UNIVERSIDADE FEDERAL DE VIÇOSA

ANALYSIS OF PROPHAGES IN *Erwinia* spp. GENOMES

Thamylles Thuany Mayrink Lima

Magister Scientiae

VIÇOSA - MINAS GERAIS

2020

THAMYLLES THUANY MAYRINK LIMA

ANALYSIS OF PROPHAGES IN *Erwinia* spp. GENOMES

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola para obtenção do título de *Magister Scientiae*.

Orientadora: Poliane Alfenas-Zerbini

Coorientador: Renan de Souza Cascardo

VIÇOSA - MINAS GERAIS

2020

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

T
L732a
2020

Lima, Thamylles Thuany Mayrink, 1996-
Analysis of prophages in *Erwinia* spp. genomes / Thamylles Thuany Mayrink Lima. – Viçosa, MG, 2020.
73 f. : il. (algumas color.) ; 29 cm.

Orientador: Poliane Alfenas-Zerbini.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Bacteriófagos. 2. Lisogenia. 3. Bactérias fitopatogênicas. 4. CRISPR (Genética). I. Universidade Federal de Viçosa. Departamento de Microbiologia. Programa de Pós-Graduação em Microbiologia Agrícola. II. Título.

CDD 22. ed. 579.26

THAMYLLES THUANY MAYRINK LIMA

ANALYSIS OF PROPHAGES IN *Erwinia* spp. GENOMES

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola para obtenção do título de *Magister Scientiae*.

APROVADA: 19 de fevereiro de 2020.

Assentimento:

Thamylles Thuany Mayrink Lima
Autora

Poliane Alfenas-Zerbini
Orientadora

Dedico

A Deus, minha força diária;

Aos meus amados pais, Ângela e Flávio;

Ao meu irmão Patrick;

À minha sobrinha Ana Cecília, meu amor mais sincero;

Ao meu noivo Igor;

À toda minha família e amigos.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me sustentado nessa longa e exaustiva jornada. Nada do que sou, do que tenho e do que faço seria possível se não fosse pela presença dEle em minha vida. Todos os dias Ele me guia, me ilumina e me ajuda a superar os maiores obstáculos;

Agradeço aos meus pais, Ângela e Flávio, razões da minha vida. Muitas vezes vocês abdicaram dos próprios sonhos para realizarem os meus e sempre acreditaram em mim. Essa vitória é mais de vocês do que minha. Vocês são minha certeza de que em momento algum estarei sozinha;

Agradeço à Ana Cecília, meu amor mais puro, por ser minha fortaleza nos dias mais difíceis. Inúmeras foram as vezes em que o desespero e o desânimo batiam e por uma simples ligação ela me dava forças para continuar;

Ao meu irmão, Patrick, pela cumplicidade e amizade e por ter me dado o maior presente da minha vida inteira;

Agradeço imensamente à minha avó, Erondina, que mesmo ausente fisicamente, está sempre presente no meu coração. Vó, sinto sua falta todos os dias. Tudo que eu queria era que você ainda estivesse aqui, para ver de pertinho tudo que venho conquistado. Obrigada por se fazer presente em todos os meus dias. Te amo infinito!

Agradeço ao meu noivo, Igor, por compartilhar essa jornada ao meu lado, me suportando nos dias em que nem eu mesma me suportava, por todo carinho, paciência, compreensão e ajuda. Você é a resposta de Deus às minhas orações. Gratidão por tudo. Amo você!

Aos meus afilhados, Bianca, Maria Eduarda e Oliver, por serem minha alegria mesmo nos dias ruins;

À toda minha família, que sempre foram meus maiores incentivadores e apoiadores de todos os meus sonhos;

Aos meus sogros, por me acolherem de forma que nunca vou conseguir explicar como verdadeiramente parte da família e por todo carinho e apoio nessa jornada;

À minha orientadora Poliane, pela orientação, incentivo e paciência e compreensão. Agradeço imensamente por todo o ensino, dedicação e por sempre me tratar com respeito e carinho;

Aos meus Co-orientadores, Renan e Tiago, que deram suporte em todas as etapas desse trabalho. Sem a ajuda de vocês, certamente a realização desse trabalho teria sido muito mais complexa;

Aos integrantes do laboratório de Vírus: Coxinha, gratidão por todas as análises que sofremos, batemos cabeça por dias, mas que no final saiu; Rafa, gratidão pelo auxílio em algumas análises e, principalmente, pela ajuda na interpretação dos resultados; Fernandinha, gratidão por sempre ter mais que um minutinho quando eu te pedia ajuda "rapidinho, 1 minutinho só, rs"; Bruna e Nataly, gratidão por estarem comigo nos vários momentos de lágrimas, angústia e desespero, por sempre me apoiarem e tornarem os dias dessa jornada mais fáceis com a companhia de vocês; Paty, gratidão pelos momentos de descontração, suas várias pérolas tornaram os dias muito mais divertidos; aos estagiários Lorhan, Luan, Sávio e Pâmela pela amizade e os vários memes trocados;

À Ana Flávia, Açucena, Fernanda, Júlia, Mísia, Paula, Raquel, Thais e Vanessa, por serem minha âncora quando tudo parecia que ia desmoronar. Vocês são as melhores amigas que eu poderia ter! Obrigada por se fazerem sempre presente, mesmo que muitas vezes eu esteja tão ausente;

À Isadora, uma das maiores amizades que 2019 trouxe para minha vida. Gratidão por me escutar repetida vezes sobre o mesmo assunto e por compreender tão perfeitamente quase tudo que se passa no meu coração;

Às Bioagirls e aos meus 2 meninos, João Marcos e Yan, por dividirem toda a tensão do mestrado, pelos almoços mais divertidos (e comunitários, rs) e pelas palavras amigas quando eu mais precisei. Vocês são o melhor presente que o mestrado poderia ter me dado!

Aos amigos do Bioagro, os quais não vou colocar nomes por medo de esquecer de citar algum. Gratidão pelas risadas, conversas jogadas fora, discussões e amparo nos momentos conflitantes;

À Jéssica, pela compreensão e apoio e por tornar nosso lar tão aconchegante;

Aos meus professores, tanto do Ensino Médio quanto da Graduação, por serem meus exemplos da profissional que devo e quero ser;

Ao Professor Mateus, pela boa vontade em me esclarecer algumas dúvidas e pela solicitude quando precisei em uma análise;

À Professora Denise, por me impulsionar para frente e por dizer várias vezes o quanto acreditava em meu potencial, por vezes me dizendo as palavras que precisava escutar naquele momento;

Aos funcionários do Departamento de Microbiologia: Gabriel, Sandra, Letícia, Sr. Paulo, Célio, Igor, Regiane, Raquel, Emília, Michelle, gratidão pelas infinitas ajudas e por sempre se mostrarem tão solícitos;

Aos funcionários da DTI, gratidão por me ajudarem a solucionar todos os meus problemas (que não foram poucos, rs) no cluster, sempre com a maior boa vontade do mundo;

À Carol, da PPG, por toda paciência nas correções da versão final;

À Universidade Federal de Viçosa e ao Departamento de Microbiologia pela infraestrutura para realização da minha dissertação;

Ao CNPq pela concessão da bolsa de estudos;

À Suzano, pelo financiamento dos nossos projetos;

A todos que torceram por mim nesta batalha e que hoje alegram-se junto comigo pela conquista de mais um objetivo;

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse uma gota”.

(Madre Teresa de Calcutá)

BIOGRAFIA

THAMYLLES THUANY MAYRINK LIMA, filha de Flávio de Souza Lima e Ângela Maria Mayrink e Lima, nasceu em Caratinga – MG, em 14 de setembro de 1996.

Em 2015 iniciou o curso de Ciências Biológicas na Universidade Federal de Viçosa, graduando em Ciências Biológicas nas modalidades Bacharelado e Licenciatura em Julho de 2018. Durante sua graduação foi bolsista do Programa de Educação Tutorial em Biologia (PET-Bio) por 3 anos, voluntária nos projetos de extensão Bioenlace também por 3 anos e Jovem Cientista por 6 meses.

Em agosto de 2018 ingressou no Programa de Pós-Graduação em Microbiologia Agrícola na Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em Fevereiro de 2020.

Durante seu mestrado, ministrou aulas na disciplina Microbiologia Geral (MBI 100) como exigência do Estágio em Ensino, onde despertou ainda mais seu amor e encanto pela docência, motivo pelo qual deseja aprimorar cada vez mais seu conhecimento e adquirir maior experiência, para assim contribuir com um ensino de qualidade para seus futuros alunos.

RESUMO

LIMA, Thamylles Thuany Mayrink Lima, M.Sc., Universidade Federal de Viçosa, fevereiro de 2020. **Análises de profagos em genomas de *Erwinia s*pp.** Orientadora: Poliane Alfenas-Zerbini. Coorientador: Renan de Souza Cascardo.

Os bacteriófagos podem ser encontrados em muitos ambientes, incluindo água, solo, sistema gastrointestinal e integrado aos genomas do hospedeiro, como profagos. Os profagos podem ter um efeito positivo, neutro ou negativo para seu hospedeiro. Podem afetar a variabilidade genética do genoma bacteriano, conferir vantagem seletiva ao hospedeiro, transportar fatores de virulência, introduzir novos genes de patogenicidade e conferir novas propriedades fenotípicas a seus hospedeiros, contribuindo para a aptidão. Bactérias do gênero *Erwinia* são patógenos que infectam um grande número de importantes culturas agrícolas. Estudos sobre profagos neste grupo de bactérias são pouco explorados. Neste trabalho, examinamos profagos nos genomas de *Erwinia* spp. Sequências de bacteriófagos na forma de profagos foram identificadas em 100% dos genomas analisados. Mais de um profago foi encontrado no mesmo genoma, caracterizando polilisogenia. A análise de profagos intactos sugere que são espécies não caracterizadas, pois mostraram baixa identidade com genomas de vírus em bancos de dados públicos. Genes bacterianos foram encontrados nos profagos integrados, sugerindo transferência horizontal de genes e genes que codificam fatores que podem conferir características especiais ao hospedeiro também foram encontrados. Detectamos sequências espaçadoras de fagos no locus CRISPR dos genomas de *Erwinia* spp. mas não encontramos espaçadores para os profagos residentes nos genomas analisados, sugerindo que os profagos residentes no genoma não foram reconhecidos pelo sistema CRISPR-Cas. Nossos resultados mostram que os bacteriófagos estão amplamente distribuídos nos genomas de Erwinia spp. e que podem estar contribuindo com as características fenotípicas do seu hospedeiro.

Palavras-chave: PHASTER; Lisogenia; CRISPR-Cas; Bacteriófagos; Bactéria fitopatogênica.

ABSTRACT

LIMA, Thamylles Thuany Mayrink Lima, M.Sc., Universidade Federal de Viçosa, February, 2020. **Analysis of prophages in genomes of *Erwinia* spp.** Adviser: Poliane Alfenas-Zerbini. Co-adviser: Renan de Souza Cascardo.

Bacteriophages can be found in many environments, including water, soil, gastrointestinal system and integrated in host genomes, such as prophages. Prophages can have a positive, neutral or negative effect on their host. They can affect the genetic variability of the bacterial genome, confer a selective advantage to the host, transport virulence factors, introduce new pathogenicity genes and confer new phenotypic properties to their hosts, contributing to their fitness. Bacteria of the genus *Erwinia* are pathogens that infect a large number of important agricultural crops. Studies on prophages in this group of bacteria are poorly explored. In this work, we examine phages in genomes of *Erwinia* spp. Phage sequences were identified in 100% of the analyzed genomes. More than one prophage was found in the same genome, characterizing polylisogeny. The analysis of intact phages suggests that they are uncharacterized species, as they showed low identity with virus genomes in public databases. Bacterial genes were found in the phages, assuming horizontal transfer of genes and genes that encode factors that may confer special characteristics to the host were also found. We detected phage spacer sequences at the CRISPR locus of the genomes of *Erwinia* spp. but we did not find spacers for the prophage residing in the analyzed genomes, suggesting that the prophage residing in the genome were not recognized by the CRISPR-Cas system. Our results show that bacteriophages are widely distributed in the genomes of *Erwinia* spp. and that may be contributing to your host's fitness.

Keywords: PHASTER. Lysogeny. CRISPR-Cas. Bacteriophages. Phytopathogenic bacteria.

SUMÁRIO

INTRODUÇÃO GERAL

Os vírus são micro-organismos que infectam todos os organismos vivos. Os vírus que infectam bactérias são denominados bacteriófagos, sendo também referidos como “fagos” (Simmonds et al., 2017). Os bacteriófagos são considerados os micro-organismos mais abundantes na Terra, podendo ser isolados em diferentes ambientes, como esgoto, fezes, solo ou água (Ul Haq et al., 2012). Eles podem ter efeito positivo, neutro ou negativo para seus hospedeiros. Podem contribuir de diversas maneiras nos diferentes ecossistemas, atuando como agente mutagênico biológico, realizando transferência horizontal de material genético, afetando os fenótipos de seus hospedeiros e conferindo plasticidade genômica (Orlova, 2012; Ramisetty and Sudhakari, 2019). Os profagos também são capazes de impactar negativamente seu hospedeiro, por meio do aumento do custo metabólico e pela interrupção de genes importantes do hospedeiro no momento da integração (Fillol-Salom et al., 2019).

Os bacteriófagos podem se multiplicar de diferentes maneiras, possuindo ciclos de replicação denominados crônico, pseudolisogênico, lítico e lisogênico (Weinbauer, 2004). Infecções crônicas ocorrem quando uma célula é infectada e a progênie viral é constantemente liberada por brotamento ou extrusão, sem que haja lise celular, como ocorre em alguns fagos filamentososos e fagos que infectam *Mycobacterium* spp., que são liberados da célula hospedeira sem destruí-la. No ciclo pseudolisogênico, a infecção ocorre em parte da população e o genoma viral não é integrado no bacteriano, podendo manter-se na forma de um plasmídeo, comumente descrito nos vírus que infectam *Pseudomonas aeruginosa* e espécies do gênero *Halobacterium* (Weinbauer, 2004).

No ciclo lítico, a expressão dos genes virais é regulada de modo a redirecionar a maquinaria de síntese do hospedeiro, favorecendo a síntese de ácidos nucleicos e proteínas virais. Os novos vírions são montados e liberados como resultados da lise (morte) celular, mediada por hidrolases e holinas, como no caso dos bacteriófagos com cauda T-pares (T2, T4 e T6) (Ackermann, 2001; Madigan et al., 2016). A lise celular pode ser visualizada como uma zona clara em placa de Petri com uma camada de células bacterianas em crescimento, convencionalmente denominadas de “placa de lise” (Gallet et al., 2011). Os bacteriófagos que adotam o ciclo lítico podem ser utilizados como ferramenta biotecnológica no controle de doenças bacterianas (Ahern *et al.*, 2014).

No ciclo lisogênico (ou temperado), os bacteriófagos se mantêm de forma estável no hospedeiro, assumindo um estado denominado de lisogenia, onde o genoma viral se integra no cromossomo do hospedeiro formando o "profago". O profago é replicado juntamente com o cromossomo do hospedeiro sem interferir no seu ciclo e pode ser transferido para as células filhas na divisão celular. Esse processo é possível, desde que a expressão dos genes que controlam as funções líticas do bacteriófago permaneça reprimida. Contudo, em determinadas condições, como a exposição a agentes mutagênicos químicos, raios ultravioletas (UV) ou ionizantes e dessecação, os fagos lisogênicos podem reverter à via lítica.

Aproximadamente 25% dos fagos na Terra existem na forma de profagos (Bondy-Denomy and Davidson, 2014), os quais podem carrear uma ampla variedade de genes que conferem vantagem para seu hospedeiro. Por exemplo, o profago Gifsy-2 de *Salmonella* Typhimurium codifica o gene *sod*C, que codifica para superóxido dismutase usada para neutralizar espécies reativas de oxigênio (Figueroa-Bossi and Bossi, 1999). Existem profagos que possuem a capacidade de aumentar a patogenicidade do seu hospedeiro, como é o caso do fago λ de *Escherichia coli* que codifica a proteína Lom, a qual aumenta a habilidade da bactéria em se aderir ao epitélio bucal de humanos (Bondy-Denomy and Davidson, 2014).

Sabe-se que os genomas de organismos celulares podem conter material genético de origem viral, sendo esses funcionais ou não. Essa porção referente aos vírus pode constituir uma porcentagem considerável do genoma do hospedeiro, por exemplo, alguns genomas bacterianos podem ter até 20% de genes de origem viral (Czajkowski, 2019). Atualmente, em muitos genomas bacterianos depositados nos bancos de dados, é possível detectar fagos integrados nos cromossomos das bactérias. Em um estudo realizado com *Flavobacterium psychrophilum* 49 cepas foram analisadas e em 80% dessas profagos foram encontrados (Castillo et al., 2014). Em bactérias não cultivadas obtidas de amostras de solo, constatou-se que aproximadamente 80% dessas bactérias eram lisogênicas (Ghosh et al., 2008). Da mesma forma, humanos também possuem vírus integrados constituindo parte do seu genoma, e estima-se que além desses vírus integrados, cada indivíduo saudável possa abrigar mais de dez infecções virais crônicas, assim como um grande número de vírus não caracterizados que podem residir em nós (Handley, 2016). Esses dados nos mostram a onipresença dos vírus, o que desperta um grande interesse em se analisar os vírus integrados que compõem o genoma dos diferentes organismos celulares.

Com o avanço das tecnologias de sequenciamento, o número crescente de genomas bacterianos sequenciados e o desenvolvimento de ferramentas de bioinformática mais poderosas, tem-se observado que vírus integrados afetam a arquitetura do genoma bacteriano e que eles representam uma fração significativa do conteúdo genético da bactéria (Czajkowski, 2019). Os profagos demonstram múltiplos e complexos mecanismos pelos quais podem influenciar a sobrevivência e patogenicidade das espécies hospedeiras (Bondy-Denomy et al., 2016), podendo causar alterações fenotípicas, levando a aumento no fitness e virulência, sendo esse efeito denominado conversão lisogênica (Bondy-Denomy and Davidson, 2014).

