UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

WESLEY OLIVEIRA DE SANTANA

GENÔMICA COMPARATIVA DE *Xylella fastidiosa*:
DIVERSIDADE DO PANGENOMA E ANÁLISE DE GENES DE PATOGENICIDADE

Versão corrigida da Tese conforme resolução CoPGr5890
O original se encontra disponível na Secretaria de Pós-Graduação do IQ-USP

São Paulo
Data de Depósito na SPG: **27/11/2012**

WESLEY OLIVEIRA DE SANTANA

**GENÔMICA COMPARATIVA DE *Xylella fastidiosa*:
DIVERSIDADE DO PANGENOMA E ANÁLISE DE GENES DE PATOGENICIDADE**

Tese apresentada ao Instituto de
Química da Universidade de São Paulo para obtenção
do Título de Doutor em Ciências, área Bioquímica

Orientadora: Profa. Dra. Aline Maria da Silva

São Paulo

2012

AUTORIZO A REPRODUÇÃO E A DIVULGAÇÃO DE TODO OU PARTE DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Santana, Wesley Oliveira de
S232g Genômica comparativa de *Xylella fastidiosa* : diversidade do pangenoma e análise de genes de patogenicidade / Wesley Oliveira de Santana. -- São Paulo, 2012.
152p.

Tese (doutorado) - Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.
Orientador: Silva, Aline Maria da

1. Genoma : Biologia molecular 2. Bactéria : Genética 3. Evolução molecular I. T. II Silva, Aline Maria da, orientador.

574.88 CDD

**Genômica comparativa de *Xylella fastidiosa*: diversidade do pangenoma e
análise de genes de patogenicidade**

Wesley Oliveira de Santana

Tese submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências, área Bioquímica.

Aprovado por:

Profa. Dra. Aline Maria da Silva
(Orientadora e Presidente)
IQ-USP

Prof. Dr. João Carlos Setúbal
IQ-USP

Prof.Dr. Carlos Takeshi Hotta
IQ-USP

Prof. Dr. Jesus Aparecido Ferro
FCAV- UNESP- Jaboticabal

Prof. Dr. Marcelo Ribeiro da Silva Briones
UNIFESP

São Paulo

04 de fevereiro de 2013

Dedico...

A minha esposa Érica Santana e aos meus filhos Arthur Santana e Pedro Santana.

Agradeço...

À Profa. Dra. Aline Maria da Silva por me aceitar em seu grupo de pesquisa e pelo suporte e incentivo que me deu durante o Doutorado.

Ao Dr. Paulo Zaini, por me apoiar no desenvolvimento das minhas ideias e pelas excelentes imagens de microscopia eletrônica de varredura.

Ao colega Paulo Pierry pela parceria neste trabalho, pela companhia nas viagens para Cordeirópolis e pela ajuda em vários experimentos.

Aos colegas Fernando Tria e Dr. Joaquim Martins Junior, pelo apoio nas análises *in silico*.

Aos demais colegas de laboratório Oséias Feitosa, Patrícia Pessoa, Rodrigo Duarte, Gustavo Chaves, Luciana Principal, Ana Paula Souza e Deibs Barbosa pela amizade e ajuda em alguns experimentos.

À Dra. Layla Farage, técnica do Centro Avançado de Tecnologias em Genômica (CATG) e aos amigos Felipe Beckedorff e Murilo Sena, pelo valioso apoio na realização do sequenciamento dos genomas da *X. fastidiosa* descrito neste trabalho.

Ao Dr. João Paulo Kitajima pela imensa atenção e dedicação, além do excelente e heróico trabalho de montagem dos genomas das cepas descritas neste trabalho.

Aos técnicos de laboratório Alexandre Sanchez, grande amigo, Marivon Pereira e Dóris Araújo, pelo imprescindível serviço prestado.

Aos Profs. Drs. Sergio Verjovski-Almeida, Walter Terra, Clélia Ferreira, Suely Lopes Gomes, Regina Baldini, Henning Ulrich, Glaucia Souza e Mauricio Baptista por possibilitarem acesso a equipamentos e reagentes para que eu pudesse realizar determinados experimentos.

Aos Drs. Marcos Machado, Helvécio Della-Coletta e Alessandra Alves de Souza, do Centro de Citricultura Sylvio Moreira, em Cordeirópolis, onde fui muito bem recebido e pude realizar parte do trabalho.

A Dra. Ana Tereza Vasconcelos e ao pesquisador Luiz Gonzaga Almeida, do Laboratório Nacional de Computação Científica (Petrópolis, RJ) pela realização do sequenciamento utilizando a estratégia *Paired End Tags*.

À Dra. Daniela Beton, pelo apoio e orientação no trabalho de clonagem e expressão de proteínas recombinantes.

Ao Renato Astolfi, que com muita paciência me passou os primeiros ensinamentos no laboratório.

Aos demais colegas de pós-graduação do IQUSP pela troca de experiências e apoio em momentos difíceis.

A todos os professores que pude ter contato em disciplinas e seminários, dentro e fora do IQUSP, pelos valiosos ensinamentos.

Aos funcionários do IQUSP por proporcionarem um ambiente agradável de trabalho e pelos serviços prestados.

Aos funcionários do Centro de Citricultura Sylvio Moreira, em Cordeirópolis, que muito me ajudaram durante minhas estadias.

Aos meus amigos de fora da USP, por todo o apoio e interesse durante o período de meu Doutorado.

A minha mãe e meus familiares por apoiarem toda a minha trajetória na realização deste sonho.

Este trabalho foi realizado com apoio financeiro da FAPESP, CNPq e FINEP. O autor recebeu bolsa de Doutorado da CAPES.

RESUMO

De Santana, Wesley Oliveira. **Genômica comparativa de *Xylella fastidiosa*: diversidade do pangenoma e análise de genes de patogenicidade**. 2012. 152p. Tese de Doutorado - Programa de Pós-Graduação Ciências, área Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

O gênero *Xylella* é composto de uma única espécie, *Xylella fastidiosa*, bactéria Gram-negativa, não flagelada, que coloniza o xilema de uma diversidade de plantas cultivadas e silvestres em várias partes do mundo. Em algumas dessas plantas, a bactéria é considerada agente causal de doenças, como a Clorose Variegada dos Citros em laranjeiras, a Doença de Pierce das videiras e escaldadura da folha de cafeeiro. Onze diferentes cepas de *X. fastidiosa*, isoladas de distintos hospedeiros, já tiveram seus genomas sequenciados, entre essas, as cepas 9a5c, isolada de laranjeira e Temecula 1, isolada de videira. Análises desses genomas indicam uma razoável variabilidade entre suas respectivas sequências e evidenciam vários genes associados a mecanismos de virulência e patogenicidade desta bactéria. No presente trabalho descrevemos o sequenciamento, a montagem e a anotação dos genomas das cepas U24d e Fb7, isoladas de laranjeiras, e da cepa 3124 isolada de cafeeiro, os quais apresentam, respectivamente 2.681.334 pb, 2.733.974 pb e 2.748.594 pb. Destas, apenas a cepa U24d apresenta um plasmídeo, o qual é idêntico ao pXF51 previamente identificado na cepa 9a5c. O genoma da cepa U24d é praticamente colinear ao genoma da cepa 9a5c enquanto que os genomas das cepas Fb7 e 3124 apresentaram maior colinearidade com a cepa Temecula1. Entre as diversas alterações encontradas nas análises comparativas destes genomas, destacamos a inserção no gene *pilQ* verificada no genoma da cepa U24d. Essa mutação causa ausência do *pilus* do tipo IV com consequente deficiência na motilidade *twitching*, sendo que plantas infectadas com a cepa U24d apresentam sintomas localizados restritos ao ponto de inoculação. Na cepa Fb7, detectamos a ausência de formação de biofilme no cultivo *in vitro*, possivelmente devido ausência da expressão dos transcritos de *mrkD* e *pspA*, que codificam respectivamente adesina do *pilus* curto e adesina similar à hemaglutinina. Postulamos que estes genes não são expressos em decorrência de um defeito na via de sinalização de DSF (Fator de Sinalização Difusível) reflexo de uma mutação em *rpfC* no genoma de Fb7. Assim como as demais cepas de *X. fastidiosa*, também os genomas de U24d, Fb7 e 3124 apresentaram elevado conteúdo de Elementos Genéticos Móveis (EGM), que aparecem em maior número nas cepas sul-americanas. Os estudos do pangenoma de *X. fastidiosa* mostraram que essa espécie tem um genoma aberto e grande parte dos genes de EGMs correspondem a genes acessórios. A grande quantidade de EGMs em *X. fastidiosa* pode estar relacionada à falta do sistema CRISPR/cas completo, um provável resultado de eventos de erosão do genoma desta espécie. A inferência filogenética

por MSLA mostrou uma clara distinção dos grupos de cepas da América do Norte em relação às do Sul, sugerindo a ocorrência de mais eventos de recombinação genética nas cepas sul-americanas, provavelmente pela falta de isolamento geográfico. Assim, é possível que as cepas norte e sul-americanas sofreram divergência alopátrica e simpátrica, respectivamente.

Palavras chave: *Xylella fastidiosa*, sequenciamento de genomas; genômica comparativa; Pangenoma; Bacteriófago; Filogenia; Clorose Variegada dos Citros.

ABSTRACT

De Santana, Wesley Oliveira. **Comparative genomics of *Xylella fastidiosa*: pan-genome diversity and analysis of pathogenicity genes**. 2012. 152p. Tese de Doutorado - Programa de Pós-Graduação Ciências, área Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

The genus *Xylella* consists of a single species, *Xylella fastidiosa*, a Gram-negative and non-flagellated bacterium that colonizes the xylem of a diversity of cultivated and wild plants in several parts of the world. In some of these plants, this bacterium is considered causal agent of diseases such as the Citrus Variegated Cholorosis in orange trees, Pierce's Disease of grapevines and coffee leaf scald. Eleven different strains of *X. fastidiosa* isolated from different hosts had their genomes sequenced, including 9a5c and Temecula1 strains, respectively isolated from orange tree and grapevine. Analyses of these genomes indicate a reasonable variability in their sequences and showed several genes associated with pathogenicity and virulence mechanisms of this bacterium. In this work we describe the genome sequencing, assembly and annotation of the strains U24d and Fb7, isolated from orange trees, and 3124 isolated from coffee, which have, respectively, 2,681,334 bp, 2,733,974 bp and 2,748,594 bp. Of these, only strain U24d has a plasmid, identical to pXF51 from strain 9a5c. The genome of U24d strain is almost collinear to the genome of strain 9a5c while the genomes of strains Fb7 and 3124 had higher collinearity to Temecula1 strain. Among many changes found in the comparative analysis of these genomes, we highlight an insertion in *pilQ* gene that was found in U24d strain genome. This mutation causes lack of type IV pilus with a consequent deficiency in twitching motility. Moreover orange trees infected with U24d strain showed localized symptoms near to the inoculation point. We verified that Fb7 strain does not form biofilm *in vitro* possibly due to the absence of expression of *mrkD* and *pspA* transcripts, which encode, respectively, a short pilus adhesin and a hemagglutinin-like adhesin. We postulate that these genes are not expressed due to a defect in the signaling pathway of DSF (Diffusible Signal Factor) reflecting a mutation on *rpfC* in the Fb7 genome. Similarly to other *X. fastidiosa* strains, the genomes of U24d, Fb7 and 3124 also showed high content of mobile genetic elements (MGE), which appear in larger numbers in South American strains. Pan genome studies of *X. fastidiosa* showed that this species has a open genome and that most of MGE genes correspond to accessory genes. The large number of MGE in *X. fastidiosa* may be related to the lack of a complete system CRISPR/cas, likely a result of erosion events of the genome of this species. The phylogenetic inference by MLSA showed a clear distinction between groups of strains from North and South America, suggesting the occurrence of more recombination events in South American strains, probably due to lack of

geographical isolation. Thus it is possible that North and South American strains underwent allopatric and sympatric divergence, respectively.

Keywords: *Xylella fastidiosa*, genome sequencing, comparative genomics, pangenome; bacteriophage; phylogeny; citrus variegated chlorosis.

LISTA DE ABREVIATURAS E SIGLAS

Abs: absorbância	LNCC: Laboratório Nacional de Computação Científica
ALS: escaldadura da folha de amendoeira	MIRA: <i>Mimicking Intelligent Read Assembly</i>
BLAST: <i>Basic Local Alignment Search Tool</i>	MLSA: Análise de sequências <i>multilocus</i>
BSA: albumina de soro bovino	MLST: Tipagem de sequências <i>multilocus</i>
CATG: Centro Avançado de Tecnologias em Genômica, IQUSP	NCBI: <i>National Center for Biotechnology Information</i>
CDS: sequência codificadora	ORF: <i>open reading frame</i>
cDNA: DNA complementar	pb: par de bases
CLS: escaldadura da folha de cafeeiro	PBS: <i>Phosphate buffered saline</i>
CRISPR: <i>Clustered Regularly Interspaced Short Palindromic Repeats</i>	PCR: reação em cadeia da polimerase
CCD: <i>charge-coupled device</i>	PD: Doença de Pierce
CT: <i>cycle threshold</i>	PET: <i>Paired End Tags</i>
CTAB: brometo de n-hexadecil trimetil amônio	PHAST: <i>Phage Search Tool</i>
CVC: Clorose Variegada dos Citros	PLS: escaldadura da folha de ameixeira
DNAse: desoxirribonuclease	PMSF: fluoreto de fenilmetilsulfonila
dNTP: desoxirribonucleotídeo 5' trifosfato	PTP: <i>PicoTiterPlate</i>
DO: densidade óptica	qPCR: PCR quantitativo
dsDNA: DNA fita-dupla	RAPD: <i>Random Amplification of Polymorphic DNA</i>
DSF: <i>Diffusible Signaling Factor</i>	RAST: <i>Rapid Annotation using Subsystem Technology</i>
DTT: ditionitreitol	RNAse: ribonuclease
EDTA: ácido etilenodiaminotetracético	rpm: rotações por minuto
EGM: elemento genético móvel	SDS: dodecil sulfato de sódio
emPCR: PCR em emulsão	SDS-PAGE: eletroforese em gel de poliacrilamida contendo SDS
EPS: Exopolissacarídeo	SSR: <i>Simple Sequence Repeats</i>
Fundecitrus: Fundo de Defesa da Citricultura	ssDNA: DNA fita-simples
IG: Ilha genômica	TE: Tris-EDTA (Tris-HCl10mM pH 7.5; EDTA 1mM)
IPTG: isopropil β -D-tiogalactopiranosídeo	Tm: temperatura de fusão
IMG/ER: <i>Integrated Microbial Genomes-Expert Review</i>	Tris: tris-(hidroximetil)-aminometano
Kpb: quilopares de bases	UV: ultravioleta

