., & Moran, N. A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, *325*(5943), 992-994. Htpps://doi.org/ 10.1126/science.1174463

Pasechnek, A., Rabinovich, L., Stadnyuk, O., Borovok, I., Sigal, N., Herskovits, A. A., Pasechnek, A., Rabinovich, L., Stadnyuk, O., Azulay, G., Mioduser, J., Argov, T., & Borovok, I. (2020). Article Active Lysogeny in Listeria Monocytogenes Is a Bacteria-Phage Adaptive Response in the Mammalian. *CellReports*, *32*(4), 107956. https://doi.org/10.1016/j.celrep.2020.107956

Pereira, F., Xavier, S., Prieto, F., Rafael, B., Rezende, R. De, Marcus, P., Vidigal, P., & Alfenas, P. (2019). Biological and molecular characterization of a bacteriophage infecting *Xanthomonas campestris* pv . *campestris*, isolated from brassica fields. *Archives of Virology*, *164*(7), 1857–1862. https://doi.org/10.1007/s00705-019-04263-4

Poplowsky, A. R., Kawalek, M. D., & Schaad, N. W. (1993). A xanthomonadin-encoding gene cluster for the identification of pathovars of *Xanthomonas campestris*. *Molecular Plant Microbe Interactions*, *6*, 545-545. https://doi.org/10.1094/MPMI-6-545

Ptashne, M. (1992). A genetic switch: phage and higher organisms. *GeorgetowUniversity*. https://doi.org/10.822/545304

Ryan, R. P., Vorhölter, F. J., Potnis, N., Jones, J. B., Van Sluys, M. A., Bogdanove, A. J., & Dow, J. M. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nature Reviews Microbiology*, *9*(5), 344-355. https://doi.org/10.1038/nrmicro2558

Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356-361. https://doi.org/10.1038/nature04160

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., Goss, E. M., & Jones, J. B. (2020). *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology*, *18*(8), 415–427. https://doi.org/10.1038/s41579-020-0361-8

Weinbauer, M. G. (2004). Minireview Are viruses driving microbial diversification and diversity ? *Environmental Microbiology, 6*, 1–11. https://doi.org/10.1046/j.1462-2920.2003.00539.x

Wilhelm, S., & Suttle, C. A. (1999). Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience*, *49*(10), 781-788. https://doi.org/10.2307/1313569

Varani, A.M., Monteiro-Vitorello, C.B., Nakaya, H.I., and Sluys, M.-A. Van (2013). The Role of Prophage in Plant-Pathogenic Bacteria. *Annual Reviews Phytopathology*. 51: 429-51. https://doi.org/10.1146/annurev-phyto-081211-173010

Yoshikawa, G., Askora, A., Blanc-Mathieu, R., Kawasaki, T., Li, Y., Nakano, M., ... & Yamada, T. (2018). *Xanthomonas citri* jumbo phage XacN1 exhibits a wide host range and high complement of tRNA genes. *Scientific reports*, *8*(1), 1-10. https://doi.org/10.1038/s41598-018-22239-3

Zhao, Y., Damicone, J. P., & Bender, C. L. (2002). Detection, survival, and sources of inoculum for bacterial diseases of leafy crucifers in Oklahoma. *Plant disease*, *86*(8), 883-888. https://doi.org/10.1094/PDIS.2002.86.8.883

CHARACTERIZATION OF PROPHAGE-LIKE ELEMENTS IN PLANT PATHOGENIC *Xanthomonas* SPECIES

Ferreira, Nataly Figueiredo; Morgan, Túlio; Souza, Flávia de Oliveira; Rezende, Rafael Reis; Mayrink, Thamylles Thuany; Alfenas-Zerbini, Poliane. Characterization of prophage-like elements in plant pathogenic *Xanthomonas* species. *Molecular Microbiology, to be submitted.*

CHARACTERIZATION OF PROPHAGE-LIKE ELEMENTS IN PLANT PATHOGENIC *Xanthomonas* SPECIES

Ferreira, Nataly Figueiredo[1]; Morgan, Túlio[1]; Souza, Flávia de Oliveira[1]; Rezende, Rafael Reis[1]; Mayrink, Thamylles Thuany[1]; Alfenas-Zerbini Alfenas-Zerbini, Poliane[1*]

[1]Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, CEP: 36570-000; Brazil.

*Corresponding author: Poliane Alfenas-Zerbini; e-mail: palfenas@ufv.br; phone: +55 31 3612-2437
Keywords: PHASTER; Lysogeny; bacteriophages; phytopathogenic bacteria.
Number of figures: 11
Supplementary Tables: 9

ABSTRACT

Bacteriophages are viruses that infect prokaryotes that are present in diverse environments. Prophages are bacteriophages with genome into the host genome and have been considered part of bacteria mobilome. For instance, bacteriophages may encode proteins that increase bacteria toxicity and virulence, confer genetic variability, and facilitate horizontal gene transfer, improving bacterial fitness. The *Xanthomonas* genus harbors many phytopathogenic bacteria that impair many crops worldwide. However, there is a scarcity of studies that describes bacteriophage distribution and diversity within this genus. Given this, our study aims to identify and characterize the presence of prophage-like elements in three species belonging to the *Xanthomonas* genus: *X. axonopodis, X. campestris* and *X. citri*. We found prophage-like sequences in 98% of the analyzed genomes, whereby more than one prophage-like was detected within the same bacterial genome, event denominated as polilysogen. The prophages-like from *Xanthomonas* spp. belonged to four viral families (*Myoviridae*, *Siphoviridae, Podoviridae, Inoviridae*), three families belonging to the *Caudovirales* order and one belonging to the *Tubulavirales* order (*Inoviridae*). Furthermore, many prophages-like harbored genes of putative bacterial origin, indicating horizontal gene transfer between the virus and host, such as genes encoding virulence factors. Taken together, these results indicated that prophages are widespread in *Xanthomonas* spp., influencing the genome diversity and fitness of these bacteria.

Keywords: Bacteriophage. PHASTER. Genome plasticity. Bacteria fitness.

INTRODUCTION

Viruses, the most abundant biological entities in the world, are obligate intracellular parasites that infect all existent types of life, from any animal cellular to bacteria, archaea and even other viruses (BERJÓN-OTERO et al., 2019). Bacteriophages, or phages, are viruses that infect prokaryotic cells. It is estimated that there are 10^{31} phage particles on Earth, the most significant number of organisms (CZAJKOWSKI, 2019). They are ubiquitous, infected bacteria and archaea, even in extreme environments. Therefore, phages are found in diverse locations, from water and soil to gastrointestinal systems and integrated into their host genomes as prophages (BONDY-DENOMY et al., 2014; LEMIRE et al., 2018).

Prophages may influence the lifestyle of their host, leading to modifications in their fitness, behavior and ecology (BONDY-DENOMY and DAVIDSON, 2014; NANDA et al., 2015). For instance, phage integration and disintegration in the host genome regulate important genes, converting nonpathogenic to pathogenic bacteria through the transfer of toxin-encoding genes, which lead to changes in host phenotype (BRUSSOW et al., 2004). The presence of these integrated elements contributes to the genetic variability of the host's genome, which may confer advantages for its survival and improve its adaptation (CZAJKOWSKI, 2019).

In phytopathogenic bacteria, prophages display different roles. For example, prophage activity may be correlated with strain differentiation and genome rearrangements in *Xanthomonas* and *Xyllella* (VARANI et al., 2013). Under conditions that simulate plant infection, some bacteria begin the prophage gene transcription. In addition, some bacteria have prophages directly involved in disease development, such as the genera *Pectobacterium, Pseudomonas, Ralstonia* and *Streptomyces* (NAKAYAMA et al., 1999; ADDY et al., 2012; BALTZ, 2012). Some studies demonstrated that phages might affect their host's virulence, fitness, ecology and behavior (CZAJKOWSKI, 2019). However, studies still need to be developed better to understand the relationship between prophages and phytopathogenic bacteria.

Xanthomonas is a Gram-negative bacteria genus that harbors approximately 35 species able to infect more than 400 plant species. (RYAN et al., 2011), including many crops with economic importance. For example, *X. oryzae* infects rice, *X. campestris* infects brassicas, *X. axonopodis* infects eucalypti, and *X. citri* infects citrus fruits (TIMILSINA et al., 2020), which may impair food security and local economies. Some prophages infect *Xanthomonas* spp. and influence their fitness, such as phage XacF1 of *X. campestris* which decreases growth rate and reduces bacterial motility (AHMAD et al., 2014). Some lytic phages have been

described and explored to be used as biological control of the caused disease, such as the phage *Xanthomonas* virus XC 2 that infects *X. campestris* (PEREIRA et al., 2019).

There are few described *Xanthomonas* prophages; thereby, little is known about how these elements affect their host's lifestyle. To develop phage-based biological control of these bacteria is essential to understand the biological features of *Xanthomonas* phages, besides their interaction and influence on the host. In this study, we analyzed prophages-like present within the genomes of *X. axonopodis, X. campestris* and *X. citri* available in the GenBank database (NCBI) and characterized the prophage-like sequences using computational tools. As a result, we described genomic features that had not yet been made for bacteria of this genus — this enables further clarification on the interactions between prophages and their hosts.

RESULTS

Detection of prophage-like elements in complete genomes of *Xanthomonas axonopodis*, *X. campestris* and *X. citri* species

We analyzed the presence of prophage-like elements in genomes of three species belonging to the *Xanthomonas* genus (*X. axonopodis*, *X. campestris* and *X. citri*) (Table S1). A total of 200 complete genomes of *X. axonopodis*, *X. campestris* and *X. citri* were downloaded, and the completeness was determined with BUSCO. Within these 200 complete genomes, 81 sequences that showed more than 96% completeness (Table S2) were subjected to PHASTER to detect the prophage-like elements. A total of 203 putative complete and incomplete prophage-like sequences were detected. A manual inspection demonstrated that 116 (57.1%) sequences could be classified as putative complete phage genome sequences, called hereafter "intact prophages-like" and 87 (42.9%) sequences were classified as defective (incomplete) prophages-like elements. One of the 116 putative intact sequences is integrated into one *X. axonopodis* genome of the two genomes analyzed in this study. In *X. campestris,* we found ten intact prophages-like in 13 genomes, and 105 intact elements were identified in 105 *X. citri* genomes. Of the 87 defective prophage-like elements, one belongs to the *X. axonopodis* genome, 14 to *X. campestris* genomes, and 71 to *X. citri* genomes (Figure 1). Only two *Xanthomonas* genomes, *X. campestris* pv. *badri* strain NEB122 (NZ_CP051651.1) and *X. citri* pv. *citri* strain LM199 (NZ_CM007619.1) did not contain any prophage-like sequences. All prophage-like sequences found were located in chromosomes of *Xanthomonas* species.

To facilitate the description of prophages-like through the text, we nominate them according to the *Xanthomonas* species they were found. The unique intact prophage-like detected in *X. axonopodis* strains was named Xax_01. The prophages-like detected in *X. campestris* were named Xc_01 to Xc_10, and the prophages-like detected in *X. citri* were named Xci_01 to Xci_60 (Table S3).

We did not detect prophages-like with high identity infecting different bacterial species. However, the same viral sequence was found in different strains of the same species, such as Xci_01, which appears in thirteen different strains of *X. citri*, Xci_02 in seven different strains, Xci_06 in eight different strains, and Xci_07 in five different strains. Also, prophage-like sequences Xci_04, Xci_05, Xci_09 and Xci_15 were detected in four different strains, Xci_16 in three different strains, and Xci_03, Xci_08, Xci_10, Xci_12 and Xci_31 each in two different strains (Figure 2).

Polylisogeny was detected in twenty-seven genomes in *X. citri* and one genome in *X. campestris*, varying from two to four elements in the same genome. The strains with the most significant number of elements belonged to *X. citri* species, where the two assemblies of strain 8ra (Assembly accessions: GCA_007567665.1 and GCA_001854145.2) harbor the viruses Xci_02, Xci_06, Xci_09 and Xci_11, and the strain CFBP 2526 harbor the viruses Xci_05, Xci_06, Xci_11 and Xci_38. Viruses integrated into the same genome are not necessarily from the same family, such as the polylisogeny detected in many *X. citri* genomes, which comprised elements belonging to different families. Furthermore, *X. campestris* strain ATCC_33913 harbored two distinct prophages-like, the Xci_06 belonging to the *Inoviridae* family and Xc_07 belonging to the *Myoviridae* family. Also, the *X. campestris* CN12 harbored two different elements that belonged to the *Inoviridae* family (Xc_09 and Xc_10) (Figure 2).

The sequence size of intact prophages-like ranged from 5.0 to 70.5 kb. On the other hand, defectives prophages-like varied from 5.0 to 50.0 kb (Figure 3. A, Table S3). The average presence of intact prophage-like in all the *Xanthomonas* genomes analyzed in this study corresponds to 7.38%., meanwhile, the average number of defective prophages corresponds to 4,52% of the analyzed sequences. The proportion of these elements, according to its respectives hosts, are presented in supplementary figures 1. A, B and C. The *X. citri* was the species with the most significant number of prophages-like found, followed by *X. campestris* and *X. axonopodis* (Figure 3. B). The host GC content had a slight variation from 64 to 65 percent; meanwhile, the GC content of most intact prophages-like ranged from 59 to 64 percent, lower than their host. On the other hand, six elements had higher GC content (Figure 3. C, Table S4).

The 116 intact prophages-like detected were classified into four families: *Inoviridae* (38), *Myoviridae* (48), *Siphoviridae* (29) and *Podoviridae* (one) (Figure 3. D and E, Table S3). After eliminating redundancies, 59 elements were considered unique, showing less than 100% of identity with each other. Thus, 31 different intact prophages-like of the *Inoviridae* family, 17 of the *Myoviridae* family, 10 of the *Siphoviridae* family and one of the *Podoviridae* family (Figure 3. F) composed the non-redundant set of prophages-like sequences.

Aiming to identify the global diversity of *X. axonopodis, X. campestris* and *X. citri* prophages-like, we analyzed the geographical distribution of these sequences based on the origin of the host bacteria strain (Figure 4. A and B). The members of the *Myoviridae* family were predominantly found in the USA (America), South Korea (Asia), and France (Europe). Elements of the *Siphoviridae* family also have a wide distribution across the continents, except for Oceania, where they are not present. The most extensive number of elements belonging to

this family was found in the USA (America), followed by China (Asia). The only member of the *Podoviridae* family identified in this work was found in Africa. This study's only viral element in Oceania belongs to the *Inoviridae* family (Figure 4. A). In the prophage-like from the USA, the myoviruses correspond to 43,33%, followed by 30% of inoviruses and 26,66% of siphoviruses. In South Korea, the prophage-like comprises 7.05% of myoviruses, 11,76% of siphoviruses, and 17,64% of inoviruses. In France, 58,82% of myoviruses were identified, 35,29% of siphoviruses and 5,88% of inoviruses. Already in Brazil, 50% belong to myoviruses, 12,5% belong to siphoviruses, and 37,5% belong to inoviruses. In Sudan, 50% of the viruses belong to the *Myoviridae* family, 33,33% of the *Siphoviridae* family and 16,66% of the *Inoviridae* family. Reunion, an island in the east of the African continent, has the only presence found in this study of a prophage-like belonging to the *Podoviridae* family, which makes up 20% of the total value of viruses found in the African continent. *Siphoviridae* includes 60% and *Myoviridae* 20% of the remaining viruses. In Burkina Faso, 75% of prophages-like belong to siphoviruses and 25% to inoviruses. Last, Thailand showed 100% of myoviruses, and New Zealand showed 100% of inoviruses (Figure 4. B). This percentage was not considered for the countries where only one was found.

To verify the integration position of each element into the genome host, we adjusted all host genomes into oriC. The genomic coordinates of the integration site in *X. citri* ranged according to the prophage-like, although some prophages-like were detected in a similar position in different host genomes (Figure 5. A). The genomic position of *X. citri* prophages-like is available in Table supplementary 9. In *X. campestris*, the integration pattern is similar for all prophages-like; even the elements did not share high sequence identity, except Xci_07. All elements have been integrated into the region, varying from 2300 kb to 2700 kb, but Xci_07 is integrated from 3500 kb to 3600 kb (Figure 5. B). Although the genomic position was similar, the predicted functionality related to those regions varied, with prophages-like integrating near genes coding for transcriptional regulators, tRNAs, hypothetical proteins, oxidoreductases, insertion elements and enzymes (Table supplementary 7).

Comparative genomic of *Xanthomonas* prophage-like elements

We analyzed each virus's gene content and genomic organization to investigate the genomic diversity of *Xanthomonas* spp. prophages-like. As expected, each phage family showed a specific genomic organization. Genomes from the *Inoviridae* family presented a marked variability in gene content and synteny, exhibiting high genetic mosaicism (Figure 6.

A, B, C and S. 1). In this case, the syntenic analysis were divided into trhee main groups, that showed the greatest similarity to each other. This strategy allowed a better visualization of which group of prophage-like sequences have the highest percentage of identity between them. In the *Myoviridae* family, 15 of 17 prophage-like genomes had a similar gene content (Figure S. 3) and synteny conservation (Figure 7. A). The synteny of the *Siphoviridae* family showed that the variability of the elements was less conservated when compared with members of the *Myoviridae* family. These viruses presented more variability in their genomic organization, and for that reason, the synteny analysis was also divided into groups containing the sequences with the highest percentage of similarity, showing us which prophages-like are most similar within the results found for this family (Figure 7. B).

To establish comparative genomic analysis within viral identified families and verify the variability between them, we aligned the complete nucleotide sequences of intact prophage-like elements. This method allows viewing and studying the conservation between the sequences diversity and is a first step to further analysis, such as phylogenyand SDT analysis, to show whether these sequences have low or high evolutionary divergence. The alignments were used to heat map each viral family (*Inoviridae*, *Myoviridae* and *Siphoviridae*) (Figure S.6). With that, we observe the formation of different clusters later supported by phylogeny and SDT analyses (Figures 9. A, B and 10). The relationship between the heatmap, phylogenic and SDT analysis serves to show us the level of evolutionary divergence within viral families in this study. With this, we were able to have a broader view of how these prophage-elements are related to each oder within their respective families.

Data were plotted on violin plot to observe the genetic variability between the elements belonging to the *Inoviridae, Myoviridae* and *Siphoviridae* families. The chance of finding new prophages is greater when there is greater genetic variability within a group. Prophages-like from the *Myoviridae* family showed less variability than the *Siphoviridae* and *Inoviridae* families (Figure 8). Although fewer prophages-like belonging to the *Siphoviridae* family were detected, they showed considerable variability, ranging from 100% to 15% of identity share among them. The *Inoviridae* family was the most diverse and abundant in the sampled *Xanthomonas* spp. genomes, with most inoviruses sequences showing approximately 20% to 50% of identity. Phylogeny and SDT analysis were made to understand more clearly the evolutionary patterns within each of the viral families founded (Figure 9. A and B). The viruses were classified evolutionarily using molecular phylogenetics from amino acid sequences of large terminase to myoviruses and siphoviruses, a much-conserved bacteriophage

protein from these groups. This analysis for the *Myoviridae* family shows that Xci_52 and Xci_53 are very related viruses and, together with Xci_12, formed a clade with great affinity. Xci_02, Xci_05 and Xci_40 formed another cluster with high identities, such as Xci_06, Xci_07 and Xci_04, and Xci_57 and Xci_26 (Figure 9. A).

Meanwhile, viruses from the *Siphoviridae* family showed that their terminase separated them into two defined clusters. The first covers the most significant number of elements and is formed by Xci_01 and Xci_17, which are very close, Xci_43, Xci_10, Xci_45 and Xci_46. Xci_11 and Xc_45 form the second cluster, viruses that share a high identity percentual terminase (Figure 9. B). We used a different approach to *Inoviridae* family analysis, where the entire genome was used to make the phylogeny. As indicated earlier, in this study, the inoviruses have high variability; they are not well conserved between them. SDT and phylogenetic analysis confirm these claims that show the heterogeneous distribution of these elements. In SDT analysis, we observe that Xc_01 and Xci_55 have low sequence identity with all the other elements, the outgroup. This result corroborates with the phylogeny, where they did not group with any other elements (Figure 10).

Prophage-like elements genes that may interact with the host

We identified genes that putatively can be responsible for modifications in host behavior or maybe the result of horizontal transfer of genes (HTG). Concerning proteins that may have some influence on the host virulence, we found three types of toxin-antitoxin (TA) (Figure 11. A). Transcription regulators were the most common bacterial genes acquired by prophages-like by HTG, followed by peptidases, DNA-methyltransferases, tRNA and sensor histidine kinase (Figure 11. B).

DISCUSSION

This work identified 203 putative prophage-like in 81 *Xanthomonas* genomes spanning three different species (Table S1). Of these, 116 were considered intact genomes and 87 were defectives (Figure 1). *In silico* characterization of the intact prophages-like is an important step because these analyses are essential for describing these viruses; therefore, in the future, allowing the characterization of the novel viruses that infect *Xanthomonas* spp..

In this study, we detected polylysogeny in some *Xanthomonas* genomes. These events are frequent in bacterial pathogens (DAVIES et al., 2016). For instance, a study showed that a polylysogenic strain of *Pseudomonas aeruginosa* with two prophages was more resistant to phage lysis than strains with one or without prophages (BURNS et al., 2014). Another study revealed that *Enterococcus faecalis* V583 harbor seven different prophages that interact with each other and have an important role in pathogenicity and bacterial infection of human cells (MATOS et al., 2013). With that, we can suggest that the polylysogenic events detected in our results could be a trait providing plasticity to the *Xanthomonas* genomes, enhancing genetic variability and adaptive advantage.