A lisogenia possui grande importância para a recombinação gênica, conferindo novas propriedades para células bacterianas como, por exemplo, virulência para bactérias patogênicas. Neste caso, os bacteriófagos codificam fatores de virulência que podem converter células não patogênicas em patogênicas (Feiner et al., 2015). Fagos também podem atuar na regulação da expressão dos genes onde se integram, atuando como elementos de ativação/inativação gênica por meio da integração e desintegração no genoma bacteriano, processo denominado lisogenia ativa (Feiner et al., 2015). Um exemplo interessante ocorre no patógeno humano *Legionella pneumophila.* A excisão e reintegração de um elemento tipo fago no genoma da bactéria é responsável pela alteração entre um estado patogênico e não patogênico dessa bactéria, devido a alteração na expressão de genes que afetam a biossíntese de lipopolissacarídeo (LPS) e do flagelo (Lüneberg et al., 2001).

Na lisogenia ativa o DNA viral excisado pode ser mantido na célula hospedeira como elemento episomal e depois ser reintegrado ou ser perdido, sendo essas formas denominadas lisogenia ativa reversíveis e irreversíveis, respectivamente. Nos dois casos a excisão do profago permite a transcrição do gene do hospedeiro anteriormente interrompido pela integração do genoma viral. Na lisogenia reversível a expressão gênica é interrompida quando o profago é novamente integrado no genoma. Como exemplo temos os genes relacionados ao escape do sistema de defesa da bactéria *Listeria monocytogenes* nas células de mamíferos, que são inativados pela presença de um profago, o qual realiza desintegração reversível. A lisogenia ativa garante que essa bactéria consiga escapar rapidamente do fagossomo do macrófago para permitir sua replicação na célula hospedeira. O processo de integração do profago no gene *com*K leva à inativação do gene, o qual é regulador mestre do sistema de competência da bactéria. Esse sistema de competência é necessário para que ocorra o escape do sistema

de defesa. Portanto, quando o fago sofre excisão esse gene é expresso e assim o sistema de competência permite a ocorrência do escape fagossômico (Feiner et al., 2015). Porém, o profago não é perdido, fica mantido na célula hospedeira como elemento episssomal e pode ser reintegrado.

A regulação das taxas de mutação em *Streptococcus pyogenes* também é um exemplo de lisogenia ativa reversível. Nesse processo são envolvidos os genes *mut*S e *mut*L, codificados pelo operon *mut*SL, envolvidos no sistema de reparo incompatível sendo responsáveis pela detecção e remoção de mutações que ocorrem de forma aleatória no genoma. Quando *S. pyogenes* encontra-se em condições de estresse, iniciando a fase estacionária de crescimento bacteriano, o fago denominado SF370.4 ou SpyCIM1 é inserido dentro do operon e interrompe a ação do sistema de reparo, levando a um aumento na taxa de mutação durante essa fase do crescimento (Scott et al., 2012). Esse fenótipo diversifica a população e com isso aumenta a chance de surgir mutações que poderiam facilitar a sobrevivência bacteriana (Jolivet-Gougeon et al., 2011). Quando a bactéria retorna à fase exponencial de crescimento bacteriano o profago é excisado e fica mantido como episssomo, restaurando assim a função do sistema e diminuindo a taxa de mutação (Feiner et al., 2015).

Outros exemplos bem estudados são a formação de endósporo em *Bacillus subtilis* e formação de heterocisto para fixação de nitrogênio em cianobactérias, sendo ambos os exemplos lisogenia ativa não-reversível. Na formação do endósporo o gene *sig*K é regulado pelo profago *skin*. Esse gene é responsável por codificar σK, o qual regula a expressão dos genes de esporulação no estágio tardio na célula mãe. Durante o crescimento vegetativo normal, o profago encontra-se inserido no gene *sig*K, deixando o mesmo inativo. Porém, na etapa de esporulação, o profago excisa e assim s*ig*K é expresso, codificando σK, que por sua vez irá ativar os genes de esporulação (Kimura et al., 2010). Já a formação do heterocisto tem atuação dos genes *nif*D, *fdx*N e *hup*L. Esses genes são necessários para que ocorra a diferenciação do heterocisto, no entanto, podem ser interrompidos por fagos tornando-os não funcionais. Em condições de limitação de nitrogênio, algumas células vegetativas se diferenciam em heterocistos para captação de nitrogênio atmosférico. Mas, para que ocorra essa diferenciação, os profagos precisam se excisar desses genes para que sejam expressos, permitindo assim a formação do heterocisto (Golden et al., 1985; Carrasco et al., 1995; Feiner et al., 2015).

Um dos mecanismos de transferência gênica horizontal é a transdução, no qual o material genético é transferido entre células bacterianas por um vírus. Esse mecanismo

desempenha um papel importante na evolução bacteriana, uma vez que os vírus podem carrear genes dos hospedeiros e, na infecção de outro hospedeiro, esses genes são introduzidos e podem ser integrados no genoma, conferindo novas características (Brussow et al., 2004). O processo de transdução pode ocorrer de duas formas: transdução generalizada e transdução especializada. A transdução generalizada envolve a formação da partícula transdutora, que é aquela formada somente por DNA do hospedeiro devido à uma falha no sistema de empacotamento genético (Tye et al., 1974). Já a transdução especializada ocorre quando o profago sofre excisão do genoma bacteriano e, nesse processo de excisão, ao invés de carrear somente genes virais, acaba carregando também uma porção do genoma bacteriano juntamente com o seu, devido à uma excisão anômala, aumentando ainda mais a variabilidade genética da bactéria receptora, uma vez que estará sendo recebido tanto DNA viral quanto DNA de outra bactéria (Brammar, 2013).

Aproximadamente metade dos genomas bacterianos sequenciados codifica o sistema CRISPR-Cas, no entanto, mesmo com esse mecanismo de defesa do hospedeiro os fagos são capazes de persistir no ambiente (Landsberger et al., 2018). CRISPR-Cas é um sistema imune de procariotos contra sequências de ácidos nucleicos estranhos, atuando como um sistema de defesa devido à inserção de fragmentos dessas sequências (espaçadores) no seu loci. Esses espaçadores armazenam sequências de memória contra o alvo invasor, levando à avaria desse alvo após seu reconhecimento por esse sistema. Há profagos que possuem genes anti-CRISPR (*acr*) que são capazes de inibir o sistema CRISPR-Cas de procariotos, como é o exemplo do profago JBD30 de *Pseudomonas aeruginosa* durante a infecção e lisogenia (Bondy-Denomy et al., 2013). Os genes anti-CRISPR são resultados da coevolução entre vírus e seu hospedeiro, o qual por sua vez impulsiona a evolução de ambas as partes (Medvedeva et al., 2019).

Dada a grande contribuição dos profagos para os genomas hospedeiros, estudos que abordem o papel que esses profagos podem desempenhar na ecologia do seu hospedeiro são de grande importância. Em *Erwinia* spp. o conhecimento acerca da ocorrência e composição dos profagos presentes nos genomas é atualmente escasso, ainda não tendo sido realizada uma análise global de vírus integrados nos genomas desse gênero.

Erwinia spp. são bácterias gram-negativas, anaeróbias facultativas, pertencentes à família *Enterobacteriaceae*. O gênero é composto por bactérias patogênicas e não patogênicas, sendo a maioria das espécies agentes patogênicos, exceto *E. billingiae*, *E.*

endophytica, E. *gerundensis* e *E. tasmaniensis* que são epífitas (Kube et al., 2010; Ramírez-Bahena et al., 2016; Rezzonico et al., 2016).

Os sintomas causados pela infecção por *Erwinia* spp. são diversos, assim como as espécies de planta que são acometidas por representantes do gênero. Uma das doenças mais destrutivas da macieira e pereira é a queima bacteriana, causada pela bactéria *Erwinia amylovora* (Araujo et al., 2016). *E. amylovora*, em 2012, estava em sétima posição na lista das dez bactérias fitopatogênicas mais importantes para a agricultura (Mansfield et al. 2012), mostrando o prejuízo econômico que pode acarretar. No Brasil, *Erwinia psidii* está entre um dos principais patógenos, causando a doença conhecida como seca dos ponteiros em *Eucalyptus* spp. *E. psidii* também é capaz de acometer goiaba (*Psidium guajava*), também pertencente às *Myrtaceae*. Há estudos que sugerem que *E. psidii* passou por uma mudança de hospedeiro, tornando-se um importante patógeno de *Eucalyptus* spp., o qual tem grande importância econômica na América do Sul sustentando indústrias de papel e celulose (Coutinho et al., 2011).

Além dessas espécies já citadas, também existem outras que causam grandes perdas, como *Erwinia mallotivora* em mamão (*Carica papaya*) (Amin et al., 2011), *Erwinia tracheiphila* em abóbora e melancia (Sanogo et al., 2011) e *Erwinia pyrifoliae* em morango (Wenneker and Bergsma-Vlami, 2015).

[i]REFERÊNCIAS

Ackermann, H.W. (2001) Frequency of morphological phage descriptions in the year 2000. *Arch Virol*. **146**: 843-857 http://link.springer.com/10.1007/s007050170120

Ahern, S.J., Das, M., Bhowmick, T.S., Young, R., and Gonzalez, C.F. (2014) Characterization of novel virulent broad-host-range phages of *Xylella fastidiosa* and *Xanthomonas*. *J Bacteriol* **196**: 459–471 http://jb.asm.org/cgi/doi/10.1128/JB.01080-13.

Amin, N.M., Bunawan, H., Redzuan, R.A., and Jaganath, I.B.S. (2011) *Erwinia mallotivora* sp., a new pathogen of papaya (*Carica papaya*) in peninsular Malaysia. *Int J Mol Sci*. **12**: 39-45 http://www.mdpi.com/1422-0067/12/1/39

Araujo, L., Medeiros, H.A. De, and Pasa, S. (2016) Doenças da macieira e da pereira. In Informe Agropecuário, Belo Horizonte. **37**: 61-74 https://www.researchgate.net/publication/309414271_Doencas_da_macieira_e_da_pereira

Bondy-Denomy, J., and Davidson, A.R. (2014) When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *J Microbiol* **52**: 235–242 http://link.springer.com/10.1007/s12275-014-4083-3.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**: 429–432 http://www.nature.com/articles/nature11723.

Bondy-Denomy, J., Qian, J., Westra, E.R., Buckling, A., Guttman, D.S., Davidson, A.R., and Maxwell, K.L. (2016) Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* **10**: 2854–2866 http://www.nature.com/articles/ismej201679.

Brammar, W.J. (2013) Specialized transduction. In *Brenner's Encyclopedia of Genetics: Second Edition*. pp. 501–502. https://www.sciencedirect.com/science/article/pii/B9780123749840014509?via%3Dihub

Brussow, H., Canchaya, C., and Hardt, W.-D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**: 560–602 http://mmbr.asm.org/cgi/doi/10.1128/MMBR.68.3.560-602.2004.

Carrasco, C.D., Buettner, J.A., and Golden, J.W. (1995) Programmed DNA rearrangement of a cyanobacterial *hup*L gene in heterocysts. *Proc Natl Acad Sci* **92**: 791–795 http://www.pnas.org/cgi/doi/10.1073/pnas.92.3.791.

Castillo, D., Espejo, R., and Middelboe, M. (2014) Genomic structure of bacteriophage 6H and its distribution as prophage in *Flavobacterium psychrophilum* strains. *FEMS Microbiol Lett* **351**: 51–58 https://academic.oup.com/femsle/article-lookup/doi/10.1111/1574-6968.12342.

Coutinho, T.A., Brady, C.L., Vaart, M. van der, Venter, S.N., Telechea, N., Rolfo, M., et al. (2011) A new shoot and stem disease of *Eucalyptus* species caused by *Erwinia psidii*. *Australas Plant Pathol* **40**: 55–60 http://link.springer.com/10.1007/s13313-010-0013-y.

Cumby, N., Edwards, A.M., Davidson, A.R., and Maxwell, K.L. (2012) The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J Bacteriol* **194**: 5012-5019 http://jb.asm.org/lookup/doi/10.1128/JB.00843-12

Czajkowski, R. (2019) May the Phage be With You? Prophage-Like Elements in the Genomes of Soft Rot *Pectobacteriaceae*: *Pectobacterium* spp. and *Dickeya* spp. *Front Microbiol* **10** https://www.frontiersin.org/article/10.3389/fmicb.2019.00138/full.

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A.A. (2015) A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol* **13**: 641–650 http://www.nature.com/articles/nrmicro3527.

Figueroa-Bossi, N., and Bossi, L. (1999) Inducible prophages contribute to *Salmonella* virulence in mice. *Mol Microbiol* **33**: 167–176 http://doi.wiley.com/10.1046/j.1365-2958.1999.01461.x.

Fillol-Salom, A., Alsaadi, A., Sousa, J.A.M. de, Zhong, L., Foster, K.R., Rocha, E.P.C., et al. (2019) Bacteriophages benefit from generalized transduction. *PLoS Pathog*. **15**: e1007888. https://doi.org/10.1371/journal.ppat.1007888

FREEMAN, V.J. (1951) Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *J Bacteriol*. **61**: 675-688 http://www.ncbi.nlm.nih.gov/pubmed/ 14850426

Fujii, N., Oguma, K., Yokosawa, N., Kimura, K., and Tsuzuki, K. (1988) Characterization of bacteriophage nucleic acids obtained from *Clostridium botulinum* types C and D. *Appl Environ Microbiol*. **54**: 69-73 https://aem.asm.org/content/54/1/69

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

Ghosh, D., Roy, K., Williamson, K.E., White, D.C., Wommack, K.E., Sublette, K.L., and Radosevich, M. (2008) Prevalence of lysogeny among soil bacteria and presence of 16S rRNA and *trz*N genes in viral-community DNA. *Appl Environ Microbiol* **74**: 495–502 http://aem.asm.org/cgi/doi/10.1128/AEM.01435-07.

Golden, J.W., Robinson, S.J., and Haselkorn, R. (1985) Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. *Nature* **314**: 419–423 http://www.nature.com/articles/314419a0.

Handley, S.A. (2016) The virome: a missing component of biological interaction networks in health and disease. *Genome Med* **8**: 32 http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0287-y.

Hofer, B., Ruge, M., and Dreiseikelmann, B. (1995) The superinfection exclusion gene (*sie*A) of bacteriophage P22: Identification and overexpression of the gene and localization of the gene product. *J Bacteriol*. **177**: 3080-6 https://jb.asm.org/content/177/11/3080

Jolivet-Gougeon, A., Kovacs, B., Gall-David, S.L. Le, Bars, H. Le, Bousarghin, L., Bonnaure-Mallet, M., *et al.* (2011) Bacterial hypermutation: Clinical implications. *J Med Microbiol*. **60**: 563-573 https://www.microbiologyresearch.org/content/journal/jmm/10.1099/ jmm.0.024083-0

Kimura, T., Amaya, Y., Kobayashi, K., Ogasawara, N., and Sato, T. (2010) Repression of *sig*K intervening (skin) element gene expression by the CI-like protein SknR and effect of SknR depletion on growth of *Bacillus subtilis* cells. *J Bacteriol* **192**: 6209–6216 http://jb.asm.org/cgi/doi/10.1128/JB.00625-10.

Kube, M., Migdoll, A.M., Gehring, I., Heitmann, K., Mayer, Y., Kuhl, H., et al. (2010) Genome comparison of the epiphytic bacteria *Erwinia billingiae* and *E. tasmaniensis* with the pear pathogen *E. pyrifoliae*. *BMC Genomics*. **11:** 393 https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-393

Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Chabas, H., *et al.* (2018) Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity. *Cell*. **174**: 908-916 https://linkinghub.elsevier.com/retrieve/pii/S0092867418307219

Lüneberg, E., Mayer, B., Daryab, N., Kooistra, O., Zähringer, U., Rohde, M., *et al.* (2001) Chromosomal insertion and excision of a 30 kb unstable genetic element is responsible for phase variation of lipopolysaccharide and other virulence determinants in *Legionella pneumophila*. *Mol Microbiol*. **39**: 1259-71 http://doi.wiley.com/10.1046/j.1365-2958.2001.02314.x

Madigan, M.T., Stahl, D. A., Buckley, D. H., Martinko, J.M and Bender, K. S. Microbiologia de Brock. 14. ed. Porto Alegre: ArtMed, 2016

McGrath, S., Fitzgerald, G.F., and Sinderen, D. van (2002) Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. *Mol Microbiol* **43**: 509–520 http://doi.wiley.com/10.1046/j.1365-2958.2002.02763.x.

Medvedeva, S., Liu, Y., Koonin, E. V., Severinov, K., Prangishvili, D., and Krupovic, M. (2019) Virus-borne mini-CRISPR arrays are involved in interviral conflicts. *Nat Commun*. **10**: 5204 http://www.nature.com/articles/s41467-019-13205-2

Nesper, J., Blaß, J., Fountoulakis, M., and Reidl, J. (1999) Characterization of the major control region of Vibrio cholerae bacteriophage K139: Immunity, exclusion, and integration. *J Bacteriol*. **181**: 2902-2913 https://jb.asm.org/content/181/9/2902

Helen E. White and Elena V. Orlova. (2019) Bacteriophages: Their structural organisation and function, bacteriophages - perspectives and future, Renos Savva, IntechOpen, DOI: 10.5772/intechopen.85484. Available from: https://www.intechopen.com/books/bacteriophages-perspectives-and-future/bacteriophages-their-structural-organisation-and-function

Plunkett, G., Rose, D.J., Durfee, T.J., and Blattner, F.R. (1999) Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene

product? *J Bacteriol*. **181**: 1767-78 https://jb.asm.org/content/181/6/1767

Ramírez-Bahena, M.H., Salazar, S., Cuesta, M.J., Tejedor, C., Igual, J.M., Fernández-Pascual, M., and Peix, Á. (2016) *Erwinia endophytica* sp. nov., isolated from potato (*Solanum tuberosum* L.) stems. *Int J Syst Evol Microbiol*. **66**: 975-981 https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.000820

Ramisetty, B.C.M., and Sudhakari, P.A. (2019) Bacterial "grounded" prophages: Hotspots for genetic renovation and innovation. *Front Genet*. **10**: 65 https://www.frontiersin.org/articles/10.3389/fgene.2019.00065/full

Rezzonico, F., Smits, T.H.M., Born, Y., Blom, J., Frey, J.E., Goesmann, A., et al. (2016) *Erwinia gerundensis* sp. nov., a cosmopolitan epiphyte originally isolated from pome fruit trees. *Int J Syst Evol Microbiol*. **66**: 1583-1592 https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.000920

Sanogo, S., Etarock, B.F., and Clary, M. (2011) First Report of Bacterial Wilt Caused by *Erwinia tracheiphila* on Pumpkin and Watermelon in New Mexico. *Plant Dis*. **95**: 1583 https://apsjournals.apsnet.org/doi/10.1094/PDIS-06-11-0507

Scott, J., Nguyen, S. V., King, C.J., Hendrickson, C., and McShan, W.M. (2012) Phage-like *Streptococcus pyogenes* chromosomal islands (SpyCi) and mutator phenotypes: control by growth state and rescue by a SpyCi-encoded promoter. *Front Microbiol*. **3**: 317 http://journal.frontiersin.org/article/10.3389/fmicb.2012.00317

Simmonds, P., Adams, M.J., Benkő, M., Breitbart, M., Brister, J.R., Carstens, E.B., *et al.* (2017) Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* **15**: 161–168 http://www.nature.com/doifinder/10.1038/nrmicro.2016.177.