Sumário

1. Introdução	14
1.1. Paineis da citricultura brasileira e da Clorose Variegada do Citros	14
1.2. O fitopatógeno <i>Xylella fastidiosa</i>	17
1.2.1. Genomas de <i>X. fastidiosa</i>	21
1.2.2. Elementos Genéticos Móveis de <i>X. fastidiosa</i>	25
1.2.3. Diversidade e evolução de <i>X. fastidiosa</i>	27
1.2.4. Virulência e Patogenicidade de <i>X. fastidiosa</i>	29
1.2.5. Sinalização celular mediada por DSF	31
1.2.6. Estratégias de controle das doenças causadas por <i>Xylella</i>	34
2. Objetivos	37
3. Procedimentos Experimentais	39
3.1. Manutenção e cultivo de <i>Xylella fastidiosa</i>	39
3.2. Microscopia Eletrônica de Varredura	39
3.3. Microscopia de colônias de <i>X. fastidiosa</i>	40
3.4. Infecção de laranja doce com <i>X. fastidiosa</i>	40
3.5. Isolamento de DNA total de <i>X. fastidiosa</i> crescida em meio PWG	40
3.6. Isolamento de DNA genômico de <i>X. fastidiosa</i> a partir de folhas de laranjeiras	41
3.7. Genotipagem de <i>X. fastidiosa</i>	41
3.8. Eletroforese de DNA	42
3.9. Pirosequenciamento de DNA	43
3.9.1. Pirosequenciamento shotgun	43
3.9.2. Sequenciamento de <i>Paired End Tags</i>	50
3.10. Sequenciamento de DNA pelo método de Sanger	51
3.11. Análises bioinformáticas	51
3.11.1. Montagem e anotação de genomas	51
3.11.2. Análises comparativas de genes e genomas	52
3.11.3. Identificação de grupos de CDS ortólogos e parálogos	52
3.11.4. Determinação do pangenoma e do <i>core-genome</i>	53
3.11.5. Reconstrução filogenética	54
3.12. Obtenção de anticorpos policlonais para PilQ e PilY1 de <i>X. fastidiosa</i>	54
3.12.1. Clonagem de segmentos de CDS de <i>X. fastidiosa</i> para expressão recombinante	54
3.12.2. Preparação de bactérias competentes para transformação	56
3.12.3. Preparação de plasmídeos e digestão com enzimas de restrição	57
3.12.4. Indução da expressão de proteínas recombinantes	57

3.12.5. Purificação das proteínas recombinantes	58
3.12.6. Produção de anticorpos policlonais	58
3.13. Eletroforese unidimensional em gel de poliacrilamida contendo SDS (SDS-PAGE)	59
3.14. Transferência eletroforética para membrana de nitrocelulose e detecção imunológica de proteínas	59
3.15. PCR quantitativa precedida de transcrição reversa (RT-qPCR)	60
3.15.1. Purificação de RNA total	60
3.15.2. RT-qPCR precedida por transcrição reversa	61
4. Resultados e Discussão.....	63
4.1. Descrição das cepas de <i>X. fastidiosa</i> selecionadas para sequenciamento do genoma	63
4.2. Avaliação do fenótipo das cepas no cultivo <i>in vitro</i>	67
4.4. Pirosequenciamento, montagem e anotação do genoma das cepas U24d, Fb7 e 3124	71
4.5. Análises comparativas dos genomas das cepas U24d, Fb7 e 3124 com genomas das cepas 9a5c e Temecula1	74
4.6. Análise de PilQ na cepa U24d	81
4.6.1. Obtenção de anticorpos policlonais anti-rPilQC e anti-rPilY1N e detecção de PilQ e PilY1 nas cepas U24d e 9a5c	83
4.6.2. Avaliação da virulência e eficiência de colonização da cepa U24d	87
4.7. Análise de <i>rpfC</i> da cepa Fb7.....	89
4.8. A heurística da dinâmica evolutiva dos genomas de <i>X. fastidiosa</i>	95
4.8.1. Pangenoma de <i>X. fastidiosa</i>	95
4.8.2. Identificação de regiões de profagos e ilhas genômicas nos 14 genomas de <i>X. fastidiosa</i>	100
4.8.3. Sistema de imunidade CRISPR-cas em <i>X. fastidiosa</i>	106
4.8.4. Filogenia das cepas Sul e Norte-Americanas de <i>X. fastidiosa</i>	109
4.9. Um novo isolado de <i>X. fastidiosa</i> em Cruz das Almas – Bahia	111
5. Conclusões	113
6. Referências	116
7. Apêndice.....	127
Súmula curricular	i

1. Introdução

1.1. Paineis da citricultura brasileira e da Clorose Variegada dos Citros

O cultivo de citros, predominantemente laranja doce, está presente em mais de 3.000 municípios em todos os Estados brasileiros. Ocupando mais de 800 mil hectares, a laranja é a fruta mais plantada no país. O Brasil detém atualmente 50% da produção mundial de suco de laranja, e exporta 98% do que produz, conseguindo incríveis 85% de participação no mercado mundial. A Flórida (EUA) e o Estado de São Paulo respondem juntos por 81% da produção mundial de suco, dos quais 53% originam-se de São Paulo. Estimativas recentes informam que as exportações de suco de laranja trazem de US\$ 1,5 a 2,5 bilhões por ano ao país. Em 2009, as exportações do complexo de citros somaram 2,9 milhões de toneladas, sendo 1,129 milhão de toneladas de suco concentrado, 939 mil toneladas de suco não concentrado e 851 mil toneladas de subprodutos (Lopes *et al.*, 2010).

Nas últimas 15 safras, a produção mundial de suco de laranja caiu 13%, sendo que a maior redução aconteceu na Flórida (quase 90%). Todavia, o PIB do setor citrícola foi de US\$ 6,5 bilhões (2009), sendo US\$ 4,39 bilhões no mercado interno e US\$ 2,15 bilhões no mercado externo. Em 2010, os pomares brasileiros tinham aproximadamente 165 milhões de árvores produzindo, enquanto que a Flórida tinha 60 milhões. Nos últimos anos a densidade de árvores por hectare aumentou bastante, passando de 250 árvores/hectare em 1980 para 357 árvores em 1990, e em 2000 atingiu 476 árvores/hectare. Atualmente, em alguns pomares, são encontrados até 850 árvores/hectare (Lopes *et al.*, 2010).

Durante a última década, quatro doenças foram responsáveis pela erradicação de 39 milhões de árvores do parque citrícola de São Paulo e Triângulo Mineiro. Com isso, a taxa anual média de mortalidade que anteriormente oscilava ao redor de 4,5% ao ano saltou para 7,3%, considerando-se um rendimento médio de duas caixas de laranja por árvore. Estima-se que o cancro cítrico, a Clorose Variegada de Citros (CVC), a Morte Súbita e o Greening foram responsáveis por uma redução anual de cerca de 78 milhões de caixas, que, comparadas com os 317 milhões de caixas colhidas na safra 2009/10, representam uma diminuição de safra da ordem de 20%. Todavia, destas doenças, a que ainda causa maior prejuízo ao produtor é a CVC. De 2000 a 2009 a CVC causou a erradicação de 26 milhões de árvores, seguida pelo Greening com 5 milhões, Cancro Cítrico com 2 milhões e Morte Súbita com 5 milhões árvores, segundo dados contabilizados pelo Centro de Pesquisa e Projetos em Marketing e Estratégia (Markestrat/<http://www.markestrat.org/>) a partir dos

percentuais anuais de erradicação de árvores divulgados pelo Fundecitrus. Desta forma, o prejuízo econômico causado pela CVC é em torno de 100 milhões de dólares por ano no país (Lopes *et al.*, 2010).

As primeiras observações de CVC foram feitas em meados de 1987 no Triângulo Mineiro e no Norte e Nordeste do Estado de São Paulo (Rossetti e De Negri, 1990; Rossetti *et al.*, 1990; Chang *et al.*, 1993; Coletta-Filho e Machado, 2003). Todavia, só em 1993 identificou-se o organismo responsável pela doença, a bactéria *Xylella fastidiosa*, e três anos depois o vetor responsável pela disseminação da bactéria nos pomares, um homóptero conhecido popularmente como cigarrinha (Chang *et al.*, 1993; Redak *et al.*, 2004). Os principais sintomas de CVC são: (i) redução drástica no tamanho da copa das plantas, com diminuição no tamanho das folhas e taxa fotossintética; (ii) áreas de clorose na superfície adaxial das folhas, muitas vezes com presença de goma na superfície abaxial e; (iii) enrijecimento das cascas dos frutos, os quais também apresentam tamanho bastante reduzido (a perda em massa dos frutos pode chegar a 75%) e amadurecimento precoce. Observa-se ainda crescimento lento da planta contaminada, ocorrendo morte dos ponteiros e queda das folhas (Lee *et al.*, 1991).

A distribuição mundial da CVC, até aonde se conhece, é restrita a América do Sul e Central. Além do Brasil, esta doença já foi detectada na Costa Rica, Paraguai e Argentina (Figura 1). No Brasil, a CVC está presente no Rio Grande do Sul, Paraná, São Paulo, Minas Gerais, Goiânia, Sergipe, Bahia e o Distrito Federal (<http://www.centrodecitricultura.br/>). No Estado de São Paulo, já foram identificadas plantas sintomáticas em Araraquara, Ariranha, Arthur Nogueira, Auriflama, Badi Bassit, Barretos, Bebedouro, Borborema, Casa Branca, Catanduva, Catiguá, Colina, Cedral, Dobrada, Eugênio de Melo, Fernando Prestes, Guapiaçu, Guaraci, Guariba, Ibirá, Irapuã, Itajobi, Itápolis, Jaborandi, Jaci, José Bonifacio, Macaubal, Matão, Mirassol, Mogi Guaçu, Monções, Monte Alto, Nova Aliança, Nova Granada, Olímpia, Pindorama, Santa Adélia, São José do Rio Preto, Tabapuã, Taiaçu, Tanabi, Taquaritinga, Terra Rocha, Uchoa, Urupês e Viradouro (De Negri, 1990). O mapeamento realizado recentemente pelo Fundecitrus constatou que a maior incidência da CVC (53%) ocorre na região norte do Estado que é seguida pela região noroeste com 42,8% (Figura 2).

Figura 1. A CVC é uma doença restrita a alguns países da América do Sul e Central. Figura cedida pelo Dr. Helvécio Della Coletta-Filho (Centro Apta Citros Sylvio Moreira, Cordeirópolis, SP).

Figura 2. Incidência de CVC nas regiões do Estado de São Paulo. Segundo o estudo realizado pelo Fundecitrus em 2011 a região Norte é a que apresenta maior incidência da doença (53%). Fonte: <http://www.fundecitrus.com.br>.

Levantamentos da incidência de CVC no Estado de São Paulo desde 1996 até 2011, realizados pelo Fu

ndecitrus, mostraram um considerável aumento na região norte, seguida pelas regiões Noroeste, Centro, Sul e Oeste (Tabela 1) ao longo desses anos. De modo geral, a incidência de CVC nos pomares paulistas passou de 35,5% em 2010 para 40,3% em 2011, sendo que número de plantas com sintomas de nível 2 (sintomas em folhas e frutos) aumentou, enquanto o número de árvores com sintomas de nível 1 (sintomas em folhas) diminuiu ou se manteve estável (<http://www.fundecitrus.com.br>).

Tabela 1. Histórico da incidência da CVC no Estado de São Paulo

Levantamento	Regiões citrícolas				
	Norte (%)	Noroeste (%)	Centro (%)	Sul (%)	Oeste (%)
1996	24,65	22,83	19,95	22,14	nr
1997	41,15	33,59	33,15	21,99	nr
1998	34,67	21,80	18,28	7,90	nr
1999	55,69	47,09	32,21	13,66	nr
2000	52,62	43,44	28,69	15,69	nr
2001	48,60	40,85	39,01	17,33	nr
2002	55,59	41,31	38,82	11,88	nr
2003	65,06	41,73	44,37	11,88	nr
2004	60,73	49,72	54,66	13,71	15,25
2005	69,74	56,25	51,96	4,72	9,50
2009	52,52	46,71	52,65	20,33	1,15
2010	59,7	42,8	38,2	25,0	4,4
2011	59,7	52,5	42,37	30,49	1,06

nr- não encontrado.

Dados obtidos junto ao Departamento Técnico do Fundecitrus, gentilmente cedidos pelo engenheiro agrônomo Cícero Massari.

1.2. O fitopatógeno *Xylella fastidiosa*

O gênero *Xylella* é composto de uma única espécie, *X. fastidiosa*, bactéria Gram-negativa, não flagelada, que coloniza o xilema de uma diversidade de plantas cultivadas e silvestres em várias partes do mundo. Em algumas dessas plantas, a bactéria é considerada agente causal de diversas doenças, como a já mencionada CVC, porém em outras não o é, devido à ausência de sintomas ou danos nos hospedeiros colonizados (Hopkins, 1989; Hopkins e Purcell, 2002; Chatterjee *et al.*, 2008a; Janse e Obradovic, 2010).

X. fastidiosa pertence a família Xanthomonadaceae, que também inclui, entre outros gêneros, *Xanthomonas* e *Stenotrophomonas*. Essa família posiciona-se filogeneticamente próxima à base da classe das gamaproteobactérias, a qual compreende um grupo de bactérias que possui grande diversidade de estilo de vida,

ecologia e metabolismo (Studholme *et al.*, 2005). *X. fastidiosa* exibe crescimento lento em meio de cultivo *in vitro*, sendo por isto considerada uma bactéria fastidiosa. Sua disseminação é feita por insetos da Ordem Homoptera (Família Cicadellidae, sub-Família Cicadellinae), popularmente conhecidos como cigarrinhas, os quais são sugadores de seiva xilemática e ocasionalmente adquirem as bactérias quando se alimentam de plantas contaminadas. No inseto vetor, a bactéria encontra-se restrita a parte anterior do esôfago (estomodeu) aderida ao forro cuticular do pré-cibário, onde se multiplica formando agregados de células associados a uma matriz, provavelmente constituída de exopolissacarídeo (Hill e Purcell, 1995; Redak *et al.*, 2004; Alves *et al.*, 2008; Backus e Morgan, 2011).

As cepas de *X. fastidiosa* são geralmente classificadas de acordo com o hospedeiro de origem, sendo que não há, até o momento, comprovação da existência de especificidade entre vetor-hospedeiro, em que diferentes linhagens de *X. fastidiosa* são transmitidas por diferentes espécies de cigarrinhas para diferentes plantas hospedeiras (Purcell e Hopkins, 1996; Redak *et al.*, 2004; Almeida *et al.*, 2005; Almeida e Purcell, 2006; Alves *et al.*, 2008).

Na figura 3 são mostradas imagens do biofilme formado por *X. fastidiosa* no xilema de plantas de citros e no pré-cibário do inseto vetor, além de imagens de diferentes espécies de cigarrinhas encontradas em pomares de citros.

Figura 3. Biofilme de *X. fastidiosa* no xilema e no pré-cibário do inseto vetor. Imagens de microscopia eletrônica de varredura do biofilme de *X. fastidiosa* no xilema de laranjeira infectada (A) e no forro cuticular do pré-cibário do inseto (B). Diferentes espécies de cigarrinhas encontradas nos pomares de citros estão mostradas em (C). Da esquerda para direita, no topo: *Bucephalogonia xanthophis*, *Dilobopterus costalimai*, e em baixo: *Acrogonia citrina* e *Oncometopia facialis*. As imagens mostradas em A e B são, respectivamente de autoria de E. W. Kitajima (ESALQ/USP/Brasil) e Rodrigo Almeida (UC Berkeley, EUA) e foram extraídas de <http://aeg.lbi.ic.unicamp.br/xf/> e http://www.genomenewsnetwork.org/articles/09_02/xylella.shtml. As imagens mostradas em (C) foram obtidas do banco de imagens do Fundecitrus.