The most significant number of detected prophage-like elements belonged to the *Myoviridae* family, followed by the *Inoviridae* family. However, the *Inoviridae* family presented the highest number of different prophages-like identified based on sequence identity. Frequently bacteriophages belonging to this family are identified in *Xanthomonas* strains (AHMAD et al., 2017; LIN et al., 1994; SOLÍS-SÁNCHEZ et al., 2020; YEH, 2017). Filamentous bacteriophages can encode secretion machinery that releases new viral particles without interfering with host life because this mechanism does not need bacteria's lysis. This way of life is interesting for both sides because without killing its host, the phage stays alive as long as its host is. The horizontal transfer of genes between them can be passed on, maintaining the acquired advantages by the bacterium. The diversity and significant amount of inoviruses in this work can be explained by the ability to keep its host alive during the lytic cycle, which is an advantage for both (virus and host).

It is expected that the GC content of mobile genetic elements like plasmids and phages tends to be similar to their host's GC content (CZAJKOWKI, 2019). However, we have seen the opposite: the GC content of the most identified prophages-like was lower, as already shown in other studies (ALMPANIS, 2018). This event may indicate that these elements have a short period of co-existence with their hosts (ROCHA and DANCHIN, 2002). The opposite was seen in Gonçalves et al., 2021, where the GC content of the studied prophages-like was

similar to its *Ralstonia* hosts. According to the authors, this indicates a long relationship between these bacteria and prophages-like (GONÇALVES et al., 2021). The proportion of phages by genome analyzed in this study was low, which also corroborated with the theory that these elements were recently acquired by their hosts, once it is known that prophages comprise 10 to 20% of their genome's host (CZAJKOWSKI, 2019).

About 25% of the bacteriophages exist on Earth in the prophage form (BONDY-DENOMY and DAVIDSON, 2014). Here, from 81 analyzed genomes, only two do not contain any prophage-like sequence, which makes us infer that prophages are strongly distributed in species of this genus. These bacteria have a wide range of hosts, making their geographic distribution very diverse. Such wide distribution corroborates with prophage-like distribution, where three of the four families detected were widespread across all continents. With that, our results suggest that this wide geographical distribution of *Xanthomonas* prophages-like influences the access to new hosts, contributing to more dispersion of genes through horizontal gene transfer, increasing genetic variability.

Prophages-like infecting *X. citri* strains are integrated into diverse genomic regions, but the same prophage-like present in different genomes usually integrates into the same region, showing an integration pattern. *X. campestris* showed a pattern of integration even though there are no equal prophages-like in different genomes. Only the Xc_07 phage integrated into a distant region from the other. Interestingly, all phages belonging to *X. campestris* genomes belonged to the *Inoviridae* family, except for the Xc_07, which belongs to the *Myoviridae* family. Inoviruses do not necessarily need an integration site to recombine into the genomes host; they can do this through transposition at a random site with their recombinase or the host recombinase. However, on these strains, the local integration may be a hotspot of mobile elements conducive to these events. On the other hand, myoviruses integrate with an integration site. In that case, Xc_07 is integrated near a tRNA and possibly does not negatively affect its host. This "guide" integration is necessary to prophage be considered tolerable the integration or even advantageous for the fitness of that bacterium. Usually, these events occur near tRNAs, intergenic regions and a few functional genes (RAMISETTY and SUDHAKARI, 2019).

Most of the phages belonging to the *Inoviridae* family showed high genetic mosaicism, except for a few elements that could be grouped into three groups by sequence similarity. It is an expected result since *Inoviridae* is a diverse viral family with only one conserved mark gene: pI (zot-protein) (ROUX et al., 2019). Besides that, genome phages

generally have modules that can be exchanged by recombination with other phages, which generates a high potential of variability between them.

Myoviruses showed syntenic blocks highly conserved, as well mostly siphoviruses. These results assume that infections by similar phages in different host strains occurred recently. The synteny results corroborate with violin plot graphic and heatmap analysis, showing that inoviruses are very diverse, followed by siphoviruses with an intermediary variability and myoviruses that possess a group of prophages-like that significantly conserve.

The vast majority of *X. axonopodis, X. campestris,* and *X. citri* genomes analyzed have integrated phages. However, we do not know how genes affect or may be associated with the adaptation and virulence of their host. In this study, we identified viral genes that may result from a horizontal transfer of genes, such as tRNA-Lys, suggesting that a recombination event between the sequences of prophage-like and host may contribute to both variabilities. In addition, genes that may be related to host virulence, such as proteins belonging to toxin-antitoxin systems, were identified. In the Xci_57 phage, we found an avirulence effector named AvrXv4, a protein belonging to a type III secretion system of the YopJ family. This effector was earlier characterized in *Xanthomonas* spp. being present in the phage genome may confer an advantage that can influence the host's virulence since it regulates interactions between the *Xanthomonas* and plant host. Furthermore, the Avr effectors may increase bacteria virulence by interfering with plant defenses (RODEN et al., 2004).

The Xci_49 harbors a toxin that belongs to a family of type II toxin-antitoxin systems named death-on-curing (Doc). This system was described on *Escherichia coli* phage P1 and is composed of the Doc toxin and the antitoxin phd (prevent host death). The hosts are highly dependent on this phage since when it is lost, the cells do not survive (LEHNHERR et al., 1993). This protein was also found in prophages-like infecting *Xanthomonas* spp. and may confer an adaptive advantage for the bacteria. Therefore, it is possible also maybe is a case of codependence. However, gene expression analysis must be performed for confirmation.

It is recurrent the presence of type II toxin-antitoxin systems in bacterial genomes. These genes are part of the bacterial mobilome that can be acquired through horizontal gene transfer (GUGLIELMINI and VAN MELDEREN, 2011). As the Doc cited above, few genes belonging to such a system were identified. Although the siphovirus Xci_35 contained the two genes of a toxin-antitoxin system, in most cases, only one gene was detected. This loss may result from the passing of generations, where one of the two genes may have been lost after integrating into the bacteria.

A high number of genes that possibly are acquired by horizontal transfer of genes (HTG) is to be expected since this process is ubiquitous between the host and mobile genetic elements (MGEs), including the prophages (BRITO, 2021). These genes could be important for contributing to phage genome plasticity, and the presence of the phage contributes to the plasticity of the host genome (CANCHAYA et al., 2004). Here we found transcriptional regulators, peptidases, methyltransferases, response regulators, tRNAs and a sensor histidine kinase in phage genomes. These genes are classified as bacterial and, possibly, were transferred to the prophages-like. For instance, methyltransferases can inhibit restriction enzymes that have some role in the hots' RM system (restriction-modification) (VARANI et al., 2013).

The presence of hypothetical proteins was recurrent in all analyzed prophages-like. There is a long way to go to characterize viral genomes entirely and understand the prophage-like host interaction and their influence on phenotypic host changes. When these events occur, new interactions may be discovered and explored (CRISPIM et al., 2018), and there be a better understanding of the role of prophages in the adaptation and evolution of their hosts.

Even though we found more intact than defective prophages-like in our analysis, there are many defective prophages-like. Intact prophages can kill their hosts if they enter their lytic life cycle. For that reason, selection may lead to the inactivation of prophages to prevent the cell from being eliminated since the element become unable to enter the lytic cycle (BOBAY et al., 2014). Thus, the significant number of defective phages detected in the sampled *Xanthomonas* spp. genomes may be due to phage domestication to avoid cell lysis.

Together, our results improved the knowledge about the genomic variability of *Xanthomonas* spp. prophages-like. In the future, *in vitro* analyses can be performed to validate our results.

METHODS

Detection of prophage-like elements in complete genomes of *Xanthomonas axonopodis*, *X. campestris* and *X. citri* genomes

Two hundred complete genomes belonging to selected *Xanthomonas* species (*X. axonopodis*, *X. campestris* and *X. citri*) were obtained from the Genome database of NCBI (https://www.ncbi.nlm.nih.gov/genome) accessed in june, 2020. To evaluate the genome assemblies quality, the genome assemblies were verified using Benchmarking Universal Single-Copy Orthologs (BUSCO) (WATERHOUSE et al., 2018) and only assemblies with completeness values greater than 96% were kept for the subsequent analysis.

We used the webserver PHASTER (PHAge Search Tool Enhanced Release) (http://phaster.ca/) (ARNDT et al., 2016) to identify prophage-like sequences in *Xanthomonas* species. Complete and incomplete prophages-like were selected and manually inspected to detect false positives. The CD-HIT v.4.8.1 was used to eliminate redundant sequences with 100% of identity (FU et al., 2012).

Genomic annotation and organization of prophage-like elements

Based on the sequence identity, the nucleotide prophage-like sequences were compared with the viruses RefSeq database from NCBI using the BLASTn algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The ORFs (Open Reading Frames) in each genome were predicted using GeneMarkS (http://exon.gatech.edu/GeneMark/genemarks.cgi) using the following parameters: Sequence type: Virus; Output format: LST; Output options format for gene prediction: Protein sequence and Gene nucleotide sequence; Advanced options: Genetic code 11 (Table S8) (BESEMER, 2001). To improve annotation accuracy, the sequences were analyzed using Rapid Annotation using Subsystem Technology (RAST) (http://rast.nmpdr.org/) (Table S8) (AZIZ et al., 2008). For functional annotation, the proteins encoded by the predicted genes were submitted to BLASTp compared with viral protein sequences belonging to NCBI's non-redundant (nr) database. The sequences were also explored manually through Geneious® 11.1.5. To visualize the prophages-like genomic organization, we used the EasyFig software system (SULLIVAN et al., 2011).

VirFam software system (http://biodev.extra.cea.fr/virfam/) was used to identify structural genes in prophages-like belonging to the order Caudovirales and classify these phages into families (LOPES et al., 2014). This software does not identify filamentous virus families.

In these cases, we identified the family by the presence of specific filamentous virus proteins, such as protein-containing Zonula occludens toxin domain (Zot).

In parallel to VirFam's analysis, we used signature genes to classify the viruses as intact or defective sequences. For elements belonging to *Myoviridae*, *Siphoviridae* and *Podoviridae* families, we used integrase, terminase, structural proteins and attachment sites to consider them as intact prophages-like. Meanwhile, for elements belonging to the *Inoviridae* family, we used Zot and replication initiation proteins. When one of these signature genes was absent, we classified them as a defective element.

Prophage-like integration sites in the genome of *X. campestris* and *X. citri* were mapped using circus plots. For *X. axonopodis,* this analysis was not performed since only one genome of this species was examined in this study. First, we downloaded a few *X. axonopodis*, *X. campestris* and *X. citri* oriC sequences from the DoriC database (http://tubic.org/doric/public/index.php). Then, we performed alignments using BLASTn to identify the most likely position of oriC in the bacterial genomes (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We used a custom Python script to adjust each genome based on the genomic coordinates that provided 100% identity and 100% query cover in the alignments. Otherwise, we used the genomic coordinates (https://blast.ncbi.nlm.nih.gov/Blast.cgi), which provided the highest number of significant alignments (identity and coverage greater than 90%) with different oriC sequences (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Finally, BLASTn alignments were performed to redefine the prophage-like positions within the adjusted genomes. The circus plots were build using "circlize" package in R platform.

Grouping intact prophage-like elements in clusters and phylogenetic analysis

Synteny analysis was performed with EasyFig software (SULLIVAN et al., 2011). In addition, we performed alignments of intact prophage-like sequences using ClustalW (https://www.ebi.ac.uk/Tools/msa/clustalo/). Heatmaps were built using Prism software to visualize the prophage-like percentage identity.

Phylogenetic analyses were performed using specific protein sequences for *Myoviridae* and *Siphoviridae* families and the total genome for the *Inoviridae* family (Table S9). First, the sequences were aligned using MAFFT v.7.471 (parameters: --local pair --retree 100), and each alignment was filtered and converted to nexus format using TrimAl v.1.2.59 (parameters: -gt 0.7 -cons 0.6 -nexus). The best evolutionary model for each protein alignment

was inferred using ModelTest-NG v.0.1.5 (parameters: -t mp -d aa -h uigf -s 11). Phylogenetic reconstructions were conducted using MrBayes v.3.2.7a according to the selected substitution models (parameter: mcmc ngen=1000000 nruns=2 nchains=4) (Table S8). For the SDT analysis, we performed a pairwise comparison by the SDT program using MUSCLE as an aligner.

Identification of putative genes that may influence bacteria ecology and fitness

The intact elements were screened for antibiotic resistance using the webserver ResFinder (https://cge.cbs.dtu.dk/services/ResFinder/). Next, the virulence factors using PHI-base (http://www.phi-base.org/) and putative genes of bacterial origin using BLASTn algorithm against NCBI virus RefSeq database of previously annotated ORFs, using default parameters.

FIGURE LEGENDS

Figure 1: Putative intact and defective prophage-like elements detected in complete genomes of *X. axonopodis, X. campestris* and *X. citri* from GenBank (NCBI).

Figure 2: Putative intact prophage-like elements distribution according to bacterial host. The first column represents prophage-like identification, the second column represents the bacterial strain, and the third column represents bacterial species.

Figure 3: Features of prophages-like found in *Xanthomonas* genomes. (A) Size range of intact and defective prophage-like elements. **(B)** Distribution of intact and defective prophage-like elements detected in host genomes. **(C)** Comparison between GC content of bacterial host and intact prophage-like elements. **(D)** Distribution of intact prophage-like families. **(E)** Distribution of viral families according to the hosts. **(F)** Unique intact prophage-like sequences distributed by family.

Figure 4: World distribution of putative intact sequences. (A) Distribution of intact prophage-like elements through continents. **(B)** Distribution of intact prophage-like elements through countries.

Figure 5: Insertion regions of the intact prophage-like elements in genomes adjusted in the origin of replication (oriC). (A) *Xanthomonas citri* adjusted genomes and integration site of intact prophages-like. **(B)** *Xanthomonas campestris* adjusted genomes and integration site of intact prophages-like.

Figure 6: Representation of ORFs and synteny of intact prophages-like from *Inoviridae* family. The line represent each genome of prophage-like sequence. The arrows are each ORF contained in that genome. The colors represent the function performed by each putative protein encoded by these ORFs.

Figure 7: Synteny of viral families belonging to Caudovirales order. (A) *Myoviridae* blocks synteny. **(B)** *Siphoviridae* blocks synteny. The line represent each genome of prophage-like

sequence. The arrows are each ORF contained in that genome. The colors represent the function performed by each putative protein encoded by these ORFs.

Figure 8: Violin plot graphic. In orange, the *Inoviridae* pair by pair comparison, in green, the *Myoviridae* pair by pair comparison and blue *Siphoviridae* pair by pair comparison, indicating the variability between viruses of the same family.

Figure 9: P**hylogenetic tree and SDT analysis of intact prophages-like from *Myoviridae* and *Syphoviridae***. **(A)** Phylogeny and SDT of TerL of prophages-like belonging to the *Myoviridae* family. **(B)** Phylogeny and SDT of TerL of prophages-like belonging to the *Syphovirdae* family.

Figure 10: Phylogenetic tree and SDT analysis of the complete genome of intact prophages-like from the *Inoviridae* family.

Figure 11: Genes that may confer adaptive advantages to the host. (A) Bar chart of putative toxin genes present in intact prophage-like elements distributed by the four viral families detected. **(B)** Bar chart of genes putatively acquired by horizontal transfer of genes.

Supplementary figure 1: Proportion between the size of intact, defective phages and host genomes. (A) *X. axonopodis* proportion. **(B)** *X. campestris* proportion. **(C)** *X. citri* proportion.
Supplementary figure 2: Genomic organization of *Inoviridae* sequence elements. Arrows indicate ORFs belonging to virus genomes. Similiar ORFs have the same color, and the legend indicates the function.

Supplementary figure 3: Genomic organization of *Myoviridae* sequence elements. Arrows indicate ORFs belonging to virus genomes. Similiar ORFs have the same color, and the legend indicates the function.

Supplementary figure 4: Genomic organization of *Podoviridae* sequence element. Arrows indicate ORFs belonging to virus genomes. Similiar ORFs have the same color, and the legend indicates the function.

Supplementary figure 5: Genomic organization of *Siphoviridae* sequence elements. Arrows indicate ORFs belonging to virus genomes. Similiar ORFs have the same color, and the legend indicates the function.

Supplementary figure 6: (A) *Siphoviridae* identity matrix – heatmap. **B)** *Myoviridae* identity matrix – heatmap. **(C)** *Inoviridae* identity matrix – heatmap. In all the three heatmaps the orange color gradient represents the sequences with the highest percentage of identity, ranging from 60 to 100%. The white color represents sequences with 50% of identity, and blue gradient represents sequences with identity percentage that ranges to 40 to 0%.

REFERENCES

Addy, H.S., Askora, A., Kawasaki, T., Fujie, M., and Yamada, T. (2012) The filamentous phage φrSS1 enhances virulence of phytopathogenic *Ralstonia solanacearum* on tomato. *Phytopathology*. 102: 244-51. http://doi.org/10.1094/PHYTO-10-11-0277

Ahmad, A. A., Askora, A., Kawasaki, T., Fujie, M., & Yamada, T. (2014). The filamentous phage XacF1 causes loss of virulence in *Xanthomonas axonopodis* pv . *citri* , the causative agent of citrus canker disease. *5(July)*, 1–11. https://doi.org/10.3389/fmicb.2014.00321

Ahmad, A. A., Kawabe, M., Askora, A., Kawasaki, T., Fujie, M., & Yamada, T. (2017). Dynamic integration and excision of filamentous phage XacF1 in *Xanthomonas citri* pv. *citri*, the causative agent of citrus canker disease. *FEBS Open Bio*, *7*(11), 1715-1721. https://doi.org/10.1002/2211-5463.12312

Almpanis, A., Swain, M., Gatherer, D., & Mcewan, N. (2018). Correlation between bacterial G + C content, genome size and the G + C content of associated plasmids and bacteriophages. *Microbial Genomics*, 0–7. https://doi.org/10.1099/mgen.0.000168

Arndt, D., Grant, J., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research.*, 2016 May 3. https://doi.org/10.1093/nar/gkw387

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., ... & Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*(1), 1-15. https://doi.org/10.1186/1471-2164-9-75

Baltz, R.H. (2012) *Streptomyces* temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms). *Journal of Industrial Microbiology & Biotechnology.* 39: 661-72. https://doi.org/10.1007/s10295-011-1069-6

Berjón-Otero, M., Koslová, A., & Fischer, M. G. (2019). The dual lifestyle of genome-integrating virophages in protists. *Annals of the New York Academy of Sciences*, *1447*(1), 97-109. https://doi.org/10.1111/nyas.14118

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, *29*(12), 2607-2618. https://doi.org/10.1093/nar/29.12.2607

Bobay, L. M., Touchon, M., & Rocha, E. P. (2014). Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences*, *111*(33), 12127-12132. https://doi.org/10.1073/pnas.1405336111

Bondy-Denomy, J., & Davidson, A. R. (2014). When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. *Journal of Microbiology*, *52*(3), 235–242. https://doi.org/10.1007/s12275-014-4083-3

Brito, I. L. (2021). Examining horizontal gene transfer in microbial communities. *Nature Reviews Microbiology*, *19*(7), 442-453. https://doi.org/10.1038/s41579-021-00534-7

Brussow, H., Canchaya, C., and Hardt, W.-D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *MMBR Reviews.* 68: 560–602. https://doi.org/10.1128/MMBR.68.3.560-602.2004

Burns, N., James, C. E., & Harrison, E. (2014). Polylysogeny magnifies competitiveness of a bacterial pathogen in vivo. *Evolutionary Applications,* 8(4), 346-351. https://doi.org/10.1111/eva.12243

Canchaya, C., Fournous, G., & Brüssow, H. (2004). The impact of prophages on bacterial chromosomes. *Molecular Microbiology*, *53*(1), 9-18. https://doi.org/10.1111/j.1365-2958.2004.04113.x

Crispim, J.S., Dias, R.S., Vidigal, P.M.P., Sousa, M.P. de, Silva, C.C. da, Santana, M.F., and Paula, S.O. de (2018) Screening and characterization of prophages in *Desulfovibrio* genomes. *Scientific Reports,* 8: 9273. https://doi.org/10.1038/s41598-018-27423-z

Czajkowski, R. (2019) May the Phage be With You? Prophage-Like Elements in the Genomes of Soft Rot Pectobacteriaceae: *Pectobacterium* spp. and *Dickeya* spp. *Frontiers Microbiology.* https://doi.org/10.3389/fmicb.2019.00138

Davies, E. V., Winstanley, C., Fothergill, J. L., & James, C. E. (2016). The role of temperate bacteriophages in bacterial infection. *FEMS Microbiology Letters*, *363*(5), 1–10. https://doi.org/10.1093/femsle/fnw015

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150-3152. https://doi.org/10.1093/bioinformatics/bts565

Guglielmini, J., & Van Melderen, L. (2011). Bacterial toxin-antitoxin systems: Translation inhibitors everywhere. *Mobile Genetic Elements*, *1*(4), 283-306. https://doi.org/10.4161/mge.18477

Lehnherr, H., Maguin, E., Jafri, S., & Yarmolinsky, M. B. (1993). Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. *Journal of Molecular Biology*, *233*(3), 414-428. https://doi.org/10.1006/jmbi.1993.1521

Lemire, S., Yehl, K.M., and Lu, T.K. (2018) Phage-based applications in synthetic biology. *Annual Reviews Virology*. 5: 453–476. https://doi.org/10.1146/annurev-virology-092917-043544.