Tye, B.K., Chan, R.K., and Botstein, D. (1974) Packaging of an oversize transducing genome by Salmonella phage P22. *J Mol Biol* **85**: 485-490 https://linkinghub.elsevier.com/retrieve/pii/0022283674903118

Ul Haq, I., Chaudhry, W.N., Akhtar, M.N., Andleeb, S., and Qadri, I. (2012) Bacteriophages and their implications on future biotechnology: A review. *Virol J*. **9**: 9 https://virologyj.biomedcentral.com/articles/10.1186/1743-422X-9-9

Vieira, O. V., Botelho, R.J., and Grinstein, S. (2002) Phagosome maturation: Aging

gracefully. Biochem J. **366**: 689-704 https://portlandpress.com/biochemj/article/366/3/689/40362/Phagosome-maturation-aging-gracefully

Vostrov, A.A., Vostrukhina, O.A., Svarchevsky, A.N., and Rybchin, V.N. (1996) Proteins responsible for lysogenic conversion caused by coliphages N15 and φ80 are highly homologous. *J Bacteriol*. **178**: 1484-6 https://jb.asm.org/content/178/5/1484

Waldor, M.K., and Mekalanos, J.J. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science.* **272**: 1910–1914 http://www.sciencemag.org/cgi/doi/10.1126/science.272.5270.1910.

Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev*. **28**: 127-81 https://academic.oup.com/femsre/article-lookup/doi/10.1016/j.femsre.2003.08.001

Wenneker, M., and Bergsma-Vlami, M. (2015) *Erwinia pyrifoliae*, a new pathogen on strawberry in the Netherlands. *J Berry Res*. **5**: 17-22 https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JBR-140086

Analysis of prophages in *Erwinia* spp. genomes

Lima, Thamylles Thuany Mayrink; Mendes, Tiago Antônio de Oliveira; Cascardo, Renan de Souza; Alfenas-Zerbini, Poliane. Analysis of prophages in genomes of *Erwinia s*pp. *Molecular Microbiology, to be submitted*.

ANALYSIS OF PROPHAGES IN *Erwinia* spp GENOMES

Lima, Thamylles Thuany Mayrink[1]; Mendes, Tiago Antônio de Oliveira[2]; Cascardo, Renan de Souza; Alfenas-Zerbini, Poliane[1*]

[1]Departamento de Microbiologia, Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Av. Peter Henry Rolfs, S/Nº, Universidade Federal de Viçosa, Viçosa, MG CEP 36570-000; Brazil

[2]Departamento de Bioquímica e Biologia Molecular, Av. Peter Henry Rolfs, S/Nº, Universidade Federal de Viçosa, Viçosa, MG CEP 36570-000; Brazil

*Corresponding authors

Poliane Alfenas-Zerbini; palfenas@ufv.br; phone: +55 31 3612-2437

Keywords: PHASTER; Lysogeny; CRISPR-Cas; bacteriophages; phytopathogenic bacteria.

Number of figures: 7

Supplementary Tables: 4

ABSTRACT

Bacteriophages can be found in many environments, including water, soil, gastrointestinal system and integrated with host genomes, such as prophages. Prophages can have a positive, neutral or negative effect on their host. They can affect the genetic variability of the bacterial genome, confer a selective advantage to the host, transport virulence factors, introduce new pathogenicity genes and confer new phenotypic properties to their hosts, contributing to their fitness. Bacteria of the genus *Erwinia* are pathogens that infect a large number of important agricultural crops. Studies on prophages in this group of bacteria are poorly explored. In this work, we examine phages in genomes of *Erwinia* spp. Phage sequences were identified in 100% of the analyzed genomes. More than one prophage was found in the same genome, characterizing polylisogeny. The analysis of intact phages suggests that they are uncharacterized species, as they showed low identity with virus genomes from public databases. Bacterial genes were found in the phages, suggesting participation in horizontal transfer mechanisms, including genes that encode factors that may confer special characteristics to the host. We detected phage spacer sequences at the CRISPR locus of the genomes of *Erwinia* spp. but no similarity between the phage spacers and prophage residing in the analyzed genomes was observed, suggesting that the prophage residing in the genome were not recognized by the CRISPR-Cas system. Our results show that bacteriophages are widely distributed in the genomes of *Erwinia* spp. and that may be contributing to host's fitness.

Keywords: PHASTER; Lysogeny; CRISPR-Cas; bacteriophages; phytopathogenic bacteria.

INTRODUCTION

Viruses are obligate intracellular parasites capable of infecting all organisms and also microorganisms including bacteria, archaea and fungi, in addition to other viruses, as virophages (Berjón-Otero et al., 2019). It is estimated that there are 10^{31} viral particles on the planet, the majority of which are viruses that infect bacteria (Czajkowski, 2019). They can be found in the most diverse environments, including water, soil, gastrointestinal system and also integrated into their host genomes, as prophages (Bondy-Denomy *et al.*, 2016; Lemire *et al.*, 2018). Bacteriophages contribute in many ways to the ecosystem and they are able to act as horizontal gene transfer agents, control microbial abundance and release organic matter in the oceans, influencing global biogeochemical cycles, and providing greater fitness and survival for their host (Breitbart, 2012).

According to infection strategies, bacteriophages can be classified as strictly lytic or temperate. Strictly lytic bacteriophages are those that once the genome enters a host cell, it is multiplied and new viral particles are produced and released through lysis of their host. On the other hand, temperate bacteriophages can integrate into the host's genome, existing as a prophage. In this case, the viral genome is replicated together with the host cell genome. Temperate bacteriophages may have the lytic cycle activated by chemical and physical stress or by spontaneous induction (Gandon, 2016; Owen *et al.*, 2017).

The presence of prophages contributes to the genetic variability of the bacterial genome and may confer selective advantage to its host by, for example, introducing new pathogenicity genes, contributing to host's adaptation to specific ecological niches and altering cellular metabolism of infected bacteria by induction or disruption of cell host genes (Boyd and Brüssow, 2002; Czajkowski, 2019). Prophages may cause phenotypic changes in their hosts, such as conversion of nonpathogenic to pathogenic bacteria by transferring toxin-producing genes to the host (Brussow *et al.*, 2004) and resistance to their host against infection by other phages. For example, the prophage ϕ80 is able to block the adsorption stage of other phages through the expression of the c*or* gene, which inactivates the membrane receptor FhuA required for the phages during the infection of the cell (Vostrov et al. 1996). Thus, other phages that require this receptor, such as phages HK022 and N15, are unable to infect cells that express the Cor protein (Cumby et al., 2012).

In phytopathogenic bacteria, the involvement of prophages in the symptoms of diseases has been demonstrated for pathogens such as *Pseudomonas*, *Ralstonia* and *Streptomyces* (Nakayama *et al.*, 1999; Addy *et al.*, 2012; Baltz, 2012). In addition, prophages can also be associated with genomic rearrangements and differentiation of strains, as in *Xanthomonas* and *Xylella* (Varani *et al.*, 2013). Recently, a study involving phages found in *Dickeya* and *Pectobacterium* demonstrated how phages can affect fitness and virulence of their hosts (Czajkowski, 2019).

The genus *Erwinia* comprises gram-negative bacteria, mostly phytopathogenic that infect agronomically important crops such as garlic, potato, onion, eucalyptus, apple, pear, tomato, strawberry, among others. *Erwinia* spp. infection leads to significant losses and the symptoms caused depend on the bacterial species and the affected host (Rhim *et al.*, 1999; Coutinho *et al.*, 2011; Arriel *et al.*, 2014; Myung *et al.*, 2016). In Brazil, *Erwinia psidii* is an important phytopathogen of *Eucalyptus* spp., causing damage to the paper and cellulose industries.

Despite the great impact that prophages can have on bacterial biology, little has been explored about the presence of these viruses in the genomes of *Erwinia*, and a global analysis of viruses integrated into the genomes of *Erwinia* spp. has not yet been performed. Thus, in this work, we analyzed phages integrated into the genomes of different species of *Erwinia* available in the GenBank database (NCBI), including the identification and characterization of prophages sequences using genomic tools, expanding the knowledge about the diversity of these viruses and how they can influence genetic and phenotypic features of the bacteria.

RESULTS

Presence of phage sequences in genomes of *Erwinia* spp.

A total of 21 complete genomes of *Erwinia* spp. and one genome scaffold of *Erwinia psidii* available in the NCBI database were subjected to analysis with BUSCO. Of the 21 complete genomes, nine genomes referring to *Candidatus* Erwinia haradaceae, presented collinearity value below 98%, indicating artifacts associated to the genome assembly and so they were removed from our data set. Thus, 12 complete genomes and the *Erwinia psidii* genome were used in the subsequent analyzes (Table S1). Analysis of the 13 genomes using PHASTER software resulted in the identification of phage sequences in 100% of the genomes (Table 1, Table S2). PHASTER was able to identify 43 phage sequences, with 13 (30.23%) of the phage sequences classified as intact phages by manual inspection (Table 1) and 30 (69.77%) as defective phages (Figure 1A).

Defectives were found on chromosomes and plasmids, unlike intact prophages, which were only found on chromosomes. The defective phages found on chromosomes varied from 7.6 to 37 kb (Figure 1B), comprising on average 0.42% of the host's chromosome (Table 1, Figure 1C). The intact phages, found in nine genomes (69.23%), ranged from 13.3 to 64.5 kb (Figure 1B), comprising on average 0.96% of the host's chromosome (Table 1, Figure 1C). In plasmids, only two defective phages were found, one with 29.7 kb in *Erwinia persicina* plasmid pEP1 (NZ_CP022726.1) constituting 20.65% of the plasmid and the other in *Erwinia pyrifoliae* strain EpK15 pEP48 (NZ_CP023568.1) with 6.4 kb and comprising 13,39% (Figure 1D). No correlation was observed between bacterial genome size and prophage genome size ($R^2 = 0.02$) (Figure 1E). The GC content of most prophages was not similar to that in *Erwinia* genomes (Figure 1F).

Defective phages were found in almost all *Erwinia* species with exception of *Erwinia gerundensis* (NZ_LN907827.1), where only a single intact phage was found. All phages classified as intact, except P4-like, had typical structural components of the order *Caudovirales*, with 69.22% belonging to the *Myoviridae* family, 15.39% belonging to the *Siphoviridae* family and 15.39% classified as P4-like (Figure 1G).

Genomes carrying more than one prophage sequence were found (Figure 1A), characterizing the condition of polylysogeny. The genomic organization and the

arrangement of the ORFs were not conserved among all 13 intact prophages, exhibiting high mosaicism, except for the prophages found in Erwinia billingiae phage 1 and Erwinia psidii phage 1, Erwinia billingiae phage 3 and Erwinia sp. QL-ZE phage 2, Erwinia billingiae phage 2, Erwinia psidii phage 2 and Erwinia tasmaniensis phage, Erwinia pyrifoliae Ep1/96 phage, Erwinia pyrifoliae EpK15 phage and Erwinia pyrifoliae DSM 12163 that showed high synteny (Figure 2).

In comparative genomic analyzes of the complete nucleotide sequence of the intact phages found in *Erwinia* genomes, we observed the formation of five different clusters in heatmap. The first cluster was formed by Erwinia persicina phage and Erwinia pyrifoliae EpK15 phage, which according to VirFam classification belong to different families. The second cluster includes Erwinia billingiae phage 3 and Erwinia sp. QL-Z3 phage 2, both P4-like. The third cluster includes Erwinia pyrifoliae Ep1/96 phage and Erwinia pyrifoliae DSM 12163 phage, both found in the same bacterial species. The fourth cluster includes Erwinia billingiae phage 1 and Erwinia psidii phage 1 and, finally, the fifth cluster by Erwinia billingiae phage 2 and Erwinia tasmaniensis phage. Erwinia sp. QL-ZE phage 1, Erwinia gerundensis phage and Erwinia psidii phage 2 were not grouped in any cluster (Figure 3).

Most of the clusters based on sequence identity analysis were maintained and sustained by phylogenetic analysis. However, Erwinia gerundesis phage and Erwinia psidii phage 2 were grouped and Erwinia sp. QL-Z3 phage 1 appeared grouped with Erwinia persicina phage in phylogeny. The grouping between Erwinia sp. QL-Z3 and Erwinia perisicina phage were only observed in phylogeny, which can be explained due to the relatively low bootstrap (Figure 4). In addition to these analyzes, a grouping by predicted protein sequences was also performed. In this grouping, most clusters from other analyzes were also maintained. Erwinia gerundensis phage and Erwinia psidii phage 2 were also grouped as in phylogeny. Erwinia persicina phage and Erwinia sp. QL-ZE phage 1 did not group, which was also observed in the sequence identity analysis. Erwinia pyrifoliae Ep1/96 phage, Erwinia pyrifoliae EpK15 phage and Erwinia pyrifoliae DSM 12163 were identified grouped as the same species only in the proteomic cluster (Figure 5), not observing this cluster with Erwinia pyrifoliae EpK15 phage in the identity and phylogeny analysis.

Despite being in the same genome, the intact phages showed low sequence conservation and they were not grouped together as observed in the heatmap (Figure 3), phylogenetic (Figure 4) and protein sequence (Figure 5) analysis.

Analysis of bacterial genes and genes encoding factors that may increase bacterial fitness in intact prophages

To understand the involvement of prophages in the pathogenicity and evolution of *Erwinia* spp., we analyzed the probable genes encoding factors that enhance host fitness in intact phage sequences.

Of the 13 intact prophages found, only two (Ewinia billingiae phage 3 and Erwinia sp. QL-ZE phage 2) had no genes with identity to genes traditionally classified as bacterial (Figure 6). The other 11 prophages contained between one (Erwinia tasmaniensis phage) to five genes (Erwinia persicina phage) apparently acquired from bacterial hosts (Figure 6). When phages excise from the bacterial genome, they can carry fragments of host DNA (Evans *et al.*, 2010). The genes found that apparently were acquired from bacterial hosts encode proteins involved in DNA regulation and modification mechanisms, such as methylases, methyltransferases, transcriptional regulators, restriction endonuclease, DNA polymerase V. We also found ribosomal protein, isocitrate dehydrogenase and metal dependent phosphohydrolases (Table 2).

Some prophages are capable of encoding potential factors that increase the host's fitness. The analysis of the coding regions of the phage genomes showed the presence of possible genes related to the host's fitness such as toxin HicA and antitoxin HicB of the type II toxin-antitoxin system, GtrA family protein and QacE protein (Table 3). No gene encoding antibiotic resistance or virulence genes was found when ResFinder and VirulenceFinder were used, respectively.

CRISPR-Cas system analysis on genomes of *Erwinia* spp.

In order to understand the acquisition of prophages, the CRISPR-Cas system of the 13 *Erwinia* species was analyzed. Using CRISPRCasFinder and CRISPRone software, we found spacers that were associated only CRISPR the type I, and four subtypes (A, B, E and F) of the system. No CRISPR locus were found in the genomes of *Erwinia billingiae*, *Erwinia gerundensis* and *Erwinia* sp. pQL-ZE (Table 4). In the genomes of *Erwinia persicina* and *Erwinia psidii*, although spacers were found, they did not hit with phages. We found 422 spacers in ten species of *Erwinia*, combining the results of the two softwares (Table 4), with 26.54% of spacers mapping with phage sequences, 13.98% with plasmids, and 52.96% are unknown sequences (Figure 7A). We

do not establish a direct relation between the number of spacer sequences and the number of prophages sequences, intacts and defectives (Figure 7B). In almost all genomes, most of the spacers found hit phages of the *Myoviridae* family, except for *Erwinia pyrifoliae* strain Ep1/96 and *Erwinia pyrifoliae* strain DSM 12163, where we found the same number of spacers for *Myoviridae* and *Siphoviridae* (Figure 7C).

No spacers were found in the host genome for the resident phage, indicating that this is probably a functional CRISPR system and that the resident phages may haves an escape mechanism from the system. However, it is not possible to classify the activity of the CRISPR-Cas system as functional or not only with *in silico* analyzes. Spacer corresponding to Enterobacteria phage HK140 was found in *Erwinia* sp. Ejp617 (NC_017445.1), and this phage was detected as the best hit of the *Erwinia* sp. QL-ZE phage 1.

DISCUSSION

A total of 43 prophages were found in our analysis. Of these, 30 were considered defective (Supplementary Table 2) and 13 intact (Table 1). The classification in families belonging to the order *Caudovirales* performed by VirFam takes into account head and tail proteins. However, since P4-type bacteriophages do not encode structural proteins, it is not possible to classify them by VirFam (Lopes *et al.*, 2014), so we call them P4-like. More than one prophage was found infecting the same bacteria. However, no prophages from different families were found in the same bacterial genome, as noted in *Vibrio campbellii* (Lorenz *et al.*, 2016).