Além da CVC, a doença de Pierce (PD) das videiras está entre as principais doenças economicamente importantes resultantes da infecção por *X. fastidiosa*. Os sintomas típicos dessas duas doenças estão mostrados na figura 4. A doença de Pierce foi identificada no estado da Califórnia, nos Estados Unidos, pela primeira vez em 1884. Entretanto, somente em 1978, a *X. fastidiosa* foi reconhecida como agente etiológico da PD e também de uma doença de pessegueiros (*phony peach disease*). A doença de Pierce também ocorre em outros Estados do sul dos EUA, como a Flórida e Texas, além do México e países da América Central. A partir de 1990, a doença agravou-se em razão do aumento de uma espécie de cigarrinhas oriunda do extremo sul dos EUA e do México. Neste mesmo ano, mais de 300 acres de videiras da região da Califórnia foram destruídas pela PD (Davis *et al.*, 1978; Wells *et al.*, 1987; Hopkins, 1989; Hopkins e Purcell, 2002). Segundo dados recentes do Programa de Controle da Doença de Pierce da Califórnia, esta doença ameaça negócios diretamente derivados da vinicultura estimados em mais de 60 bilhões de dólares (<http://www.cdfa.ca.gov/pdcp/>).

Figura 4. Sintomas de doenças causadas por *X. fastidiosa*. Cloroses assimétricas e necrose no tecido foliar na região adaxial da folha de laranjeira doce (A) e redução do tamanho dos frutos (B) são sintomas típicos de CVC. Folhas murchas e clorose nas folhas de videiras (C) e frutos secos são sintomas típicos de PD.

1.2.1. Genomas de *X. fastidiosa*

Onze diferentes cepas de *X. fastidiosa*, isoladas de distintos hospedeiros, já tiveram seus genomas sequenciados: cepa 9a5c, isolada de laranjeira doce e agente etiológico da CVC; cepa Temecula1, isolada de videira na Califórnia e agente etiológico da PD; cepas Dixon, M12 e M23, isoladas de amendoeiras com sintomas de escaldadura; cepa Ann-1, isolada de espiroideira com sintomas de escaldadura; cepa GB514, obtida de uma videira com sintomas de PD no Texas; cepa EB-92.1, originalmente isolada de sabugueiro mas que é capaz de colonizar videira e sobreviver por muitos anos sem causar sintomas, atuando assim como biocontrole contra a PD; cepa J1a12, isolada de laranjeira e reconhecidamente não patogênica; cepa Pr8x, isolada de ameixeira e cepa Hib4, isolada de hibisco (Simpson *et al.*, 2000b; Bhattacharyya *et al.*, 2002a; Bhattacharyya *et al.*, 2002b; Van Sluys *et al.*, 2002; Van Sluys *et al.*, 2003; Chen *et al.*, 2010; Schreiber *et al.*, 2010; Zhang *et al.*, 2011; Pierry, 2012). Vale destacar que apenas os genomas das cepas 9a5c, Temecula1, M12, M23, GB514 e J1a12 estão completos e montados com uma única sequência contígua e completa. Os demais são rascunhos (*draft genome sequence*). Além disso, foi recentemente reportado que a preparação de DNA utilizada como sendo de uma única cepa (Ann-1) estava contaminada com DNA de outra cepa não identificada, e que, portanto, a sequência genômica originalmente descrita para Ann-1 corresponde aos genomas de duas cepas (Nunney *et al.*, 2012). Outra observação relevante é que a cepa Temecula1 sequenciada em 2003 (Van Sluys *et al.*, 2003) não tem correspondência com a cepa distribuída como “Temecula1” pela ATCC (*American Type Tissue Collection*) (Nunney *et al.*, 2012).

Os genomas de *X. fastidiosa* já elucidados estão listados na Tabela 2. Análises comparativas das sequências e dos dados de anotação dos genomas dessas cepas revelaram razoável variabilidade, o que certamente justifica sua separação em distintos patovares. A comparação dos genomas das cepas Temecula1 e 9a5c mostraram que das 2.066 ORFs identificadas em Temecula1, 2.025 (98%) também estão presentes em 9a5c, sendo que 94.5% das ORFs compartilhadas entre essas linhagens apresentam ao menos 80% de identidade entre seus resíduos de aminoácidos (Van Sluys *et al.*, 2003). Essas ORFs estão relacionadas com o metabolismo basal deste fitopatógeno. A análise mostrou que 41 ORFs (1.9%) são específicas da cepa Temecula1, assim como 152 ORFs (6.8%) são específicas da cepa 9a5c. Dentre essas, grande parte são hipotéticas e outras estão relacionadas com Elementos Genéticos Móveis (EGMs), principalmente regiões de profagos. Entre as diferenças observadas, merece destaque o fato de que o gene que codifica a

poligalacturonase em 9a5c e em outras cepas isoladas de citros e isolados de cafeeiros apresenta a inserção de um nucleotídeo que gera alteração de fase de leitura e uma ORF truncada, contrariamente ao verificado para cepa Temecula1 e para as outras cepas norte-americanas já sequenciadas. Postula-se que essa diferença explique a maior agressividade da cepa Temecula1 (Van Sluys *et al.*, 2003; Chatterjee *et al.*, 2008a). Outra diferença observada entre as cepas Temecula1 e 9a5c é a ausência do plasmídeo pXF51 na cepa Temecula1 (Van Sluys *et al.*, 2003).

Apesar do alto índice de identidade de ORFs compartilhadas entre as cepas Temecula1 e 9a5c, a colinearidade entre elas não é conservada. O alinhamento dos genomas a partir da origem de replicação evidencia três principais eventos de rearranjo cromossômico (translocações e inversões), sendo todos flanqueados em suas bordas por integrases de fagos (Bhattacharyya *et al.*, 2002a) como esquematizado na Figura 5 (Van Sluys *et al.*, 2003).

Análises comparativas do genoma da cepa EB-92.1 mostraram alta similaridade e sintonia com o genoma da cepa Temecula1. Nestas análises, foi detectada a ausência de 10 genes na cepa EB92-1 que podem eventualmente explicar seu fenótipo não-virulento. Entre os genes ausentes, estão CDSs que codificam uma serina-protease, uma lipase secretada e dois genes que codificam toxina de zonula occludens (Zot), além de seis genes de hemaglutininas (Zhang *et al.*, 2011).

O recente sequenciamento do genoma da cepa J1a12, isolada de citros, revelou a ausência de alguns genes que correlacionam com seu fenótipo menos agregativo e menos virulento (Koide *et al.*, 2004; Pierry, 2012). Entre os genes ausentes estão uma cópia do gene da adesina da fimbria e 2 cópias dos 3 genes que codificam serina proteases da família de autotransportadores. Além disso, a cepa J1a12 parece ser deficiente na expressão de seu arsenal de toxinas proteicas e também no processamento de DSF (*Diffusible Signaling Factor*) que participa do sistema de sinalização célula-célula dependente da densidade celular, pois apresenta, respectivamente alterações em genes relacionados à produção de microcinas e no gene *rpfB*, aparentemente associado a síntese de DSF (Pierry, 2012).

Tabela 2. Genomas de *Xylella fastidiosa*

Cepa	Hospedeiro	Origem	Doença ^a	Replicons	pb	% G+C	Referência
9a5c	Laranjeira doce	Brasil	CVC	3	2679306	52,67	(Simpson <i>et al.</i> , 2000a)
					51158	49,57	
					1285	55,56	
Ann-1	Espirradeira	EUA	OLS	2	2617604	51,92	Van Sluys <i>et al.</i> , 2003
					30 270	49,47	
Dixon	Amendoeira	EUA	ALS	2	2629797	52,22	Bhattacharyya <i>et al.</i> , 2002 ^a
M12			ALS	1	31.572		
M23			ALS	2	2475130	51,92	Chen <i>et al.</i> , 2010
Temecula1	Videira	EUA	PD	2	2535690	51,76	Chen <i>et al.</i> , 2010
					38297	49,20	
GB514			PD	2	2519802	51,78	Van Sluys <i>et al.</i> , 2003
EB92-1 ^b	Sabugueiro	EUA	Não virulenta	2	1346	53,79	Schreiber <i>et al.</i> , 2010
					2491203	51,79	
					26180	49,34	
J1a12	Laranjeira doce	Brasil	Não virulenta	3	2478730	-	Zhang <i>et al.</i> , 2011
					1346	-	
					2788789	52.9	
Hib4	Hibisco	Brasil	HLS	2	51180	49.6	Da Silva, 2011 Pierry, 2012
					27268	49.4	
Pr8x	Ameixeira	Brasil	PLS	2	2813297	52.7	Pierry, 2012
					64251	52.2	
Pr8x	Ameixeira	Brasil	PLS	2	2666242	52.6	Pierry, 2012
					39580	49.5	

^aCVC, Clorose Variegada dos Citros; OLS, escaldadura da folha de espirradeira; ALS, escaldadura da folha de amendoeira; PD, Doença de Pierce; HLS, escaldadura da folha de hibisco; PLS, escaldadura da folha de ameixeira.

^b Genoma incompleto

Figura 5. Representação esquemática do alinhamento dos cromossomos das cepas Temecula1 (PD) e 9a5c (CVC). As diferentes cores representam as principais regiões colineares em ambos os cromossomos. A orientação das setas indicam sua orientação relativa, evidenciando rearranjos (inversão e translocação) na estrutura de ambos os genomas. Os triângulos em preto representam as regiões de prófagos e ilhas genômicas. Figura extraída de Van Sluys et al., 2003.

Diferenças relevantes também foram detectadas nos genomas da cepa Pr8x (de ameixeira) e Hib4 (de hibisco). Assim como no genoma da cepa J1a12, também em Pr8x foi observada a deleção do gene da arginina deaminase. A cepa Hib4, isolada de hibisco, tem o maior cromossomo entre as cepas já sequenciadas e também o maior plasmídeo (pXF64 com 64.251 pb) já descrito para *X. fastidiosa*. Além disso, entre as CDS ausentes no genoma de Hib4, está uma cópia do gene da adesina da fimbria e que também está ausente na cepa J1a12, como mencionado acima. O gene de uma adesina não-fimbrial (XF1529) (Caserta *et al.*, 2010) também apresenta uma extensa deleção na sua porção central no genoma da cepa Pr8x. Se por um lado esta alteração parece não afetar o fenótipo agregativo da cepa Pr8x, especula-se que ausência de duas adesinas possa estar correlacionada com o fenótipo de menor adesão observado no cultivo de Hib4. Interessantemente os genomas de Pr8x e Hib4 codificam a proteína precursora de poligalacturonase completa, ao contrário do verificado em diversas cepas sul-americanas isoladas de citros e cafeeiros (Pierry, 2012).

As análises dos genomas já sequenciados de *X. fastidiosa* mostraram a variedade e diversidade de regiões relacionadas a bacteriófagos e de plasmídeos que

co-existem nas diferentes cepas deste fitopatógeno, sendo que as discrepâncias no conteúdo flexível do genoma são o que mais diferencia as cepas de *X. fastidiosa*, como iremos abordar mais adiante neste trabalho. A *X. fastidiosa* pertence a um grupo de organismos com genomas de tamanho relativamente pequeno e pode-se observar que os prováveis eventos de deleção genômica parecem ter sido mais abundantes do que os eventos de ganho por duplicação ou transferência lateral, ainda que sequências de fagos tenham se inserido em seu genoma. Esta espécie carece de alguns genes e vias importantes, fato que provavelmente reflete no crescimento destas bactérias no xilema de seus respectivos hospedeiros e da dependência de um vetor animal para sua propagação (Bhattacharyya *et al.*, 2002a; Bhattacharyya *et al.*, 2002b; Van Sluys *et al.*, 2002; Moreira *et al.*, 2005; Chatterjee *et al.*, 2008a).

A reconstrução filogenética de *X. fastidiosa* com espécies do gênero *Xanthomonas*, mostrou que a *Xanthomonas albilineans* e *X. fastidiosa* cepa 9a5c agrupam juntas e derivam, provavelmente, de um grupo de *Stenotrophomonas*. Entre constituintes desta filogenia, o cromossomo de *X. albilineans* e da cepa 9a5c são os menores do que qualquer outra *Xanthomonas* que possua o sistema de secreção tipo 3, sugerindo assim, que ambas as espécies surgiram de um progenitor através de sucessivas erosões dos seus genomas. Essa hipótese é suportada pela perda de grande quantidade de genes verificada nos genomas de *X. albilineans* e *X. fastidiosa* em relação ao grupo de *Xanthomonas*. Assim, é provável que a redução do genoma causada por eventos de erosão tenha selecionado indivíduos limitados a determinados tipo de *habitat* (Pieretti *et al.*, 2009).

1.2.2. Elementos Genéticos Móveis de *X. fastidiosa*

Os Elementos Genéticos Móveis (EGMs) são segmentos de DNA que codificam enzimas e outras proteínas que medeiam o movimento de DNA intragenoma e entre genomas bacterianos. Os EGMs podem ser plasmídeos, transposons, integrons, profagos e ilhas genômicas. A principal característica comum a eles é a presença de um segmento de DNA que codifica enzimas com atividade de integrase, responsáveis pela integração dos elementos móveis ao cromossomo bacteriano. Em função da sua atividade móvel, EGMs não são considerados partes integrantes do 'genoma central' de um organismo ao contrário, são considerados como DNA exógeno, que são adquiridos por Transferência Genética Lateral (TGL). Os EGMs garantem diversidade genética às bactérias necessária a sua adaptação e sobrevivência no ambiente sendo importantes para evolução de uma determinada espécie (Frost *et al.*, 2005; Binnewies *et al.*, 2006).

Como já mencionado, a comparação dos genomas das cepas Temecula1 e 9a5c mostrou que 98% dos genes são compartilhados entre essas cepas, com uma média de 95,7% de identidade de aminoácidos entre as proteínas preditas. As principais diferenças encontradas entre essas cepas derivam de regiões de profagos, sendo os mesmos responsáveis pelos rearranjos e deleções cromossômicas entre as linhagens de *X. fastidiosa* (Simpson *et al.*, 2000b; Van Sluys *et al.*, 2003).

Análises comparativas por hibridização subtrativa de genomas de seis cepas sul-americanas de *X. fastidiosa*, tendo como referência a cepa 9a5c, sugere que os EGMs contribuem com a dinâmica e a evolução desses genomas, uma vez que os produtos contidos nesses elementos, particularmente ilhas genômicas (IGs), são potencialmente envolvidos com adaptação e patogenicidade ao hospedeiro. Interessantemente, esses elementos encontrados nestas cepas sul-americanas estão ausentes nas cepas norte-americanas e apresentam intensa atividade transposicional (da Silva *et al.*, 2007). A análise global da expressão gênica da cepa 9a5c submetida a situações de estresse mostrou um aumento do nível de transcritos de regiões de bacteriófagos, tais como do gene anti-repressor, sugerindo a mudança dos fagos do ciclo lisogênico para o lítico (Koide *et al.*, 2006; da Silva Neto *et al.*, 2008).

De Mello Varani e colaboradores (de Mello Varani *et al.*, 2008), através de análises detalhadas de EGMs de *X. fastidiosa*, identificaram e caracterizaram várias regiões de profagos e de profagos remanescentes nos genomas de cepas da América do Norte e do Sul. Esses autores mostraram que os EGMs representam em média 13% dos genomas das cepas analisadas. Mais da metade da CDS preditas (51%) estão relacionadas à estrutura do bacteriófago, pertencente à família *Siphoviridae*. Esses profagos também transportam genes com funções não diretamente relacionadas com a propagação do fago, mas com potencial função de virulência e patogenicidade. A maior parte da outra metade das CDS preditas nessas regiões de profagos codificam proteínas de função desconhecida e vários destes genes são específicos para cada cepa e possivelmente relacionados com a geração de novas variantes de cepas bacterianas (de Mello Varani *et al.*, 2008). Assim, conjuntamente, estas observações são indicativas da importância dos EGMs como mediadores na evolução deste fitopatógeno.