Lin, N. T., You, B. Y., Huang, C. Y., Kuo, C. W., Wen, F. S., Yang, J. S., & Tseng, Y. H. (1994). Characterization of two novel filamentous phages of *Xanthomonas*. *Journal of General Virology*, *75*(9), 2543-2547. https://doi.org/10.1099/0022-1317-75-9-2543

Lopes, A., Tavares, P., Petit, M. A., Guérois, R., & Zinn-Justin, S. (2014). Automated classification of tailed bacteriophages according to their neck organization. *BMC Genomics*, *15*(1), 1-17. https://doi.org/10.1186/1471-2164-15-1027.

Matos, R. C., Lapaque, N., Rigottier-gois, L., Debarbieux, L., Meylheuc, T., Gonzalez-zorn, B., Repoila, F., & Lopes, M. D. F. (2013). *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. *Plos Genetics. 9*(6). https://doi.org/10.1371/journal.pgen.1003539

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y., and Hayashi, T. (1999) The complete nucleotide sequence of ΦCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: Implications for phage evolution and horizontal gene transfer via bacteriophages. *Molecular Microbiology.* 31: 399-419. https://doi.org/10.1046/j.1365-2958.1999.01158.x

Nanda, A.M., Thormann, K., and Frunzke, J. (2015) Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *Journal of Bacteriology.* 197: 410–419. https://doi.org/10.1128/JB.02230-14

Pereira, F., Xavier, A. S., Prieto, F., Rafael, B., Rezende, R. De, Marcus, P., Vidigal, P., & Alfenas-Zerbini, P. (2019). Biological and molecular characterization of a bacteriophage infecting *Xanthomonas campestris* pv . *campestris* , isolated from brassica fields. *Archives of Virology*, *164*(7), 1857–1862. https://doi.org/10.1007/s00705-019-04263-4

Ramisetty, B. C. M., & Sudhakari, P. A. (2019). Bacterial "grounded" prophages: Hotspots for genetic renovation and innovation. *Frontiers in Genetics*, *10*(FEB), 1–17. https://doi.org/10.3389/fgene.2019.00065

Rocha, C., & Danchin, A. (2002). Base composition bias might result from competition for. *Trends in Genetics, 18*(6), 291–294. https://doi.org/10.1016/S0168-9525(02)02690-2

Roden, J., Eardley, L., Hotson, A., Cao, Y., & Mudgett, M. B. (2004). Characterization of the *Xanthomonas* AvrXv4 effector, a SUMO protease translocated into plant cells. *Molecular plant-microbe interactions*, *17*(6), 633-643. https://doi.org/10.1094/MPMI.2004.17.6.633

Roux, S., Krupovic, M., Daly, R. A., Borges, A. L., Nayfach, S., Schulz, F., Sharrar, A., Carnevali, P. B. M., Cheng, J., Ivanova, N. N., Bondy-denomy, J., Wrighton, K. C., Woyke, T., Visel, A., Kyrpides, N. C., & Eloe-fadrosh, E. A. (2019). bacteria and archaea across Earth ' s biomes. *Nature Microbiology*, *4*(November). https://doi.org/10.1038/s41564-019-0510-x

Ryan, R. P., Vorhölter, F. J., Potnis, N., Jones, J. B., Van Sluys, M. A., Bogdanove, A. J., & Dow, J. M. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nature Reviews Microbiology*, *9*(5), 344-355. https://doi.org/10.1038/nrmicro2558

Solís-Sánchez, G. A., Quiñones-Aguilar, E. E., Fraire-Velázquez, S., Vega-Arreguín, J., & Rincón-Enríquez, G. (2020). Complete Genome Sequence of XaF13, a Novel Bacteriophage of *Xanthomonas vesicatoria* from Mexico. *Microbiology Resource Announcements*, *9*(5). https://doi.org/ 10.1128/MRA.01371-19

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics*, *27*(7), 1009-1010. https://doi.org/10.1093/bioinformatics/btr039

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., Goss, E. M., & Jones, J. B. (2020). *Xanthomonas* diversity, virulence and plant–pathogen interactions. *Nature Reviews Microbiology*, *18*(8), 415–427. https://doi.org/10.1038/s41579-020-0361-8

Varani, A.M., Monteiro-Vitorello, C.B., Nakaya, H.I., and Sluys, M.-A. Van (2013). The Role of Prophage in Plant-Pathogenic Bacteria. *Annual Reviews Phytopathology*. 51: 429-51. https://doi.org/10.1146/annurev-phyto-081211-173010

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., ... & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution*, *35*(3), 543-548. https://doi.org/10.1093/molbev/msx319

Yeh, T. Y. (2017). Complete nucleotide sequence of a new filamentous phage, Xf109, which integrates its genome into the chromosomal DNA of *Xanthomonas oryzae*. *Archives of Virology*, *162*(2), 567-572. https://doi.org/ 10.1007/s00705-016-3105-3

SUPPLEMENTARY TABLES

Supplementary table 1: Identification and GenBank accession numbers of *Xanthomonas* spp. used in this work.

Species	Strain	Accession no.	Taxonomy ID	References
Xanthomonas axonopodis pv. *vasculorum*	NCPPB796	NZ_CP053649.1	325777	Teixeira, A. C. et al., 2020
X. axonopodis	Xac29-1	NC_020800.1	1304892	Chen, G. et al., 2013
Xanthomonas campestris	CFBP5825R	NZ_CM002638.1	339	Roux, B. et al., 2015
Xanthomonas campestris pv. *badri*	NEB122	NZ_CP051651.1	149696	Fomenkov, A. et al., 2020
X. campestris pv. *campestris*	17	NZ_CP011946.1	340	Liu,Y. C. et al., 2015
X. campestris pv. *campestris*	3811	NZ_CP025750.1	340	Lv, H., 2018
X. campestris pv. *campestris*	B100	NC_010688.1	509169	Vorholter, F. J. et al., 2008
X. campestris pv. *campestris*	ICMP 4013	NZ_CP012146.1	340	Desai, D. et al., 2015
X. campestris pv. *campestris*	ICMP21080	NZ_CP012145.1	340	Desai, D. et al., 2015
X. campestris pv. *campestris*	8004	NC_007086.1	314565	Qian, W. et al., 2005
X. campestris pv. *campestris*	ATCC 33913	NC_003902.1	190485	da Silva, A. C. et al., 2002
X. campestris pv. *campestris*	CFBP1869	NZ_CM002545.1	1357999	Bolot, S. et al., 2015
X. campestris pv. *campestris*	CFBP 5817	NZ_CM002673.1	1358009	Bolot, S. et al., 2015
X. campestris pv. *campestris*	CN03	NZ_CP017308.1	1358004	Szurek, B. et al., 2018
X. campestris pv. *campestris*	CN12	NZ_CP017310.1	1358017	Noel, L. D. et al., 2016
Xanthomonas citri pv. phaseoli var. fuscans	CFBP 6166	NZ_CP021001.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP 6167	NZ_CP021018.1	473423	Ruh, M. et al., 2017
X. citri pv. *anacardii*	CFBP 2913	NZ_CP024057.1	1437881	Gagnevin, L. and Bolot,S., 2017
X. citri pv. *aurantifolii*	1566	NZ_CP012002.1	76802	Varani, A. M. et al., 2015
X. citri pv. *aurantifolii*	FDC 1559	NZ_CP011160.1	76802	Varani, A. M. et al., 2015
X. citri pv. *aurantifolii*	FDC 1609	NZ_CP011163.1	76803	Varani, A. M. et al., 2016
X. citri pv. *citri*	03-1638-1-1	NZ_CP023285.1	611301	Gochez, A. M. et al., 2018

X. citri pv. *citri*	5208	NZ_CP009028.1	1308542	Zhang, Y. et al., 2014
X. citri pv. *citri*	AW13	NZ_CP009031.1	611301	Zhang, Y. et al., 2015
X. citri pv. *citri*	AW14	NZ_CP009034.1	611301	Zhang, Y. et al., 2015
X. citri pv. *citri*	AW15	NZ_CP009037.1	611301	Zhang, Y. et al., 2015
X. citri pv. *citri*	gd3	NZ_CP009016.1	611301	Zhang, Y. et al., 2014
X. citri pv. *citri*	jx4	NZ_CP009013.1	611301	Zhang, Y. et al., 2014
X. citri pv. *citri*	jx5	NZ_CP009010.1	611301	Zhang, Y. et al., 2015
X. citri pv. *citri*	jx-6	NZ_CP011827.2	611301	Chen, F. et al., 2016
X. citri pv. *citri*	LH201	NZ_CP018858.1	611301	Lefeuvre, P., 2016
X. citri pv. *citri*	LH276	NZ_CP018854.1	611301	Lefeuvre, P., 2016
X. citri pv. *citri*	LJ207-7	NZ_CP018850.1	611301	Lefeuvre, P., 2016
X. citri pv. *citri*	LL074-4	NZ_CP018847.1	611301	Lefeuvre, P., 2016
X. citri pv. *Citri*	LM180	NZ_CM007622.1	611301	Lefeuvre, P., 2016
X. citri pv. *citri*	mf20	NZ_CP009007.1	611301	Zhang, Y. et al., 2015
X. citri pv. *citri*	Sample (SAMEA2752396)	NZ_LN606820.1	611301	Register, A., 2014
X. citri pv. *citri*	Sample (SAMEA2755371)	NZ_LN590509.1	611301	Register, A., 2014
X. citri pv. *citri*	306	NC_003919.1	190486	da Silva, A. C., et al., 2002
X. citri pv. *citri*	TX160197	NZ_CP020889.1	611301	Munoz-Bodnar, A. et al., 2017
X. citri pv. *citri*	Xcc29-1	NZ_CP023661.1	611301	Liu, X. and Wang, X., 2017
X. citri pv. *citri*	Xcc49	NZ_CP023662.1	611301	Liu, X. and Wang, X., 2017
X. citri pv. *citri*	LM199	NZ_CM007619.1	611301	Lefeuvre, P., 2016
X. citri pv. *fuscans*	4834-R	NC_022541.1	1240239	Darrasse, A. et al., 2013
X. citri pv. *fuscans*	CFBP 6988	NZ_CP026331.1	1365651	Bolot, S. et al., 2017
X. citri pv. *fuscans*	ISO12C3	NZ_CP012055.1	366649	Xie, W. et al., 2015
X. citri pv. *fuscans*	ISO118C1	NZ_CP012053.1	366649	Xie, W. et al., 2015
X. citri pv. *fuscans*	ISO118C5	NZ_CP012051.1	366649	Xie, W. et al., 2015
X. citri pv. *glycines*	8ra Sample (SAMN12340633)	NZ_CP041781.1	1401257	Kang, I. J. et al., 2019
X. citri pv. *glycines*	1018	NZ_CP041961.1	473421	Kang, I. J. et al., 2019

X. citri pv. *glycines*	1157	NZ_CP041963.1	473421	Kang, I. J. et al., 2019
X. citri pv. *glycines*	2098	NZ_CP041965.1	473421	Kang, I. J. et al., 2019
X. citri pv. *glycines*	CFBP 2526	NZ_CM002268.1	1365648	Darrasse, A. et al., 2013
X. citri pv. *glycines*	CFBP 7119	NZ_CM002264.1	1365649	Darrasse, A. et al., 2013
X. citri pv. *glycines*	EB08	NZ_CP026334.1	473421	Carpenter, S. C. D. et al., 2019
X. citri pv. *glycines*	K2	NZ_CP041967.1	473421	Kang, I. J. et al., 2019
X. citri pv. *glycines*	8ra Sample (SAMN05722944)	NZ_CP017188.2	1401257	Seong, H. J. et al., 2018
X. citri pv. *glycines*	12.-2	NZ_CP015972.1	1150615	Carpenter, S. C. D. et al., 2016
X. citri pv. *malvacearum*	AR81009	NZ_CP023155.1	86040	Phillips, A. Z., 2017
X. citri pv. *malvacearum*	HD-1	NZ_CP046019.1	86040	Liu, F., 2019
X. citri pv. *malvacearum*	MS14003	NZ_CP023159.1	86040	Phillips, A. Z., 2017
X. citri pv. *malvacearum*	MSCT	NZ_CP017020.1	86040	Showmaker, K. C. et al., 2016
X. citri pv. *malvacearum*	X18	NZ_CM002136.1	1220027	Cunnac, S. et al., 2013
X. citri pv. *malvacearum*	X20	NZ_CM002029.1	1220028	Cunnac, S. et al., 2013
X. citri pv. *malvacearum*	XcmN1003	NZ_CP013006.1	86040	Bogdanove, A. J. et al., 2015
X. citri pv. *phaseoli var. fuscans*	CFBP4885	NZ_CP020992.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6165	NZ_CP020998.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6975	NZ_CP021006.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6988R	NZ_CP020979.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6989	NZ_CP020981.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6990	NZ_CP020983.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6991	NZ_CP021015.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6992	NZ_CP020985.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6994R	NZ_CP020987.1	473423	Ruh, M. et al., 2017
X. citri pv. *phaseoli var. fuscans*	CFBP6996R	NZ_CP020989.1	473423	Ruh, M. et al., 2017
X. citri pv. *vignicola*	CFBP7111	NZ_CP022263.1	473426	Ruh, M. et al., 2017
X. citri pv. *vignicola*	CFBP7112	NZ_CP022267.1	473426	Ruh, M. et al., 2017
X. citri pv. *vignicola*	CFBP7113	NZ_CP022270.1	473426	Ruh, M. et al., 2017

X. citri subsp. *citri*	A306	NZ_CP006857.1	1308541	Jalan, N. and Wang, N., 2013
X. citri subsp. *citri*	Aw12879	NC_020815.1	1137651	Jalan, N. et al., 2013
X. citri subsp. *citri*	UI6	NZ_CP008992.1	1308548	Zhang, Y. et al., 2015

Supplementary table 2: Genome completeness results by BUSCO.

Specie	Strain	Collinearity
Xanthomonas axonopodis pv. *vasculorum*	NCPPB796	96.5%
X. axonopodis	Xac29-1	96.0%
Xanthomonas campestris	CFBP5825R	96.5%
X. campestris pv. *badri*	NEB122	96.5%
X. campestris pv. *campestris*	17	96.5%
X. campestris pv. *campestris*	3811	96.5%
X. campestris pv. *campestris*	B100	96.2%
X. campestris pv. *campestris*	ICMP 4013	96.5%
X. campestris pv. *campestris*	ICMP21080	96.5%
X. campestris pv. *campestris*	8004	96.5%
X. campestris pv. *campestris*	ATCC 33913	96.5%
X. campestris pv. *campestris*	CFBP1869	96.5%
X. campestris pv. *campestris*	CFBP 5817	96.5%
X. campestris pv. *campestris*	CN03	96.5%
X. campestris pv. *campestris*	CN12	96.7%
Xanthomonas citri pv. *phaseoli var. fuscans*	CFBP 6166	96.7%
X. citri pv. *phaseoli var. fuscans*	CFBP 6167	96.7%
X. citri pv. *anacardii*	CFBP 2913	96.5%
X. citri pv. *aurantifolii*	1566	96.2%
X. citri pv. *aurantifolii*	FDC 1559	96.2%
X. citri pv. *aurantifolii*	FDC 1609	96.2%
X. citri pv. *citri*	03-1638-1-1	96.2%
X. citri pv. *citri*	5208	96.5%
X. citri pv. *citri*	AW13	96.5%
X. citri pv. *citri*	AW14	96.5%
X. citri pv. *citri*	AW15	96.5%

X. citri pv. *citri*	gd3	96.5%
X. citri pv. *citri*	jx4	96.5%
X. citri pv. *citri*	jx5	96.5%
X. citri pv. *citri*	jx-6	96.5%
X. citri pv. *citri*	LH201	96.5%
X. citri pv. *citri*	LH276	96.5%
X. citri pv. *citri*	LJ207-7	96.5%
X. citri pv. *citri*	LL074-4	96.5%
X. citri pv. *Citri*	LM180	96.0%
X. citri pv. *citri*	mf20	96.5%
X. citri pv. *citri*	Sample (SAMEA2752396)	96.5%
X. citri pv. *citri*	Sample (SAMEA2755371)	96.5%
X. citri pv. *citri*	306	96.5%
X. citri pv. *citri*	TX160197	96.5%
X. citri pv. *citri*	Xcc29-1	96.5%
X. citri pv. *citri*	Xcc49	96.2%
X. citri pv. *citri*	LM199	96.7%
X. citri pv. *fuscans*	4834-R	96.5%
X. citri pv. *fuscans*	CFBP 6988	96.7%
X. citri pv. *fuscans*	ISO12C3	96.7%
X. citri pv. *fuscans*	ISO118C1	96.7%
X. citri pv. *fuscans*	ISO118C5	96.5%
X. citri pv. *glycines*	8ra Sample (SAMN12340633)	96.5%
X. citri pv. *glycines*	1018	96.5%
X. citri pv. *glycines*	1157	96.5%
X. citri pv. *glycines*	2098	96.5%
X. citri pv. *glycines*	CFBP 2526	96.5%
X. citri pv. *glycines*	CFBP 7119	96.5%

X. citri pv. *glycines*	EB08	96.5%
X. citri pv. *glycines*	K2	96.2%
X. citri pv. *glycines*	8ra Sample (SAMN05722944)	96.5%
X. citri pv. *glycines*	12.-2	96.5%
X. citri pv. *malvacearum*	AR81009	96.5%
X. citri pv. *malvacearum*	HD-1	96.7%
X. citri pv. *malvacearum*	MS14003	96.7%
X. citri pv. *malvacearum*	MSCT	96.7%
X. citri pv. *malvacearum*	X18	96.5%
X. citri pv. *malvacearum*	X20	96.5%
X. citri pv. *malvacearum*	XcmN1003	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP4885	96.2%
X. citri pv. *phaseoli var. fuscans*	CFBP6165	96.7%
X. citri pv. *phaseoli var. fuscans*	CFBP6975	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP6988R	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP6989	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP6990	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP6991	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP6992	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP6994R	96.5%
X. citri pv. *phaseoli var. fuscans*	CFBP6996R	96.5%
X. citri pv. *vignicola*	CFBP7111	96.2%
X. citri pv. *vignicola*	CFBP7112	96.9%
X. vignicola	CFBP7113	96.5%
X. citri subsp. *citri*	A306	96.5%
X. citri subsp. *citri*	Aw12879	96.5%
X. citri subsp. *Citri*	UI6	96.5%
X. citri pv. *citri*	TX160042	96.5%

Supplementary table 3: Intact prophage-like elements sequences identification, size and family classification.