It is known that phages can contribute a significant proportion of the bacterial genome. However, we found a low proportion in *Erwinia g*enomes, corresponding to a maximum of 1.6% of the chromosome and 20.65% of the plasmid. This suggests that viruses and these bacteria may have a short period of coexistence. In addition, we speculate that perhaps the defense mechanisms against viral infection in *Erwinia* spp. are quite effective, providing better protection for the bacteria. The probable short period of co-evolution between the bacteria and the virus, may contribute to insufficient evolution of mechanisms of defense suppression. Our speculation can be corroborated by the analysis of the GC content. The GC content of most phages was lower than that of their hosts, as observed in previous studies (Almpanis *et al.*, 2018). This suggests that these prophages have probably become part of the bacterial genome more recently, supporting our hypothesis of the short period of co-existence.

Althought all the complete genomes of *Erwinia* spp. available in the NCBI database have prophage sequences, little is known about how these viruses affect *Erwinia* spp. and how they may be associated with virulence and adaptation of the host bacteria. Some genes that maybe related to virulence, such as HicA and HicB, cell surface polysaccharide synthesis and quaternary ammonium resistance were found in phage sequences.

The HicB antitoxin gene was identified in three phage sequences. However, only in one was also found the HicA toxin, forming the toxin-antitoxin pair, components of the type II toxin-antitoxin (TA) system that may contribute to bacterial virulence. This toxin-antitoxin system, although classified into bacterial genes, is often detected in phages and plasmids, as well as chromosomes, due to extensive horizontal gene transfer.

Type II toxin-antitoxin systems are abundant in bacterial genomes, forming part of the mobilome, which move through horizontal gene transfer (Guglielmini and Van Melderen, 2011). The fact that we found only HicB antitoxin in two phage sequences can be related to several factors: (1) the bacterial cell toxin-antitoxin system, at the time of phage excision, may have been incorporated into the viral genome through transduction, thus carrying the genes for HicA and HicB. In previous studies of *Pseudomonas aeruginosa* genomes, the *hic*AB locus was found and, interestingly, this *hic*AB locus is linked to sequences that encode phage-related proteins, supporting our hypothesis (Li *et al.*, 2016). Due to this, we hypothesized that the HicA toxin has been lost over the generations after the integration of the virus into the bacteria; (2) Or it may be that only the gene corresponding to the HicB antitoxin has been carried together with viral genome, since the HicA toxin was not found; (3) Although we have not analyzed the genes that make up the genome of host bacteria, another possibility is that the HicA toxin can be encoded in the bacterial genome, supporting the fact that we find HicB in the prophages. This antitoxin would neutralize the effects of HicA on the host's genome and, HicB being found in the prophages, a more stable relationship could be formed between the bacterium and the prophages, because in the possible loss of this prophages carrying the antitoxin. The bacteria would suffer the effects of its toxin itself, as occurs in some plasmids that carry toxins (Li *et al.*, 2016; Rocker and Meinhart, 2016); (4) Finally, we also speculate that this antitoxin may provide protection against toxins from other systems, thereby conferring a competitive advantage about other bacteria (Guglielmini and Van Melderen, 2011).

The presence of HicA in Erwinia pyrifoliae EpK15 phage can confer an adaptive advantage for its host, since this toxin can make the bacteria a better competitor. Previous studies have shown that bacteria that carry genes for toxins encoded by phages are more virulent and therefore have an advantage in infecting mammalian cells (Feiner *et al.*, 2015).

In addition, we detected a protein from the GtrA family. Members of this family are integral membrane proteins involved in synthesis of cell surface polysaccharide. It is involved in the modification of antigen O by Shigella flexneri bacteriophage X (SfX), but does not determine the specificity of glycosylation. Its function remains unknown, but it may play a role in translocation of undecaprenyl phosphate-bound glucose (UndP-Glc) across the cytoplasmic membrane (Guan *et al.*, 1999). Due to involvement in the synthesis of polysaccharides on the cell surface, this protein can cause changes in the

bacterial cell surface, which could interfere with the recognition of the bacteria by plant, being able to act as an evasion mechanism of the plant's defense system (D'Haeze and Holsters, 2004).

We found in an intact prophage of *Erwinia psidii* the protein QacE, a quaternary ammonium resistance protein (QAC) and strong cationic surfactant, which may confer niche adaptation for this bacteria. This protein is part of an efflux system that mediates the extrusion of QAC cations. QACs are mainly used in disinfectants and antiseptic formulations used in residences, agriculture and industry and may result in the emergence and spread of resistance to these compounds among different bacterial genera (Tezel and Pavlostathis, 2015).

Methyltransferases have been found associated with phages in the genomes of Erwinia billingiae phage 1, Erwinia psidii phage 1, Erwinia psidii phage 2, Erwinia pyrifoliae Ep1/96, Erwinia pyrifoliae EpK15 and Erwinia pyrifoliae DSM 12163. Although traditionally classified as bacterial genes it has been described that they can act as inhibitors of restriction enzymes, which will fight the RM system (restriction-modification) of bacteria (Varani *et al.*, 2013).

Many hypothetical proteins have been found in the sequences of the prophages. Other interesting genes can be discovered through the analysis of these hypothetical proteins, in order to unravel what each one of them encodes. This will allow, for example, a better understanding of the interactions between phages and their hosts (Crispim *et al.*, 2018) and the important roles these phages can play in genome evolution and in the adaptation of *Erwinia* spp.

In our results, most of the prophages found were considered defective (69.77%) and many of these apparently do not contain genes coding for all structural proteins needed to assemble their particles and genes for interaction of the bacteriophage with its host, such as integrases. In addition, 30% did not have the binding sites (*att*) and 70% did not have lytic proteins. However, absence of lytic proteins was not used to classify a phage as defective. The absence indicates that are not temperate phages and therefore cannot be induced (Golding *et al.*, 2019).

Regardless of the contributions that prophages to their hosts, temperate intact prophages can kill bacteria when their lytic cycle is activated. Because intact prophages are likely to kill the cell upon induction of the lytic cycle, it is believed that there should be a strong selection for mutations leading to prophage inactivation. The high number of defective prophages found in our analysis may lead us to assume that these

prophages may have been subjected to inactivation in the bacterial genome, a phenomenon known as phage domestication (Bobay *et al.*, 2014). This phenomenon would be a way for bacterial genomes to be cured of exogenous/toxic genetic material (Czajkowski, 2019).

Some defective prophages may be gene transfer agents (GTAs) that transfer random pieces of the bacterial chromosome to another cell. This hypothesis is based on the fact that GTAs have an almost complete set of structural and lysis genes, while genes related to regulation, replication, integration/excision are absent (Bobay et al., 2014). Generally, the size of GTAs can vary from 14 to 30 kb, according to Bobay et al. (2014). Most of identified defective prophages showed absence of structural and lysis genes that are associated to large deletions within the prophage genomes, causing an unbalanced gene exclusion, once genes encoded in the central part of the prophage (such as lysis or packaging genes) are much more frequently excluded than genes at the edges (such as integrase).

We also found P4-type prophages, all of them classified as incomplete by PHASTER software. Only two were classified as intact, despite having no structural genes, such as Enterobacteria fago P4 (NC_001609.1) - a defective virus that needs the Enterobacteria fago P2 helper - these had the other signature genes analyzed. This P4 bacteriophage may be in the lysogenic state in the absence of its helper, but it needs P2 for its particle formation, DNA packaging, and lytic growth, since it has no structural and lytic proteins (Six, 1975; Christie and Calendar, 2011; Howard-Varona *et al.*, 2017). Among all analyzed genomes, the coexistence of P2-like and P4-like prophages, is exclusive to *Erwinia billingiae* Ebb61 (NC_014306.1)

Despite having the signature genes adopted with exception of structural protein, two other P4-type prophages were not classified as intact. These phages did not show similarity to Enterobacteria phage P4 when BLASTn algorithm was performed, suggesting to be another phage. By manual inspection, we found that, compared to the complete P4 genome (NC_001609.1), the two intact phage have the absence of some proteins, such as the gop protein and the β protein in Erwinia billingiae phage 3 and in the gop protein, β and polarity-suppressing protein in Erwinia sp. QL-ZE phage 2. The gop protein kills *Escherichia coli* if β is absent (Christie and Calendar, 1990). Gop and β are proteins related to *E. coli*, and since P4-like prophages were found in the genomes of *Erwinia* spp., we believe it is an indication that these phages lost these proteins when they started to infect *Erwinia* spp. and in this process of infection, these proteins may

not be relevant. Polarity suppression protein, on the other hand, is a P4 capsid decoration protein (Isaksen *et al.*, 1992; Isaksen *et al.*, 1993). However, even with the absence of these proteins, we classify them as intact P4-like, since they are non-essential for the activity of P4 (Shen *et al.*, 2020).

Together, our results showed that there is a small diversity of viral families infecting *Erwinia* spp., with only phages of the families *Myoviridae* and *Siphoviridae*. The predominance of these families has also been observed for other phytopathogenic bacteria, such as *Pectobacterium* spp. and *Dickeya* spp. (Czajkowski, 2019) and also for the sulfate-reducing bacteria *Desulfovibrio* spp. (Crispim *et al.*, 2018). Despite the small diversity in relation to families, the viral sequence classified in the same family are quite different. The low diversity of viral family found may also be associated with the possible short period of coexistence of phages and bacteria.

Most of the phages showed high genetic mosaicism, except for the phage identified in *Erwinia pyrifoliae*. This species of *Erwinia* has three different strains, with strain EpK/15 having a larger plasmid than plasmids of strains Ep1/96 and DSM 12163 and three plasmids reported in Ep1/96 and DSM 12163 were absent in the draft genome of EpK/15 (Lee *et al.*, 2018). The difference between strain Ep1/96 and DSM 12163 is in the slightly different size of plasmids pEP2.6 and pEP36 and the presence of two additional small plasmids in DSM 12163, plasmids pEP5 and pEP3 (Smits *et al.*, 2010). The phages found in *Erwinia pyrifoliae* showed high synteny, which led us to suppose that the differentiation of the strains may have occurred after infection by the prophage. In fact they can be the same prophage due to high conservation in the nucleotide level. This hypothesis is also supported protein sequence analysis, since we found that all the intact phages found in the genomes of *E. pyrifoliae* belong to the same species. Based on phylogenetic analysis, 12 different species of intact phage were identified, since the phage found in *E. pyrifoliae* Ep1/96 and *E. pyrifoliae* DSM 12163 belong to the same species (Supplementary table 3). In the other phages we find a high mosaicism, which is already expected, since it is known that the phages are highly diverse and that their genomes have modules that can be exchanged between different phages by recombination (Czajkowski 2019).

Some phages showed a certain synteny as Erwinia billingiae phage 1 and Erwinia psidii phage 1, Erwinia billingiae phage 3 and Erwinia sp. QL-ZE phage 2, Erwinia billingiae phage 2, Erwinia psidii phage 2 and Erwinia tasmaniensis phage, in addition to the three phages found in *Erwinia pyrifoliae*. In the cluster analyzes, these

phages were grouped together. This led us to suppose that these phages were probably acquired by *Erwinia* before differentiation into species, justifying the fact that we found very similar phages in different bacterial species.

Most prophages found (53.49%) do not contain genes that code for lytic enzymes, such as holin, lysozyme and spanin. Only 1 (2.33%) prophage found in *Erwinia* sp. QL-ZE (NZ_CP037950.1) contains genes for these three enzymes. 6.97% contains genes for only one of them and 37.21% contains genes for two, which may be a combination of holin with lysozyme or holina with spanin. Most bacteriophages lyse their hosts producing at least two proteins, holins and endolysins, which are responsible for the disruption of the host cell wall thus allowing the release of viral particles into the environment. Holins are membrane proteins responsible for lysis and endolysins disturb the peptidoglycan. In addition to these two essential lysis proteins, phages infecting Gram-negative hosts generally synthesize a third class of lysis proteins, the spanins, needed to overcome the bacterial outer membrane (Catalao *et al.*, 2011; Young, 2014; Joshi *et al.*, 2017). In addition, seven (16.28%) phopages have a lysis regulatory protein, LysB. The presence of genes encoding these enzymes characterizes these prophages as probably infectious and indicates that they may possibly be induced (Nanda *et al.*, 2015). However, only with *in silico* analyzes not can be asserted about the infectivity of these prophages. Propagation induction experiments will be conducted later to better understand their role in the genomes of *Erwinia* spp., studying phage-bacterial interactions to confirm their infectivity. If indeed they are infectious, they can be used to control diseases caused by *Erwinia* spp. (Fenton *et al.*, 2010), since genes encoding lytic proteins can be used to develop bacterial control tools, for example by industrially expressing these enzymes for later environmental application (Malnoy *et al.*, 2005). The absence of lysozyme, spanin and/or holin in the phage genome has already been reported. Dickeya spp. bacteriophage LIMEstone1 is an infectious phage and has no gene for holin, as is Pectobacterium carotovorum subsp. carotovorum phage PP1 which is absent of lysozyme (Czajkowski, 2019).

Our analyzes identified 13 probable new phages in the genomes of *Erwinia* spp. The data shows how virus diversity in *Erwinia* spp. is little explored, as we find in most cases the best hit phages infecting other hosts such as *Salmonella* spp., *E. coli* and *Klebsiella* spp., instead of *Erwinia* spp., although all of these bacteria are from the same *Enterobacteriaceae* family. We have tried to classify them, but identity and coverage values gives the idea that it may be a new virus. None of our prophages identified as

best hit are classified by ICTV, probably because current ICTV classification procedures are not keeping up with viral discoveries. For example, of the 4,400 viral genomes deposited in the NCBI in viral RefSeq, only 43% had been classified by the ICTV in 2015 (Bolduc *et al.*, 2017).

Phages and bacteria live in a constant evolutionary arms race, in which one develops different strategies to combat the other (Koskella and Brockhurst, 2014). One of the strategies developed by bacteria to defend against viral infection is the CRISPR-Cas system. The number of spacers in the CRISPR locus is a measure of the number of sequences directed by the system and presumably their ability to provide phage protection. Although we found a significant portion of spacers, we found no spacers in the genome for the phage resident in that genome, leading us to think that the phage may have mechanisms to suppress the activity of this system. In addition, these results lead us to suppose that these prophages found were probably not recognized by the CRISPR-Cas system because the system of these bacteria may have a small matrix of spacers (Touchon *et al.*, 2016) or that the infection of these prophages occurred before the acquisition of the CRISPR-Cas system by the bacterium. We did not observe a relation between the number of phages and the number of spacers. Contrary to our results, in previous studies, *Cronobacter sakazakii* has already observed a relationship of low content of prophages with higher number of spacers, thus demonstrating the likely involvement of the CRISPR system in protection against bacteriophages (Zeng *et al.*, 2017). 26,54% spacers mapped with plasmids, however, we did not analyze whether these plasmids have integrated phages.

We found a spacer for Enterobacteria phage HK140 in *Erwinia* sp. Epj617 and, interestingly, this phage was found to be the best hit of one of our intact prophages, Erwinia sp. QL-ZE phage 1. This shows us that, possibly, this Enterobacteria phage HK140, or other phage related to it, tried to infect one species of *Erwinia* and probably was not successful in this, but managed to infect another species of the same genus. This may be due to different infection strategies, thereby interfering with the host range of the phage.

In previous studies, phages of the *Podoviridae* family have been characterized, infecting *E. amylovora* (Schwarczinger et al., 2017). We found spacers for the *Podoviridae* family in five genomes of *Erwinia* in our analyzes, however, we did not find intact prophage belonging to that family, leading us to assume that the *Podoviridae*

prophages found in the spacers failed to suppress the CRISPR-Cas system from defense of these bacteria.

This study presents the results of prophages found in *Erwinia* spp., their effects on the host's genome, providing genomic plasticity, increasing the host's fitness and also on the diversity of phages that infect the genus *Erwinia*. However, we only perform bioinformatics analyzes without experimental evidence, therefore, the classification of the prophages as intact or defective and their impact on the host genome are based only on these *in silico* analyzes.

EXPERIMENTAL PROCEDURES

Identification of prophages in genomes of Erwinia spp.

One hundred genome assemblies were available in the NCBI database in July 2019. However, most of these genomes are in the form of scaffolds or contigs, with only 21 (21%) of these being in the form of the complete genome, including *Candidatus* Erwinia genomes. These 21 complete genomes were analyzed with Universal Benchmarking Single-Copy Orthologs (BUSCO) (parameter in Supplementary table 4) to verify the quality of the genome assembly (Waterhouse *et al.*, 2018). Genomes that showed collinearity values greater than 98% were used in this study, except for the genome of *Erwinia psidii*. The genome of *Erwinia psidii* was assembled as scaffold, however, it was used in our study because it is the only isolate we have for further analysis. In total, 13 genomes belonging to nine species of the genus *Erwinia* were used in this work (Supplementary table 1). Phage identification was performed using the PHASTER (PHAge Search Tool Enhanced Release) web server (http://phaster.ca/). The sequences identified by PHASTER were manually inspected using the integrase position and the last phage-related gene as criteria for demarcation of viral genome boundaries. The genomes of the prophages were automatically annotated using Rapid Annotation using Subsystem Technology (RAST) (Supplementary Table 4 and the other parameters were according to the default) (Aziz *et al.*, 2008) (http://rast.nmpdr.org/), the Open Reading Frame (ORF) were predicted by GeneMarkS web server (parameters in Supplementary Table 4) (http://exon.gatech.edu/GeneMark/genemarks.cgi) and compared to protein sequences deposited in the database non-redundant protein sequences (nr) using the Basic Local Alignment Search Tool for proteins (BLASTp) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with Organism Viruses and default parameters (Altschul *et al.*, 1990a).

For the characterization of the sequences identified as prophages attachment sites (*att*L and *att*R) were identified using PHASTER. Structural genes were verified by VirFam (parameters in Supplementary Table 4) (http://biodev.cea.fr/virfam/) (Lopes *et al.*, 2014).