Outro tipo de EGM encontrado em *X. fastidiosa* são os plasmídeos. A cepa 9a5c possui dois, um maior pXF51 e menor pXF1.3, ausente e presente respectivamente na cepa norte-americana Temecula1 (Simpson *et al.*, 2000b; Van Sluys *et al.*, 2003). O pXF51 em especial, não parece ter um papel direto na patogenicidade da cepa para com o hospedeiro, mas possivelmente participa na transferência lateral de genes, por ser um plasmídeo conjugativo. Interessantemente,

esse plasmídeo é encontrado em outras cepas sul-americanas, como a cepa J1a12 e Pr8x, e não encontrado na cepa Hib4, que aparentemente possui outro tipo de plasmídeo (Pierry, 2012). Na cepa J1a12 foi encontrado também um novo plasmídeo pXF27, presente em outros isolados de citros do Estado de São Paulo (Pierry, 2012). Esse plasmídeo possui extensa similaridade com o plasmídeo pVEIS01 de *Verminephrobacter eiseniae* (Pinel *et al.*, 2008) e com plasmídeos de cepas de *X. fastidiosa* que infectam amoreira (Lee *et al.*, 2010; Stenger *et al.*, 2010) e videira (Schreiber *et al.*, 2010).

1.2.3. Diversidade e evolução de *X. fastidiosa*

X. fastidiosa é um patógeno adaptado ou ao cibário do inseto vetor ou aos vasos mortos na planta, o xilema (Chatterjee *et al.*, 2008a), sendo que a adaptação a esses dois ambientes se deve, provavelmente, aos eventos de erosão dos genomas ocorridos no decorrer da história evolutiva dessa espécie (Pieretti *et al.*, 2009). No entanto, esta erosão não faz deste fitopatógeno um organismo com baixa variabilidade genética (Bhattacharyya *et al.*, 2002a; Van Sluys *et al.*, 2003). Pelo contrário, análises filogenéticas usando marcadores moleculares do tipo RAPD (*Random Amplification of Polymorphic DNA*) e rRNA16S de cepas de citros, videira e amendoeiras, evidenciaram uma ampla diversidade genética, propiciando a separação dessas cepas em três grupos clonais: grupo de isolados de citros, causadores da CVC, grupo de isolados de videiras, causadores da PD e o grupo de isolados de amoreira, causadores da escaldadura da folha de amoreira (Chen *et al.*, 2002).

Um estudo recente realizado com 93 isolados de *X. fastidiosa* obtidos de videiras em Napa Valley, USA, mostrou uma baixa ou ausência de estrutura genética em escala geográfica, classificando esses isolados como uma população em *panmixia*. Essa hipótese da *panmixia* populacional é suportada pelo fato da mata ciliar se encontrar adjacente às plantações de videiras, as quais funcionam como reservatório de *Xylella*, estabilizando assim a estrutura genética da população (Coletta-Filho *et al.*, 2011). No entanto, esses dados contrastam com isolados de citros que apresentam uma clara estrutura genética geográfica (Coletta-Filho e Machado, 2003).

A reconstrução filogenética de 26 isolados norte-americanos utilizando marcadores moleculares do tipo *housekeeping genes*, possibilitou definir três subespécies de *X. fastidiosa*, sendo elas: subespécie *fastidiosa*, a qual compreende a cepa Temecula1; subespécie *multiplex*, cepas Dixon e M12 e subespécie *sandy*, formada pela cepa Ann-1. Interessantemente, essas subespécies foram

evolutivamente definidas há mais de 15.000 anos atrás (Schuenzel *et al.*, 2005). Mais recentemente, *X. fastidiosa* foi isolada da planta ornamental chitalpa, encontrada no Novo México, nos Estados Unidos. A filogenia desse isolado feita utilizando rRNA16S e a região espaçadora intergênica 16S-23S a classificou em uma nova subespécie, *subespecie tashke* (Randall *et al.*, 2009). No entanto, ainda não está totalmente esclarecida a origem de *X. fastidiosa* na América do Norte. Alguns estudos sugerem que possivelmente esse fitopatógeno foi introduzido naquele ecossistema por uma planta não nativa (Schuenzel *et al.*, 2005).

Como mostra a Figura 6, atualmente *X. fastidiosa* está dividida filogeneticamente em cinco subespécies: *pauca*, que inclui a cepa 9a5c, causadora de CVC; *perci*, que inclui as cepas Temecula1, GB514, M23; *multiplex*, que inclui as cepas Dixon, M12; *sandy*, que inclui a cepa Ann-1 e *tashke*, inclui o isolado da chitalpa (Randall *et al.*, 2009).

Figura 6. Reconstrução filogenética de cepas Norte e Sul-Americanas de *X. fastidiosa* a partir de rRNA 16S e da região espaçadora intergênica 16S-23S. A filogenia obtida com o método da máxima verossimilhança destaca as taxa integrantes das respectivas cinco subespécies. *Xanthomonas campestris* foi utilizada como grupo externo. Figura extraída de Randall et al., 2009.

Outros estudos que comprovam a variabilidade genética da *X. fastidiosa* foram realizados através do método de Tipagem de Sequência Multilocus (MLST) com 26 isolados de citros e 20 isolados de cafeeiro de diferentes regiões do Brasil, evidenciando uma clara distinção genética entre estes grupos de isolados, ainda que todos sejam da subespécie *pauca* (Almeida *et al.*, 2008).

Recentemente, Nunney e colaboradores desenvolveram um teste de hibridação introgressiva para determinar se as diferenças genéticas encontradas nos estudos por MLST de 55 isolados de citros e 22 isolados de cafeeiro da subespécie *pauca*, foram devidas da introgressão via recombinação homóloga. Neste estudo, foram observadas sete regiões de introgressão, totalizando 2.172 pb de 4.161 pb (52%). Este alto nível de introgressão sugere a hipótese de que *X. fastidiosa*, subsp. *pauca*, tornou-se patogênica em citros e café (cultivados no Brasil há centenas de anos), apenas recentemente, depois que ganhou variabilidade genética através de recombinação interespecífica, facilitando assim uma mudança de hospedeiros nativos. Nesse trabalho, foi sugerido que um possível candidato que tenha originado as cepas de citros e café seria a subespécie que infecta ameixa, presente na região desde 1935 (possivelmente *X. fastidiosa* subsp *multiplex*) (Nunney *et al.*, 2012).

1.2.4. Virulência e Patogenicidade de *X. fastidiosa*

Estudos em escala genômica conjuntamente aos estudos que utilizam abordagens de genética molecular, de bioquímica e de biologia celular têm corroborado a hipótese de que virulência e patogenicidade de *X. fastidiosa* compreendem múltiplos fatores, entre estes a adesão das bactérias aos tecidos da planta e do inseto vetor e que resulta na formação de um biofilme. Outros fatores relevantes para o estabelecimento da infecção por *X. fastidiosa* incluem sua capacidade de migração ao longo do vaso do xilema contra o fluxo da seiva pelo movimento do tipo contração (*twitching motility*) e também entre os vasos do xilema proporcionada pela expressão de enzimas capazes de degradar as pontuações de membrana que os separam. Além destes fatores, também tem sido considerados relevantes a captação de ferro e outros metais de transição por *X. fastidiosa*, possivelmente causando sua depleção no xilema das plantas infectadas, e a expressão e secreção de proteínas e/ou toxinas que promovam a sobrevivência da bactéria nos tecidos (Simpson *et al.*, 2000b; Koide *et al.*, 2004; Meng *et al.*, 2005; Roper *et al.*, 2007; Chatterjee *et al.*, 2008a; Zaini *et al.*, 2008; Caserta *et al.*, 2010; Killiny *et al.*, 2010; Perez-Donoso *et al.*, 2010; Silva *et al.*, 2011; Sun *et al.*, 2011).

Como já mencionado, a transmissão de *X. fastidiosa* ocorre através de um inseto vetor (cigarrinha) que ocasionalmente adquire as bactérias quando se alimenta de plantas contaminadas. Uma vez infectado, o inseto injetará a bactéria no xilema quando se alimentar da seiva bruta da planta. Nos vasos xilemáticos, a baixa concentração inicial do patógeno resulta em menor concentração da molécula sinalizadora DSF (ver adiante), refletindo na baixa expressão de adesinas afimbriais e fimbriais e alta expressão de proteínas do *pilus* IV e enzimas extracelulares. Esta pode ser chamada de fase exploratória, que consiste na migração e invasão do xilema. Se individualizadas, as células poderiam se deslocar no vaso através de motilidade *twitching*, mediada principalmente por *pilus* do tipo IV (De La Fuente *et al.*, 2007). Para disseminação sistêmica, a *X. fastidiosa* dispõe de um arsenal de enzimas extracelulares, como beta-1,4 endoglucanase, xilanase, xilosidase e poligalacturonase, para degradar membranas dos poros que separam vasos do xilema. Na cepa Temecula1, a poligalacturonase é considerada crítica para virulência e patogenicidade em videira (Roper *et al.*, 2007; Perez-Donoso *et al.*, 2010; Sun *et al.*, 2011).

Acompanhando o aumento da população bacteriana, ocorre acúmulo da concentração de DSF e, como consequência, a adesão da bactéria às paredes dos vasos é aumentada pela expressão de adesinas fimbriais e afimbriais e, também, pelo aumento da produção de exopolissacarídeo (EPS), resultando na formação de um biofilme maduro (Meng *et al.*, 2005; Chatterjee *et al.*, 2008a; Caserta *et al.*, 2010; Chatterjee *et al.*, 2010; Silva *et al.*, 2011).

Durante a formação do biofilme maduro, ocorrem mudanças temporais no padrão de expressão de proteínas fimbriais, como PilC e PilA2, e adesinas não-fimbriais, como XadA1 e XadA2 (Caserta *et al.*, 2010). Uma análise proteômica do biofilme de *X. fastidiosa* mostrou que 37% (ou 144 proteínas) das proteínas que foram detectadas no biofilme são diferentes das que foram detectadas em células planctônicas (Silva *et al.*, 2011). Uma vez formado o biofilme maduro na planta, sugere-se que ocorra o processo de dispersão ao longo dos vasos no xilema. Este processo já foi demonstrado no cultivo *in vitro*, sendo que a dispersão pode ocorrer como agregados celulares ou células individuais (Caserta *et al.*, 2010).

Hipotetiza-se que com o aumento da população bacteriana no xilema acompanhada da formação do biofilme, ocorre a oclusão dos vasos, resultando no estresse hídrico que é fundamental para o desenvolvimento de sintomas na planta infectada. Neste estágio, a planta torna-se uma fonte de aquisição da bactéria pelo inseto-vetor (Chatterjee *et al.*, 2008a; Zaini *et al.*, 2009; Chatterjee *et al.*, 2010). No inseto, a bactéria coloniza o cibário, e pela presença da atividade quitinolítica de enzimas produzidas por *X. fastidiosa*, provavelmente se alimentaria no início da

colonização de um dos principais componentes do exoesqueleto do inseto, a quitina (Killiny *et al.*, 2010). Uma vez, colonizada, a cigarrinha injeta *Xylella* ao picar as plantas na região apical para sugar a seiva bruta. Acredita-se que uma cigarrinha inocule ~100 bactérias por picada na planta (Dr. Elaine A. Backus, comunicação pessoal). Estudos recentes mostraram que a inoculação de bactérias ocorre junto com a saliva da cigarrinha, a qual contém enzimas, como a β -1,4-endoglucanase, possibilitando a quebra de ligações glicosídicas dos polissacarídeos encontrados na parede da planta e no EPS (Backus *et al.*, 2012).

1.2.5. Sinalização celular mediada por DSF

A sinalização extracelular em *X. fastidiosa* relacionada a percepção de quorum (*quorum sensing*) parece ser mediada principalmente por DSF (*Difusible Signaling Factor*), uma pequena molécula sinalizadora cuja síntese acompanha o aumento da densidade celular (Scarpari *et al.*, 2003; Wang *et al.*, 2004; Colnaghi Simionato *et al.*, 2007; Chatterjee *et al.*, 2008a; Dow, 2008; Newman *et al.*, 2008). Em bactérias patogênicas, o mecanismo de *quorum sensing* é capaz de sincronizar a expressão de genes de virulência de acordo com a densidade celular (Wang *et al.*, 2004; Jayaraman e Wood, 2008; Guo *et al.*, 2012; Pollumaa *et al.*, 2012).

A síntese, processamento e regulação de DSF estão associados a genes do cluster *rpf*, o qual, até o momento, parece ser exclusivo dos gêneros *Xanthomonas* e *Xylella*. Em *Xanthomonas*, o cluster *rpf* compreende nove genes (*rpfA-I*). O *rpfA* codifica uma aconitase implicada em homeostase de ferro; *rpfB* codifica uma ligase de acil-CoA de cadeia longa, e junto com *rpfF*, que codifica uma enoil CoA-hidratase, estão envolvidos na síntese e processamento de DSF. O produto de *rpfF* parece ser o principal responsável pela síntese de DSF, que, em *Xanthomonas campestris* *pv. campestris*, foi estruturalmente caracterizado como ácido cis-11-metil-2-dodecenóico. Em baixa densidade celular, são produzidos níveis basais de DSF, enquanto que em alta densidade celular, o DSF acumula-se no meio extracelular, levando a ativação de um sistema de dois componentes constituído de RpfC, uma histidina quinase transmembranar (sensor), e RpfG, uma fosfodiesterase de c-di-GMP (regulador da resposta), para percepção e transdução do sinal extracelular. Algumas evidências sugerem que RpfC liga-se a RpfF na ausência de DSF. Todavia, o acúmulo de DSF no meio extracelular induz a autofosforilação de RpfC e liberação de RpfF, resultando em sua ativação (Slater *et al.*, 2000; He *et al.*, 2006; Guo *et al.*, 2012; Reddy *et al.*, 2012).

Em *X. campestris* *pv. campestris*, o DSF modula positivamente a expressão de genes de virulência. Assim, a ativação de RpfG (fosfodiesterase de di-cGMP) resulta

na diminuição dos níveis de c-di-GMP. Mudanças nos níveis de c-di-GMP modulam a atividade do regulador transcricional Clp (*cAMP receptor-like protein*), o qual está diretamente envolvido no controle da expressão de genes de virulência nesta bactéria. Altos níveis de c-di-GMP promovem a dissociação de Clp da região promotora e consequente modulação negativa do *regulon* de Clp. Além disso, aumento do nível de c-di-GMP, pode tanto inibir a própria expressão de *clp* quanto à atividade de diguanilato ciclase (enzima responsável pela síntese de c-di-GMP) por inibição alostérica (Ryan *et al.*, 2006; He *et al.*, 2007; Leduc e Roberts, 2009; Zhang, 2010; Ryan *et al.*, 2011).

X. fastidiosa compartilha também a capacidade de coordenar a expressão de genes relacionados com virulência e patogenicidade via mudança de densidade celular no meio, usando o DSF como mediador (Chatterjee *et al.*, 2008a). Análises do genoma de *Xylella* revelaram a presença de um *cluster* de *rpf*, com sintenia similar aos encontrados nas cepas de *Xanthomonas* (Chatterjee *et al.*, 2008a). Entretanto, um rearranjo no genoma de *Xylella* separa os genes *rpfB* e *rpfF*, dividindo o *cluster* em duas partes com a deleção dos genes *rpfH*, *rpfD* e *rpfI*. Assim, ao contrário das *Xanthomonas*, *rpfB* e *rpfA* de *Xylella* estão fora do *cluster rpf* (Dow e Daniels, 2000; Simpson *et al.*, 2000b; Slater *et al.*, 2000).