Specie	Strain	Prophage-like identification	Prophage-like size (kb)	Family classification
X. axonopodis	NCPPB796	Xax_1	30.9	*Inoviridae*
X. campestris	CFBP5825R	Xc_01	9.7	*Inoviridae*
	17	Xc_02	13.2	*Inoviridae*
	ICMP4013	Xc_03	10.9	*Inoviridae*
	ICMP21080	Xc_04	13.5	*Inoviridae*
	8004	Xc_05	42.7	*Inoviridae*
	ATCC_33913	Xc_06	42.3	*Inoviridae*
		Xc_07	38.9	*Myoviridae*
	CFBP1869	Xc_08	28.9	*Inoviridae*
	CN12	Xc_09	11.6	*Inoviridae*
		Xc_10	47.1	*Inoviridae*
	CFBP6166	Xci_18	9.4	*Inoviridae*
	CFBP6167	Xci_19	22.1	*Inoviridae*
	FDC1559	Xci_20	7.8	*Inoviridae*
	FDC1609	Xci_21	36.9	*Inoviridae*
	03-1638-1-1	Xci_01	46.2	*Siphoviridae*
	5208	Xci_01	46.2	*Siphoviridae*
X. citri	AW13	Xci_06	37.2	*Myoviridae*
		Xci_15	10.7	*Inoviridae*
		Xci_16	46.6	*Myoviridae*
	AW14	Xci_04	51.5	*Myoviridae*
		Xci_15	10.7	*Inoviridae*
		Xci_16	33	*Myoviridae*
	AW15	Xci_04	51.5	*Myoviridae*
		Xci_15	10.7	*Inoviridae*
		Xci_16	33	*Myoviridae*

gd3	Xci_01	46.2	*Siphoviridae*
jx4	Xci_01	46.2	*Siphoviridae*
jx5	Xci_01	46.2	*Siphoviridae*
LH201	Xci_17	27.9	*Siphoviridae*
LH276	Xci_01	46.2	*Siphoviridae*
LJ207-7	Xci_01	46.2	*Siphoviridae*
LL074-4	Xci_17	27.9	*Siphoviridae*
LM180	Xci_24	8.7	*Inoviridae*
	Xci_01	46.2	*Siphoviridae*
mf20	Xci_01	46.2	*Siphoviridae*
SAMEA2752396	Xci_26	33.7	*Myoviridae*
SAMEA2755371	Xci_27	14.4	*Inoviridae*
	Xci_28	61.1	*Myoviridae*
306	Xci_01	46.2	*Siphoviridae*
TX160197	Xci_04	58.4	*Myoviridae*
	Xci_30	37.1	*Inoviridae*
Xcc29-1	Xci_01	46.2	*Siphoviridae*
Xcc49	Xci_01	46.2	*Siphoviridae*
4834-R	Xci_09	8.8	*Inoviridae*
6988	Xci_31	42.5	*Myoviridae*
	Xci_32	41	*Podoviridae*
ISO12C3	Xci_33	25.4	*Inoviridae*
ISO118C1	Xci_08	25.5	*Inoviridae*
ISO118C5	Xci_08	25.5	*Inoviridae*
	Xci_06	37.2	*Myoviridae*
8ra Sample (SAMN12340633)	Xci_11	33.4	*Siphoviridae*
	Xci_09	7.9	*Inoviridae*
	Xci_02	38.7	*Myoviridae*

1018	Xci_06	37.2	*Myoviridae*
	Xci_02	38.7	*Myoviridae*
1157	Xci_11	33.4	*Siphoviridae*
	Xci_02	38.7	*Myoviridae*
	Xci_05	55	*Myoviridae*
2098	Xci_35	34.6	*Siphoviridae*
	Xci_36	30.9	*Inoviridae*
	Xci_12	41.1	*Myoviridae*
CFBP 2526	Xci_06	37.2	*Myoviridae*
	Xci_38	36	*Siphoviridae*
	Xci_11	33.7	*Siphoviridae*
	Xci_05	51.9	*Myoviridae*
CFBP 7119	Xci_06	37.2	*Myoviridae*
	Xci_02	38.8	*Myoviridae*
	Xci_05	55.5	*Myoviridae*
EB08	Xci_06	37.2	*Myoviridae*
	Xci_11	33.4	*Siphoviridae*
	Xci_07	39.5	*Myoviridae*
K2	Xci_06	37.2	*Myoviridae*
	Xci_40	36.1	*Myoviridae*
	Xci_02	38.7	*Myoviridae*
	Xci_05	53.6	*Myoviridae*
8ra Sample (SAMN05722944)	Xci_06	37.2	*Myoviridae*
	Xci_11	33.4	*Siphoviridae*
	Xci_09	7.8	*Inoviridae*
12.-2	Xci_02	38.7	*Myoviridae*
	Xci_06	37.2	*Myoviridae*
	Xci_02	38.7	*Myoviridae*

	Xci_12	40.9	*Myoviridae*
AR81009	Xci_41	16.8	*Inoviridae*
HD-1	Xci_09	8.5	*Inoviridae*
MS14003	Xci_10	29.4	*Siphoviridae*
	Xci_42	29.3	*Inoviridae*
MSCT	Xci_10	29.4	*Siphoviridae*
X18	Xci_43	26.4	*Siphoviridae*
	Xci_44	26.6	*Inoviridae*
X20	Xci_45	23.2	*Siphoviridae*
XcmN1003	Xci_46	16.3	*Siphoviridae*
CFBP4885	Xci_47	10.7	*Inoviridae*
CFBP6165	Xci_48	35.1	*Inoviridae*
CFBP6975	Xci_49	58.1	*Myoviridae*
	Xci_50	21.8	*Inoviridae*
CFBP6988R	Xci_31	44.8	*Myoviridae*
CFBP6989	Xci_07	38.3	*Myoviridae*
CFBP6990	Xci_07	38.3	*Myoviridae*
CFBP6991	Xci_07	38.3	*Myoviridae*
CFBP6992	Xci_52	29.1	*Myoviridae*
CFBP6994R	Xci_53	37.3	*Myoviridae*
CFBP6996R	Xci_53	37.3	*Myoviridae*
CFBP7111	Xci_55	27	*Inoviridae*
	Xci_56	43	*Siphoviridae*
CFBP7112	Xci_57	34.3	*Myoviridae*
	Xci_58	26.1	*Inoviridae*
CFBP7113	Xci_59	41.9	*Myoviridae*
	Xci_60	10.9	*Inoviridae*
A306	Xci_01	46.2	*Siphoviridae*

	Xci_06	11.7	*Myoviridae*
Aw12879	Xci_15	10.7	*Inoviridae*
	Xci_07	38.4	*Myoviridae*
UI6	Xci_01	46.2	*Siphoviridae*

Supplementary table 4: GC content of intact prophages-like and their hosts.

Specie	Strain	Host GC (%)	Prophage-like GC (%)	Prophage-like identification
X. axonopodis	NCPPB796	64,09	61,14	Xax_1
X. campestris	CFBP5825R	65,1	62,29	Xc_01
	17	65,1	62,46	Xc_02
	ICMP4013	65,3	61,74	Xc_03
	ICMP21080	65,3	62,65	Xc_04
	8004	65	63,33	Xc_05
	ATCC_33913	65,1	63,28	Xc_06
			61,97	Xc_07
	CFBP1869	65,09	61,74	Xc_08
	CN12	65,1	59,2	Xc_09
			61,52	Xc_10
X. citri	CFBP6166	64,68	61,11	Xci_18
	CFBP6167	64,55	59,78	Xci_19
	FDC1559	64,65	60,28	Xci_20
	FDC1609	64,73	60,72	Xci_21
	03-1638-1-1	64,47	60,27	Xci_01
	5208	64,74	60,28	Xci_01
	AW13	64,66	62,18	Xci_06
			59,42	Xci_15
			62,89	Xci_16
	AW14	64,66	62,05	Xci_04
			59,42	Xci_15
			63,82	Xci_16
	AW15	64,66	62,05	Xci_04
			59,42	Xci_15

		63,82	Xci_16
gd3	64,74	60,28	Xci_01
jx4	64,74	60,27	Xci_01
jx5	64,74	60,27	Xci_01
LH201	64,52	62,83	Xci_17
LH276	64,51	59,16	Xci_01
LJ207-7	64,5	60,34	Xci_01
LL074-4	64,52	59,12	Xci_17
LM180	64,37	62,83	Xci_24
		59,15	Xci_01
mf20	64,74	60,27	Xci_01
SAMEA2752396	64,69	60,28	Xci_25
	64,69	58,68	Xci_26
SAMEA2755371	64,54	60,48	Xci_27
	64,54	59,38	Xci_28
306	64,74	60,22	Xci_01
TX160197	64,65	60,28	Xci_04
		62,18	Xci_30
Xcc29-1	64,72	60,71	Xci_01
Xcc49	64,74	60,27	Xci_01
4834-R	64,71	60,42	Xci_09
6988	64,6	61,08	Xci_31
		61,52	Xci_32
ISO12C3	64,56	61,52	Xci_33
ISO118C1	64,56	59,43	Xci_08
ISO118C5	64,56	59,42	Xci_08
8ra Sample (SAMN12340633)	64,56	59,4	Xci_06

			62,67	Xci_11
			64,95	Xci_09
			60,43	Xci_02
	1018	64,68	61,64	Xci_06
			62,67	Xci_02
	1157	64,57	61,63	Xci_34
			62,7	Xci_11
			64,95	Xci_02
			61,64	Xci_05
	2098	65,8	62,4	Xci_35
			62,65	Xci_36
			59,65	Xci_12
	CFBP 2526	64,67	61,99	Xci_06
			63,04	Xci_38
			64,84	Xci_11
			64,74	Xci_05
	CFBP 7119	64,34	61,38	Xci_06
			62	Xci_02
			61,63	Xci_05
	EB08	64,48	61,69	Xci_06
			62,67	Xci_11
			64,95	Xci_07
	K2	64,58	62,32	Xci_06
			62,04	Xci_40
			62,27	Xci_02
			61,64	Xci_05
	8ra Sample (SAMN05722944)	64,58	61,86	Xci_06

		62,67	Xci_11
		64,95	Xci_09
		60,45	Xci_02
12,-2	64,38	61,64	Xci_06
		62,67	Xci_02
		61,64	Xci_12
AR81009	64,54	62,17	Xci_41
HD-1	64,57	60,98	Xci_09
MS14003	64,65	60,91	Xci_10
		61,23	Xci_42
MSCT	64,64	59,78	Xci_10
X18	64,73	61,23	Xci_43
		60,89	Xci_44
X20	64,47	61,63	Xci_45
XcmN1003	64,47	62,06	Xci_46
CFBP4885	64,65	62,67	Xci_47
CFBP6165	64,76	61,26	Xci_48
CFBP6975	64,68	60,64	Xci_49
		62,65	Xci_50
CFBP6988R	64,65	59,78	Xci_31
CFBP6989	64,65	61,6	Xci_07
CFBP6990	64,65	62,62	Xci_07
CFBP6991	64,63	61,6	Xci_07
CFBP6992	64,56	62,62	Xci_52
CFBP6994R	64,66	62,32	Xci_53
CFBP6996R	64,76	62,54	Xci_53
CFBP7111	64,46	62,51	Xci_55

			60,73	Xci_56
	CFBP7112	64,65	64,26	Xci_57
			63,22	Xci_58
	CFBP7113	64,7	60,17	Xci_59
			62,76	Xci_60
	A306	64,74	59,49	Xci_01
	Aw12879	64,66	60,28	Xci_06
			59,7	Xci_15
			59,42	Xci_07
	UI6	64,74	63,88	Xci_01

Supplementary table 5: Genomic coordinates of intact prophage-like elements in the bacterial genome and their putative attachment region.

Species	Strain	Prophage-like identification	Coordinates in the bacterial genome	The putative phage attachment region
X. axonopodis	NCPPB796	Xax_1	2327358-2346160	CGGCCTGCTCAA
X. campestris	CFBP5825R	Xc_01	2514197-2523322	Nonexistent
	17	Xc_02	2364589-2375263	Nonexistent
	ICMP4013	Xc_03	2439182-2448493	Nonexistent
	ICMP21080	Xc_04	2410503-2419120	GCGGTGGTCGCGC
	8004	Xc_05	2531332-2545098	CGTCGCGCAGCG
	ATCC_33913	Xc_06	2434843-2448511	CGCTGCGCGACG (1) GCAGTTCTGCAG (2)
		Xc_07	3529052-3560152	TGGTGGGCCGTGATGGATTCGAACCATCGACCAAAAGATTAAAAG
	CFBP1869	Xc_08	2585935-2603393	AAGTTGATCAGTA (1) TATACATTATGCGAA (2)
	CN12	Xc_09	2547214-2557531	GCGCCGCCGCCC
		Xc_10	2606824-2632247	CCATCGCCAGCG (1) AGGGGAGGGGGCTTC (2) CTTGTTCGACGAC (3) GACCGCGCCGAC (4)
X. citri	CFBP6166	Xci_18	1682544-1689132	Nonexistent
	CFBP6167	Xci_19	4275516-4289978	TGCCTGCCTGTT
	FDC1559	Xci_20	2115849-2121584	Nonexistent
	FDC1609	Xci_21	2865619-2882391	TCCATGAAGAAGTCCCGCTTTACCGACAGCCAGATCATCGCCGTGCTCAAGCAGGCCCAGGCCGGTGC GCCCGTGCCGGA
	03-1638-1-1	Xci_01	3221016-3235361	CTTGACGACGGACGGC

5208	Xci_01	3105244-3119783	CTTGACGACGGACGGC
	Xci_06	1284479-1309268	ATCAGGGATCAC
AW13	Xci_15	1775277-1781270	Nonexistent
	Xci_16	5054091-5087470	AACCAAAGGATTATGAGTC
	Xci_04	1286084-1309248	ACTCTTAATCAATAGGTCCAAGGTTCGAATCCT
AW14	Xci_15	1774876-1781268	Nonexistent
	Xci_16	5054131-5087486	Nonexistent
	Xci_04	1269257-1309246	ACTCTTAATCAATAGGTCCAAGGTTCGAATCCT
AW15	Xci_15	1774875-1781254	Nonexistent
	Xci_16	5054125-5076411	Nonexistent
gd3	Xci_01	3083272-3108859	CTTGACGACGGACGGC
jx4	Xci_01	3094063-3108412	CTTGACGACGGACGGC
jx5	Xci_01	3094078-3108427	CTTGACGACGGACGGC
LH201	Xci_17	509315-523655	TCGCCTTTGCGCG
LH276	Xci_01	3429380-3456109	CTTGACGACGGACGGC
LJ207-7	Xci_01	951656-965997	CTTGACGACGGACGGC
LL074-4	Xci_17	2370074-2384443	TCGCCTTTGCGCG
LM180	Xci_24	301694-309325	Nonexistent
	Xci_01	915034-940615	CTTGACGACGGACGGC

mf20	Xci_01	3105193-3119589	CTTGACGACGGACGGC
SAMEA2752396	Xci_26	3100817-3119689	ATCCTTGGCAAT
SAMEA2755371	Xci_27	2614234-2627683	TGCGCCGTGGCTG
	Xci_28	2888638-2934293	GGGTTGGTAACG
306	Xci_01	3094009-3119596	CTTGACGACGGACGGC
TX160197	Xci_04	3005738-3050720	ACTCTTAATCAATAGGTCCAAGGTTCGAATCCT
	Xci_30	4574301-4597578	GTGCAGCGCAGA (1) GCTGTGTTGGAAGCC (2) ACGCGTCACGAT (3)
Xcc29-1	Xci_01	3093557-3119140	CTTGACGACGGACGGC
Xcc49	Xci_01	3082856-3108447	CTTGACGACGGACGGC
4834-R	Xci_09	2575876-2582191	Nonexistent
6988	Xci_31	1200791-1232843	CGGGAAAATTAC
	Xci_32	2827949-2867268	CCCTTCCGCTCCACCA
ISO12C3	Xci_33	2656197-2674945	GTCCGGTCCGCT
ISO118C1	Xci_08	2655245-2675011	GTCCGGTCCGCT
ISO118C5	Xci_08	2655228-2674959	GTCCGGTCCGCT
8ra Sample (SAMN12340633)	Xci_06	1260165-1291767	ATCAGGGATCAC
	Xci_11	2063519-2094124	CGCTGCAGGCGA
	Xci_09	2595182-2600441	Nonexistent
	Xci_02	3848293-3880191	CTTTTAATCTTTTGGTCGATGGTTCGAATCCATCACGGCCCACCA

1018	Xci_06	1262486-1294088	ATCAGGGATCAC
	Xci_02	3637208-3669074	CTTTTAATCTTTTGGTCGATGGTTCGAATCCATCACGGCCCACCA
	Xci_11	2056579-2087218	CGCTGCAGGCGA
1157	Xci_02	3683121-3715019	CTTTTAATCTTTTGGTCGATGGTTCGAATCCATCACGGCCCACCA
	Xci_05	4516155-4547351	AGCCAACTGCCT (1) ATTTTTCCCGGC (2)
	Xci_35	2010270-2031420	CTGCGCACCGCG
2098	Xci_36	2594731-2622293	CGCCATCACCGC
	Xci_12	4381539-4412556	AACCAAAGGATTATGAG
CFBP 2526	Xci_06	1225995-1257546	ATCAGGGATCAC
	Xci_38	1937129-1958676	GCCCGCCGGCGTGGA
	Xci_11	2226492-2247430	CGCTGCAGGCGA
	Xci_05	4504643-4542555	TTATGGTGGGCCCA
CFBP 7119	Xci_06	1254779-1303535	ATCAGGGATCAC
	Xci_02	3903917-3935829	CTTTTAATCTTTTGGTCGATGGTTCGAATCCATCACGGCCCACCA
	Xci_05	4714791-4743771	TTATGGTGGGCCCA (1) CAATTTTCCCGA (2)
EB08	Xci_06	1259548-1291111	ATCAGGGATCAC
	Xci_11	2062045-2081790	CGCTGCAGGCGA
	Xci_07	4615172-4647813	AACCAAAGGATTATGAGT
K2	Xci_06	1289575-1336882	ATCAGGGATCAC

	Xci_40	3113972-3144873	TGGCGCCCGAAGTTGGACTCGAACCAACGACCCCCTGATTAACAGT
	Xci_02	3734430-3766288	CTTTTAATCTTTTGGTCGATGGTTCGAATCCATCACGGCCCACCA
	Xci_05	4557167-4595924	TTATGGTGGGCCCA (1) CAATTTTCCCGA (2)
	Xci_06	1260164-1291804	ATCAGGGATCAC
8ra Sample (SAMN05722944)	Xci_11	2063553-2084060	CGCTGCAGGCGA
	Xci_09	2593035-2600884	Nonexistent
	Xci_02	3848367-3880185	CTTTTAATCTTTTGGTCGATGGTTCGAATCCATCACGGCCCACCA
	Xci_06	1260045-1291647	ATCAGGGATCAC
12.-2	Xci_02	3854556-3886411	CTTTTAATCTTTTGGTCGATGGTTCGAATCCATCACGGCCCACCA
	Xci_12	4679572-4707099	CAATTTTCCCGA
AR81009	Xci_41	2690504-2705778	Nonexistent
HD-1	Xci_09	3702554-3710049	Nonexistent
MS14003	Xci_10	1229399-1242643	TCGTCTGTTTGAG
	Xci_42	2620566-2643136	GATTGATCGCGT (1) GGTGATGCGCCGC (2)
MSCT	Xci_10	1229369-1242644	TCGTCTGTTTGAG
X18	Xci_43	1221294-1233570	TCGTCTGTTTGAG
	Xci_44	2586880-2598581	CGGTCGGCACGC
X20	Xci_45	1241985-1263944	ACTCTTAATCAATAGGTCCAAGGTTCGAATCCTTGACGGCCCACCAA
XcmN1003	Xci_46	1172768-1186836	ACTCTTAATCAATAGGTCCAAGGTTCGAATCCTTGACGGCCCACCAA

CFBP4885	Xci_47	1988521-1995103	Nonexistent
CFBP6165	Xci_48	3930859-3950998	AACAGGCAGGCA (1) CGGTTACACAAAAT (2)
CFBP6975	Xci_49	1516300-1546067	TTGGCGCCAGAA (1) AACCAAAGGATTATGAGT (2)
	Xci_50	4870957-4887378	GTTGACCATCGT
CFBP6988R	Xci_31	1736104-1769210	CGGGAAAATTAC
CFBP6989	Xci_07	2834481-2857602	TCGCGGCGGCATT
CFBP6990	Xci_07	995959-1029066	TCGCGGCGGCATT
CFBP6991	Xci_07	1494631-1517792	TCGCGGCGGCATT
CFBP6992	Xci_52	1706692-1731545	GACTGGGACAAG
CFBP6994R	Xci_53	3700416-3731348	GACTGGGACAAG
CFBP6996R	Xci_53	2145678-2176571	TCGCGGCGGCATT
CFBP7111	Xci_55	2794256-2812748	CTGCCCAGCGCC
	Xci_56	3409072-3443860	CAGCGCCCAGAC
CFBP7112	Xci_57	1220248-1253222	GCAGCAGGCTGC
	Xci_58	2555720-2569277	TCCACTTCTGGCG
CFBP7113	Xci_59	1233671-1262762	TCCCGAATCCCG (1) GTTGTTGCGAGG (2)
	Xci_60	2507533-2513916	GCACCAACGTGC
A306	Xci_01	3094003-3119542	CTTGACGACGGACGGC
Aw12879	Xci_06	1269219-1278485	ACGGGCTGGGGC

	Xci_15	1774011-1780924	Nonexistent
	Xci_07	5054420-5078963	TGCGCACCGATT
UI6	Xci_01	3081848-3107484	CTTGACGACGGACGGC

Supplementary table 6: The most similar bacteriophage with the identified sequences in this study and their similarity percentage.