In addition, a manual check for the presence of signature genes was performed. The signature genes included: genes encoding structural viral proteins, integrase, terminase

and genes encoding lytic proteins. Prophages were classified as intact when all of these signature genes and the attachment sites were found, as used by Czajkowski (2019). Those that lacked any of these signature genes, except lytic proteins, were classified as defective. This work didn't analyze the presence of non-integrase-based forms of lysogenesis, such as transposable phages or plasmid-based replication.

Search for bacterial genes

Bacterial genes were searched in the phage sequences by BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with database non-redundant protein sequences (nr) (Altschul *et al.*, 1990). To classify the genes as being of bacterial the criteria used were: (1) whether the gene is traditionally present in the bacterial genome; (2) not required for bacteriophage life cycle; (3) encodes proteins with enzymatic activity not required for the virus to interact with its host (Czajkowski, 2019).

Virulence genes and antibiotic resistance genes

The presence of putative virulence-associated genes was analyzed using VirulenceFinder 2.0 with default parameters (https://cge.cbs.dtu.dk/services/VirulenceFinder/) as perfomed by Czajkowski (2019). The genes that confer antibiotic resistance in the viral genomes were analyzed using ResFinder (parameters in Supplementary Table 4) (https://cge.cbs.dtu.dk/services/ResFinder/) as done also by Czajkowski (2019) and The Comprehensive Antibiotic Resistance Database (CARD) (parameter in Supplementary Table 4) (https://card.mcmaster.ca/analyze/blast), respectively.

CRISPR-Cas system analysis and spacer determination

The CRISPR-Cas system was detected in *Erwinia* spp. using CRISPCasFinder (https://crisprcas.i2bc.paris-saclay.fr) (Couvin *et al.*, 2018) and CRISPRone (Zhang and Ye, 2017). All spacer sequences were extracted for comparison with the NCBI virus database by BLASTn version 2.6.0+. Only viral sequences that demonstrated at least 84% identity with spacer sequences, as suggested by Zeng et al (2017) and E-value $\leq 10^{-1}$ were considered for probable identification of viral targets. Plasmid comparison was performed using the A CLAssification of Mobile genetic Elements (ACLAME) database with default parameters (https://aclame.ulb.ac.be) (Leplae *et al.*, 2010).

Grouping of prophages

All sequences of the 13 intact phages were opened at the integrase position and aligned with Muscle using Geneious version 9.1. Synteny analysis was performed with Mauve, also using Geneious version 9.1. The heatmap was performed based on the identity of the intact prophage sequences by the R program and the phylogenetic tree inferred using the Genome-BLAST Distance Phylogeny method (GBDP) available in VICTOR server (https://ggdc.dsmz.de/victor.php#) (Meier-Kolthoff and Göker, 2017) (parameter in Supplementary Table 4). The ViPTree server (https://www.genome.jp/viptree/) (Nishimura *et al.*, 2017) (parameters in Supplementary Table 4) with nucleotide data was used to generate a dendogram of the genomic sequences for all intact phages based on the similarities of sequences across the genome calculated by tBLASTx.

TABLES

Table 1: Features of intact prophages present in genomes of *Erwinia s*pp. available on GenBank (NCBI).

Number of prophage	Putative prophage	Host (GenBank acession)	Prophage size (kb)*	GC%*	Coordinates in bacterial genome*	Putative phage attachment region*	% host genome	Classification VirFam	Best hit (BLASTn)	Query (%)	E-value	Identity (%)
1	Erwinia billingiae phage 1	*Erwinia billingiae* Eb661 (NC_014306.1	38.2	52.94	103.545 - 141.825	CACCATCCCTGTCTT	0.75	*Myoviridae*	Salmonella phage SEN5 (NC_028701.2)	43	0.0	69.56
2	Erwinia billingiae phage 2	*Erwinia billingiae* Eb661 (NC_014306.1	35.6	53.23	1.245.185 - 1.280.792	CTGTCGCCATTTTGTCGCCA	0.7	*Myoviridae*	Klebsiella phage ST512-KPC3phi13.6 (MK433577.1)	60	0.0	77.88
3	Erwinia billingiae phage	*Erwinia billingiae* Eb661 (NC_014306.1)	21.8	49.52	4.924.727 - 4.946.598	ACCGCTGGCGGCC	0.43	Unclassified	Enterobacteria phage P4 (NC_001609.1)	32	0.0	71.08
4	Erwinia gerundensis phage	*Erwinia gerundesis* (NZ_LN907827.1)	34.8	54.34	2.197.490 - 2.232.332	GTATTCGGTCTTTTTTT	0.93	*Myoviridae*	Erwinia phage ENT90 (HQ110084.1)	65	0.0	82.53
5	Erwinia persicina phage	*Erwinia persicina* BC4 (NZ_CP022725.1)	47.6	50.10	546.851 - 594.465	GGAATCGAACCTG	0.99	*Siphoviridae*	Erwinia phage vB_EhrS_59 (MH443101)	39	0.0	87.2
6	Erwinia sp. QL-ZE phage 1	*Erwinia* sp. QL-ZE (NZ_CP037950.1)	56	49.48	4.807.467 - 4.863.515	ATAAAAAAACCC	1.14	*Siphoviridae*	Enterobacteria phage HK140 (JQ086370.1)	26	0.0	72.16
7	Erwinia sp. QL-ZE phage 2	*Erwinia* sp. QL-ZE (NZ_CP037950.1)	13.3	49.04	752.268 - 765.579	TGGTGCCGAAGGCCGGA	0.27	Unclassified	Enterobacteria phage P4 (NC_001609.1)	50	0.0	76.52
8	Erwinia psidii phage 1	*Erwinia psidii* (NZ_RHHM00000000.1)	36.2	53.57	3.501.407 - 3.537.701	AAAGGCGACTCTCAGGT	0.81	*Myoviridae*	Enterobacter phage phiT5282H (MG589387.1)	60	0.0	82.72
9	Erwinia psidii phage 2	*Erwinia psidii* (NZ_RHHM00000000.1)	59.4	50.72	3.789.756 - 3.849.156	CGCGCTCCTGAA; GCGTTCCTCACA; GCGACACTTTTG	1.32	*Myoviridae*	Erwinia phage ENT90 (HQ110084.1)	33	0.0	84.32
10	Erwinia pyrifoliae Ep1/96 phage	*Erwinia pyrifoliae* Ep1/96 (NC_012214.1)	62	51.84	2.273.980 - 2.336.023	TGCAGCGGCCAG; GGGTTTTGATAAC	1.54	*Myoviridae*	Salmonella typhimurium phage ST64B (AY055382.1)	30	0.0	69.99
11	Erwinia pyrifoliae EpK15 phage	*Erwinia pyrifoliae* EpK15 (NZ_CP023567.1)	64.5	51.83	1.647.330 - 1.711.850	CTGGCCGCTGCA; CATCCCGCCTTAA	1.6	*Myoviridae*	Salmonella typhimurium phage ST64B (AY055382.1)	28	0.0	70.10
12	Erwinia pyrifoliae DSM 12163 phage	*Erwinia pyrifoliae* DSM 12163 (NC_017390.1)	45.2	51.39	2.281.145 - 2.326.440	GGGTTTTGATAAC	1.12	*Myoviridae*	Salmonella typhimurium phage ST64B (AY055382.1)	29	0.0	69.99

13	Erwinia tasmaniensis phage	*Erwinia tasmaniensis* Et1/99 (NC_010694.1)	34.6	52.92	625.327 - 659.938	GACCAAGCGATTATGAGT	0.89	*Myoviridae*	Klebsiella phage ST13-OXA48phi12.1 (MK422453.1)	74	0.0	78.45

*Information provided by PHASTER

TABLE 2: Genes found in phages genomes that are typically classified as bacterial.

Host (GenBank acession)	Prophage	Gene	BLASTp parameters
Erwinia billingiae Eb661 (NC_014306.1)	Erwinia billingiae phage 1	Dam family site-specific DNA-(adenine-N6) - methyltransferase (WP_041691845.1)	Query: 100% E-value: 0.0 Identity: 100%
		ImmA/IrrE family metallo-endopeptidase (WP_013200140.1)	Query: 100% E-value: 0.0 Identity: 100%
	Erwinia billingiae phage 2	DNA adenine methylase (WP_013201112.1)	Query: 100% E-value: 0.0 Identity: 100%
		ATP-binding protein (WP_013201130.1)	Query: 100% E-value: 0.0 Identity: 100%
Erwinia gerundesis (NZ_LN907827.1)	Erwinia gerundensis phage	Restriction endonuclease (WP_067435379.1)	Query: 100% E-value: 0.0 Identity: 100%
		TraR/DksA family transcriptional regulator (WP_067431361.1)	Query: 100% E-value: 7e-46 Identity: 100%
Erwinia persicina BC4 (NZ_CP022725.1)	Erwinia persicina phage	GIY-YIG nuclease family protein (WP_118663616.1)	Query: 100% E-value: 1e-130 Identity: 100%
		S-adenosylmethionine-binding protein (WP_118663623.1)	Query: 100% E-value: 8e-145 Identity: 100%
		50S ribosomal protein L13 (WP_118663646.1)	Query: 100% E-value: 2e-32 Identity: 100%
		Glycosyltransferase family 2 protein (WP_118663660.1)	Query: 100% E-value: 0.0 Identity: 100%
		DNA polymerase V (WP_118665678.1)	Query: 100% E-value: 4e-51 Identity: 100%
Erwinia *psidii* (NZ_RHHM00000000.1)	Erwinia psidii phage 1	3'-5' exoribonuclease (RQM37229.1)	Query: 100% E-value: 7e-137 Identity: 100%
		DNA adenine methylase (MGJ32935.1)	Query: 99% E-value: 0.0 Identity: 88.26%
		DNA cytosine methyltransferase (WP_124234088.1)	Query: 100% E-value: 0.0 Identity: 100%
Erwinia *psidii* (NZ_RHHM00000000.1)	Erwinia psidii phage 2	TraR/DksA family transcriptional regulator (WP_124234367.1)	Query: 100% E-value: 1e-42 Identity: 100%
		DNA adenine methylase (WP_124234368.1)	Query: 100% E-value: 0.0 Identity: 100%
		DNA (cytosine-5-)-methyltransferase (WP_12423432.1)	Query: 100% E-value: 0.0 Identity: 100%
Erwinia pyrifoliae Ep1/96 (NC_012214.1)	Erwinia pyrifoliae Ep1/96 phage	N-Acyltransferase superfamily (WP_012668325.1)	Query: 100% E-value: 0.0 Identity: 100%
		Isocitrate dehydrogenase (CAY74567.1)	Query: 100% E-value: 5e-28 Identity: 100%
		site-specific DNA-methyltransferase (WP_012668355.1)	Query: 100% E-value: 0.0 Identity: 100%
		DNA N-6-adenine-methyltransferase (Dam) (WP_012668361.1)	Query: 100% E-value: 0.0 Identity: 100%
		PerC family transcriptional regulator (WP_012668362.1)	Query: 100% E-value: 1e-107 Identity: 100%

TABLE 2: Genes found in phages genomes that are typically classified as bacterial. (cont.)

Erwinia pyrifoliae EpK15 (NZ_CP023567.1)	Erwinia pyrifoliae EpK15 phage	3'-5' exoribonuclease (WP_104945056.1)	Query: 100% E-value: 2e-139 Identity: 100%
		Metal dependent phosphohydrolases with conserved 'HD' motif (CAY74617.1)	Query: 100% E-value: 7e-130 Identity: 100%
		PerC family transcriptional regulator (WP_012668362.1)	Query: 100% E-value: 1e-107 Identity: 100%
		DNA N-6-adenine-methyltransferase (Dam) (WP_012668361.1)	Query: 100% E-value: 0.0 Identity: 100%
		site-specific DNA-methyltransferase (WP_012668355.1)	Query: 100% E-value: 0.0 Identity: 100%
Erwinia pyrifoliae DSM 12163 (NC_017390.1)	Erwinia pyrifoliae DSM 12163 phage	N-Acyltransferase superfamily (WP_012668325.1)	Query: 100% E-value: 0.0 Identity: 100%
		Isocitrate dehydrogenase (CAY74567.1)	Query: 100% E-value: 5e-28 Identity: 100%
		site-specific DNA-methyltransferase (WP_012668355.1)	Query: 100% E-value: 0.0 Identity: 100%
		DNA N-6-adenine-methyltransferase (Dam) (WP_012668361.1)	Query: 100% E-value: 0.0 Identity: 100%
		PerC family transcriptional regulator (WP_012668362.1)	Query: 100% E-value: 1e-107 Identity: 100%
Erwinia sp. QL-ZE (NZ_CP037950.1)	Erwinia sp. QL-ZE phage 1	DNA polymerase V (WP_134824957.1)	Query: 100% E-value: 2e-53 Identity: 100%
		Acyltransferase (WP_134824712.1)	Query: 100% E-value: 0.0 Identity: 100%
		TonB family protein (WP_134824658.1)	Query: 100% E-value: 1e-64 Identity: 100%
		Serine acetyltransferase (WP_134824679.1)	Query: 100% E-value: 4e-42 Identity: 100%
Erwinia tasmaniensis Et1/99 (NC_010694.1)	Erwinia tasmaniensis phage	Prokaryotic dksA/traR C4-type zinc finger (WP_012440312.1)	Query: 100% E-value: 7e-65 Identity: 100%
		Adenine-specific DNA Methyltransferase (WP_042958617.1)	Query: 100% E-value: 1e-48 Identity: 100%

TABLE 3: Found genes that possibly give special characteristics to the host.

Host (GenBank acession)	Prophage	ORF	Function	BLASTp parameters
Erwinia persicina BC4 (NZ_CP022725.1)	Erwinia persicina phage 1	GtrA Family protein	Involved in the synthesis of cell surface polysaccharides	Query: 100% E-value: 5e-79 Identity: 100%
Erwinia *psidii* (NZ_RHHM00000000.1)	Erwinia psidii phage 2	QacE	Involved in the resistance to bacteriocidal quaternary ammonium compounds	Query: 100% E-value: 2e-84 Identity: 100%
Erwinia pyrifoliae Ep1/96 (NC_012214.1)	Erwinia pyrifoliae Ep1/96 phage	Type II toxin-antitoxin system HicB family antitoxin	Neutralizes the effects of the HicA toxin	Query: 100% E-value: 1e-97 Identity: 100%
Erwinia pyrifoliae EpK15 (NZ_CP023567.1)	Erwinia pyrifoliae EpK15 phage	Type II toxin-antitoxin system HicB family antitoxin	Neutralizes the effects of the HicA toxin	Query: 100% E-value: 1e-97 Identity: 100%
		Type II toxin-antitoxin system HicA family toxin	Inhibition of cell growth and reduces the number of colony-forming cells	Query: 100% E-value: 3e-35 Identity: 100%
Erwinia pyrifoliae DSM 12163 (NC_017390.1)	Erwinia pyrifoliae DSM 12163 phage	Type II toxin-antitoxin system HicB family antitoxin	Neutralizes the effects of the HicA toxin	Query: 100% E-value: 1e-97 Identity: 100%

TABLE 4: CRISPR system detected in genomes of *Erwinia s*pp.

Host (GenBank acession)	Number of CRISPR array informed by CRISPRone	Type/subtype of CRISPR informed by CRISPRone	Number of spacers	Genes cas informed by CRISPRone
Erwinia amylovora E-2 (NZ_CP024970.1)	3	Type: I Subtype: I-B and I-E	74	10
Erwinia amylovora 1430 (NC_013961.1)	3	Type: I Subtype: I-B and I-E	75	10
Erwinia amylovora 49946 (NC_013971.1)	3	Type: I Subtype: I-B and I-E	72	10
Erwinia billingiae (NC_014306.1)	0	0	0	0
Erwinia gerundensis (NZ_LN907827.1)	0	0	0	3
Erwinia persicina (NZ_CP022725.1)	1[a]	Type: I Subtype: I-A	2[a]	3
Erwinia psidii (NZ_RHHM00000000.1)	1[b]	-	3[b]	0
Erwinia pyrifoliae strain Ep1/96 (NC_012214.1)	4	Type: I Subtype: I-E and I-F	42	14
Erwinia pyrifoliae strain EpK15 (NZ_CP023567.1)	4	Type: I Subtype: I-E and I-F	41	14
Erwinia pyrifoliae strain DSM 12163 (NC_017390.1)	4	Type: I Subtype: I-E and I-F	42	14
Erwinia sp. strain Ejp617 (NC_017445.1)	2	Type: I Subtype:I-E	30	8
Erwinia sp. strain pQL-ZE (NZ_CP037950.1)	0	0	0	0
Erwinia tasmaniensis Et1/99 (NC_010694.1)	2	Type: I Subtype: I-A and I-F	41	9

[a]Detected only by CRISPRone; [b]Detected only by CRISPRCasFinder.

FIGURE LEGENDS

Figure 1: Features of prophages found in *Erwinia* genomes. (A) Distribution of prophages defectives and intacts found in *Erwinia* genomes detected using PHASTER. (B) Number of prophages related to the size range of the sequence found. (C) Contribution of prophage content to genomic composition of chromosomes and plasmids (D) of *Erwinia* spp. divided into defective and intact. (E) Correlation between the genome size of *Erwinia* and prophages. According with the Pearson correlation coefficient, plots has no linear relationship. (F) Comparation of GC content between bacterial genomes and prophages. (G) Distribution of intact phage in viral families. Two phages are P4-like and have not been classified by VirFam in the family because they do not have structural genes.

Figure 2: Genomic organization of intact prophages. The arrows indicate open reading frames (ORFs) in the genomes of the prophages predicted by the RAST server. Equal colors determine ORFs with the same function. Their designations are labeled at the end of figure.

Figure 3: Heatmap of the identity of intact prophages, demonstrating similarity between two pairs of sequence.

Figure 4: Phylogenomic tree built using a sequence of nucleotides inferred by the VICTOR server. All pairwise comparisons of the nucleotide sequences were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method under settings recommended for prokaryotic viruses. The scale bars represent the calculated distance metric, branch support values at the VICTOR trees were calculated from 100 pseudobootstrap replicates.