Estudos do sistema de sinalização de DSF em *X. fastidiosa* sugerem que, nesta bactéria, a regulação de genes de virulência e patogenicidade envolve duas diferentes vias, sendo uma no nível da membrana citoplasmática e outra intracelular. Assim, o DSF produzido na célula difunde-se para o meio extracelular e sua percepção ocorre através do sensor RpfC, que provavelmente está associado a membrana celular. Essa ativação de RpfC, aparentemente, regula negativamente a produção de DSF. Assim como em *Xanthomonas*, RpfC de *X. fastidiosa* deve interagir com RpfF e com outras proteínas, incluindo um putativo repressor e modulador da produção de DSF. O DSF que é produzido por RpfF poderia acumular-se dentro da célula, e ser detectado por outro potencial sistema de dois componentes, sensor de DSF intracelular (Chou *et al.*, 1997; Chatterjee *et al.*, 2008a).

Em *X. fastidiosa*, DSF modula positivamente a formação de biofilme. Postula-se que esta modulação positiva ocorra pela ativação de RpfC/RpfG que resulta em uma suposta diminuição dos níveis de di-cGMP e ativação de Clp. Alternativamente, a formação de biofilme poderia ser também modulada pela produção de c-di-GMP catalizada pela di-guanilato ciclase, CgsA. A deleção de *cgsA* em *X. fastidiosa* cepa Temecula1 resulta em aumento dos níveis de transcritos de adesinas fimbriais e afimbriais e do operon *Gum* (produção de exopolissacarídeo) o que reflete diretamente no aumento na formação de biofilme e maior adesividade. O modelo atualmente

Embora DSF esteja claramente implicado na modulação da virulência em *X. fastidiosa*, a deleção de *rpfB*, cujo produto parece atuar no processamento de DSF, não afetou a colonização na planta. Entretanto, foi observada uma redução na colonização e transmissão pelo inseto vetor. Assim, é possível que *X. fastidiosa* coordene a colonização no hospedeiro vetor e na planta, variando as proporções de diferentes formas de DSF via RpfB (Almeida *et al.*, 2012).

Os dados acumulados até o momento indicam claramente que *Xanthomonas campestris* pv. *campestris* e *X. fastidiosa* compartilham o sistema de sinalização celular mediado por DSF que utiliza o segundo mensageiro c-di-GMP, ainda que as respostas finais sejam distintas. Em *Xanthomonas campestris* pv. *campestris*, a mutação de *rpfF* resulta em um fenótipo de deficiência na virulência, na produção de exopolissacarídeo e endonucleases. Todavia, em *X. fastidiosa*, o mutante de *rpfF* apresenta um fenótipo oposto, exibindo hipervirulência, baixa produção de biofilme e deficiência na transmissão e colonização da bactéria pelo inseto. Já a mutação de *rpfC* resultou em deficiência na virulência para ambas as espécies. Interessantemente, enquanto que em *X. campestris* foi observada diminuição na formação de agregados celulares e de biofilme, em *X. fastidiosa*, a mutação de *rpfC* resultou em um aumento considerável do biofilme e da agregação celular. Ademais, como mencionado acima, em *X. fastidiosa*, a mutação em *cgsA* e consequente diminuição dos níveis de di-cGMP resulta em aumento do biofilme, maior produção de exopolissacarídeo e no fenótipo mais agregativo. Contrariamente, a mutação em *rpfG* de *X. campestris* e consequente aumento nos níveis de di-cGMP resulta em um fenótipo hiperagregativo e deficiência na síntese de fatores de virulência (EPS e enzimas extracelulares). Assim, a atuação de c-di-GMP na regulação final dos fatores de virulência e patogenicidade são, aparentemente, opostos em *X. fastidiosa* e *X. campestris* (Barber *et al.*, 1997; Slater *et al.*, 2000; Dow *et al.*, 2003; Newman *et al.*, 2004; Ryan *et al.*, 2006; Torres *et al.*, 2007; Chatterjee *et al.*, 2008a; Ryan *et al.*, 2011; Sondermann *et al.*, 2012).

1.2.6. Estratégias de controle das doenças causadas por *Xylella*

Uma vez que a ocorrência de *X. fastidiosa* é restrita ao continente americano, a principal estratégia de controle das doenças causadas por esse fitopatógeno é implementar procedimentos de quarentena e fitossanitários de modo a prevenir a introdução do patógeno em regiões não afetadas (Janse e Obradovic, 2010).

Em muitos hospedeiros, sintomas causados por *X. fastidiosa* se desenvolvem com maior severidade em plantas enfraquecidas ou malcuidadas, uma vez que fatores

de estresse como seca, presença de outras doenças, superprodução de frutos e senescência normal da planta agravam os sintomas da doença. Assim boas práticas de manejo são importantes para manutenção do pomar saudável, tais como inspecções regulares dos pomares, podas após remoção de árvores doentes, controle de ervas-daninhas dentro e ao redor dos pomares e a eliminação de árvores lenhosas, especialmente carvalhos perto dos pomares (Hopkins e Purcell, 2002; Janse e Obradovic, 2010). Outros fatores ambientais, como temperatura e características do solo também podem afetar a susceptibilidade à doença. O estresse pela exposição de videiras a temperatura muito baixas resulta em mudanças fisiológicas, incluindo a produção de compostos fenólicos, que comprometem a sobrevivência de *X. fastidiosa* no tecido do xilema e podem resultar na cura da doença (Meyer, 2010).

O controle das populações de insetos vetores também é muito importante no manejo das doenças causadas por *X. fastidiosa*. Porém, a variedade de espécies capazes de transmitir a bactéria dificulta o seu controle, uma vez que diferem quanto a suscetibilidades a inseticidas. Além disso, a frequência de visitas às árvores das diferentes espécies de insetos também pode influenciar na dosagem da aplicação dos inseticidas. Um agravante, é que muitas espécies de plantas nas vizinhanças visitadas por estes insetos podem ser reservatórios de *X. fastidiosa*, dificultando ainda mais o controle do vetor. Existe um potencial para controle das populações de insetos vetores pela manipulação das condições que influenciem a química do fluido xilemático, gerando plantas inadequadas para alimentação dos insetos, o que resultaria na diminuição da densidade populacional (Redak *et al.*, 2004; Janse e Obradovic, 2010).

Uma alternativa vislumbrada para o controle de doenças causadas por *X. fastidiosa* é a utilização de linhagens bacterianas menos virulentas ou avirulentas como biocontrole. Tais linhagens atenuadas poderiam ser isoladas naturalmente ou geradas pela mutação de genes de virulência (Hopkins e Purcell, 2002). Um estudo recente mostrou que o pré-tratamento de videiras como uma linhagem biocontrole, EB92-1, não influenciou a incidência de sintomas de PD durante o primeiro ano após o tratamento, mas forneceu bom controle após o segundo e terceiro ano (Hopkins *et al.*, 2010). Outra possibilidade considerada é utilizar como biocontrole cepas de bactérias endofíticas. Araújo e colaboradores (2002) sugeriram que pode existir uma interação entre *Curtobacterium flaccumfaciens*, espécies de *Methylobacterium* e *X. fastidiosa* (Araújo *et al.*, 2002). Se por um lado o crescimento *in vitro* de alguns isolados de *X. fastidiosa* foi estimulado por isolados de *Methylobacterium*, foi verificado que sobrenadante do cultivo de *C. flaccumfaciens* resultou na inibição do crescimento (Lacava *et al.*, 2004).

A busca de variedades naturais e/ou geneticamente selecionadas de plantas que sejam mais resistentes a infecção também se coloca entre as alternativas para prevenção de doenças causadas por *X. fastidiosa*. No caso de videiras, quase todos os cultivares mais populares são susceptíveis à Doença de Pierce. Porém, espécies de videira resistentes a PD já foram descritas, sendo que tais genótipos exibem diferenças significativas no perfil de expressão gênica relativamente a plantas suscetíveis, possibilitando a identificação de genes potencialmente relacionados à resistência e susceptibilidade (Hopkins e Purcell, 2002; Yang *et al.*, 2011). No caso de citros, já foram descritas variedades de laranjas azedas e doces que não apresentaram sintomas em folhas até 27 meses após a inoculação (de Souza *et al.*, 2006). A expectativa é que análises genômicas e transcritômicas comparativas entre variedades de citros suscetíveis e resistentes à *X. fastidiosa* venham facilitar a identificação de genes relacionados a resistência e que possam guiar estratégias de melhoramento genético em citros (Gmitter Jr *et al.*, 2012).

O desenvolvimento de estratégias para o controle das doenças causadas por *X. fastidiosa* tem sido motivado pela importância econômica das culturas afetadas por este patógeno, como a vinicultura na Califórnia (EUA) e a citricultura no Estado de São Paulo. Como já mencionado, dados recentes do Fundecitrus indicam aumento da incidência e da severidade desta doença, visto que a proporção de árvores doentes no Estado de São Paulo passou de 35,5% em 2010 para 40,3% em 2011 (<http://www.fundecitrus.com.br>). Este aumento é preocupante e revela que muitos citricultores deixaram, aparentemente, de adotar as medidas regulares de prevenção.

2. Objetivos

O objetivo desse trabalho foi investigar genes potencialmente associados à adaptação ao hospedeiro, virulência e patogenicidade de *X. fastidiosa* através de um estudo de genômica comparativa associado à análise dos fenótipos exibidos por diferentes cepas de *X. fastidiosa*.

Para tal, nossos objetivos específicos foram:

1. Realizar o sequenciamento, a montagem e a anotação dos genomas das cepas U24d, Fb7 e 3124. U24d é uma cepa virulenta isolada de uma laranjeira com sintomas de CVC e representante de um haplótipo atualmente predominante no Estado de São Paulo. Fb7, também virulenta foi isolada de laranjeira com sintomas de CVC em Bella Vista, Corrientes na Argentina. A cepa 3124 foi isolada de cafeeiro com sintomas de requeima do cafeeiro, em São José do Rio Preto, São Paulo;
2. Realizar análises comparativas destes três genomas com os genomas das cepas de referência, 9a5c, virulenta isolada de laranjeira doce em sintomas de CVC no Estado de São Paulo, Brasil (Li *et al.*, 1999; Simpson *et al.*, 2000b), e Temecula-1, agente causal da doença de Pierce em videiras e isolada na Califórnia, Estados Unidos da América (Van Sluys *et al.*, 2003);
3. Avaliar o fenótipo das cepas U24d, Fb7 e 3124 *in planta* e no cultivo *in vitro* e estabelecer correlações com as diferenças verificadas a partir da análise de seus respectivos genomas;
4. Validar experimentalmente algumas das alterações verificadas nos genomas sequenciados que potencialmente se relacionam aos fenótipos destas cepas;
5. Realizar análises comparativas das sequências de todos os genomas de *X. fastidiosa* disponíveis (cepas Sul e Norte-Americanas) para descrever o pangenoma desta espécie. Além dos genomas de cepas 9a5c e Temecula1 já mencionados, outros genomas de *X. fastidiosa* já foram sequenciados completa ou parcialmente de cepas isoladas de distintos hospedeiros na América do Norte: GB514 de videira; Dixon, M12 e M23 de amendoeira; Ann-1 de espiroleira; EB92-1, isolada de sabugueiro e utilizada como biocontrole para Doença de Pierce de videiras (Simpson *et al.*, 2000b; Bhattacharyya *et al.*, 2002b; Van Sluys *et al.*, 2003; Chen *et al.*, 2005; Schreiber *et al.*, 2010; Zhang *et al.*, 2011) e na América do Sul: J1a12, não-virulenta isolada de laranjeira doce, Pr8x, isolada de ameixeira e Hib4, isolada de hibisco (da Silva, 2010; Pierry, 2012).

6. Realizar a reconstrução filogenética de todas as cepas sul e norte-Americanas por Análises de Sequências Multilocus (MLSA);
7. Identificar regiões de profagos e ilhas genômicas presentes nos genomas das cepas U24d, Fb7 e 3124 e realizar análises comparativas destes Elementos Genéticos Móveis (EGMs) entre os demais genomas de *X. fastidiosa*;
8. Iniciar estudos para compreender a aquisição dos EGMs por *X. fastidiosa*.

3. Procedimentos Experimentais

3.1. Manutenção e cultivo de *Xylella fastidiosa*

A manutenção das culturas das cepas utilizadas neste trabalho foi feita em placas de meio PWG-ágar (PWG-contendo ágar 1,5%) com repiques semanais e incubação a 28°C. O meio complexo PWG deriva do meio PW (Davis et al., 1981) sendo constituída de fitona peptona 4g/L; peptona de caseína com digestão trípica 1g/L; cloreto de hemina 0,001%; K₂HPO₄ 1,2g/L; KH₂PO₄ 1g/L; MgSO₄.7H₂O 0,4g/L; glutamina 0,4% e glicose 0,5%. Após 30 passagens, no máximo, novo cultivo era iniciado a partir de estoques de células em diversas passagens mantidos em PWG contendo 50% de glicerol em freezer a -80°C.

Para o cultivo em meio líquido, as bactérias após 7 dias de cultivo em meio sólido eram transferidas para 50 mL de meio PWG e mantidas a 28°C com 100 rpm de agitação por mais 7 dias. Após esse período, eram adicionados 50 mL de meio PWG fresco às culturas que permaneciam a 28°C e 100 rpm, por mais 2 dias.

As cepas de *X. fastidiosa* utilizadas neste trabalho foram: 9a5c, (Li et al., 1999; Simpson et al., 2000b); U24d, Fb7 e 3124 (cedidas pelo Dr. Helvécio Della Coletta Filho do Centro APTA Citros - Sylvio Moreira, Cordeirópolis, SP).

3.2. Microscopia Eletrônica de Varredura

As cepas de *X. fastidiosa* foram cultivadas durante 7 dias em meio PWG contendo lâminas de borossilicato submersas sobre as quais as células cresceram e desenvolveram o biofilme. As amostras foram fixadas com gluteraldeído a 2,5% durante 1 h e lavadas 4 vezes com tampão 0,1 M catecolato pH 7,2 por 10 min cada lavagem. Após a fixação, as laminas foram então suavemente colocadas em tetróxido de ósmio 1% em tampão catecolato durante 1 h, lavadas novamente 3 vezes com tampão catecolato e tratadas com ácido tânico 1% durante 30 min. Em seguida, as lâminas foram lavadas duas vezes com água deionizada. Após outra impregnação com tetróxido de ósmio durante 30 min, as amostras foram novamente lavadas com água deionizada 3 vezes e em seguida desidratadas em concentrações crescentes, (iniciando a 50% e incrementando 10% a cada passagem) de etanol durante 10 min cada. Depois da secagem ao ponto crítico das amostras, essas foram revestidas por uma camada de ~25nm de ouro por pulverização catódica e mantidas desidratadas até serem analisadas num microscópio eletrônico de varredura modelo FEG Quanta FEI 250 no Centro de Microscopia Eletrônica da Universidade Federal de São Paulo.