Specie	Strain	Prophage-like identification	The phage with the highest number of proteins is most similar a to those found in the prophage-like (GenBank accession, no. similar proteins)	Percentage of similarity with the most similar prophage-like
X. axonopodis	NCPPB796	Xax_1	Xanthomonas phage XacF1 DNA	71%
X. campestris	CFBP5825R	Xc_01	Xanthomonas phage phiLf2	22%
	17	Xc_02	Xanthomonas phage phiLf2	99%
	ICMP4013	Xc_03	Xanthomonas phage phiLf2	89%
	ICMP21080	Xc_04	Xanthomonas phage phiLf UK variant	93%
	8004	Xc_05	Xanthomonas phage phiLf2	83%
	ATCC_33913	Xc_06	Xanthomonas phage phiLf UK variant	96%
		Xc_07	ND	
	CFBP1869	Xc_08	Xanthomonas phage phiXv2	84%
	CN12	Xc_09	Xanthomonas phage XacF1 DNA	22%
		Xc_10	Xanthomonas phage phiXv2	24%
	CFBP6166	Xci_18	Xanthomonas phage phiLf2	75%
	CFBP6167	Xci_19	Xanthomonas phage XacF1 DNA	23%
	FDC1559	Xci_20	Xanthomonas phage XaF13	39%
	FDC1609	Xci_21	Xanthomonas phage XaF13	42%
X. citri	03-1638-1-1	Xci_01	Stenotrophomonas phage Smp131	14%
	5208	Xci_01	Stenotrophomonas phage Smp131	14%
	AW13	Xci_06	Stenotrophomonas phage Smp131	19%
		Xci_15	Xanthomonas phage Cf2	99%
		Xci_16	Stenotrophomonas phage Smp131	25%
	AW14	Xci_04	Stenotrophomonas phage Smp131	19%
		Xci_15	Xanthomonas phage Cf2	99%
		Xci_16	Stenotrophomonas phage Smp131	20%
	AW15	Xci_04	Stenotrophomonas phage Smp131	19%
		Xci_15	Xanthomonas phage Cf2	99%

	Xci_16	Stenotrophomonas phage Smp131	25%
gd3	Xci_01	Stenotrophomonas phage Smp131	15%
jx4	Xci_01	Stenotrophomonas phage Smp131	15%
jx5	Xci_01	Stenotrophomonas phage Smp131	15%
LH201	Xci_17	Stenotrophomonas phage Smp131	15%
LH276	Xci_01	Stenotrophomonas phage Smp131	15%
LJ207-7	Xci_01	Stenotrophomonas phage Smp131	15%
LL074-4	Xci_17	Stenotrophomonas phage Smp131	13%
LM180	Xci_24	Xanthomonas phage XacF1 DNA	61%
	Xci_01	Stenotrophomonas phage Smp131	15%
mf20	Xci_01	Stenotrophomonas phage Smp131	15%
SAMEA2752396	Xci_26	Stenotrophomonas phage Smp131	20%
SAMEA2755371	Xci_27	Xanthomonas phage Cf2	47%
	Xci_28	Stenotrophomonas phage Smp131	16%
306	Xci_01	Stenotrophomonas phage Smp131	15%
TX160197	Xci_04	Stenotrophomonas phage Smp131	19%
	Xci_30	Xanthomonas phage phiXv2	29%
Xcc29-1	Xci_01	Stenotrophomonas phage Smp131	15%
Xcc49	Xci_01	Stenotrophomonas phage Smp131	15%
4834-R	Xci_09	Xanthomonas phage phiLf2	82%
6988	Xci_31	Stenotrophomonas phage Smp131	27%
	Xci_32	Stenotrophomonas phage S1	2%
ISO12C3	Xci_33	Xanthomonas phage phiLf UK variant	34%
ISO118C1	Xci_08	Xanthomonas phage phiLf UK variant	34%
ISO118C5	Xci_08	Xanthomonas phage phiLf UK variant	34%
8ra Sample (SAMN12340633)	Xci_06	Stenotrophomonas phage Smp131	32%
	Xci_11	ND	
	Xci_09	Xanthomonas phage phiLf2	78%

1018	Xci_02	Stenotrophomonas phage Smp131	9%
	Xci_06	Stenotrophomonas phage Smp131	32%
	Xci_02	Stenotrophomonas phage Smp131	9%
	Xci_11	ND	
1157	Xci_02	Stenotrophomonas phage Smp131	9%
	Xci_05	Stenotrophomonas phage Smp131	21%
	Xci_35	ND	
2098	Xci_36	Xanthomonas phage XacF1 DNA	14%
	Xci_12	Stenotrophomonas phage Smp131	18%
	Xci_06	Stenotrophomonas phage Smp131	32%
CFBP 2526	Xci_38	Rhizobium phage RHEph06	1%
	Xci_11	ND	
CFBP 7119	Xci_05	Stenotrophomonas phage Smp131	23%
	Xci_06	Stenotrophomonas phage Smp131	32%
	Xci_02	Stenotrophomonas phage Smp131	9%
EB08	Xci_05	Stenotrophomonas phage Smp131	23%
	Xci_06	Stenotrophomonas phage Smp131	32%
	Xci_11	ND	
K2	Xci_07	Stenotrophomonas phage Smp131	31%
	Xci_06	Stenotrophomonas phage Smp131	32%
	Xci_40	Stenotrophomonas phage Smp131	35%
	Xci_02	Stenotrophomonas phage Smp131	9%
8ra Sample (SAMN05722944)	Xci_05	Stenotrophomonas phage Smp131	23%
	Xci_06	Stenotrophomonas phage Smp131	32%
	Xci_11	ND	
	Xci_09	Xanthomonas phage phiLf2	78%
12.-2	Xci_02	Stenotrophomonas phage Smp131	9%
	Xci_06	Stenotrophomonas phage Smp131	32%

	Xci_02	Stenotrophomonas phage Smp131	9%
	Xci_12	Stenotrophomonas phage Smp131	23%
AR81009	Xci_41	Xanthomonas phage phiLf2	61%
HD-1	Xci_09	Xanthomonas phage phiLf2	67%
MS14003	Xci_10	Stenotrophomonas phage Smp131	17%
	Xci_42	Xanthomonas phage phiXv2	17%
MSCT	Xci_10	Stenotrophomonas phage Smp131	17%
X18	Xci_43	Stenotrophomonas phage Smp131	18%
	Xci_44	Xanthomonas phage phiLf2	37%
X20	Xci_45	Stenotrophomonas phage Smp131	14%
XcmN1003	Xci_46	Stenotrophomonas phage Smp131	21%
CFBP4885	Xci_47	Xanthomonas phage phiLf2	82%
CFBP6165	Xci_48	Xanthomonas phage phiXv2	30%
CFBP6975	Xci_49	Stenotrophomonas phage Smp131	27%
	Xci_50	Xanthomonas phage phiLf2	33%
CFBP6988R	Xci_31	Stenotrophomonas phage Smp131	26%
CFBP6989	Xci_07	Stenotrophomonas phage Smp131	31%
CFBP6990	Xci_07	Stenotrophomonas phage Smp131	31%
CFBP6991	Xci_07	Stenotrophomonas phage Smp131	31%
CFBP6992	Xci_52	Stenotrophomonas phage Smp131	32%
CFBP6994R	Xci_53	Stenotrophomonas phage Smp131	33%
CFBP6996R	Xci_53	Stenotrophomonas phage Smp131	33%
CFBP7111	Xci_55	Xanthomonas phage phiLf2	11%
	Xci_56	ND	
CFBP7112	Xci_57	Stenotrophomonas phage Smp131	22%
	Xci_58	Xanthomonas phage XacF1 DNA	38%
CFBP7113	Xci_59	Stenotrophomonas phage Smp131	15%
	Xci_60	Xanthomonas phage Cf2	63%

A306	Xci_01	Stenotrophomonas phage Smp131	14%
	Xci_06	Stenotrophomonas phage Smp131	32%
Aw12879	Xci_15	Xanthomonas phage Cf2	99%
	Xci_07	Stenotrophomonas phage Smp131	29%
UI6	Xci_01	Stenotrophomonas phage Smp131	14%

Supplementary table 7: Bacterial ORFs flanking the detected prophages-like.

Specie	Strain	Prophage-like identification	ORFs in the host genome flanking the integrated prophage-like sequence, L: left flank, R: right flank
X. axonopodis	NCPPB796	Xax_1	L: IS3 family transposase; DotA/TraY family protein IS5 family transposase R: IS3 family transposase; Gfo/Idh/MocA family oxidoreductase; beta-lactamase family protein;
X. campestris	CFBP5825R	Xc_01	L: hypothetical protein; hypothetical protein; cyclic beta 1-2 glucan synthetase R: NfuA family Fe-S biogenesis protein; 4a-hydroxytetrahydrobiopterin dehydratase; energy transducer TonB
	17	Xc_02	L: DNA-binding protein; avirulence protein; hypothetical protein R: hypothetical protein; NfuA family Fe-S biogenesis protein; 4a-hydroxytetrahydrobiopterin dehydratase
	ICMP4013	Xc_03	L: NfuA family Fe-S biogenesis protein; 4a-hydroxytetrahydrobiopterin dehydratase; energy transducer TonB R: hypothetical protein; hypothetical protein; DUF3693 domain-containing protein
	ICMP21080	Xc_04	L: NfuA family Fe-S biogenesis protein; 4a-hydroxytetrahydrobiopterin dehydratase; energy transducer TonB R: hypothetical protein; hypothetical protein; DUF3693 domain-containing protein
	8004	Xc_05	L: conjugal transfer protein TrbP; hypothetical protein; NfuA family Fe-S biogenesis protein R: NdvB; PAS domain-containing sensor histidine kinase; response regulator
	ATCC_33913	Xc_06	L: hypothetical protein; cyclic beta 1-2 glucan synthetase; PAS domain-containing sensor histidine kinase R: repetitive transmembrane protein; NfuA family Fe-S biogenesis protein; 4a-hydroxytetrahydrobiopterin dehydratase
		Xc_07	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein R: tRNA-Lys; 7-cyano-7-deazaguanine synthase QueC; tol-pal system protein YbgF
	CFBP1869	Xc_08	L: hypothetical protein; hypothetical protein; 4a-hydroxytetrahydrobiopterin dehydratase R: trypsin-like serine protease; alpha/beta hydrolase; sensor histidine kinase
	CN12	Xc_09	L: hypothetical protein; DUF3693 domain-containing protein; hypothetical protein R: hypothetical protein; type II secretion pathway, component ExeA; hypothetical protein
		Xc_10	L: Avirulence protein AvrBs3; transposase; hypothetical protein R: Tn3 family transposase; conjugal transfer protein; repetitive transmembrane protein
X. citri	CFBP6166	Xci_18	L: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase R: IS256-like element ISXax1 family transposase; IS256-like element ISXax1 family transposase; hypothetical protein
	CFBP6167	Xci_19	L: hypothetical protein; hypothetical protein; hypothetical protein R: GMC family oxidoreductase; alpha/beta hydrolase; glycosyltransferase
	FDC1559	Xci_20	L: hypothetical protein; transcriptional regulator; hypothetical protein R: DUF2523 domain-containing protein; hypothetical protein; methyltransferase
	FDC1609	Xci_21	L: DUF2523 domain-containing protein; hypothetical protein; methyltransferase R: glutamine-hydrolyzing GMP synthase; IMP dehydrogenase; bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase FolD
	03-1638-1-1	Xci_01	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: IS3 family transposase; type IV pilin protein; pilus assembly protein
	5208	Xci_01	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R:IS3 family transposase; type IV pilin protein; pilus assembly protein

	Xci_06	L: N-acetylglucosamine-6-phosphate deacetylase; toprim domain-containing protein; hypothetical protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; IS21-like element ISXci1 family helper ATPase IstB
AW13	Xci_15	L: hypothetical protein; hypothetical protein; hypothetical protein R: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase
	Xci_16	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: hypothetical protein; RNA polymerase sigma factor RpoD; D-tyrosyl-tRNA(Tyr) deacylase
	Xci_04	L: helix-turn-helix transcriptional regulator; DNA-binding protein; hypothetical protein; R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; IS21-like element ISXci1 family helper ATPase IstB
AW14	Xci_15	L: hypothetical protein; hypothetical protein; hypothetical protein R: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase
	Xci_16	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: tRNA-Ile; hypothetical protein; RNA polymerase sigma factor RpoD
	Xci_04	L: hypothetical protein; lysozyme; hypothetical protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
AW15	Xci_15	L: hypothetical protein; hypothetical protein; hypothetical protein. R: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase
	Xci_16	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: transcriptional regulator; hypothetical protein; hypothetical protein
gd3	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; type IV pilin protein; pilus assembly protein
jx4	Xci_01	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: IS3 family transposase; type IV pilin protein; pilus assembly protein
jx5	Xci_01	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: IS3 family transposase; type IV pilin protein; pilus assembly protein
LH201	Xci_17	L: IS3 family transposase; type IV pilin protein; pilus assembly protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; IS3-like element ISXac2 family transposase
LH276	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; type IV pilin protein; pilus assembly protein
LJ207-7	Xci_01	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: IS3 family transposase; type IV pilin protein; pilus assembly protein
LL074-4	Xci_17	L: IS3 family transposase; type IV pilin protein; pilus assembly protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; IS3-like element ISXac2 family transposase
LM180	Xci_24	L: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase R: IS3 family transposase; oxidoreductase; DedA family protein
	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; hypothetical protein; type IV pilin protein

mf20	Xci_01	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: IS3 family transposase; type IV pilin protein; pilus assembly protein
SAMEA2752396	Xci_26	L: tRNA-Asn; excinuclease ABC subunit UvrB; tRNA-Val R: hypothetical protein; IS3 family transposase; transposase
SAMEA2755371	Xci_27	L: hypothetical protein; hypothetical protein; hypothetical protein R: GMC family oxidoreductase; alpha/beta hydrolase; glycosyltransferase
	Xci_28	L: tRNA-Asn; excinuclease ABC subunit UvrB; tRNA-Val R: transposase; transposase; hypothetical protein
306	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; type IV pilin protein; pilus assembly protein; pilus assembly protein
TX160197	Xci_04	L: hypothetical protein; lysozyme; hypothetical protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; IS21-like element ISXci1 family helper ATPase IstB
	Xci_30	L: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase R: IS3-like element ISXac3 family transposase; oxidoreductase; DedA family protein
Xcc29-1	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; type IV pilin protein; pilus assembly protein
Xcc49	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; type IV pilin protein; pilus assembly protein
4834-R	Xci_09	L: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase R: hypothetical protein; IS256-like element ISXax1 family transposase; IS256-like element ISXax1 family transposase
6988	Xci_31	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
	Xci_32	L: tRNA-Ala; H-NS histone family protein; DUF3606 domain-containing protein R: hypothetical protein; hypothetical protein; hypothetical protein
ISO12C3	Xci_33	L: hypothetical protein; hypothetical protein; hypothetical protein R: hypothetical protein; hypothetical protein; hypothetical protein
ISO118C1	Xci_08	L: thioesterase; MbtH family NRPS accessory protein; efflux RND transporter periplasmic adaptor subunit R: hypothetical protein; hypothetical protein; hypothetical protein
ISO118C5	Xci_08	L: thioesterase; MbtH family NRPS accessory protein; efflux RND transporter periplasmic adaptor subunit R: hypothetical protein; hypothetical protein; hypothetical protein
8ra Sample (SAMN12340633)	Xci_06	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
	Xci_11	L: hypothetical protein; hypothetical protein; hypothetical protein R: PAS domain S-box protein; pirin family protein; LysR family transcriptional regulator
	Xci_09	L: hypothetical protein; DUF3693 domain-containing protein; NAD(P)H-dependent oxidoreductase R: hypothetical protein; SEC-C domain-containing protein; helix-turn-helix domain-containing protein
1018	Xci_02	L: 7-cyano-7-deazaguanine synthase QueC; PAS domain-containing protein; response regulator R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
	Xci_06	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase

	Xci_02	L: tRNA-Lys; 7-cyano-7-deazaguanine synthase QueC; PAS domain-containing protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
	Xci_11	L: hypothetical protein; hypothetical protein; hypothetical protein R: PAS domain S-box protein; pirin family protein; LysR family transcriptional regulator
1157	Xci_02	L: tRNA-Lys; 7-cyano-7-deazaguanine synthase QueC; PAS domain-containing protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
	Xci_05	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; AAA family ATPase R: tRNA-Ile; hypothetical protein; RNA polymerase sigma factor RpoD
	Xci_35	L: DUF1441 family protein; hypothetical protein; hypothetical protein R: hypothetical protein, hypothetical protein; hypothetical protein
2098	Xci_36	L: three-Cys-motif partner protein TcmP; hypothetical protein; IS1595 family transposase R: repetitive transmembrane protein; NfuA family Fe-S biogenesis protein; 4a-hydroxytetrahydrobiopterin dehydratase
	Xci_12	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; LysR family transcriptional regulator R: tRNA-Ile; hypothetical protein; RNA polymerase sigma factor RpoD
	Xci_06	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
CFBP 2526	Xci_38	L: hypothetical protein; hypothetical protein; hypothetical protein R: transfer-messenger RNA; S8 family peptidase; SsrA-binding protein SmpB
	Xci_11	L: hypothetical protein; hypothetical protein; hypothetical protein R: hypothetical protein; hypothetical protein; hypothetical protein
	Xci_05	L: AlpA family phage regulatory protein; hypothetical protein; ogr/Delta-like zinc finger family protein R: tRNA-Ile; hypothetical protein; RNA polymerase sigma factor RpoD
	Xci_06	L: hypothetical protein; hypothetical protein; hypothetical protein R: site-specific DNA-methyltransferase; GNAT family N-acetyltransferase; site-specific DNA-methyltransferase
CFBP 7119	Xci_02	L: tRNA-Lys; 7-cyano-7-deazaguanine synthase QueC; PAS domain-containing protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
	Xci_05	L: AlpA family phage regulatory protein; hypothetical protein; ogr/Delta-like zinc finger family protein R: transcriptional regulator; DNA-binding protein; hypothetical protein
	Xci_06	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
EB08	Xci_11	L: hypothetical protein; hypothetical protein; hypothetical protein R: hypothetical protein; DUF1833 family protein; hypothtical protein
	Xci_07	l: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein R: tRNA-Ile; hypothetical protein; RNA polymerase sigma factor RpoD
	Xci_06	L: hypothetical protein; hypothetical protein; hypothetical protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
K2	Xci_40	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: tRNA-Asn; type IV pilin protein; pilus assembly protein

8ra Sample (SAMN05722944)	Xci_02	L: tRNA-Lys; 7-cyano-7-deazaguanine synthase QueC; PAS domain-containing protein R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
	Xci_05	L: hypothetical protein; ogr/Delta-like zinc finger family protein; hypothetical protein R: tRNA-Ile; hypothetical protein; RNA polymerase sigma factor RpoD
	Xci_06	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
	Xci_11	L: hypothetical protein; hypothetical protein; hypothetical protein R: hypothetical protein; hypothetical protein; hypothetical protein
	Xci_09	L: thioesterase; MbtH family NRPS accessory protein; efflux RND transporter periplasmic adaptor subunit R: hypothetical protein; SEC-C domain-containing protein; helix-turn-helix domain-containing protein
12.-2	Xci_02	L: 7-cyano-7-deazaguanine synthase QueC; PAS domain-containing protein; response regulator R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
	Xci_06	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
	Xci_02	L: 7-cyano-7-deazaguanine synthase QueC; PAS domain-containing protein; response regulator R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; DUF4411 family protein
	Xci_12	L: hypothetical protein; hypothetical protein; integrase arm-type DNA-binding domain-containing protein R: hypothetical protein; hypothetical protein; toprim domain-containing protein
AR81009	Xci_41	L: SDR family oxidoreductase; hypothetical protein; hypothetical protein R: hypothetical protein; IS3-like element ISXac2 family transposase; oxidoreductase
HD-1	Xci_09	L: IS3-like element ISXac2 family transposase; oxidoreductase; DedA family protein R: SDR family oxidoreductase; hypothetical protein; hypothetical protein
MS14003	Xci_10	L: hypothetical protein; tRNA-Lys; tRNA-His R: recombinase family protein; IS3 family transposase; glycogen debranching protein GlgX
	Xci_42	L: hypothetical protein; hypothetical protein; hypothetical protein R: hypothetical protein; hypothetical protein; SDR family oxidoreductase
MSCT	Xci_10	L: hypothetical protein; tRNA-Lys; tRNA-His R: IS3 family transposase; glycogen debranching protein GlgX; 30S ribosomal protein S6--L-glutamate ligase
X18	Xci_43	L: hypothetical protein; tRNA-Lys; tRNA-His R: recombinase family protein; IS3 family transposase; glycogen debranching protein GlgX
	Xci_44	L: IS3-like element ISXc8 family transposase; FAD-dependent monooxygenase; DedA family protein R: hypothetical protein; hypothetical protein; SDR family oxidoreductase
X20	Xci_45	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; hypothetical protein; PAS domain-containing sensor histidine kinase
XcmN1003	Xci_46	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; IS3-like element ISXac2 family transposase; hypothetical protein
CFBP4885	Xci_47	L: IS256-like element ISXax1 family transposase; IS256-like element ISXax1 family transposase; IS256-like element ISXax1 family transposase R: SDR family oxidoreductase; GMC family oxidoreductase; alpha/beta hydrolase

CFBP6165	Xci_48	L: hypothetical protein; hypothetical protein; hypothetical protein R: hypothetical protein; hypothetical protein; hypothetical protein
CFBP6975	Xci_49	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: tRNA-Ile; hypothetical protein; RNA polymerase sigma factor RpoD
	Xci_50	L: hypothetical protein; hypothetical protein; hypothetical protein R: DUF2523 domain-containing protein; IS256-like element ISXax1 family transposase; IS256-like element ISXax1 family transposase
CFBP6988R	Xci_31	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
CFBP6989	Xci_07	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; GNAT family N-acetyltransferase R: helix-turn-helix transcriptional regulator; DNA-binding protein; hypothetical protein
CFBP6990	Xci_07	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
CFBP6991	Xci_07	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; GNAT family N-acetyltransferase R: helix-turn-helix transcriptional regulator; DNA-binding protein; hypothetical protein
CFBP6992	Xci_52	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
CFBP6994R	Xci_53	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; site-specific DNA-methyltransferase
CFBP6996R	Xci_53	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; GNAT family N-acetyltransferase R: hypothetical protein; tRNA-Lys; tRNA-Arg
CFBP7111	Xci_55	L: hypothetical protein; hypothetical protein; hypothetical protein R: hypothetical protein; SDR family oxidoreductase; GMC family oxidoreductase
	Xci_56	L: PAS domain-containing methyl-accepting chemotaxis protein; pirin family protein; LysR family transcriptional regulator R: helix-turn-helix domain-containing protein; S24 family peptidase; hypothetical protein
CFBP7112	Xci_57	L: hypothetical protein; tRNA-Lys; tRNA-His R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein
	Xci_58	L: ribbon-helix-helix domain-containing protein; putative toxin-antitoxin system toxin component, PIN family; Tn3 family transposase R: helix-turn-helix transcriptional regulator; hypothetical protein; hypothetical protein
CFBP7113	Xci_59	L: tRNA-Lys; tRNA-His; tRNA-Arg R: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; GNAT family N-acetyltransferase
	Xci_60	L: hypothetical protein; ATP-binding protein; IS3-like element ISXc8 family transposase R: hypothetical protein; SDR family oxidoreductase; GMC family oxidoreductase
A306	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; type IV pilin protein; pilus assembly protein
Aw12879	Xci_06	L: lysozyme; hypothetical protein; IS21 family transposase R: hypothetical protein; hypothetical protein; hypothetical protein
	Xci_15	L: hypothetical protein; IS3-like element ISXac3 family transposase; oxidoreductase R: hypothetical protein; SDR family oxidoreductase; GMC family oxidoreductase

	Xci_07	L: Com family DNA-binding transcriptional regulator; site-specific DNA-methyltransferase; hypothetical protein R: toprim domain-containing protein; IS21-like element ISXci1 family helper ATPase IstB; IS21-like element ISXci1 family transposase
UI6	Xci_01	L: tRNA-Asn; GspH/FimT family pseudopilin; excinuclease ABC subunit UvrB R: IS3 family transposase; type IV pilin protein; pilus assembly protein

Supplementary table 8: Program parameters used in genomic analysis.