Figure 5: Dendogram among intact prophages sequences and related genomes determined by ViPTree server based on genome-wide sequence similarities computed by tBLASTx. The arrows represent, from left to right: E. tasmaniensis phage, E. billingiae phage 2, E. gerundensis phage, E. psidii phage 2, E. billingiae phage 1, E. psidii phage 1, E. billingiae phage 3, E. sp. QL-ZE phage 2, E. sp. QL-ZE phage 1, E. pyrifoliae phage Ep1 / 96 phage, E. pyrifoliae EpK15 phage, E. pyrifoliae DSM 12163 phage and E. persicina phage.

Figure 6: Distribution of bacterial genes found in intact prophages according to the criteria used. The description of each gene is in Table 2.

Figure 7: Relation between the CRISPR-Cas system and putative prophages targeted sequences. (A) Percentage of spacer sequences that are targets of CRISPR-Cas of *Erwinia* spp. (B) Distribution of prophages and number of CRISPR spacers in *Erwinia g*enomes. (C) Proportion of spacers corresponding to tailed phage genomes. The genomes *Erwinia billingiae*, *Erwinia gerundensis* and *Erwinia* sp. QL-ZE were removed from the figure because they have no CRISPR locus and in the genomes *Erwinia persicina* and *Erwinia psidii*, no spacers were found that hit with phages, and were also removed from the figure.

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://apsjournals.apsnet.org/doi/10.1094/PHYTO-10-11-0277

Almpanis, A., Swain, M., Gatherer, D., and McEwan, N. (2018) Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microb Genomics* **4** https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000168.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990b) Basic local alignment search tool. *J Mol Biol* **215**: 403–410 https://linkinghub.elsevier.com/retrieve/pii/S0022283605803602.

Arriel, D.A.A., Fonseca, N.R., Guimarães, L.M.S., Hermenegildo, P.S., Mafia, R.G., Borges Júnior, N., *et al.* (2014) Wilt and die-back of *Eucalyptus* spp. caused by *Erwinia psidii* in Brazil. *For Pathol* **44**: 255–265 http://doi.wiley.com/10.1111/efp.12087.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al.* (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**: 75 http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-9-75.

Baltz, R.H. (2012) Streptomyces temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms). *J Ind Microbiol Biotechnol*. **39**: 661-72 http://link.springer.com/10.1007/s10295-011-1069-6

Berjón-Otero, M., Koslová, A., and Fischer, M.G. (2019) The dual lifestyle of genome-integrating virophages in protists. *Ann N Y Acad Sci*. **1447**: 97-109 https://nyaspubs.onlinelibrary.wiley.com/doi/full/10.1111/nyas.14118

Bobay, L.-M., Touchon, M., and Rocha, E.P.C. (2014) Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci* **111**: 12127–12132

http://www.pnas.org/cgi/doi/10.1073/pnas.1405336111.

Bolduc, B., Jang, H. Bin, Doulcier, G., You, Z.-Q., Roux, S., and Sullivan, M.B. (2017) vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ.* **5**: e3243 https://peerj.com/articles/3243.

Bondy-Denomy, J., Qian, J., Westra, E.R., Buckling, A., Guttman, D.S., Davidson, A.R., and Maxwell, K.L. (2016) Prophages mediate defense against phage infection through diverse mechanisms. *ISME J* **10**: 2854–2866 http://www.nature.com/articles/ismej201679.

Boyd, E.F., and Brüssow, H. (2002) Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol* **10**: 521–529 https://linkinghub.elsevier.com/retrieve/pii/S0966842X02024599.

Breitbart, M. (2012) Marine Viruses: Truth or Dare. *Ann Rev Mar Sci.* **4**: 425–448 http://www.annualreviews.org/doi/10.1146/annurev-marine-120709-142805.

Brussow, H., Canchaya, C., and Hardt, W.-D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev.* **68**: 560–602 http://mmbr.asm.org/cgi/doi/10.1128/MMBR.68.3.560-602.2004.

Catalao, M.J., Gil, F., Moniz-Pereira, J., and Pimentel, M. (2011) Functional analysis of the holin-like proteins of Mycobacteriophage Ms6. *J Bacteriol* **193**: 2793–2803 http://jb.asm.org/cgi/doi/10.1128/JB.01519-10.

Christie, G.E., and Calendar, R. (1990) Interactions between satellite cacteriophage P4 and its helpers. *Annu Rev Genet* **24**: 465–490 http://www.annualreviews.org/doi/10.1146/annurev.ge.24.120190.002341.

Christie, G.E., and Calendar, R.L. (2011) P4-like Satellite Viruses. In *The Springer Index of Viruses*. Springer New York, New York, NY. pp. 2023–2027 http://link.springer.com/10.1007/978-0-387-95919-1_330.

Coutinho, T.A., Brady, C.L., Vaart, M. van der, Venter, S.N., Telechea, N., Rolfo, M., *et al.* (2011) A new shoot and stem disease of *Eucalyptus* species caused by *Erwinia psidii*. *Australas Plant Pathol* **40**: 55–60 http://link.springer.com/10.1007/s13313-010-0013-y.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., *et al.* (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* **46**: 246–251 https://academic.oup.com/nar/article/46/W1/W246/5001162.

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. *Sci Rep* **8**: 9273 http://www.nature.com/articles/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. *Front Microbiol* **10** https://www.frontiersin.org/article/10.3389/fmicb.2019.00138/full.

D'Haeze, W., and Holsters, M. (2004) Surface polysaccharides enable bacteria to evade plant immunity. *Trends Microbiol*. **12:** 555-61 https://linkinghub.elsevier.com/retrieve/pii/S0966842X04002410

Evans, T.J., Coulthurst, S.J., Komitopoulou, E., and Salmond, G.P.C. (2010) Two mobile *Pectobacterium atrosepticum* prophages modulate virulence. *FEMS Microbiol Lett*. **304**: 195-202 https://academic.oup.com/femsle/article-lookup/doi/10.1111/j.1574-6968.2010.01901.x

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A.A. (2015) A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol* **13**: 641–650 http://www.nature.com/articles/nrmicro3527.

Fenton, M., McAuliffe, O., O'Mahony, J., and Coffey, A. (2010) Recombinant

bacteriophage lysins as antibacterials. *Bioeng Bugs* **1**: 9–16 http://www.tandfonline.com/doi/abs/10.4161/bbug.1.1.9818.

Gandon, S. (2016) Why Be Temperate: Lessons from Bacteriophage λ. *Trends Microbiol* **24**: 356–365 https://linkinghub.elsevier.com/retrieve/pii/S0966842X16000445.

Golding, I., Coleman, S., Nguyen, T.V.P., and Yao, T. (2019) Decision making by temperate phages. In *Reference Module in Life Sciences*. Elsevier, https://linkinghub.elsevier.com/retrieve/pii/B9780128096338209692.

Guan, S., Bastin, D.A., and Verma, N.K. (1999) Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX. *Microbiology* **145**: 1263–1273 https://www.microbiologyresearch.org/content/journal/micro/10.1099/13500872-145-5-1263.

Guglielmini, J., and Melderen, L. Van (2011) Bacterial toxin-antitoxin systems: Translation inhibitors everywhere. *Mob Genet Elements*. **1**: 283-290 http://www.tandfonline.com/doi/abs/10.4161/mge.18477

Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., and Sullivan, M.B. (2017) Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J* **11**: 1511–1520 http://www.nature.com/articles/ismej201716.

Isaksen, M.L., Dokland, T., and Lindqvist, B.H. (1993) Characterization of the capsid associating activity of bacteriophage p4's psu protein. *Virology*. **194**: 674-81 https://linkinghub.elsevier.com/retrieve/pii/S0042682283713073

Isaksen, M.L., Rishovd, S.T., Calendar, R., and Lindqvist, B.H. (1992) The polarity suppression factor of bacteriophage P4 is also a decoration protein of the P4 capsid. *Virology*. **188**: 831-9 https://linkinghub.elsevier.com/retrieve/pii/004268229290538Z

Jorgensen, M.G., Pandey, D.P., Jaskolska, M., and Gerdes, K. (2009) HicA of

Escherichia coli defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *J Bacteriol* **191**: 1191–1199 http://jb.asm.org/cgi/doi/10.1128/JB.01013-08.

Joshi, H., Seniya, S.P., Suryanarayanan, V., Patidar, N.D., Singh, S.K., and Jain, V. (2017) Dissecting the structure-function relationship in lysozyme domain of mycobacteriophage D29-encoded peptidoglycan hydrolase. *FEBS Lett* **591**: 3276–3287 http://doi.wiley.com/10.1002/1873-3468.12848.

Koskella, B., and Brockhurst, M.A. (2014) Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev.* **38**: 916-931 https://academic.oup.com/femsre/article-lookup/doi/10.1111/1574-6976.12072

Lee, G.M., Oh, E.J., Ko, S., Park, J., Park, D.H., Kim, D., and Oh, C.S. (2018) Draft genome sequence of a bacterial plant pathogen *Erwinia pyrifoliae* strain EpK1/15 isolated from an apple twig showing black shoot blight. *Korean J Microbiol*. **55**: 69-70 http://test-kjm.inforang.com/journal/view.html?uid=10&vmd=Full&

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

Leplae, R., Lima-Mendez, G., and Toussaint, A. (2010) ACLAME: A CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* **38**: 57–61 https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkp938.

Li, G., Shen, M., Lu, S., Le, S., Tan, Y., Wang, J., *et al.* (2016) Identification and characterization of the HicAB toxin-antitoxin system in the opportunistic pathogen *Pseudomonas aeruginosa*. *Toxins (Basel)*. **8**: 113 http://www.mdpi.com/2072-6651/8/4/113

Lopes, A., Tavares, P., Petit, M.A., Guérois, R., and Zinn-Justin, S. (2014) Automated classification of tailed bacteriophages according to their neck organization. *BMC*

Genomics. **15:** 1027 http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-1027

Lorenz, N., Reiger, M., Toro-Nahuelpan, M., Brachmann, A., Poettinger, L., Plener, L., *et al.* (2016) Identification and initial characterization of prophages in *Vibrio campbellii*. *PLoS One*. **1**: e0156010 https://dx.plos.org/10.1371/journal.pone.0156010

Malnoy, M., Faize, M., Venisse, J.S., Geider, K., and Chevreau, E. (2005) Expression of viral EPS-depolymerase reduces fire blight susceptibility in transgenic pear. *Plant Cell Rep*. **23**: 632-8 http://link.springer.com/10.1007/s00299-004-0855-2

Meier-Kolthoff, J.P., and Göker, M. (2017) VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*. **33:** 3396-3404 https://academic.oup.com/bioinformatics/article/33/21/3396/3933260

Myung, I.-S., Lee, J.-Y., Yun, M.-J., Lee, Y.-H., Lee, Y.-K., Park, D.-H., and Oh, C.-S. (2016) Fire blight of apple, caused by *Erwinia amylovora*, a new disease in Korea. *Plant Dis* **100**: 1774–1774 http://apsjournals.apsnet.org/doi/10.1094/PDIS-01-16-0024-PDN.

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. *Mol Microbiol*. **31**: 399-419 http://doi.wiley.com/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. *J Bacteriol* **197**: 410–419 http://jb.asm.org/lookup/doi/10.1128/JB.02230-14.

Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., and Goto, S. (2017) ViPTree: The viral proteomic tree server. *Bioinformatics*. **33**: 2379-2380 https://academic.oup.com/bioinformatics/article/33/15/2379/3096426

Owen, S. V., Wenner, N., Canals, R., Makumi, A., Hammarlöf, D.L., Gordon, M.A., *et al.* (2017) Characterization of the prophage repertoire of African *Salmonella* Typhimurium ST313 reveals high levels of spontaneous induction of novel phage BTP1. *Front Microbiol*. **8**: 235 http://journal.frontiersin.org/article/10.3389/fmicb.2017.00235/full

Rhim, Volksch, Gardan, Paulin, Langlotz, Kim, and Geider (1999) *Erwinia pyrifoliae*, an *Erwinia* species different from *Erwinia amylovora*, causes a necrotic disease of Asian pear trees. *Plant Pathol* **48**: 514–520 http://doi.wiley.com/10.1046/j.1365-3059.1999.00376.x.

Rocker, A., and Meinhart, A. (2016) Type II toxin: antitoxin systems. More than small selfish entities? *Curr Genet*. **62**: 287-290 http://link.springer.com/10.1007/s00294-015-0541-7

Shen, J., Zhou, J., Xu, Y., and Xiu, Z. (2020) Prophages contribute to genome plasticity of *Klebsiella pneumoniae* and may involve the chromosomal integration of ARGs in CG258. *Genomics*. **112**: 998-1010 https://linkinghub.elsevier.com/retrieve/pii/S0888754318307225

Six, E.W. (1975) The helper dependence of satellite bacteriophage P4: Which gene functions of bacteriophage P2 are needed by P4? *Virology* **67**: 249–263 https://linkinghub.elsevier.com/retrieve/pii/0042682275904225.

Smits, T.H.M., Jaenicke, S., Rezzonico, F., Kamber, T., Goesmann, A., Frey, J.E., and Duffy, B. (2010) Complete genome sequence of the fire blight pathogen *Erwinia pyrifoliae* DSM 12163T and comparative genomic insights into plant pathogenicity. *BMC Genomics*. **11**: 2 http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-2

Schwarczinger, I., Kolozsváriné Nagy, J., Künstler, A., Szabó, L., Geider, K., Király, L., and Pogány, M. (2017) Characterization of *Myoviridae* and *Podoviridae* family bacteriophages of *Erwinia amylovora* from Hungary - potential of application in

biological control of fire blight. *Eur J Plant Pathol*. **149**: 639-652 http://link.springer.com/10.1007/s10658-017-1214-9

Tezel, U., and Pavlostathis, S.G. (2015) Quaternary ammonium disinfectants: microbial adaptation, degradation and ecology. *Curr Opin Biotechnol* **33**: 296–304 https://linkinghub.elsevier.com/retrieve/pii/S0958166915000622.

Touchon, M., Bernheim, A., and Rocha, E.P.C. (2016) Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J*. **10**: 2744-2754 http://www.nature.com/articles/ismej201647

Varani, A.M., Monteiro-Vitorello, C.B., Nakaya, H.I., and Sluys, M.-A. Van (2013) The Role of Prophage in Plant-Pathogenic Bacteria. *Annu Rev Phytopathol*. **51:** 429-51 http://www.annualreviews.org/doi/10.1146/annurev-phyto-081211-173010

Vostrov, A.A., Vostrukhina, O.A., Svarchevsky, A.N., and Rybchin, V.N. (1996) Proteins responsible for lysogenic conversion caused by coliphages N15 and φ80 are highly homologous. *J Bacteriol*. **178**: 1484-6 https://jb.asm.org/content/178/5/1484

Waterhouse, R.M., Seppey, M., Simao, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., *et al.* (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. **35**: 543-548 https://academic.oup.com/mbe/article/35/3/543/4705839

Young, R. (2014) Phage lysis: Three steps, three choices, one outcome. *J Microbiol* **52**: 243–258 http://link.springer.com/10.1007/s12275-014-4087-z.

Zeng, H., Zhang, J., Li, C., Xie, T., Ling, N., Wu, Q., and Ye, Y. (2017) The driving force of prophages and CRISPR-Cas system in the evolution of *Cronobacter sakazakii*. *Sci Rep*. **7,** 40206 http://www.nature.com/articles/srep40206

Zhang, Q., and Ye, Y. (2017) Not all predicted CRISPR-Cas systems are equal: Isolated cas genes and classes of CRISPR like elements. *BMC Bioinformatics*. **18**: 92 http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1512-4

SUPPLEMENTARY TABLES

Supplementary Table 1: Genomes of *Erwinia* spp. used in this study.

Species	Strain	Acession no.	Taxonomy ID	Reference
Erwinia amylovora	E-2	NZ_CP024970.1	552	Lagonenko et al., 2008
Erwinia amylovora	1430	NC_013961.1	665029	Smits et al., 2010
Erwinia amylovora	49946	NC_013971.1	716540	Sebaihia et al., 2010
Erwinia billingiae	Eb661	NC_014306.1	634500	Kube et al., 2010
Erwinia gerundensis	EM 595	NZ_LN907827.1	1619313	Rezzonico et al., 2016
Erwinia persicina	BC4	NZ_CP022725.1	55211	Unpublished
Erwinia psidii	IBSBF 435	NZ_RHHM00000000.1	69224	Hermenegenildo et al., 2019
Erwinia pyrifoliae	Ep1/96	NC_012214.1	634499	Kube et al., 2010
Erwinia pyrifoliae	EpK1/15	NZ_CP023567.1	79967	Lee et al., 2018
Erwinia pyrifoliae	DSM 12163	NC_017390.1	644651	Smits et al., 2010
Erwinia sp.	Ejp617	NC_017445.1	215689	Park et al., 2011
Erwinia sp.	QL-Z3	NZ_CP037950.1	2547962	Unpublished
Erwinia tasmaniensis	Et1/99	NC_010694.1	465817	Kube et al., 2008

Supplementary Table 2: Defective prophages present in genomes of *Erwinia s*pp. available on GenBank (NCBI) with the reasons to classify as defective prophages.