3.3. Microscopia de colônias de *X. fastidiosa*

As microscopias de colônias das cepas 9a5c, U24d, Fb7e 3124 foram realizadas em microscópio óptico invertido *Axiovert 200* (*Carl Zeiss MicroImaging*) com aumento de 10x. As cepas foram crescidas durante 5 dias em PWG-ágar. Regiões do ágar contendo colônias foram cuidadosamente transferidas para lâminas com auxílio de um estilete. As imagens foram registradas com o programa Scopeimage 9.0.

3.4. Infecção de laranjeira doce com *X. fastidiosa*

Os experimentos de infecção em laranjeiras foram realizados no Centro Apta Citros Sylvio Moreira (Cordeirópolis). As mudas (plântulas) de laranja doce variedade Pera *Citrus sinensis* L. Osbeck (sem porta-enxerto) com dois meses de vida e altura de 10-15 cm em tubetes de 50 mL contendo substrato HS Citros-5 foram inoculadas com as cepas U24d e 9a5c de *X. fastidiosa*. Para tal, foram feitas perfurações superficiais no caule da planta com uma agulha fina e em seguida foi aplicado 20 µL da suspensão de células em PBS a concentração de 305 UFC/ µL para a cepa U24d e 239 UFC/ µL para a cepa 9a5c. As plântulas inoculadas ou não (controle) foram mantidas em casa de vegetação com temperatura controlada para não ultrapassar 30°C. As irrigações das plântulas foram feitas em dois períodos diariamente. Após três meses, as plântulas foram transferidas para vasos de plástico de 500 mL contendo o mesmo substrato usado nos tubetes, os quais são mantidos em casa de vegetação sem controle de temperatura. A partir dessa fase as irrigações foram feitas diariamente, uma única vez.

3.5. Isolamento de DNA total de *X. fastidiosa* crescida em meio PWG

Para purificação de DNA genômico foi utilizado o *Wizard Genomic DNA Kit* (Promega) seguindo as instruções do fabricante com pequena modificação introduzida após alguns testes. Tendo em vista que *X. fastidiosa* forma biofilme adensado por exopolissacarídeo foi necessário aumentar 4 vezes o volume de solução de lise adicionado ao precipitado de bactérias no primeiro passo do protocolo (lise celular). Assim, partindo de 25 mL de cultura em meio PWG líquido, coletamos as células por centrifugação a 3.220xg por 5 min a 4°C e em seguida adicionamos 3,0 mL de solução de lise do *Wizard Genomic DNA Kit* ao precipitado de células. Após homogeneização, a suspensão foi distribuída em seis microtubos de 1,5 mL, sendo os procedimentos seguintes exatamente ao descrito no protocolo proposto pelo fabricante do *kit*. A quantificação e análise da pureza do DNA isolado foram avaliadas através da medida

das absorvâncias (Abs) a 230nm, 260nm e 280nm de 1,5 µL da amostra no espectrofotômetro ND-1000 (*NanoDrop Technologies*) e por eletroforese em gel de agarose e/ou eletroforese capilar no *BioAnalyzer* (*Agilent*).

3.6. Isolamento de DNA genômico de *X. fastidiosa* a partir de folhas de laranjeiras

Folhas de *Citrus sinensis* L. Osbeck infectadas com *X. fastidiosa* foram coletadas e lavadas com água corrente, seguindo-se remoção do pecíolo para trituração. A trituração foi feita com 250mg do material fresco com duas esferas metálicas (Stainless Steel beads-*Qiagen*) no triturador *TissueLyser* (*Qiagen*) por 2 min com 30Hz. Em seguida foram adicionados 700 µL do tampão CTAB-Sarcosyl (Tris-HCl pH 7,5 1M; EDTA 0,5M; NaCl 5%; CTAB 10%; Sarcosyl 10%; beta-mercaptoetanol 140 mM), seguindo-se incubação por 30 min a 60°C em banho-maria. A amostra foi então resfriada até a temperatura ambiente previamente à adição de 700 µL de clorofórmio: álcool isoamílico (24:1), seguindo-se centrifugação a 1.000 rpm por 8 min. A fase superior foi transferida para microtubo de 1,5mL, seguindo-se adição de 35 µL de CTAB 10%, vertendo-se o tubo várias vezes até completa homogeneização. Ao homogeneizado, foram acrescentados 400 µL de tampão de precipitação CTAB (CTAB 10%; Tris HCl pH 7,5 1M; EDTA 0,5M). A suspensão foi misturada gentilmente e, após em repouso por 15 min, centrifugada por 4 min a 1.000 rpm. O sobrenadante foi descartado e o precipitado foi ressuspenso em 400 µL de TE alto sal (Tris – HCl pH 7,5 1M; EDTA 0,5M; NaCl 5M) a 65°C em banho-maria. O DNA foi precipitado pela adição de 2 volumes de etanol absoluto gelado e centrifugação por 5 min a 12.000 rpm. O precipitado foi lavado sequencialmente com etanol 70% e etanol 100% por 5 min, e o líquido foi drenado pela inversão do tubo por 15 min. O precipitado foi dissolvido em 50 µL de TE contendo RNase.

3.7. Genotipagem de *X. fastidiosa*

As genotipagens das cepas de *X. fastidiosa* foram realizadas através de ensaios de PCR que detectam diferenças genômicas previamente identificadas entre as cepas 9a5c e J1a12, utilizando-se iniciadores para as regiões externas às CDS XF0077, que codifica um precursor de adesina de fímbria, e XF1250, que codifica uma arginina desaminase, as quais estão ausentes no genoma de J1a12 (Koide *et al.*, 2004). Estes pares de iniciadores também possibilitam distinguir a cepa 3124.

Também utilizamos os iniciadores CVC-1/272-2 classicamente utilizados na identificação de cepas associadas à CVC (Pooler e Hartung, 1995) e um par de iniciadores que definimos para genotipagem da cepa U24d.

O planejamento dos pares de iniciadores foi realizado com o programa PRIMER3 (Rozen e Skaletsky, 2000). As sequências dos iniciadores e o tamanho esperado dos *amplicons* estão apresentados na tabela 3.

Nos ensaios de PCR foram utilizados 100ng de DNA total purificado das cepas em estudo, 1 µL de solução de dNTPs 10mM, 1,5 µL de MgCl₂ 25mM, 10 µL de tampão 5x da *Taq* DNA Polimerase (Invitrogen), 1 µL de cada iniciador 10pmol/µL e água MilliQ estéril para o volume final da reação de 50 µL. As reações foram submetidas à desnaturação a 95°C, 5 min e a 35 ciclos de 95°C, 45 seg; 60°C, 30 seg; 72°C, 2 min. Ao final destes ciclos, as reações foram incubadas a 72°C, 10 min e armazenadas a 4°C até sua análise por eletroforese em gel de agarose 1%.

Tabela 3. Iniciadores para genotipagem de cepas de *X.fastidiosa* por PCR

Primer	Sequência	Tamanho do Amplicon nas cepas			
		9a5c	3124	J1a12	U24d
Δ0077 F	5'-ACTAAGCGCGACAATGAGGA	1599 pb	549 pb	549 pb	1599 pb
Δ0077 R	5'-GTGGTGTTTGGGCAAGCTAT				
Δ1250 F	5'-AATGGCAAGGTGGAAATCAG	1122 pb	1122 pb	159 pb	1122 pb
Δ1250 R	5'-CAACGCACATCCCCTAAGTC				
CVC-1 F	5'-AGATGAAAACAATCATGCAAA	506 pb	506 pb	506 pb	506 pb
272-2 R	5'-GCCGCTTCGGAGAGCATTCCT				
<i>pilQ</i> -InsF	5'- TAGTCACGTGTAGTCACG	---	---	---	347 pb
<i>pilQ</i> -InsR	5'-GGTCACACGCAACTCAA				

3.8. Eletroforese de DNA

A amostra de DNA (aproximadamente 200ng) foi previamente misturada com 0,1 volumes de tampão de amostra (azul de bromofenol 0,25% e sacarose 40%) e a eletroforese foi realizada em gel de agarose (0,7-1,2%) contendo 0,5 ug/mL de brometo de etídio, a 80-90 V em TBE (Tris-HCl 89 mM, ácido bórico 89 mM, EDTA 2

mM, pH 8,0). A corrida foi interrompida após o corante ter percorrido aproximadamente 4/5 da distância entre o local da aplicação e o limite inferior do gel. Ao final da corrida, o gel foi fotografado em um transiluminador com luz UV e o tamanho dos fragmentos foi estimado com base na comparação à mobilidade de fragmentos de DNAs com tamanhos já conhecidos (1Kpb Ladder - Invitrogen).

Alternativamente as amostras de DNA foram analisadas por eletroforese capilar no *BioAnalyzer* (Agilent), onde a matriz do gel aplicada nos capilares do *chip* contém uma mistura de fluoróforos e o marcador de tamanho. A ligação dos fluoróforos ao marcador de tamanho e ao DNA fita dupla resulta na emissão de fluorescência que possibilita a detecção do DNA.

3.9. Pirosequenciamento de DNA

Para o sequenciamento dos genomas de *X. fastidiosa* utilizamos a tecnologia de pirosequenciamento de DNA na plataforma 454 Genome Sequencer FLX (454 Life Sciences/Roche), com a metodologias de *shotgun* e/ou *Paired End Tags*.

O pirosequenciamento não requer a construção prévia de bibliotecas genômicas, seleção de clones e isolamento de plasmídeos (Margulies *et al.*, 2005; Ahmadian *et al.*, 2006). Na metodologia de *shotgun* o DNA genômico total é rompido mecanicamente em fragmentos de 500 a 800 pares de bases que serão diretamente pirosequenciados. Na metodologia *Paired End Tags* (Fullwood *et al.*, 2009), o DNA genômico é rompido mecanicamente em fragmentos maiores (3, 8 ou 20 kpb) que são purificados, ligados a adaptadores (*tags*) e circularizados com DNA ligase. Em seguida, procede-se fragmentação do DNA por nebulização e as etapas regulares do protocolo de pirosequenciamento *shotgun*. As sequências das *tags* (~40 nucleotídeos) demarcam extremidades pareadas (*paired-ends*) dos fragmentos de DNA de tamanho pré-determinado, orientando a montagem computacional da sequência genômica pela conectividade das sequências das *tags* pareadas.

3.9.1. Pirosequenciamento shotgun

O pirosequenciamento *shotgun* foi realizado no CATG (Centro Avançado de Tecnologia em Genômica) do Instituto de Química/USP (IQUSP) segundo as recomendações dos protocolos da 454 Life Sciences/Roche. Foram utilizados os reagentes e *kits* da série titânio, sendo a maioria dos sequenciamentos realizados com o método *Rapid Library Preparation*.

3.9.1.1. Quantificação e análise qualitativa do DNA genômico

A quantificação da amostra de DNA bacteriano total purificado foi realizada com o reagente “*Quant-it PicoGreen dsDNA Reagent*” (Invitrogen). Para a curva padrão são utilizadas concentrações diferentes de DNA de fago lambda fornecido pelo fabricante (50 ng, 25 ng, 12,5 ng, 6,25 ng, 3,125 ng, 1,563 ng e 0,781 ng). A amostra de DNA foi previamente diluída para concentrações menores que 50 ng, com base em medições prévias da quantidade de DNA no espectrofotômetro NanoDrop. Para a quantificação, 100 µL da solução do reagente fluorescente “*Quant-it Picogreen dsDNA Reagent*”, previamente diluído 1:200 em TE, são misturados a 100 µL das amostras de DNA em uma placa de 96 poços. A placa foi deixada em repouso à temperatura ambiente durante 5 min, antes da leitura da fluorescência em um espectrofluorímetro (excitação em 480nm e a intensidade da emissão medida a 520nm).

O DNA bacteriano total purificado teve sua integridade analisada pelo “*High sensitivity DNA kit*” (Agilent) utilizando o equipamento *Bioanalyzer* (Agilent). Antes de iniciar a análise das amostras de DNA, os reagentes a serem utilizados foram equilibrados à temperatura ambiente durante 30 min. Alíquotas do gel para uma corrida foram filtradas antes da adição de 1,5 µL do “*dye*” seguindo-se centrifugação por 10 min a 4000 rpm. A solução “*gel-dye mix*”, devidamente protegida da luz, foi utilizada para a montagem do *chip* com o auxílio de um suporte e de uma seringa. Foram adicionados 5 µL do marcador que demarca os limites de tamanho de amostras de DNA de cada tipo de kit utilizado no *Bioanalyzer* (no caso do *High Sensitivity*, usado neste ensaio, os limites são 50-7.000 pb). Em seguida, em um dos poços foi adicionado 1 µL do *ladder* e nos poços restantes foi adicionado 1 µL das amostras de DNA total. O *chip* foi vigorosamente misturado em seu suporte por 1 min a 2400 rpm e, em seguida, rapidamente inserido no “*Agilent 2100 Bioanalyzer*” para início da eletroforese em no máximo 5 min.

3.9.1.2. Construção da biblioteca de DNA fita-simples

As amostras de DNA genômico foram utilizadas na construção de bibliotecas de fragmentos de DNA fita-simples (ssDNA) para o pirosequenciamento com o protocolo “*GS FLX Titanium Series - Rapid Library Preparation Method*” da 454 Roche. Amostras de 500ng de DNA genômico de *X. fastidiosa* purificado e previamente avaliado e quantificado como descrito acima foram transferidas para microtubos de 1,5 mL e o volume final para 100 µL foi completado com tampão TE. A amostra foi então

depositada no fundo do aparato de nebulização (*nebulizer cup*) seguindo-se adição de 500 µL do tampão de nebulização e homogeneização da mistura.

A nebulização foi realizada em fluxo laminar com pressão de nitrogênio gasoso de 30 psi (2,1 bar) e com tempo de 1 min. A fragmentação ideal gera fragmentos de 500 a 800 pares de base.

Após a nebulização, 2,5 mL de tampão PBI (*MinElute PCR Purification Kit*, Qiagen) foram adicionados à amostra no aparato de nebulização, seguindo-se homogeneização e posterior filtração em colunas *MinElute*. A purificação consistiu na adição sucessiva de 750 µL da amostra de DNA nebulizado em uma coluna de purificação e, a cada adição o tubo era centrifugado por 15 segundos a 13000 rpm. Após a passagem de toda a amostra, o tubo com a coluna foi centrifugado por 1 min a 13000 rpm. Em seguida, a coluna foi lavada por centrifugação a 13000 rpm por um min com 750 µL de tampão PE. Após nova centrifugação por mais 1 min, a coluna foi invertida e centrifugada por mais 1 min, a 13000 rpm. Os fragmentos purificados foram eluídos com 16 µL de tampão TE através de centrifugação por 1 min a 13000 rpm. Em seguida, foi realizada a etapa de “*Fragment End Repair*” ou polimento das extremidades dos fragmentos pela adição de 2,5 µL de tampão PNK 10x, 2,5 µL de ATP, 1 µL de dNTP, 1 µL de T4 DNA polimerase, 1 µL de PNK (polinucleotídeo quinase) e 1 µL de Taq polimerase e incubação em termociclador (Applied Biosystems) por 20 min a 25°C e 20 min a 72°C. A reação foi armazenada a 4°C.