PROGRAM	PARAMETER
BUSCO	Database: Gammaproteobacteria (452 sequences)
RAST	Taxonomy ID: NCBI taxonomy-ID Domain: Bacteria Genetic code: 11
GeneMarkS	Sequence type: Virus Output format: LST Output options format for gene prediction: Protein sequence and Gene nucleotide sequence Advanced options: Genetic code 11
MrBayes v.3.2.7[a] *according to substitution models and optimizers described below	mcmc ngen=1000000 nruns=2 nchains=4
MAFFT v.7.471	--localpair --retree 100
TrimAl v.1.2.59	-gt 0.7 -cons 0.6 -nexus
ModelTest-NG v.0.1.5	-t mp -d aa -h uigf -s 11

Supplementary table 9: Proteins and substitution models used in phylogenetic analysis.

Myoviridae	Protein	Substitution models
	Terminase ATPase subunit	JTT+I+G4+F

Siphoviridae	Protein	Substitution models and optimizers
	Terminase ATPase subunit	BLOSUM62+F

FIGURES
Figure 1

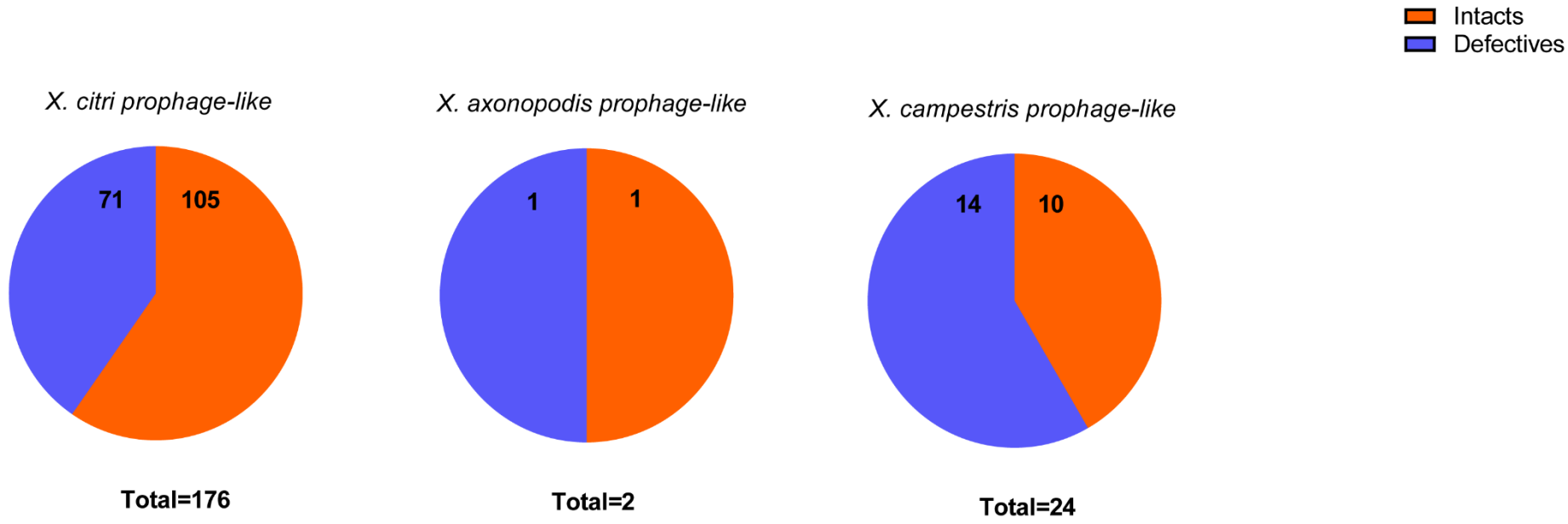


Figure 2

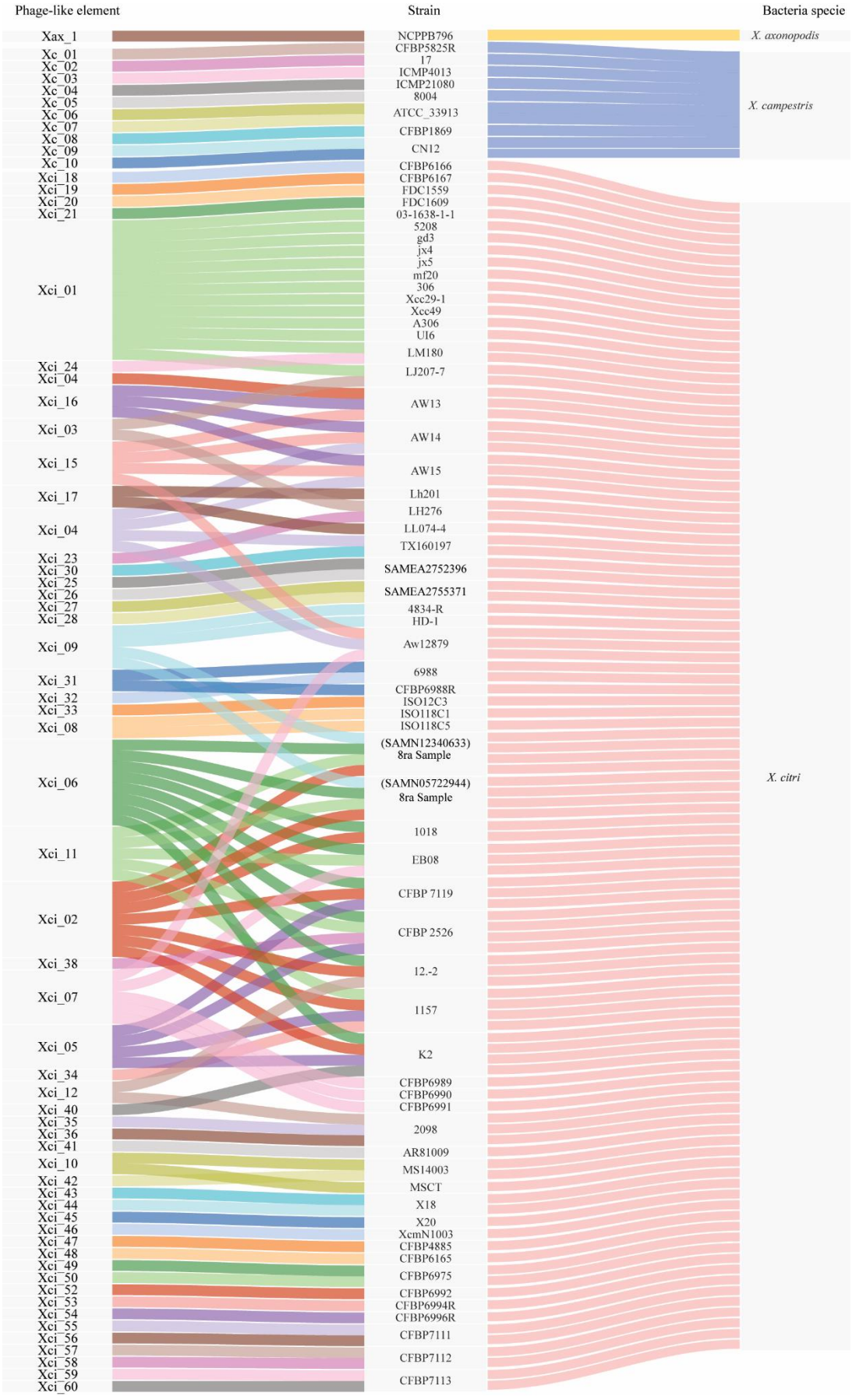

Phage-like element
Strain
Bacteria specie
Xax_1
Xc_01
Xc_02
Xc_03
Xc_04
Xc_05
Xc_06
Xc_07
Xc_08
Xc_09
Xc_10
Xci_18
Xci_19
Xci_20
Xci_21
Xci_01
Xci_24
Xci_04
Xci_16
Xci_03
Xci_15
Xci_17
Xci_04
Xci_23
Xci_30
Xci_25
Xci_26
Xci_27
Xci_28
Xci_09
Xci_31
Xci_32
Xci_33
Xci_08
Xci_06
Xci_11
Xci_02
Xci_38
Xci_07
Xci_05
Xci_34
Xci_12
Xci_40
Xci_35
Xci_36
Xci_41
Xci_10
Xci_42
Xci_43
Xci_44
Xci_45
Xci_46
Xci_47
Xci_48
Xci_49
Xci_50
Xci_52
Xci_53
Xci_54
Xci_55
Xci_56
Xci_57
Xci_58
Xci_59
Xci_60
NCPPB796
CFBP5825R
17
ICMP4013
ICMP21080
8004
ATCC_33913
CFBP1869
CN12
CFBP6166
CFBP6167
FDC1559
FDC1609
03-1638-1-1
5208
gd3
jx4
jx5
mf20
306
Xcc29-1
Xcc49
A306
UI6
LM180
LJ207-7
AW13
AW14
AW15
Lh201
LH276
LL074-4
TX160197
SAMEA2752396
SAMEA2755371
4834-R
HD-1
Aw12879
6988
CFBP6988R
ISO12C3
ISO118C1
ISO118C5
(SAMN12340633)
8ra Sample
(SAMN05722944)
8ra Sample
1018
EB08
CFBP 7119
CFBP 2526
12.-2
1157
K2
CFBP6989
CFBP6990
CFBP6991
2098
AR81009
MS14003
MSCT
X18
X20
XcmN1003
CFBP4885
CFBP6165
CFBP6975
CFBP6992
CFBP6994R
CFBP6996R
CFBP7111
CFBP7112
CFBP7113
X. axonopodis
X. campestris
X. citri

Figure 3

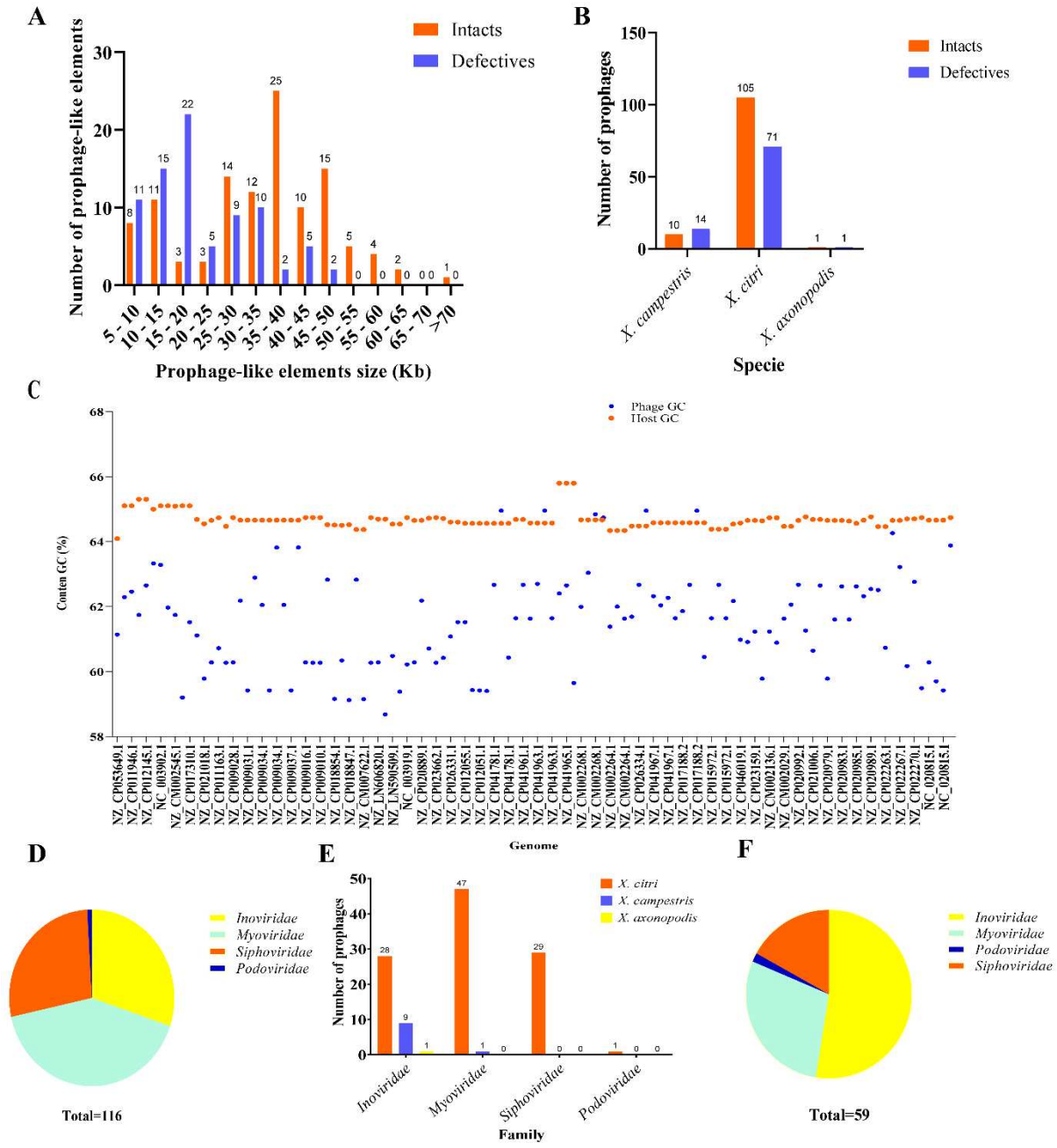

A
Intacts
Defectives
Number of prophage-like elements
Prophage-like elements size (Kb)
B
Intacts
Defectives
Number of prophages
X. campestris
X. citri
X. axonopodis
Specie
C
Phage GC
Host GC
Conten GC (%)
Genome
D
Inoviridae
Myoviridae
Siphoviridae
Podoviridae
Total=116
E
X. citri
X. campestris
X. axonopodis
Number of prophages
Inoviridae
Myoviridae
Siphoviridae
Podoviridae
Family
F
Inoviridae
Myoviridae
Podoviridae
Siphoviridae
Total=59

Figure 4

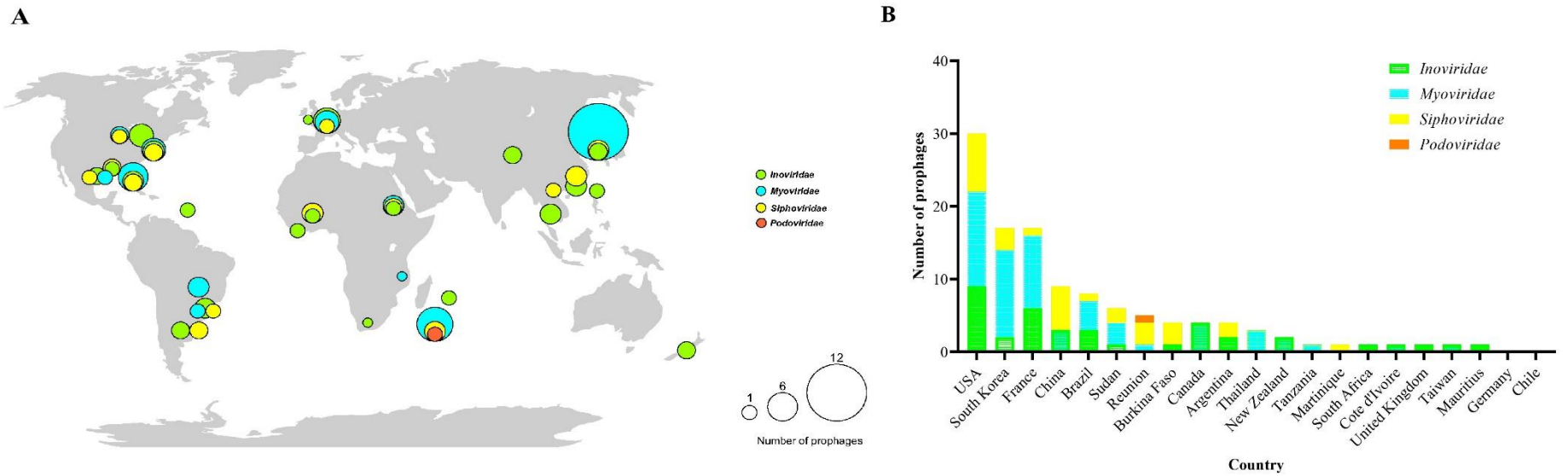
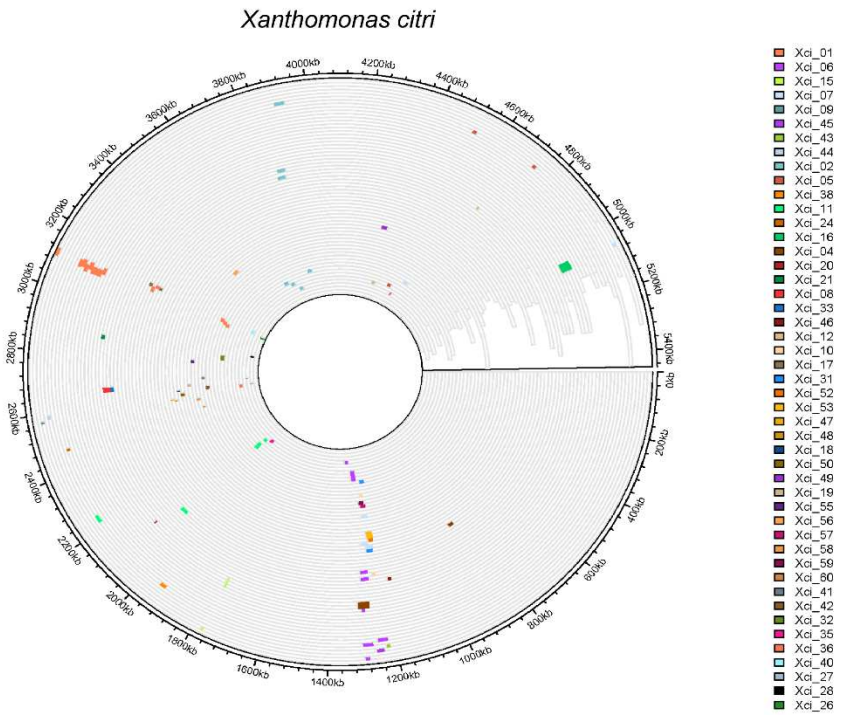

A
Inoviridae
Myoviridae
Siphoviridae
Podoviridae
1
6
12
Number of prophages
B
Inoviridae
Myoviridae
Siphoviridae
Podoviridae
Number of prophages
0
10
20
30
40
USA
South Korea
France
China
Brazil
Sudan
Reunion
Burkina Faso
Canada
Argentina
Thailand
New Zealand
Tanzania
Martinique
South Africa
Cote d'Ivoire
United Kingdom
Taiwan
Mauritius
Germany
Chile
Country

Figure 5

A


Xanthomonas citri
Xci_01
Xci_06
Xci_15
Xci_07
Xci_09
Xci_45
Xci_43
Xci_44
Xci_02
Xci_05
Xci_38
Xci_11
Xci_24
Xci_16
Xci_04
Xci_20
Xci_21
Xci_08
Xci_33
Xci_46
Xci_12
Xci_10
Xci_17
Xci_31
Xci_52
Xci_53
Xci_47
Xci_48
Xci_18
Xci_50
Xci_49
Xci_19
Xci_55
Xci_56
Xci_57
Xci_58
Xci_59
Xci_60
Xci_41
Xci_42
Xci_32
Xci_35
Xci_36
Xci_40
Xci_27
Xci_28
Xci_26

B

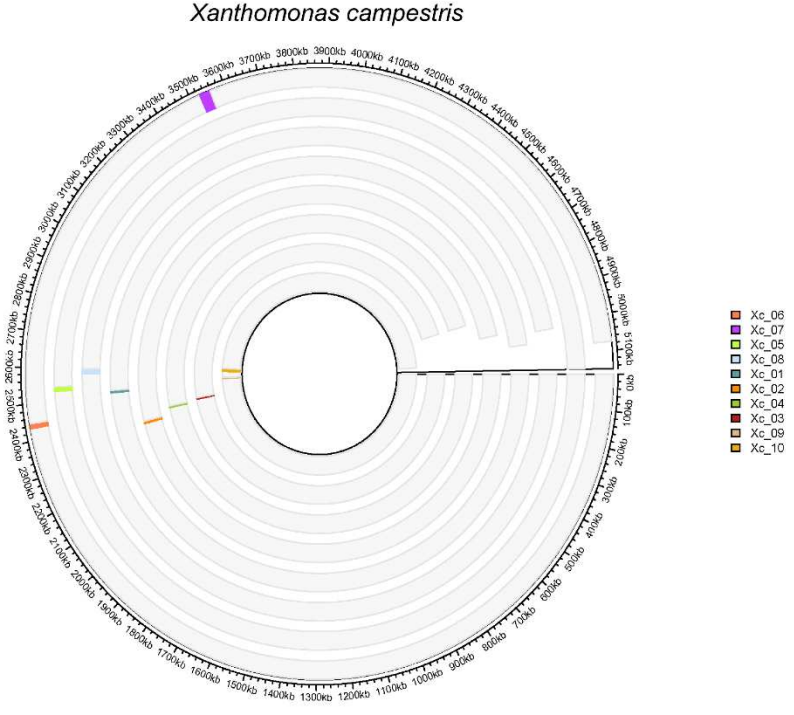

Xanthomonas campestris
Xc_06
Xc_07
Xc_05
Xc_08
Xc_01
Xc_02
Xc_04
Xc_03
Xc_09
Xc_10