No.	Host (GenBank acession)	Prophage size (kb)*	GC %*	Coordinates in bacterial genome*	% Genome of host	Classification according to VirFam	Reason(s) to classify the prophage as putative defective virus
1	*Erwinia amylovora* E-2 (NZ_CP024970.1)	**29.6**	50.17	340.663 - 370.288	**0.78**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent.
2	*Erwinia amylovora* E-2 (NZ_CP024970.1)	**8.2**	45.90	1.931.595 - 1.939.84	**0.22**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except two proteins of tail); terminase - absent; proteins of lysis - absent.
3	*Erwinia amylovora* 1430 (NC_013961.1)	**9.5**	46.96	1.894.799 - 1.904.356	**0.25**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except protein of baseplate and one protein of tail); terminase - absent; proteins of lysis - absent.
4	*Erwinia amylovora* 1430 (NC_013961.1)	**13**	45.36	3.418.865 – 3.431.925	**0.34**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except two proteins of tail).
5	*Erwinia amylovora* 49946 (NC_013971.1)	**10.9**	45.20	339.739 - 350.732	**0.29**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except protein of baseplate and one protein of tail); attachment sites - absent.
6	*Erwinia amylovora* 49946 (NC_013971.1)	**9.5**	46.94	1.932.326 - 1.941.883	**0.25**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except two proteins of tail); terminase - absent; proteins of lysis - absent.
7	*Erwinia billingiae* Eb661 (NC_014306.1)	**19.3**	49.85	3.552.551 - 3.571.881	**0.38**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except protein major capsid and protein of portal); integrase - absent; attachment sites - absent.
8	*Erwinia* sp. Ejp617 (NC_017445.1)	**7.6**	52.95	3.474.113 - 3.481.770	**0.20**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent; integrase - absent.
9	*Erwinia* sp. Ejp617 (NC_017445.1)	**16.1**	47.35	3.781.420 - 3.797.548	**0.41**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except one protein of tail); terminase - absent; proteins of lysis - absent.
10	*Erwinia persicina* BC4 (NZ_CP022725.1)	**17.5**	43.71	578.831 - 596.380	**0.37**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except some proteins of tail); terminase - absent; proteins of lysis - absent.

11	*Erwinia persicina* BC4 (NZ_CP022725.1)	**14.6**	43.71	1.446.987 - 1.461.675	**0.31**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except protein portal); terminase - absent; proteins of lysis - absent; integrase - absent; attachment sites - absent.
12	*Erwinia persicina* BC4 (NZ_CP022725.1)	**13.1**	57.53	3.306.924 - 3.320.076	**0.27**	***Myoviridae***	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except major tail tube protein, protein sheath and one protein of tail); terminase - absent; integrase - absent; attachment sites - absent.
13[a]	*Erwinia persicina* BC4 (NZ_CP022726.1)	**29.7**	53.28	67.450 - 97.242	**20.65**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent.
14	*Erwinia* sp. QL-ZE (NZ_CP037950.1)	**10.7**	51.12	1.283.392 - 1.294.091	**0.22**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent.
15	*Erwinia psidii* (NZ_RHHM00000000.1)	**30.7**	50.72	3.019.226 - 3.050.019	**0.68**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except one protein of tail); terminase - absent.
16	*Erwinia psidii* (NZ_RHHM00000000.1)	**12.7**	46.88	3.651.818 - 3.664.572	**0.28**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; proteins of lysis - absent; integrase - absent.
17	*Erwinia psidii* (NZ_RHHM00000000.1)	**12.5**	49.71	3.660.280 - 3.672.819	**0.28**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent.
18	*Erwinia pyrifoliae* Ep1/96 (NC_012214.1)	**15.6**	53.41	74.217 - 89.839	**0.39**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent (except one protein of tail); terminase - absent; proteins of lysis - absent.
19	*Erwinia pyrifoliae* Ep1/96 (NC_012214.1)	**10.8**	52.98	1.059.276 - 1.070.142	**0.27**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent; integrase - absent.
20	*Erwinia pyrifoliae* Ep1/96 (NC_012214.1)	**28.8**	49.71	1.976.290 - 2.005.154	**0.72**	***Myoviridae***	PHASTER score: incomplete Manual inspection: structural phage proteins – absent (protein major capsid, protein of portal and protein major tail tube); terminase - absent; attachment sites - absent.

21	*Erwinia pyrifoliae* Ep1/96 (NC_012214.1)	**9.6**	51.67	2.591.863 - 2.601.464	**0.24**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent; integrase - absent; attachment sites - absent.
22	*Erwinia pyrifoliae* Ep1/96 (NC_012214.1)	**37**	58.32	3.176.080 - 3.213.140	**0.92**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent.
23	*Erwinia pyrifoliae* EpK15 (NZ_CP023567.1)	**34.7**	50.33	1.977.932 - 2.012.672	**0.86**	***Myoviridae***	PHASTER score: incomplete Manual inspection: structural phage proteins – absent (protein major capsid, protein major tail tube and protein of portal); terminase - absent; integrase - absent.
24	*Erwinia pyrifoliae* EpK15 (NZ_CP023567.1)	**10**	54.18	3.907.773 - 3.917.827	**0.25**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent.
25[a]	*Erwinia pyrifoliae* EpK15 (NZ_CP023568.1)	**6.4**	47.02	29.346 - 35.834	**13.39**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent; integrase - absent; attachment sites - absent.
26	*Erwinia pyrifoliae* DSM 12163 (NC_017390.1)	**9.8**	52.97	1.059.308 - 1.069.187	**0.24**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent; integrase - absent; attachment sites - absent.
27	*Erwinia pyrifoliae* DSM 12163 (NC_017390.1)	**28.8**	49.71	1.976.341 - 2.005.206	**0.72**	***Myoviridae***	PHASTER score: incomplete Manual inspection: some structural phage proteins – absent (protein major capsid, protein major tail tube and protein of portal); terminase - absent; attachment sites - absent.
28	*Erwinia pyrifoliae* DSM 12163 (NC_017390.1)	**9.6**	51.67	2.591.921 - 2.601.522	**0.24**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent; integrase - absent; attachment sites - absent.
29	*Erwinia pyrifoliae* DSM 12163 (NC_017390.1)	**37**	58.25	3.176.004 - 3.213.065	**0.92**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent.

30	*Erwinia tasmaniensis* Et1/99 (NC_010694.1)	**17.1**	53.64	2.728.403 - 2.745.594	**0.44**	Unclassified	PHASTER score: incomplete Manual inspection: structural phage proteins - absent; terminase - absent; proteins of lysis - absent.

*Information provided by PHASTER. [a]Phages found in plasmids.

Supplementary Table 3: Results of phylogenetic analysis by VICTOR server.

Genomes	Species	Genus	Family
E. billingiae phage 1	1	5	2
E. pyrifoliae Ep1/96 phage	3	1	1
E. pyrifoliae EpK15 phage	2	1	1
E. pyrifoliae DSM 12163 phage	3	1	1
E. tasmaniensis phage	4	5	2
E. billingiae phage 2	5	5	2
E. billingiae phage 3	6	4	2
E. gerundensis phage	7	5	2
E. persicina phage	8	2	1
E. sp. QL-ZE phage 1	9	3	1
E. sp. QL-Z3 phage 2	10	4	2
E. psidii phage 1	11	5	2
E. psidii phage 2	12	5	2

Supplementary Table 4: Parameters of the programs used in the analyzes

PROGRAM	PARAMETER	VALUE
BUSCO	Database	Gammaproteobacteria
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
VirFam	Type of analysis	Head-neck-tail module genes
ResFinder	Option	Acquired antimicrobial resistance genes
	Type of reads	Assembled Genome/Contigs
CARD	Data type	BLASTn
VICTOR	Data type	Nucleotide
VipTree	Nucleic acid types of viruses	dsDNA
	Host categories of viruses	Prokaryote
	Genetic code for gene prediction	Prokaryotic (table 11)

FIGURES

Figure 1

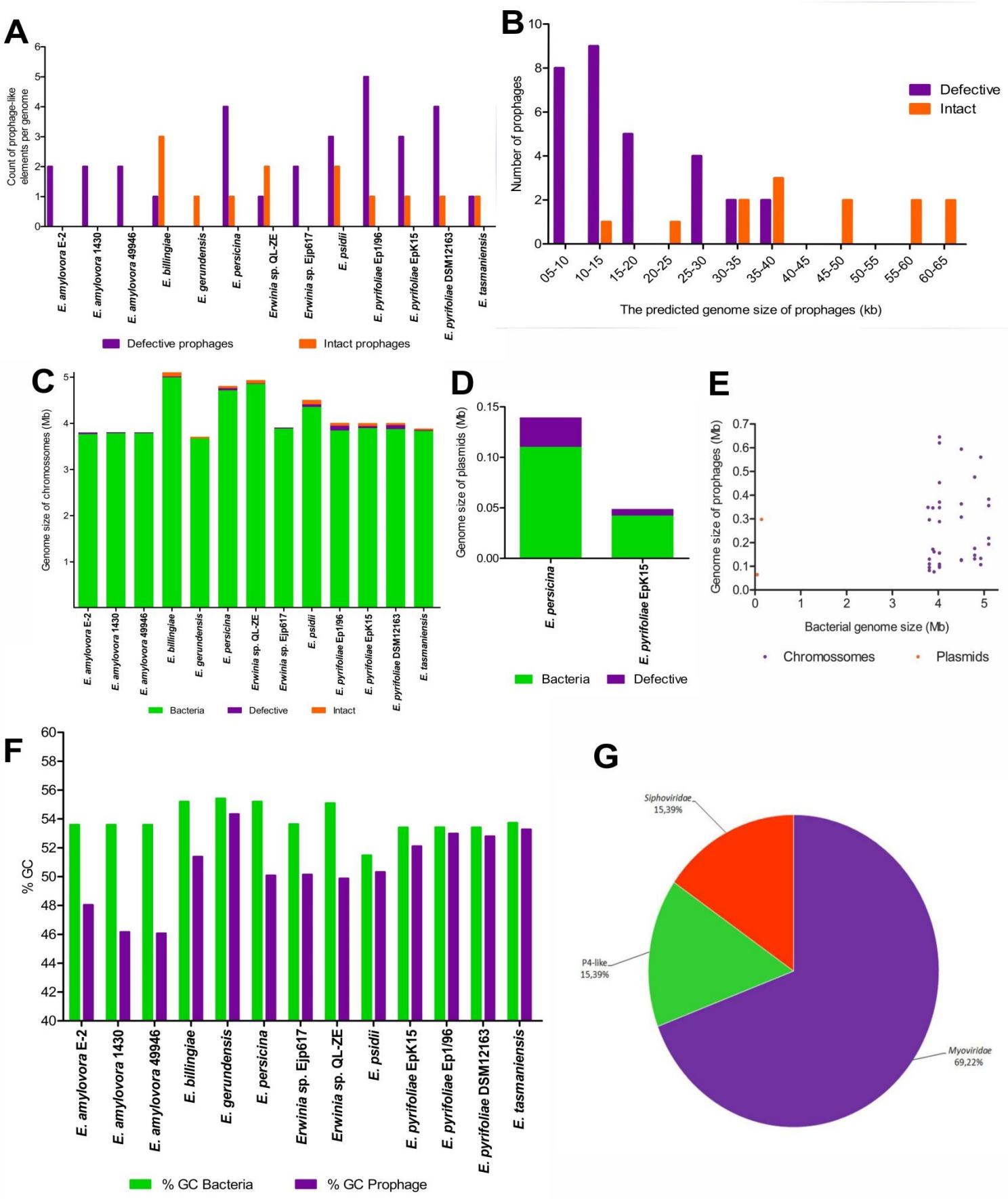


Figure 2

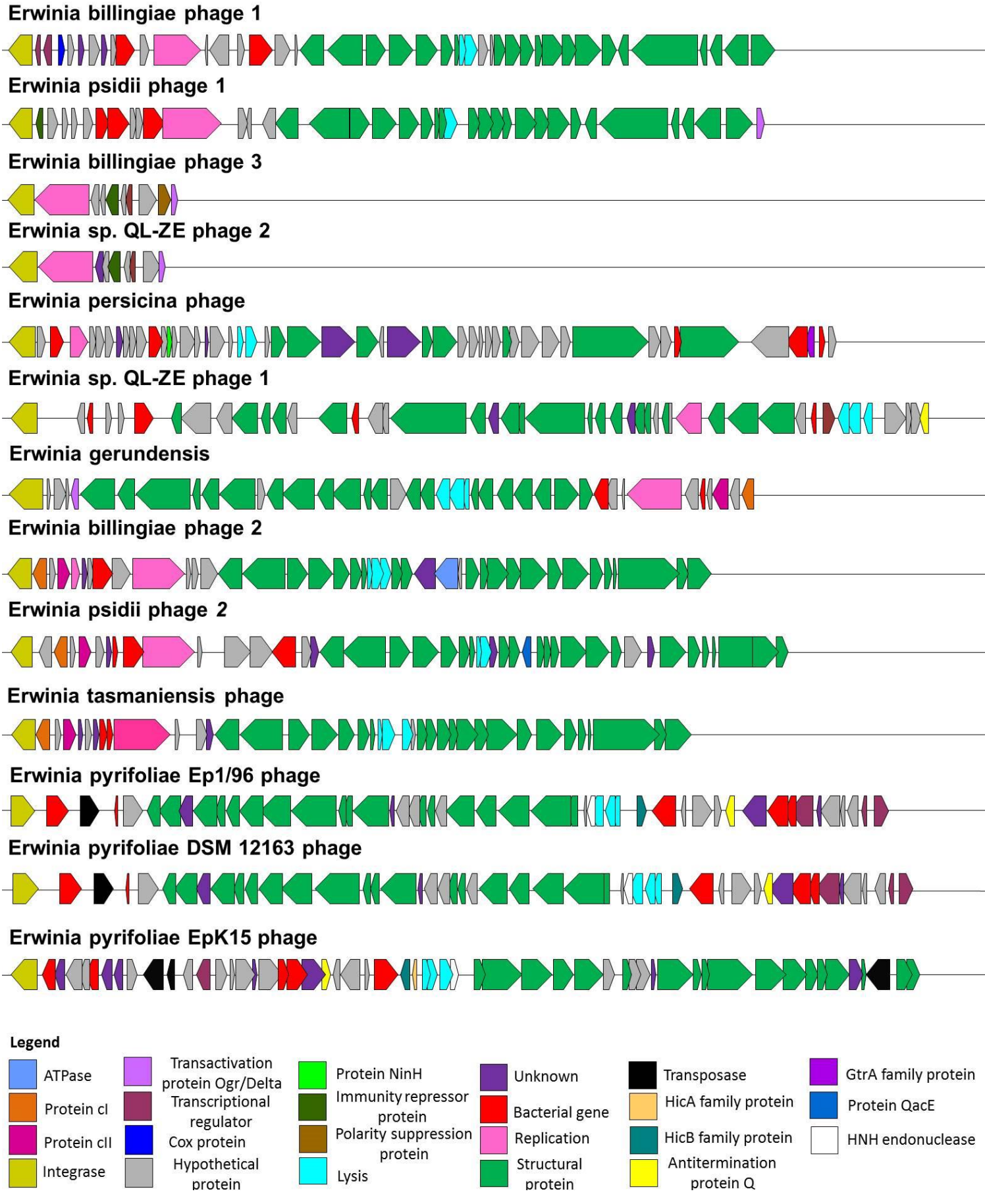

Erwinia billingiae phage 1
Erwinia psidii phage 1
Erwinia billingiae phage 3
Erwinia sp. QL-ZE phage 2
Erwinia persicina phage
Erwinia sp. QL-ZE phage 1
Erwinia gerundensis
Erwinia billingiae phage 2
Erwinia psidii phage 2
Erwinia tasmaniensis phage
Erwinia pyrifoliae Ep1/96 phage
Erwinia pyrifoliae DSM 12163 phage
Erwinia pyrifoliae EpK15 phage
Legend
ATPase
Protein cI
Protein cII
Integrase
Transactivation protein Ogr/Delta
Transcriptional regulator
Cox protein
Hypothetical protein
Protein NinH
Immunity repressor protein
Polarity suppression protein
Lysis
Unknown
Bacterial gene
Replication
Structural protein
Transposase
HicA family protein
HicB family protein
Antitermination protein Q
GtrA family protein
Protein QacE
HNH endonuclease

Figure 3

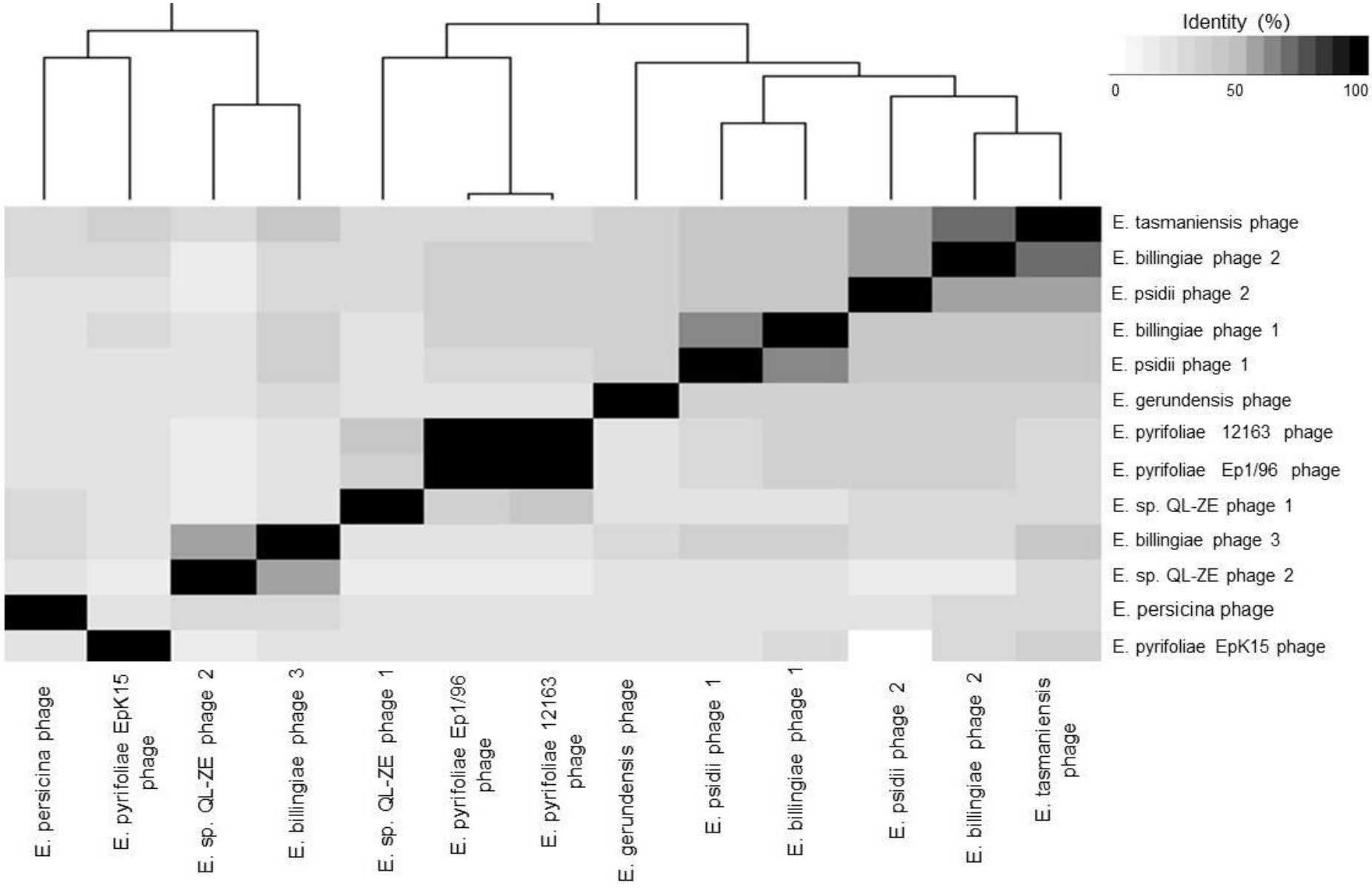

Identity (%)
0
50
100
E. tasmaniensis phage
E. billingiae phage 2
E. psidii phage 2
E. billingiae phage 1
E. psidii phage 1
E. gerundensis phage
E. pyrifoliae 12163 phage
E. pyrifoliae Ep1/96 phage
E. sp. QL-ZE phage 1
E. billingiae phage 3
E. sp. QL-ZE phage 2
E. persicina phage
E. pyrifoliae EpK15 phage
E. persicina phage
E. pyrifoliae EpK15 phage
E. sp. QL-ZE phage 2
E. billingiae phage 3
E. sp. QL-ZE phage 1
E. pyrifoliae Ep1/96 phage
E. pyrifoliae 12163 phage
E. gerundensis phage
E. psidii phage 1
E. billingiae phage 1
E. psidii phage 2
E. billingiae phage 2
E. tasmaniensis phage