A etapa seguinte consiste na ligação dos adaptadores A e B às duas extremidades dos fragmentos com DNA ligase. A reação consistiu na incubação dos reagentes (ligase e adaptadores) com a amostra de fragmentos purificada por 10 min a 25°C. Estes adaptadores são um par de oligonucleotídeos dupla-fita que possibilitam a imobilização dos fragmentos de DNA nas microesferas, além de gerar regiões iniciadoras para a PCR em emulsão (emPCR) e para o sequenciamento. Eles possuem ainda uma sequência palindrômica de quatro nucleotídeos que será utilizado pelo programa de análise de dados do sequenciador 454 para reconhecer e selecionar os *reads*. O adaptador B possui biotina na sua extremidade 5' que garante a imobilização da biblioteca de dsDNA nas *beads* de imobilização (que possuem estreptavidina acoplada) e o subsequente isolamento da biblioteca de ssDNA que servirá de molde para o sequenciamento. Os adaptadores A e B não fosforilados são ligados ao DNA em reações que contém excesso molar de adaptadores para maximizar sua ligação ao DNA dupla-fita e evitando a formação de concatâmeros.

Em seguida, os fragmentos foram recuperados pelo método *Double SPRI* (*Solid Phase Reversible Immobilization*) que consiste na utilização de *beads* magnéticas (*AMPure beads*, Agencourt AMPure System, Beckman Coulter) que se

ligam preferencialmente a fragmentos de DNA de tamanhos maiores. A faixa de tamanho desejada dos fragmentos é de 400 a 800pb, com tamanho médio de ~700pb. Após todo o processo de ligação dos fragmentos às *beads*, lavagens são realizadas para a retirada dos fragmentos que não contém os adaptadores e por isso não ligaram às *beads*. As *beads* com os fragmentos são lavadas com etanol 70%, incubadas a 37°C por 1min. A biblioteca de ssDNA é ressuspensa em 50 µL de tampão TE e separada das *beads* magnéticas utilizando o *Magnetic Particle Concentrator*. A biblioteca de ssDNA é recuperada no sobrenadante e será então submetida a emPCR com *primers* complementares aos adaptadores A e B.

Após obtenção da biblioteca purificada é necessário quantificar o número de moléculas de DNA e analisar a distribuição dos tamanhos dos fragmentos de DNA. A quantificação é realizada em placas de 96 poços e lida em espectrofluorímetro. Uma curva padrão é confeccionada a partir da leitura de uma diluição seriada de uma amostra de DNA fornecida pelo fabricante. Em cada poço, deve-se adicionar 50 µL das amostras, sendo que no poço com a amostra da biblioteca a ser medida deve-se aplicar todo o seu conteúdo. Após a medição, a biblioteca é recuperada e transferida para um novo tubo. A curva padrão e o cálculo da concentração da biblioteca foram obtidos com o uso da “*Rapid Library Quantitation Calculator*” disponível em www.454.com/my454 e amostra é diluída para 1×10^7 moléculas/µL. A análise da distribuição de tamanhos dos fragmentos na biblioteca foi realizada no equipamento *Agilent 2100 Bioanalyzer* (como descrito no item anterior) utilizando o *High sensitivity DNA chip* com 1µL da biblioteca.

3.9.1.3. Titulação da biblioteca e emPCR em pequena escala

Uma vez quantificada e analisada quanto a sua qualidade, a biblioteca de DNA é utilizada para a amplificação dos fragmentos por PCR em emulsão. A reação de emPCR requer que cada *bead* esteja ligada a apenas uma molécula molde de ssDNA amplificável. Para garantir o sucesso do sequenciamento, a quantidade ótima de DNA adicionada à solução de *beads* deve ser determinada uma vez que, se pouco DNA é adicionado, a reação de emPCR resultaria em um grande número de *beads* que não estariam ligadas a DNA molde. Por outro lado, se for adicionado DNA em excesso, algumas *beads* conteriam múltiplos moldes amplificáveis o que é inadequado para o sequenciamento. A quantidade de biblioteca de ssDNA que resulta em uma única e efetiva molécula de DNA molde ligada a cada *bead* depende da qualidade do DNA, da composição da biblioteca, bem como de outros fatores não identificados.

Inicialmente foi realizada a titulação da biblioteca utilizando o protocolo de emPCR em pequena escala, “GS FLX Titanium Series emPCR Method– Lib-L SV” e o “GS FLX Titanium Small Volume emPCR Kit” para determinar a porcentagem de *beads* enriquecidas (o valor ideal é de 8% de *beads* enriquecidas). Para tal, volumes diferentes de cada biblioteca de ssDNA na concentração de 1×10^6 e 2×10^6 moléculas de DNA/ μL (diluídas da amostra com 1×10^7 moléculas de DNA/ μL) foram misturados a um número fixo de *capture beads* de modo a obtermos 0,5, 2, 4, 8 cópias por *bead*.

Nesta etapa, os fragmentos que constituem a biblioteca de ssDNA são imobilizados nas *capture beads* para permitir sua separação na emulsão. O objetivo é garantir que cada *capture bead* esteja ligada a apenas uma molécula de DNA e que cada microrreator aquoso possua apenas uma *capture bead*. A imobilização às *capture beads* envolve o anelamento do adaptador B, que está ligado à extremidade de cada molécula molde que constitui a biblioteca, aos iniciadores de captura imobilizados nessas *beads*. Para reação de emPCR, as *capture beads* com a molécula de ssDNA imobilizada são ressuspensas com os reagentes de amplificação e óleo, seguindo-se vigorosa agitação mecânica no equipamento *TissueLyser* (Qiagen) a 25 Hz por 2 minutos, para formação da emulsão.

Após a emulsificação, a mistura para amplificação contendo as *beads* com fragmentos foi distribuída em uma placa de 96 poços, sendo 100 μL em cada poço. A emPCR seguiu-se com o seguinte programa de amplificação: 94°C por 4 minutos, 50 ciclos de 94°C por 30 segundos, 58°C por 4,5 minutos e 68°C por 30 segundos, sendo mantidas ao final a 10°C até serem retiradas do termociclador. As emulsões e os microreatores foram então quebrados, para a recuperação das *beads* contendo a biblioteca de ssDNA amplificada, com o auxílio de uma seringa, seguindo-se diversas lavagens com isopropanol.

O processo de amplificação gera uma proporção de *beads* cuja amplificação não foi bem sucedida. Desta forma, o passo seguinte consiste no enriquecimento (separação) da população de *beads* da biblioteca de ssDNA efetivamente amplificada. Esta etapa baseia-se no anelamento do *primer* de enriquecimento, o qual é biotinilado, ao adaptador A de cada molde amplificado, e a posterior ligação desse *primer* a *beads* magnéticas contendo estreptavidina. As *beads* contendo a biblioteca de ssDNA amplificada são então separadas dessas *beads* magnéticas por desnaturação dos produtos de amplificação do *primer* de enriquecimento utilizando NaOH. A proporção final entre a quantidade de *beads* obtidas com o enriquecimento e a quantidade inicial de *beads* utilizada na emPCR corresponde à porcentagem de *beads* enriquecidas e é um indicativo da qualidade da biblioteca amplificada. Para a contagem de *beads*, 3 μL

das amostras são adicionadas a 10 mL de Isoton e avaliadas no equipamento *Beckman Coulter*. Com os valores das porcentagens de enriquecimento em cada diluição da titulação, foram construídas as curvas de titulação para ambas as amostras e a quantidade ótima de biblioteca de DNA a ser utilizada na emPCR em grande escala é aquela que resulta na porcentagem de 8% de *beads* enriquecidas, segundo o fabricante. Este enriquecimento corresponde à quantidade que resultou, supostamente, em uma única molécula de DNA utilizada como molde para a amplificação.

3.9.1.4. emPCR definitiva

A partir da titulação descrita acima, é definida a quantidade da biblioteca de ssDNA que será utilizada na emPCR definitiva, isto é número de cópias por *bead*, para que a porcentagem de enriquecimento ao final do processo fique idealmente entre 5% e 20%. Valores de enriquecimento muito acima de 20% indicam alta probabilidade de existirem *beads* mistas, ou seja, que contenham mais de um fragmento de ssDNA ligados a uma única *bead* antes do processo de emPCR, resultando na perda destas sequências ao final do processo.

O procedimento para realização da emPCR em larga escala se assemelha ao procedimento de emPCR em pequena escala, exceto pelo maior volume inicial. O processo de ligação às *capture beads*, formação da emulsificação, mistura de amplificação e reação de amplificação são os mesmos que o processo de emPCR em pequena escala. A reação é realizada em *cups* e não mais microtubos. Ao final do processo, cada biblioteca gerará duas placas de 96 poços inteiras para a etapa de amplificação.

Após a amplificação, a emulsão e os microrreatores são quebrados para separação das *beads* amplificadas, neste caso com um aparato interligado a dois tubos de 50 mL ligado a uma bomba a vácuo, e não manualmente com uma seringa. Novamente as lavagens são realizadas diversas vezes com isopropanol. O processo de enriquecimento das *beads* assemelha-se a emPCR em pequena escala descrito acima, incluindo a c ontagem de *beads* no equipamento *Beckman Coulter* para estimativa do enriquecimento obtido.

3.9.1.5. Preparação da PTP e pirosequenciamento

A etapa de sequenciamento é dividida em três fases: o preparo da *PicoTiter Plate* (PTP), a pré-lavagem do sequenciador e a corrida de sequenciamento. É importante sincronizar o preparo da placa e do sequenciador, para assegurar que a

corrida de sequenciamento seja iniciada imediatamente após o término do preparo da placa. O aparelho *Genome Sequencer FLX* deve ser cuidadosamente lavado antes de cada corrida de sequenciamento. Após a amplificação, as *beads* recobertas com fragmentos de DNA são depositadas e imobilizadas na PTP que contem 1,6 milhões de poços, os quais têm a dimensão exata para comportar uma única microesfera. A PTP foi previamente carregada com outras *beads* que garantem o empacotamento e a formação de camadas de apoio. Em seguida são aplicadas as *beads* ligadas a enzimas que atuam na reação de pirosequenciamento e a placa pronta é inserida no sequenciador, seguindo-se processamento automatizado para adição sucessiva dos reagentes de sequenciamento.

A reação de pirosequenciamento ocorre em ciclos, paralelamente em cada um dos milhões de poços da PTP, sendo que cada um deles contém, idealmente, uma única *bead* com moléculas de ssDNA idênticas resultantes da etapa de amplificação na emPCR. A reação é iniciada a partir de um *primer* de sequenciamento que é anelado a extremidade de cada molécula de DNA imobilizada nas *beads*.

A cada ciclo da reação de sequenciamento cada um dos desoxirribonucleotídeos (dGTP, dTTP, dCTP e dATP) é adicionado isoladamente e de forma alternada. Caso ocorra a reação de incorporação do nucleotídeo pela DNA polimerase, a partir da sequência do DNA molde, uma molécula de pirofosfato é liberada; esta é utilizada na síntese de ATP pela ATP sulfúrilase, juntamente com o substrato adenosina fosfossulfato (APS). O ATP liberado é hidrolisado pela luciferase na reação com luciferina com produção de luz. O dNTP residual é hidrolisado pela enzima apirase antes da liberação do fluxo do dNTP seguinte. Desta forma a cada ciclo, uma câmera CCD (*Charge Coupled Device*) captura imagens para cada fluxo de um dos 4 dNTPs, onde pontos luminosos significam que ocorreu a incorporação do nucleotídeo naquele poço da placa.

Os dados brutos de uma corrida de sequenciamento consistem de uma série de imagens digitais capturadas pela câmera CCD, representando a superfície da PTP durante as reações de sequenciamento em tempo real. A quantidade de luz emitida é proporcional ao número de nucleotídeos que são adicionados pela DNA polimerase. Em seguida, os dados brutos serão filtrados quanto à presença de *beads* mistas e os dados brutos serão processados. *Beads* mistas são detectadas pelo próprio sequenciador a partir da presença de dois ou mais sinais luminosos distintos e simultâneos no mesmo poço durante o sequenciamento o que significa a presença de mais de um fragmento de ssDNA ligado a uma *bead*, resultado de uma diluição errada da biblioteca para a realização da emPCR em larga escala.

Após ~10-12 h de corrida, são obtidos dados brutos das imagens dos sinais luminosos gerados a partir da PTP e detectados por uma câmera CCD. Após ~10 h adicionais de análise computacional, as imagens são processadas pelo *software* que acompanha o 454 *Genome Sequencer* que converte o sinal de fluorescência em sequências de nucleotídeos (*reads*), sendo cada sequência derivada das imagens de um poço da placa. O processamento dos dados envolve etapas de correção automática dos dados para compensar efeitos óticos e alguma ineficiência química do processo, sendo atribuído a cada nucleotídeo sequenciado um *score* de qualidade ou confiabilidade. As sequências processadas são então submetidas a filtros de qualidade que identificam e excluem aquelas que apresentam baixa qualidade.

Com as versões dos *kits* e reagentes que utilizamos, são obtidas, geralmente, aproximadamente 1 milhão de sequências cada uma com mais de 400 bases com 99% de confiança (*Phred score* > 20).

Os dados brutos, metadados e dados métricos são armazenados em planilhas e arquivos texto, a partir dos quais informações sobre o sequenciamento podem ser extraídas. Um dos tipos de arquivos que é gerado contém a sequência de cada *read* associada à qualidade de cada base, o qual é utilizado como entrada nos programas de montagem dos *contigs*, tais como o *Newbler*, como será detalhado adiante.

3.9.2. Sequenciamento de *Paired End Tags*

Na tentativa de facilitar a montagem, os genomas das cepas U24d e Fb7 foram também sequenciados utilizando-se a metodologia *Paired End Tags* da plataforma 454/Roche (Fullwood *et al.*, 2009) o qual foi realizado na Unidade de Genômica Computacional Darcy Fontoura no Laboratório Nacional de Computação Científica (LNCC), em Petrópolis.

O DNA bacteriano foi submetido à quebra mecânica no equipamento *Hydroshear* (Digilab Inc) para obtenção de fragmentos de ~3 kpb. Os fragmentos tiveram suas extremidades aparadas para ligação dos adaptadores (*tags*) e posterior circularização com DNA ligase utilizando-se o kit *GS FLX Titanium Library Paired End Adaptors*. Em seguida, procede-se fragmentação do DNA circularizado por nebulização e as etapas regulares do protocolo de pirosequenciamento na plataforma 454/Roche descritas anteriormente.

A biblioteca de *shotgun* gerada a partir da biblioteca de 3Kpb *Paired End Tags* contém fragmentos com a sequência do adaptador de 44 nucleotídeos flanqueada por sequências originalmente separadas por 3 kpb, sendo que as sequências destes

fragmentos de extremidades pareadas serão identificadas computacionalmente na etapa de montagem do genoma.

3.10. Sequenciamento de DNA pelo método de Sanger

O sequenciamento de DNA pelo método de Sanger foi realizado no aparelho ABIPRISM™ 3100 (Applied Biosystems) conforme recomendações do fabricante, no serviço de sequenciamento de DNA do Departamento de Bioquímica do Instituto de Química-USP. A reação de sequenciamento foi realizada utilizando-se 2 µL da mistura da reação *Big Dye terminator mix* (Applied Biosystems), 100-200ng de DNA, 9,6 pmol de cada oligonucleotídeo, Tris-HCl 200mM pH 9,0 e MgCl₂ 5mM em um volume final de 15 µL. A reação foi incubada por 2 min a 95°C, seguida de 35 ciclos de 45 segundos a 96°C, 30 segundos a 50°C e 4 min a 60°C. Terminada a reação, as amostras foram precipitadas com etanol 100% e glicogênio 1mg/mL, lavadas com etanol 70%, suspensas em formamida, desnaturadas a 94°C por 5 min e aplicadas no sequenciador automatizado. A qualidade das sequências geradas foi analisada no programa Chromas lite versão 2.01.