Figure 6

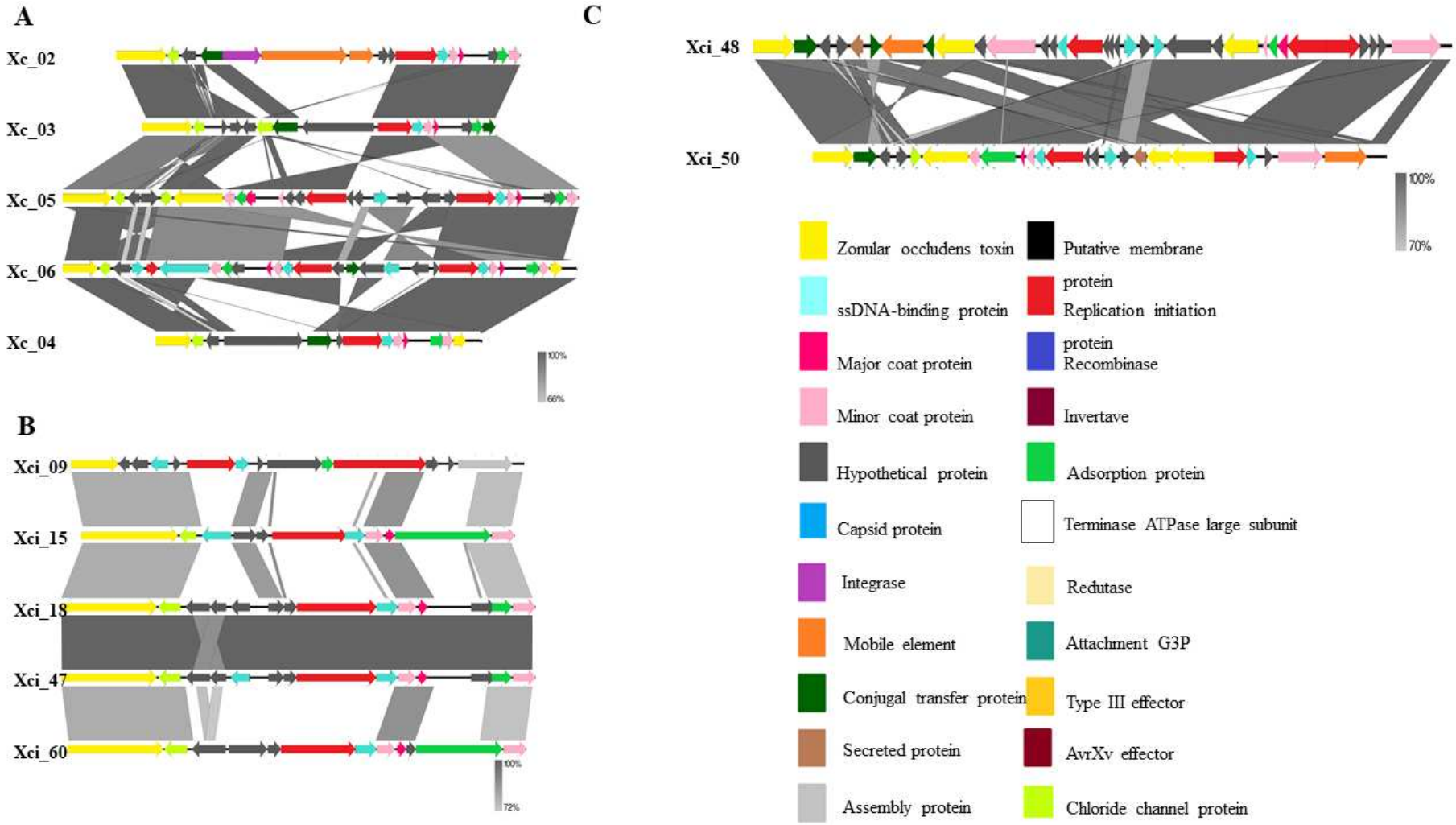

A
Xc_02
Xc_03
Xc_05
Xc_06
Xc_04
100%
66%
B
Xci_09
Xci_15
Xci_18
Xci_47
Xci_60
100%
72%
C
Xci_48
Xci_50
100%
70%
Zonular occludens toxin
Putative membrane protein
ssDNA-binding protein
Replication initiation protein
Major coat protein
Recombinase protein
Minor coat protein
Invertave
Hypothetical protein
Adsorption protein
Capsid protein
Terminase ATPase large subunit
Integrase
Redutase
Mobile element
Attachment G3P
Conjugal transfer protein
Type III effector
Secreted protein
AvrXv effector
Assembly protein
Chloride channel protein

Figure 7

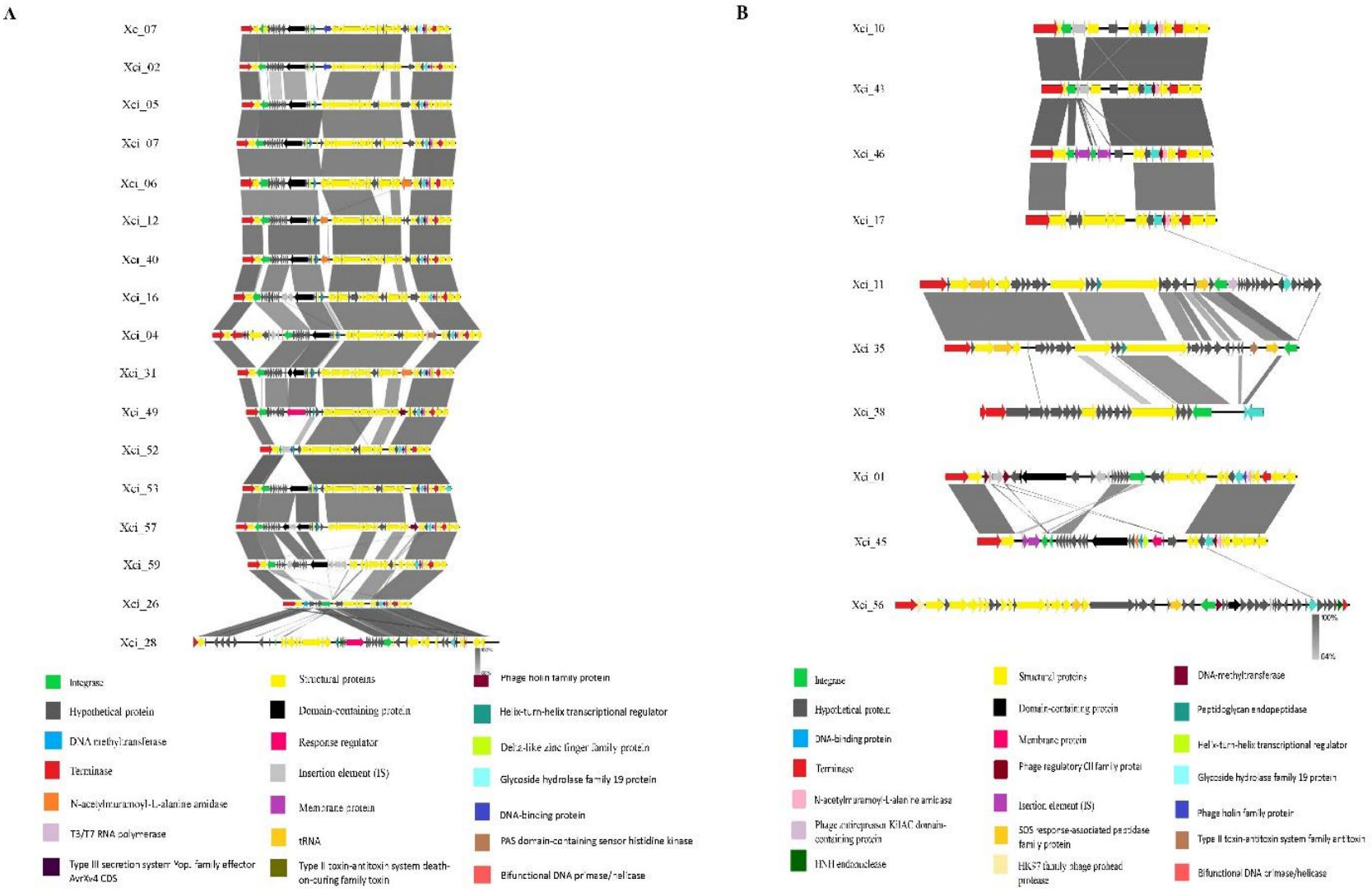

A
Xc_07
Xci_02
Xci_05
Xci_07
Xci_06
Xci_12
Xci_40
Xci_16
Xci_04
Xci_31
Xci_49
Xci_52
Xci_53
Xci_57
Xci_59
Xci_26
Xci_28
100%
60%
Integrase
Hypothetical protein
DNA methyltransferase
Terminase
N-acetylmuramoyl-L-alanine amidase
T3/T7 RNA polymerase
Type III secretion system YopJ family effector AvrXv4 CDS
Structural proteins
Domain-containing protein
Response regulator
Insertion element (IS)
Membrane protein
tRNA
Type II toxin-antitoxin system death-on-curing family toxin
Phage holin family protein
Helix-turn-helix transcriptional regulator
Delta-like zinc finger family protein
Glycoside hydrolase family 19 protein
DNA-binding protein
PAS domain-containing sensor histidine kinase
Bifunctional DNA primase/helicase
B
Xci_10
Xci_43
Xci_46
Xci_17
Xci_11
Xci_35
Xci_38
Xci_01
Xci_45
Xci_56
100%
64%
Integrase
Hypothetical protein
DNA-binding protein
Terminase
N-acetylmuramoyl-L-alanine amidase
Phage antirepressor KilAC domain-containing protein
HNH endonuclease
Structural proteins
Domain-containing protein
Membrane protein
Phage regulatory CII family protein
Insertion element (IS)
SDS response-associated peptidase family protein
HK97 family phage prohead protease
DNA-methyltransferase
Peptidoglycan endopeptidase
Helix-turn-helix transcriptional regulator
Glycoside hydrolase family 19 protein
Phage holin family protein
Type II toxin-antitoxin system family antitoxin
Bifunctional DNA primase/helicase

Figure 8

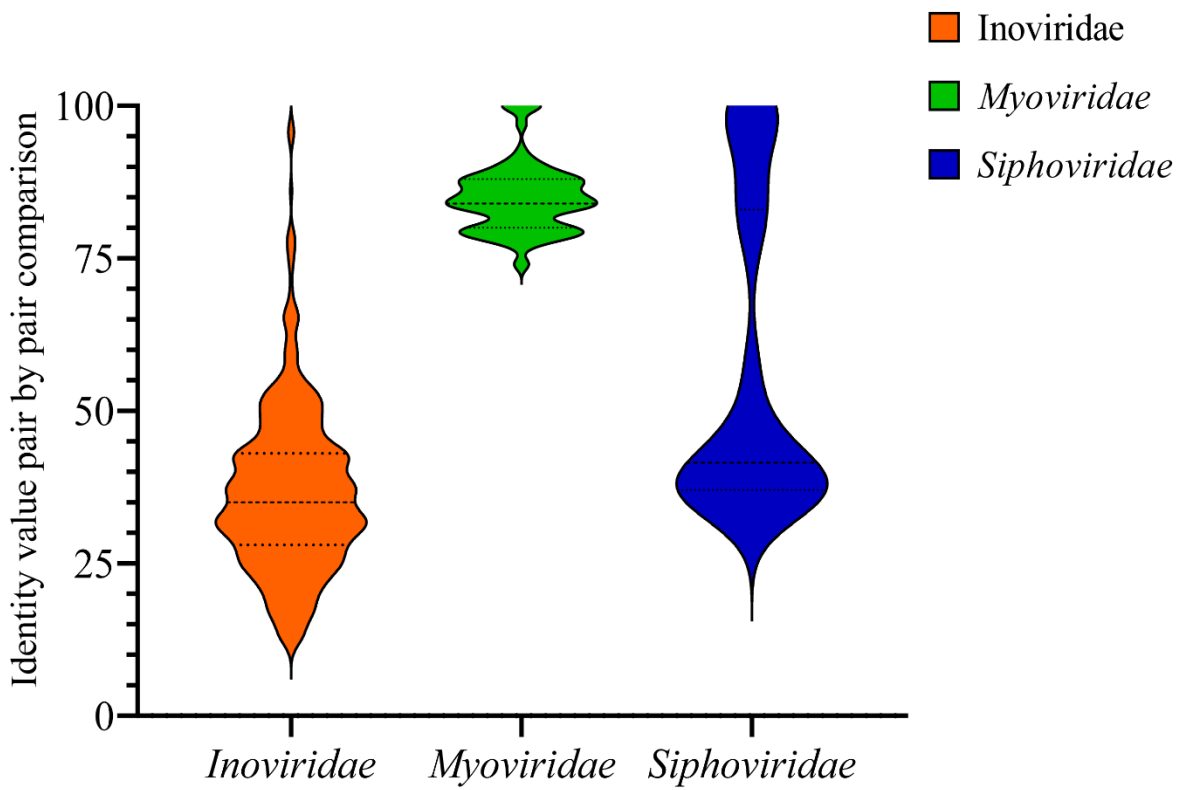
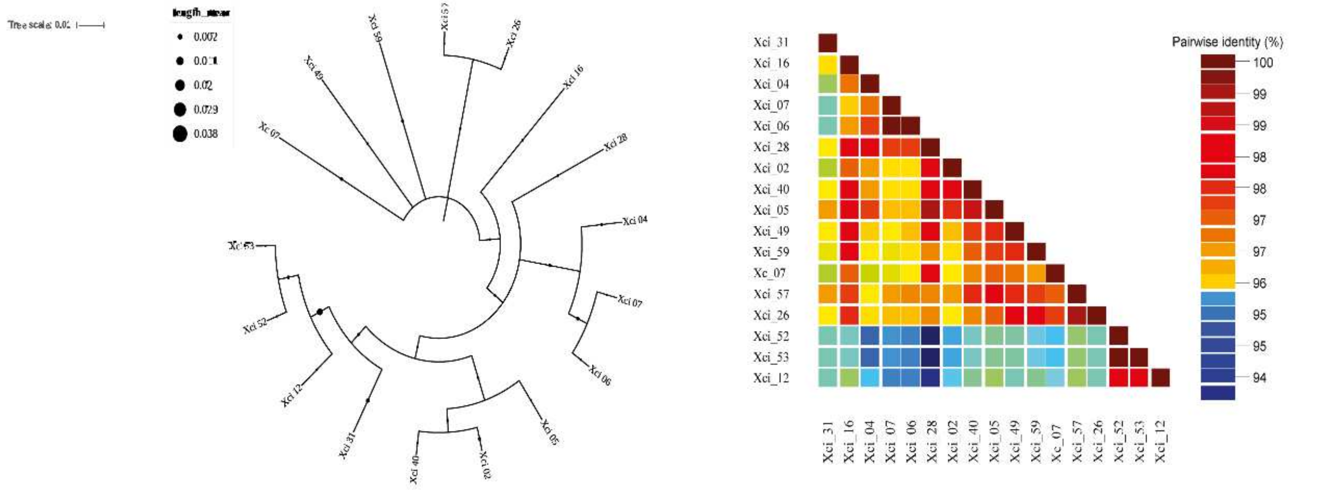

Inoviridae
Myoviridae
Siphoviridae
Identity value pair by pair comparison
100
75
50
25
0
Inoviridae
Myoviridae
Siphoviridae

Figure 9

A


Pairwise identity (%)
100
99
99
98
98
97
97
96
95
95
94
Xci_31
Xci_16
Xci_04
Xci_07
Xci_06
Xci_28
Xci_02
Xci_40
Xci_05
Xci_49
Xci_59
Xc_07
Xci_57
Xci_26
Xci_52
Xci_53
Xci_12

B

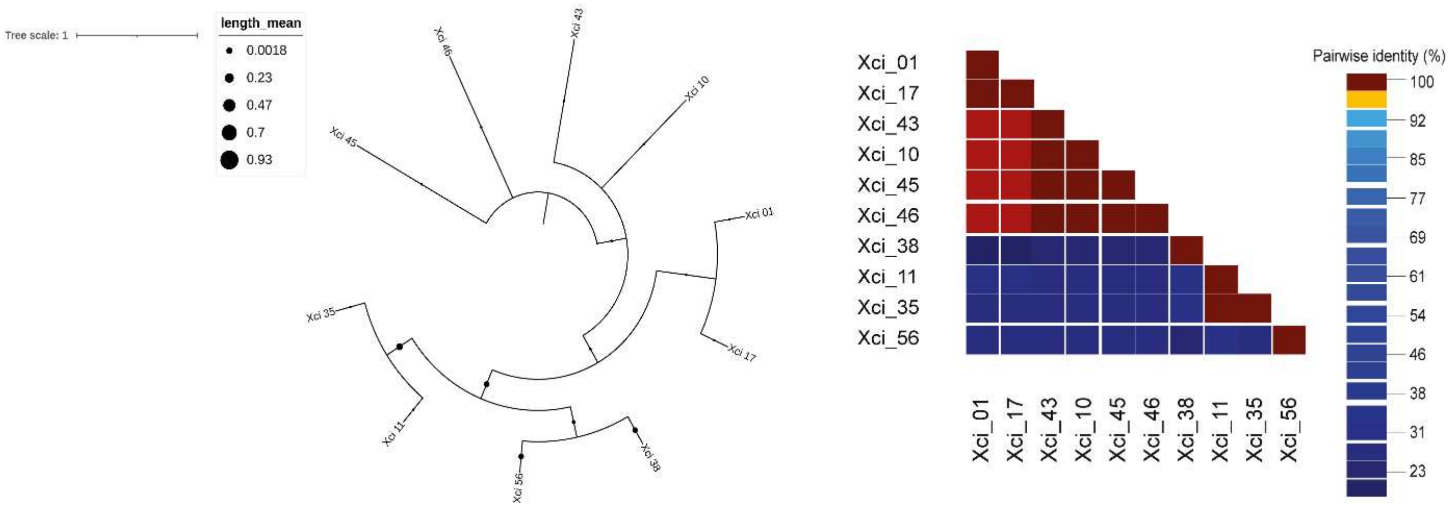

Tree scale: 1
length_mean
0.0018
0.23
0.47
0.7
0.93
Xci 43
Xci 46
Xci 10
Xci 45
Xci 01
Xci 35
Xci 17
Xci 11
Xci 56
Xci 38
Pairwise identity (%)
100
92
85
77
69
61
54
46
38
31
23
Xci_01
Xci_17
Xci_43
Xci_10
Xci_45
Xci_46
Xci_38
Xci_11
Xci_35
Xci_56