Figure 4

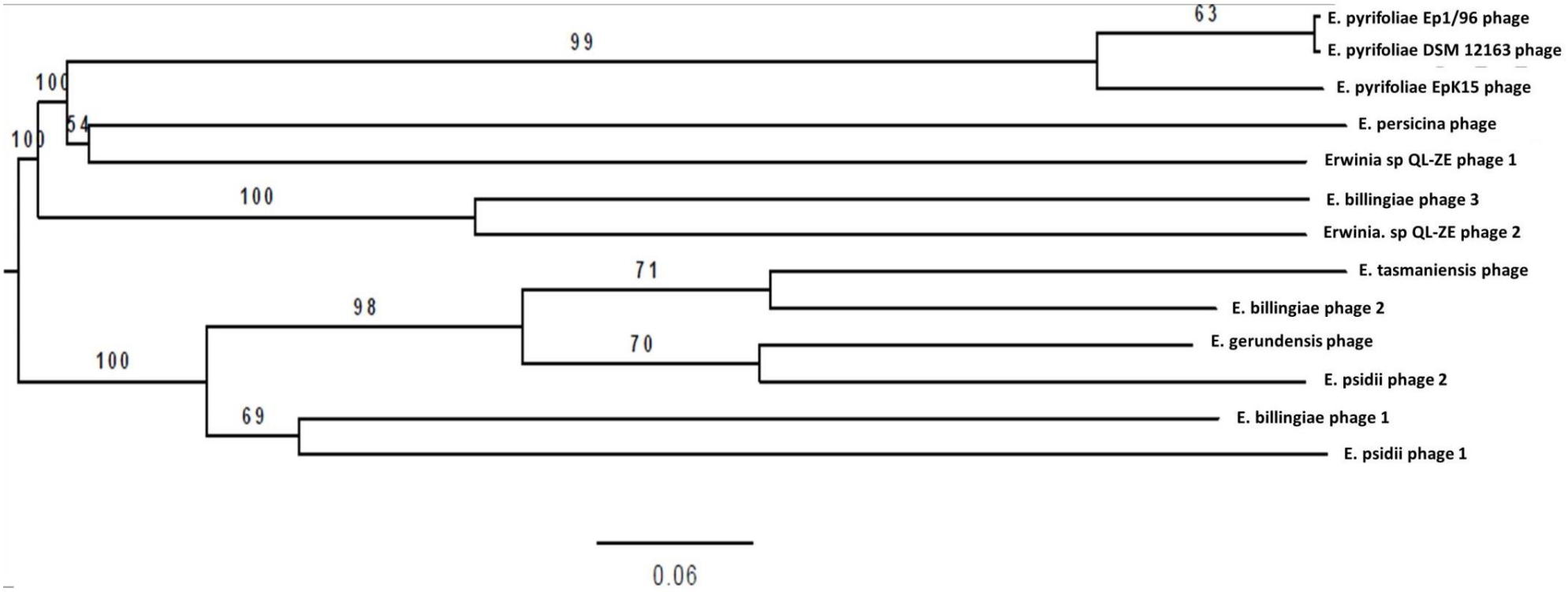

63
99
100
54
100
100
71
98
70
100
69
E. pyrifoliae Ep1/96 phage
E. pyrifoliae DSM 12163 phage
E. pyrifoliae EpK15 phage
E. persicina phage
Erwinia sp QL-ZE phage 1
E. billingiae phage 3
Erwinia. sp QL-ZE phage 2
E. tasmaniensis phage
E. billingiae phage 2
E. gerundensis phage
E. psidii phage 2
E. billingiae phage 1
E. psidii phage 1
0.06

Figure 5

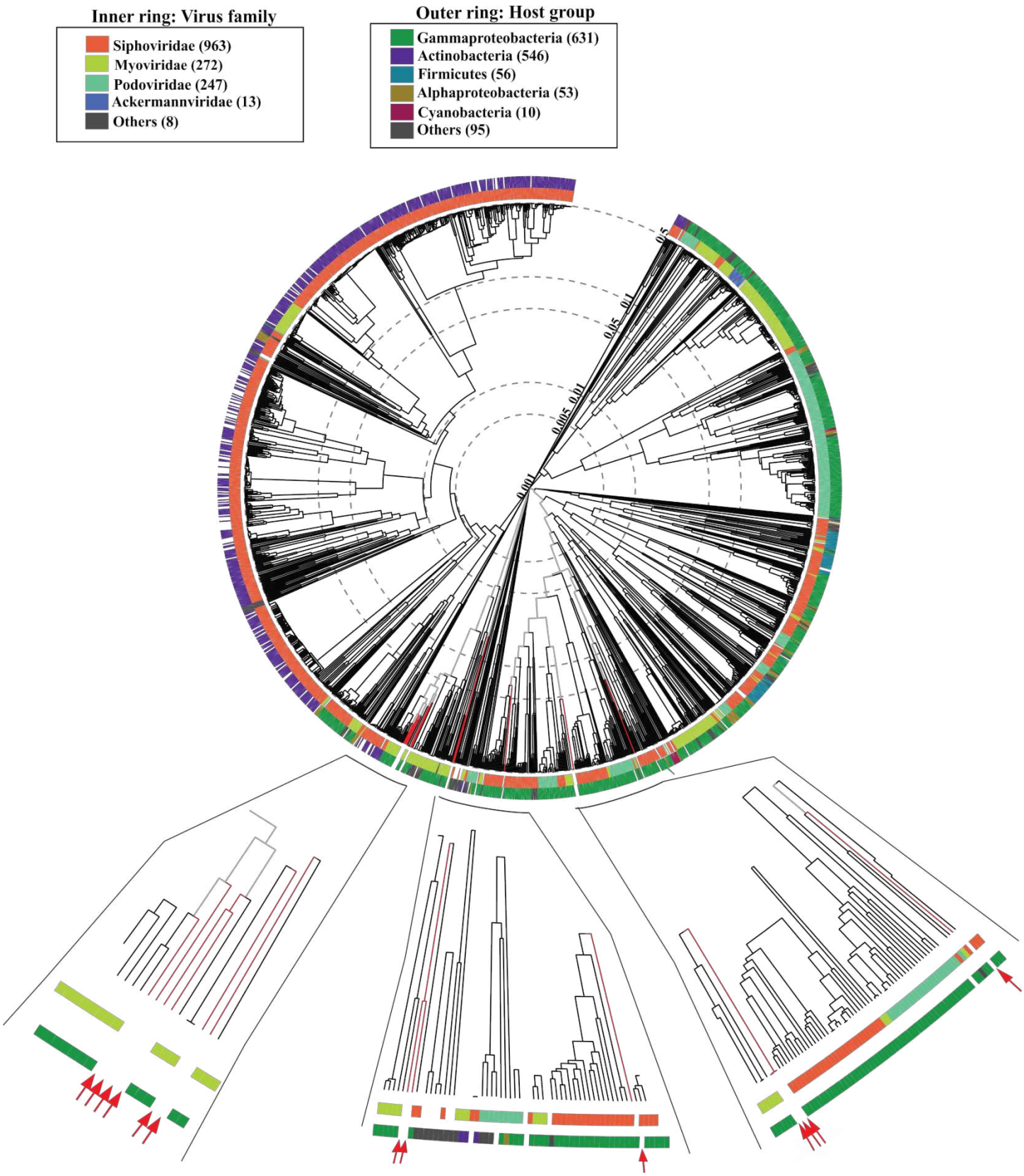

Inner ring: Virus family
Siphoviridae (963)
Myoviridae (272)
Podoviridae (247)
Ackermannviridae (13)
Others (8)
Outer ring: Host group
Gammaproteobacteria (631)
Actinobacteria (546)
Firmicutes (56)
Alphaproteobacteria (53)
Cyanobacteria (10)
Others (95)
0.5
0.1
0.05
0.01
0.005
0.001

Figure 6

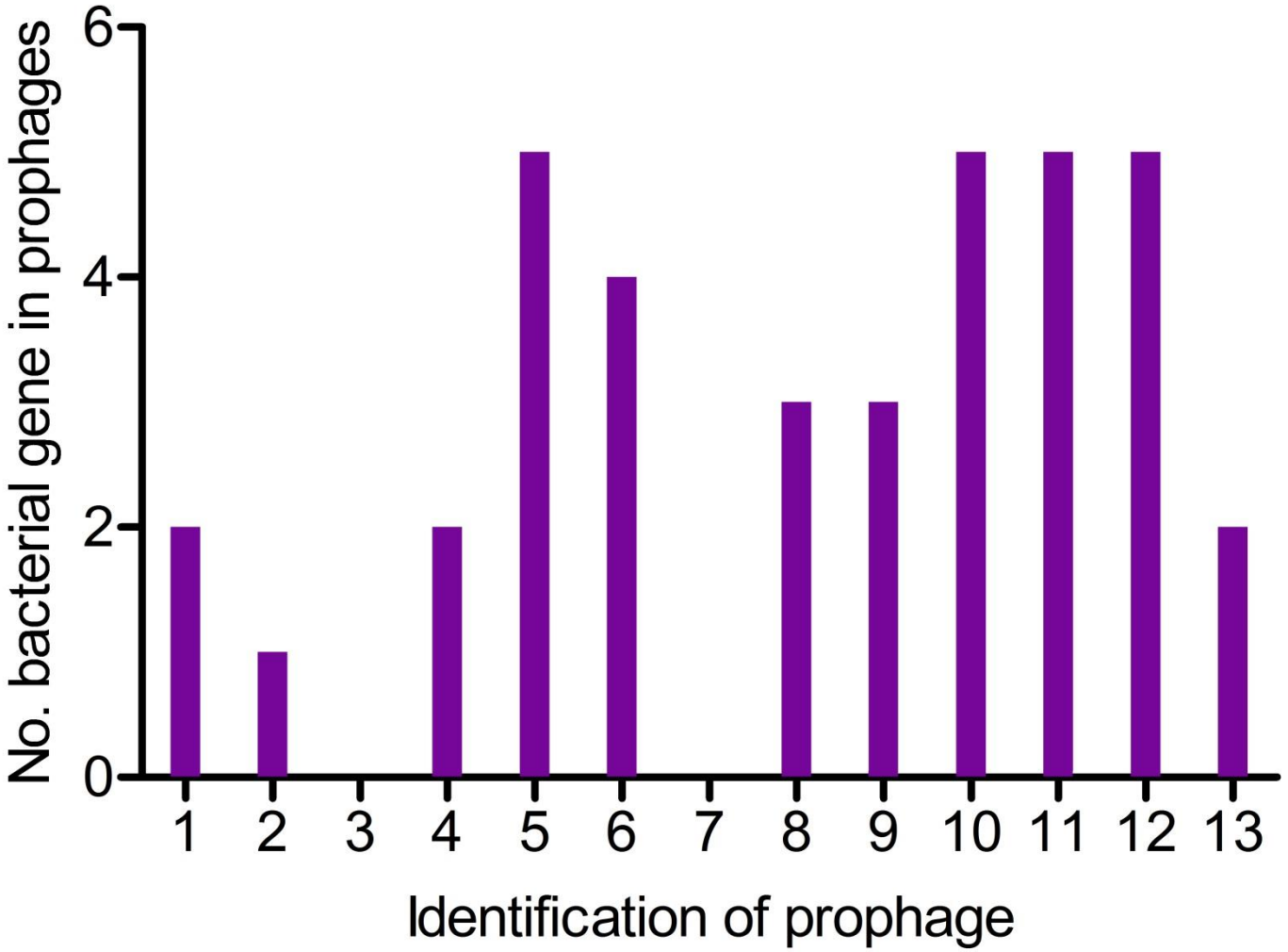
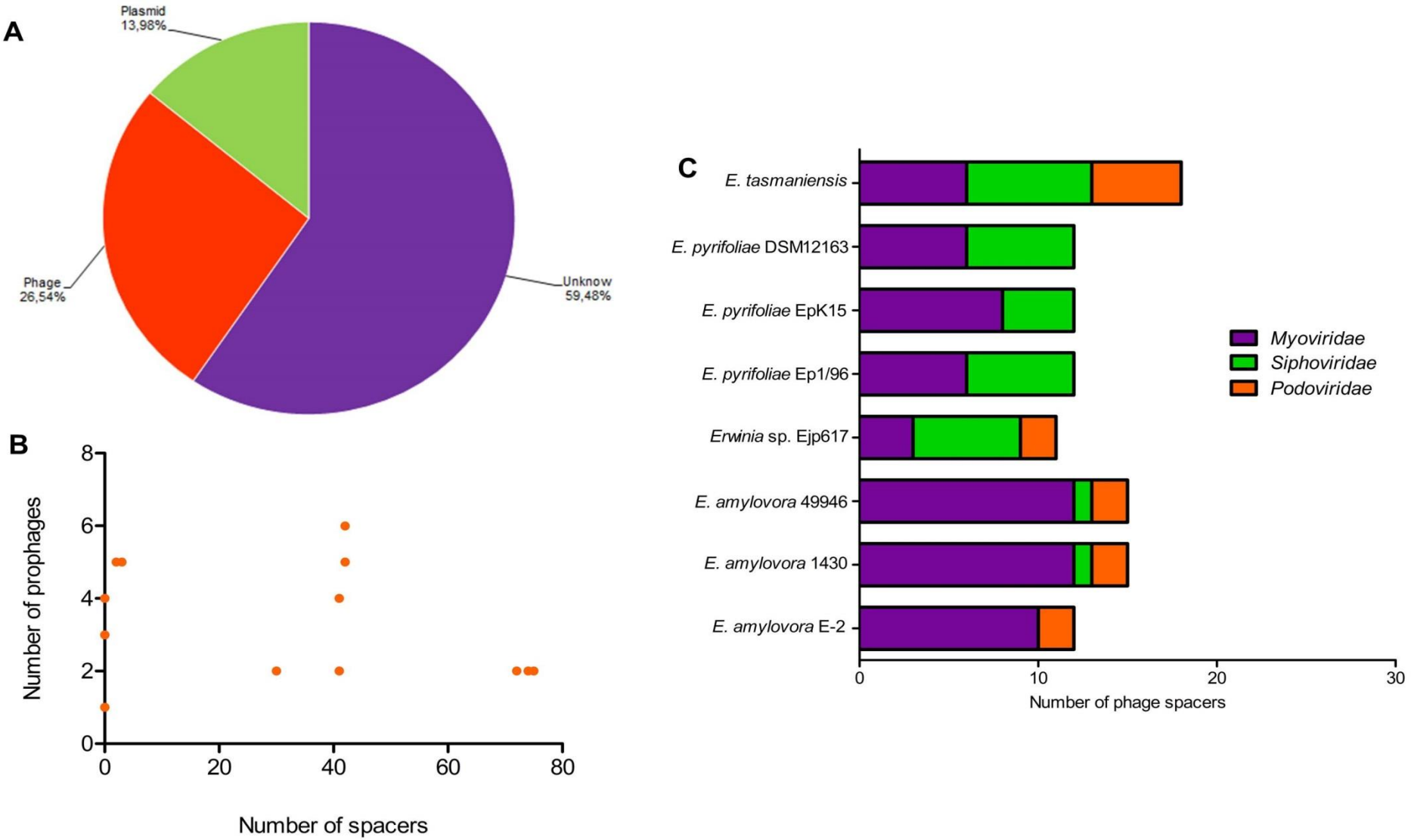

No. bacterial gene in prophages
0
2
4
6
1
2
3
4
5
6
7
8
9
10
11
12
13
Identification of prophage

Figure 7


A
Plasmid
13,98%
Phage
26,54%
Unknow
59,48%
B
Number of prophages
Number of spacers
0
2
4
6
8
0
20
40
60
80
C
E. tasmaniensis
E. pyrifoliae DSM12163
E. pyrifoliae EpK15
E. pyrifoliae Ep1/96
Erwinia sp. Ejp617
E. amylovora 49946
E. amylovora 1430
E. amylovora E-2
Myoviridae
Siphoviridae
Podoviridae
0
10
20
30
Number of phage spacers