Figure 10

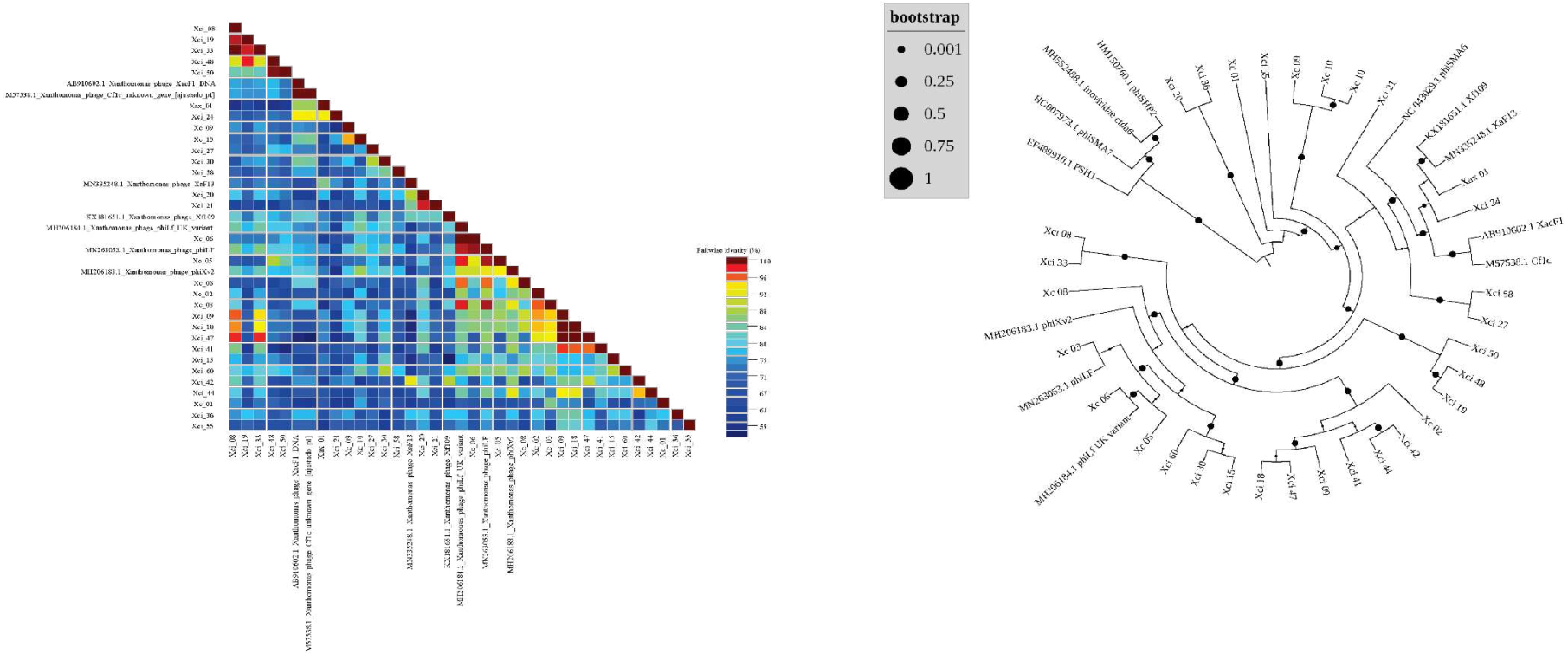
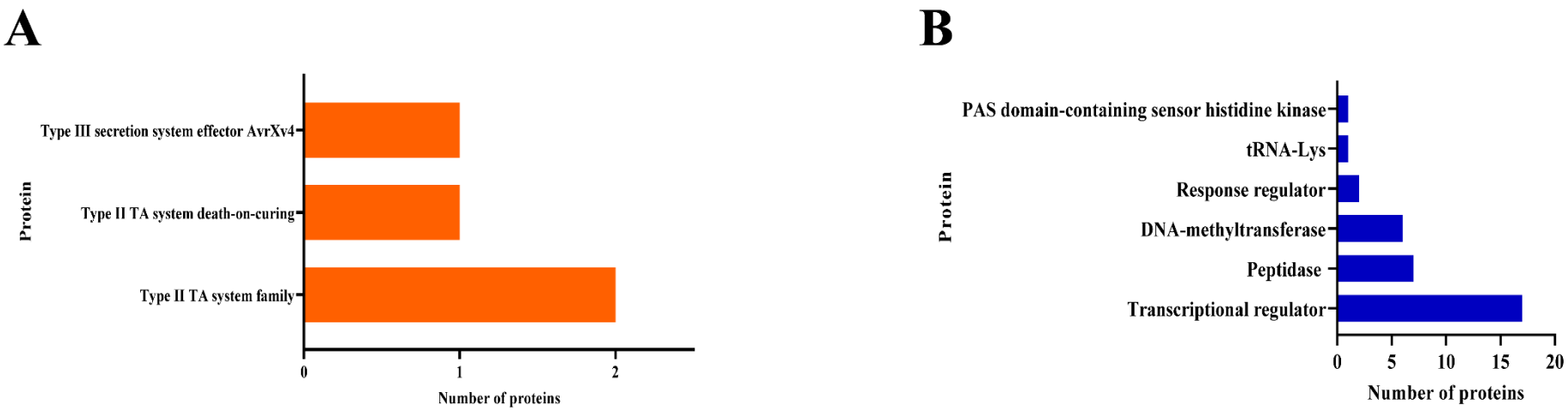
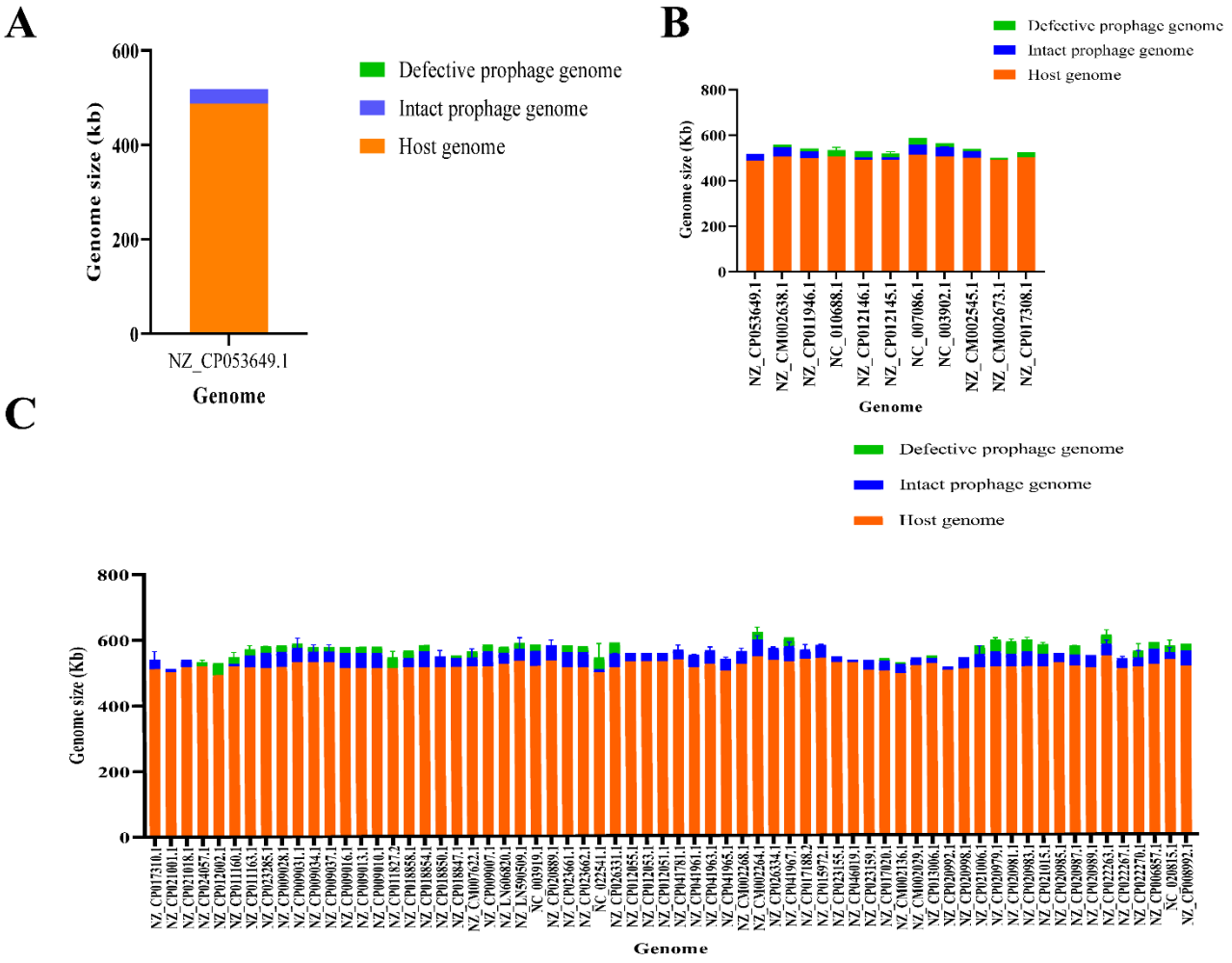

bootstrap
0.001
0.25
0.5
0.75
1
Pairwise identity (%)
100
96
92
88
84
80
75
71
67
63
59
Xci_08
Xci_19
Xci_33
Xci_48
Xci_50
AB910602.1_Xanthomonas_phage_XacF1_DNA
M57538.1_Xanthomonas_phage_CflC_unknown_gene_[ajuzado_pl]
Xax_01
Xci_24
Xc_09
Xc_10
Xci_27
Xci_30
Xci_58
MN335248.1_Xanthomonas_phage_XafF13
Xci_20
KX181651.1_Xanthomonas_phage_Xf109
MH206183.1_Xanthomonas_phage_phiLf_UK_variant
Xc_06
MN263053.1_Xanthomonas_phage_phiLF
Xc_05
MH206183.1_Xanthomonas_phage_phiXv2
Xc_08
Xc_02
Xc_03
Xci_09
Xci_18
Xci_47
Xci_41
Xci_15
Xci_60
Xci_42
Xci_44
Xc_01
Xci_36
Xci_55
HG007973.1 phiSMA7
EF489910.1 PSH1
MH552408.1 brevidoe cluster
HM150760.1 phiSHP2
NC 042829.1 phiSMA6
KX181651.1 Xf109
MN335248.1 XafF13
AB910602.1 XacF1
M57538.1 CflC
MH206184.1 phiLf UK variant

Figure 11

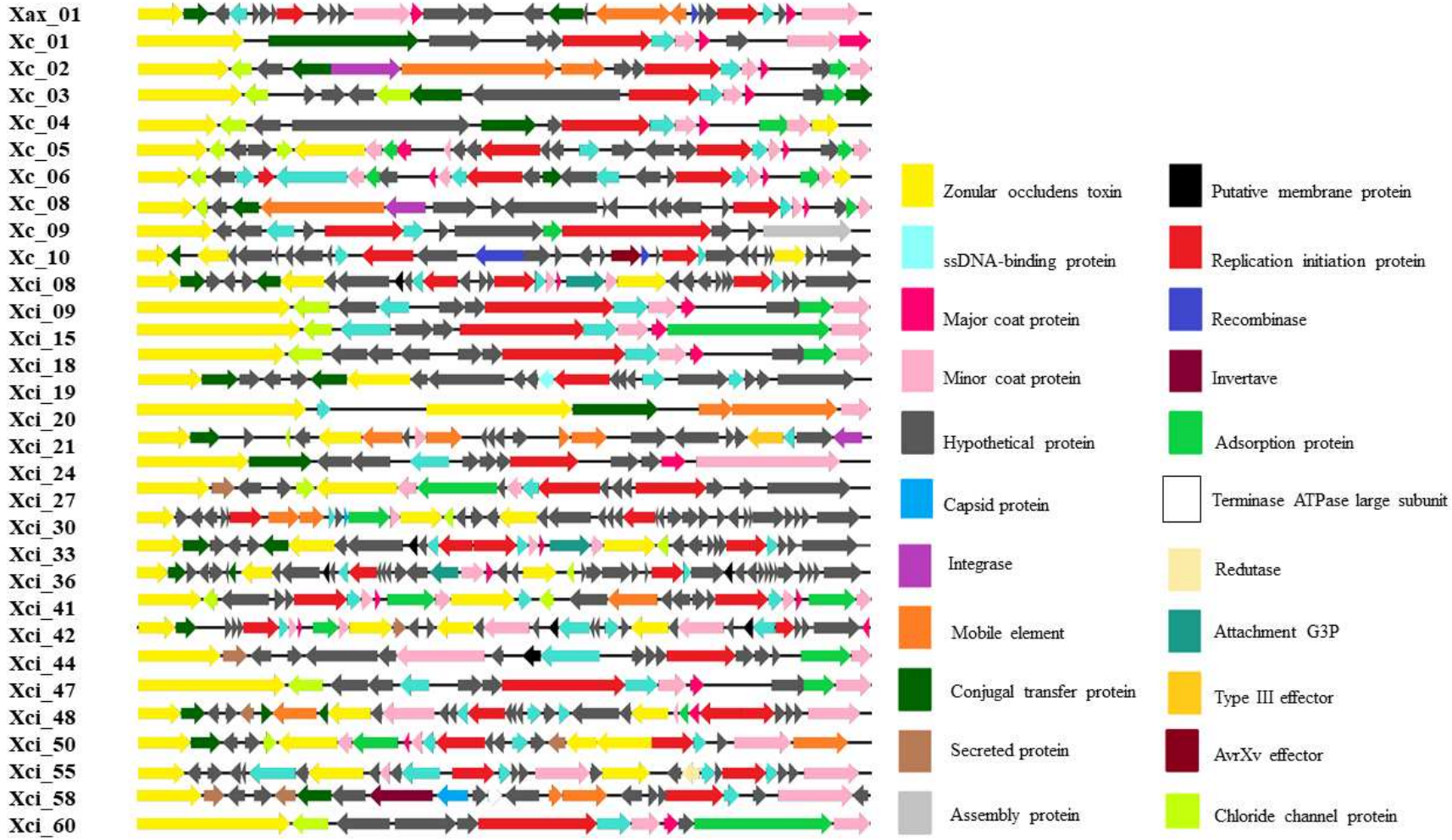

A
Type III secretion system effector AvrXv4
Type II TA system death-on-curing
Type II TA system family
Protein
0
1
2
Number of proteins
B
PAS domain-containing sensor histidine kinase
tRNA-Lys
Response regulator
DNA-methyltransferase
Peptidase
Transcriptional regulator
Protein
0
5
10
15
20
Number of proteins

SUPPLEMENTARY FIGURES

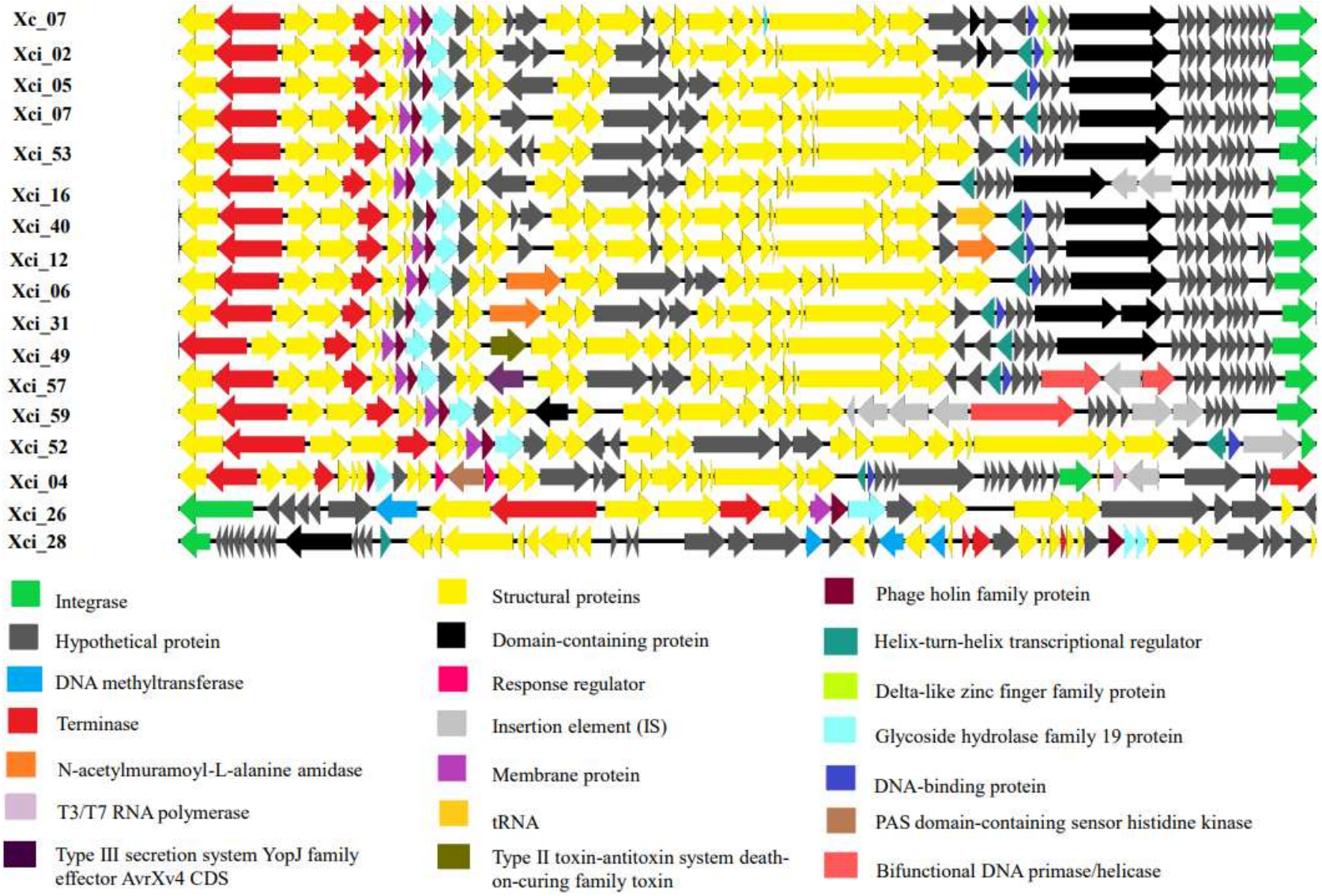
Supplementary figure 1



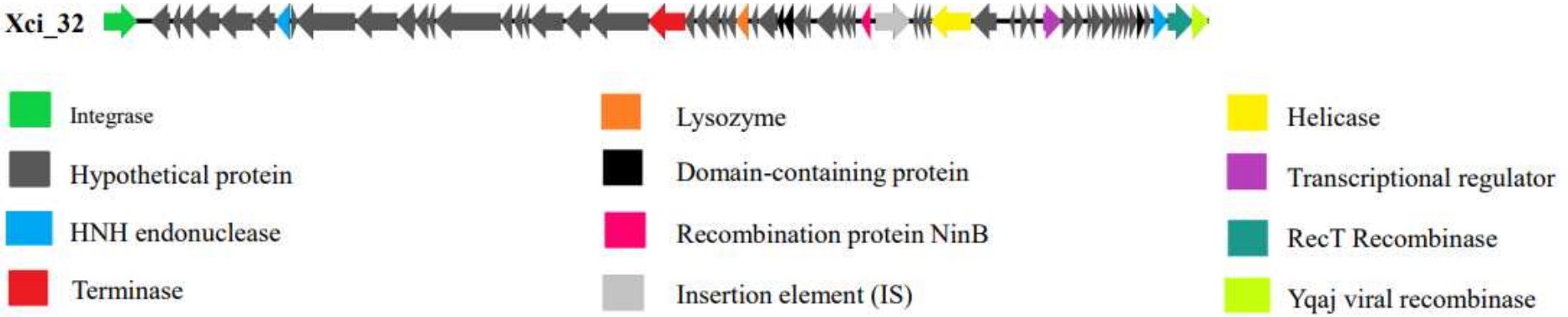
Supplementary figure 2



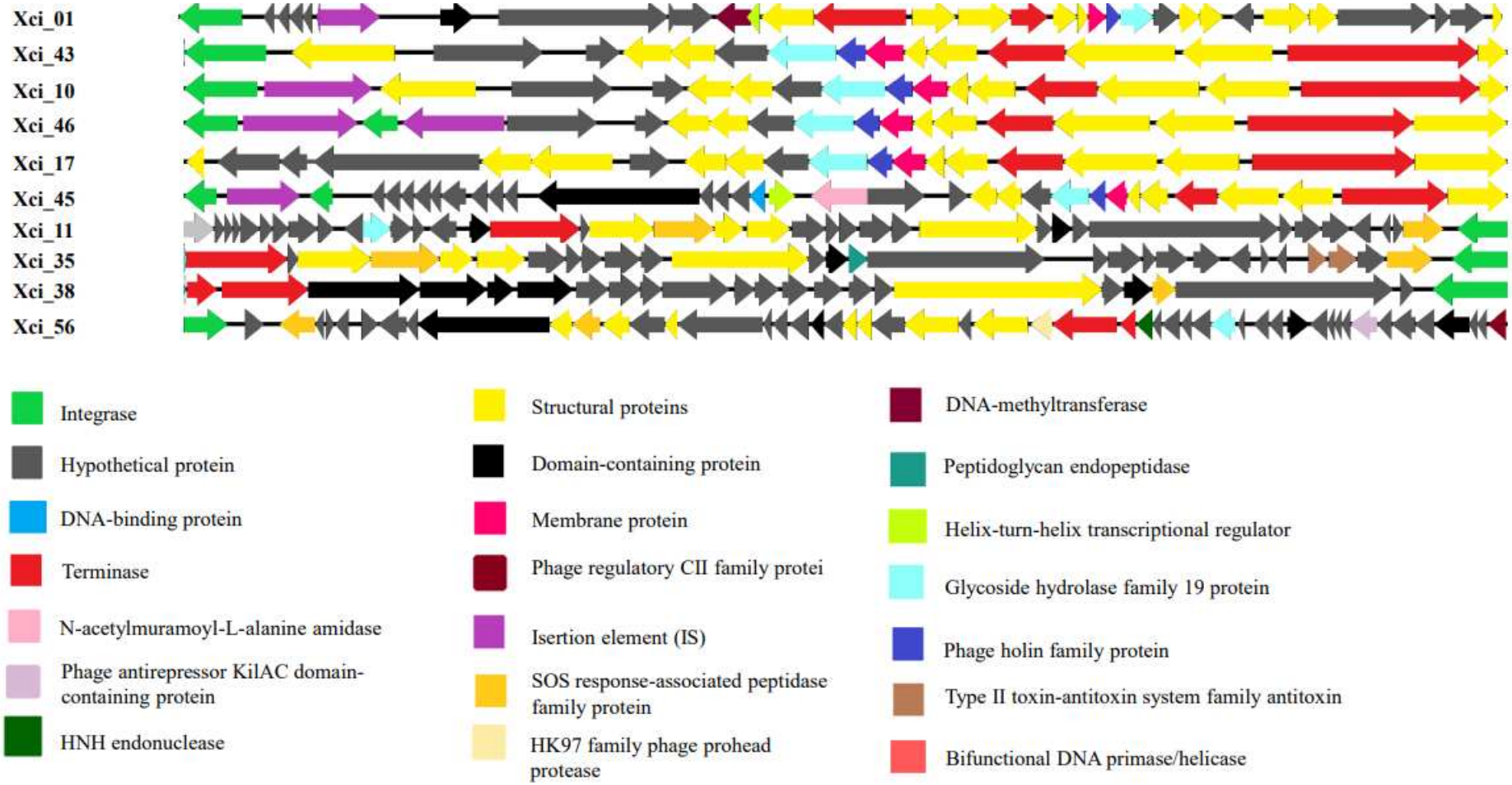
Supplementary figure 3



Supplementary figure 4



Supplementary figure 5



Supplementary figure 6