3.11. Análises bioinformáticas

3.11.1. Montagem e anotação de genomas

A montagem de *contigs* e *scaffolds* dos genomas que foram sequenciados (cepas U24d, Fb7 e 3124) foi realizada *ab initio* com as sequências obtidas no sequenciador 454 utilizando-se os Programas *Newbler* (454 Life Sciences/Roche), MIRA (http://chevreux.org/projects_mira.html) e Cross_Match (<http://www.phrap.org/phredphrap/general.html>), aliados a cuidadosas intervenções manuais. As montagens destes 3 genomas foram realizadas pelo Dr. João Paulo Kitajima. Algumas etapas de verificação das montagens contaram com o apoio de Luiz Gonzaga Paula de Almeida (Laboratório Nacional de Computação Científica, Petrópolis, RJ).

A montagem do genoma da cepa EB92.1 foi realizada a partir de dados publicamente disponíveis (Zhang *et al.*, 2011). Os 164 *contigs* referentes ao genoma desta cepa foram ancorados às sequências de genomas publicamente disponíveis de cepas de mesma subespécie, Temecula1, GB514 e M23 (Van Sluys *et al.*, 2003; Chen *et al.*, 2010; Schreiber *et al.*, 2010). Com intervenções manuais foi possível montar um único *scaffold*, ainda que 500 *contigs* tenham sido excluídos da montagem. No entanto, vale ressaltar que essa montagem não foi usada para realizar os estudos de

pangenoma, mas para as análises de rearranjo cromossômico e identificação de região de profagos.

A anotação automática e análises comparativas das CDS/ORFs dos genomas de *X. fastidiosa* foram realizadas utilizando-se a plataforma IMG/ER (*Integrated Microbial Genomes-Expert Review*) (Markowitz *et al.*, 2012), disponível em <http://img.jgi.doe.gov/>. Esta plataforma realiza a predição de genes e sua anotação automática integrando informações e termos de bancos de dados tais como Nomenclatura de Enzimas (Bairoch, 2000), COG (Tatusov *et al.*, 2003), Pfam (Punta *et al.*, 2012), TIGRfams (Selengut *et al.*, 2007), InterPro (Hunter *et al.*, 2012), e *Kegg Ortholog* (Kanehisa *et al.*, 2012), entre outros.

Além disso, o IMG/ER disponibiliza diversas funcionalidades para análises comparativas diversas como BLAST (*Basic Local Alignment Search Tool*) (Altschul *et al.*, 1997), Artemis (Carver *et al.*, 2012) e além da visualização gráfica de mapas cromossômicos (*Genome browser*) e ferramentas de anotação manual pelo usuário.

3.11.2. Análises comparativas de genes e genomas

As análises comparativas para anotação foram realizadas em bases públicas de dados tais como GenBank do NCBI (*National Center for Biotechnology Information*, <http://www.ncbi.nlm.nih.gov>), *Xylella fastidiosa* Genome Project (<http://aeg.lbi.ic.unicamp.br/xf/>) e *Xylella fastidiosa*/Pierce's Disease Strain Genome Project (<http://bioinfo03.ibi.unicamp.br/lbi/index.html>). Para buscas por identidade e similaridade com sequências de nucleotídeos ou de aminoácidos nestas bases de dados foram usadas as ferramentas Blastn, Blastp e Blastx (*Basic Local Alignment Search Tool*) (Altschul *et al.*, 1990; Altschul *et al.*, 1997). Para identificação de genes supostamente ortólogos, consideramos como parâmetros: identidade $\geq 80\%$, E-valor $\leq 10^{-10}$ e cobertura $\geq 90\%$.

Para o alinhamento de sequências de aminoácidos ou de nucleotídeos utilizamos os programas ClustalW2 (Thompson *et al.*, 2002).

Para alinhamento global das sequências genômicas utilizamos os programas Mummer (Delcher *et al.*, 2003) e MAUVE (Darling *et al.*, 2010). Para confecção de atlas cromossômicos das cepas utilizamos o programa GenomeViz (Ghai *et al.*, 2004).

3.11.3. Identificação de grupos de CDS ortólogas e parálogas

Para identificação de grupos de CDS ortólogas e parálogas de forma mais acurada, foi necessário submeter todos os genomas comparados a mesma estratégia

de anotação. Desta forma, além dos novos genomas sequenciados neste trabalho, os demais genomas incluídos nas análises comparativas foram reanotados na plataforma IMG-ER como descrito acima. Estas reanotações são de uso restrito do nosso grupo para evitar conflitos com anotações publicamente disponíveis.

A ferramenta INPARANOID (Remm *et al.*, 2001) foi aplicada a cada par possível dos 14 genomas de *X. fastidiosa* incluídos na comparação, utilizando-se como entrada dois arquivos fastas contendo o conjunto completo de proteínas do par de genomas analisado (os genes plasmidiais não foram incluídos). Esta ferramenta identifica grupos de proteínas ortólogas entre dois genomas distintos, através da estratégia de melhor *hit* Blastp bidirecional. Os resultados obtidos foram combinados com a ferramenta MultiParanoid (Alexeyenko *et al.*, 2006), o que possibilitou a identificação de grupos de proteínas ortólogas presentes em todas as combinações dos pares possíveis dos 14 genomas analisados.

Para determinar a diversidade do genoma de cada isolado, isto é, o número de grupos parálogos em um dado genoma, foi feita uma rodada de Blastall (blastp) do proteoma contra ele mesmo e os *hits* com *query coverage* e identidade maior ou igual a 60% foram utilizados para compor cada grupo de parálogos, os quais foram utilizados para contabilizar a diversidade dos pangenomas individuais. Esta estratégia foi utilizada como alternativa ao INPARANOID tendo em vista que características intrínsecas da etapa de agrupamento (*clustering*) desta ferramenta poderia levar a superestimação do número de grupos parálogos se aplicada em um proteoma contra ele mesmo.

3.11.4. Determinação do pangenoma e do *core-genome*

O número de genes compartilhados, o número de genes exclusivos e a diversidade do pangenoma são medidas que sofrem influência do número de genomas comparados. A inclusão sequencial de cada um dos 14 genomas de *X. fastidiosa* na análise foi simulada em todas as combinações possíveis e os resultados obtidos para cada uma das medidas dada por $n!/(n-i)!.i$, onde n é o número total de genomas e i o valor do eixo X (combinações dos genomas), para o pangenoma, core-genoma e genes novos foram plotados em função do número de genomas sequencialmente adicionados. As medianas foram utilizadas como referência para inferir as tendências de cada uma das medidas. No gráfico de pangenoma, os valores iniciais representam o número de genes em cada uma das 14 cepas isoladamente.

3.11.5. Inferência filogenética

A inferência filogenética foi feita usando o método de Máxima Verossimilhança com 1000 réplicas de *bootstrap*. Os modelos de evolução estimados a partir das sequências concatenadas foi o GTR+G, recomendados pelo jModelTest v0.1.1 (Posada, 2008) considerando-se 11 tipos de substituições.

Sete *loci*, correspondentes a genes *housekeeping*, foram usados nas análises filogenéticas: *leuA* (XF1818), *petC* (XF0910), *malF* (XF2447), *cysG* (XF0832), *hoIC* (XF0136), *nuoL* (XF0316) e *gltT* (XF0656). A escolha desses genes foi baseada na inferência filogenética de isolados de *X. fastidiosa* descrita recentemente (Nunney *et al.*, 2012).

As sequências completas de nucleotídeos de CDS ortólogas para todas as cepas incluídas na filogenia foram recuperadas de bases de dados públicas GenBank e de dados gerados pelo sequenciamento de genomas realizado neste trabalho e em nosso grupo de pesquisa. As sequências de nucleotídeos destas CDS genes foram previamente concatenadas e submetidas ao alinhamento utilizando-se o programa ClustalW incluído na plataforma BioEdit (*Biological Sequence Alignment Editor*) (Hall, 1999; Chenna *et al.*, 2003), disponível em <http://www.mbio.ncsu.edu/bioedit/bioedit.html>.

A partir do alinhamento foi calculada a filogenia usando o PAUP4b10 (Wilgenbusch e Swofford, 2003). A topologia das árvores foi construída empregando-se o método de máxima verossimilhança com 1000 buscas heurísticas e utilizando-se como algoritmo de substituição de ramos BS (*Branch swapping*) o TBR (*Tree Bisection and Recognition*). Para verificar o suporte dos ramos, foram conduzidas análises de *bootstrap* com 100 réplicas, cada uma com 10 pseudoréplicas. *Xanthomonas albilineans* foi usada como grupo externo.

3.12. Obtenção de anticorpos policlonais para PilQ e PilY1 de *X. fastidiosa*

Anticorpos policlonais para as proteína PilQ e PilY1 de *X. fastidiosa* foram preparados pela imunização de coelhos com as proteínas recombinantes correspondentes obtidas como descrito a seguir.

3.12.1. Clonagem de segmentos de CDS de *X. fastidiosa* para expressão recombinante

Os segmentos de interesse de *pilQ* e *pilY1* de *X. fastidiosa* foram amplificados por PCR com os pares de iniciadores listados na tabela 4. O planejamento destes

pares de oligonucleotídeos foi realizado com o programa PRIMER3 (Rozen e Skaletsky, 2000), sendo incluídos sítios de restrição para *Xho*I ou *Eco*RI para posterior clonagem no vetor pAE (Ramos *et al.*, 2004).

As PCRs para amplificação do segmento N-terminal de *pilQ* e *pilY1* e C-terminal de *pilQ* foram realizadas com 100ng de DNA total purificado da cepa 9a5c, 1 µL de solução de dNTPs 10mM, 1,5 µL de MgCl₂ 25mM, 10 µL de tampão 5x da *Taq* DNA Polimerase (Invitrogen), 1 µL de cada iniciador 10 pmol/ µL e água MilliQ estéril para o volume final da reação de 50 µL. As reações de PCR para N-terminal de *PilQ* e *PilY1* foram submetidas à desnaturação a 95°C, 5 min e a 25 ciclos de 95°C, 45 seg; 64°C, 30 seg; 72°C, 40 seg. Para amplificação do segmento C-terminal de *pilQ* as reações foram submetidas à desnaturação a 95°C, 5 min e a 25 ciclos de 95°C, 45 seg; 60°C, 30 seg; 72°C, 40 seg. Ao final dos ciclos, as reações foram incubadas a 72°C, 10 min e armazenadas a 4°C.

Os amplicons obtidos foram analisados em eletroforese em gel de agarose 1% e purificados utilizando-se o *kit WizardSV Gel and PCR Clean-up System* (Promega) de acordo com o protocolo sugerido pelo fabricante. A quantificação do DNA extraído foi estimada por análise em eletroforese em gel de agarose e comparado com fragmento de DNA de tamanho e quantidade conhecida.

Tabela 4. Iniciadores para amplificação de segmentos de *pilQ* e *pilY1*

CDS	Primer	Sequência	Amplicon (9a5c)
<i>pilQC</i> -terminal XF0373	<i>pilQC</i> -F	5'- <u>CTCGAGG</u> CATTGCAGCAGGAATCCCG <i>Xho</i> I	616 pb
	<i>PilQC</i> -R	5'- <u>GAATTC</u> AGCCACACGCAACACCTTAG <i>Eco</i> RI	
<i>pilQN</i> -terminal XF0373	<i>pilQN</i> -F	5'- <u>CTCGAG</u> CTGAAAACAAGCAAGCGATTTC <i>Xho</i> I	583 pb
	<i>PilQN</i> -R	5'- <u>GAATTC</u> ATTAGCACGCAAAACGATGTC <i>Eco</i> RI	
<i>pilY1N</i> XF0478	<i>pilY1N</i> -F	5'- <u>CTCGAG</u> CGTGATCACAAGTTGTACTACGAGGC <i>Xho</i> I	504 pb
	<i>PilY1N</i> -R	5'- <u>GAATTC</u> CACAGCTGACGCGCTTATTA <i>Eco</i> RI	

Os amplicons purificados foram ligados ao vetor pGEM-T utilizando-se o *Kit pGEM-T Vector Systems*, seguindo-se transformação de bactérias competentes XL-1 Blue e seleção com carbenicilina 100µg/mL.

As construções no vetor pGEM-T foram digeridas com as enzimas *Xho*I e *Eco*RI e os insertos foram separados em gel de agarose 1%, seguindo-se eluição com

o *Kit WizardSV Gel and PCR Clean-up System* (Promega) e ligação ao vetor de expressão pAE previamente digerido com estas enzimas, utilizando-se a T4 DNA ligase (1U/ μ g DNA) por 12-16h a 16°C. As reações de ligação foram utilizadas na transformação de bactérias competentes XL-1 Blue.

No vetor pAE, a proteína recombinante é expressa com a região amino-terminal fusionada a uma sequência de seis resíduos de histidina o que facilita sua posterior purificação por cromatografia de afinidade em resina de Ni-NTA agarose.

3.12.2. Preparação de bactérias competentes para transformação

Para a clonagem e preparação das construções plasmidiais foi utilizada *Escherichia coli* cepa XL1 Blue (*supE44 hsdR17 recA1 endA1gyrA46thiA1 lacF'*[*proAB⁺lac^qlacZ Δ M15 Tn10(tet^r)*]) e para expressão de proteínas recombinantes foi utilizada a cepa BL21 (DE 3) *Gold* (*F⁻ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻) λ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7nin5])*).

Estoques das culturas de *E. coli* XL1 Blue e BL21(DE3) foram mantidos em meio LB contendo glicerol 20% (v/v) a -80°C. Alíquotas destes estoques foram semeadas em placas de meio LB [triptona 1%; extrato de levedura 0,5%; NaCl 1% e ágar 1,5%, pH 7,5 contendo tetraciclina 15 μ g/mL no caso de XL1 Blue. Para obtenção de culturas em meio líquido, uma colônia foi inoculada em 3 mL de meio LB líquido e crescida a 37°C por uma noite a 200rpm. No dia seguinte, uma fração do pré-inóculo foi diluída 1:100 em meio LB fresco, seguindo-se crescimento (200 rpm, 37°C) até $DO_{600nm} = 0,5$.

Para preparação de células competentes, 40 mL da cultura com $DO_{600nm} = 0,5$ foi transferida para garrafa de polipropileno estéril, resfriada em banho de gelo por 10 min, seguindo-se centrifugação (3.000 rpm por 10 min a 4°C). O sedimento foi suspenso em 10 mL de solução de cálcio gelada ($CaCl_2$ 60mM, PIPES 10mM pH 7 e glicerol 15%) e mantido em banho de gelo por 10 min, seguindo-se centrifugação na mesma condição anterior. O sedimento de células foi novamente ressuspensionado em 10 mL de solução de cálcio gelada e mantido em banho de gelo por mais 30 min. Após centrifugação, o sedimento foi delicadamente ressuspensionado em 2 mL da solução de $CaCl_2$ gelada, aliqotado, imediatamente congelado em banho de gelo seco/etanol e armazenado a -80°C.

Para transformação, 0,2 μ g de DNA (plasmídeos ou produtos de reações de ligação de DNA) foram adicionados a 100 μ L de bactérias competentes e a mistura foi mantida em banho de gelo por 30 min. As células foram submetidas a choque térmico

a 42°C por 90 seg e imediatamente incubadas em banho de gelo por 5 min. Em seguida foram adicionados 500 µL de meio LB e a suspensão bacteriana foi transferida para tubo de cultura, seguindo-se incubação a 37°C por 1 h a 200rpm. Após esta etapa de recuperação alíquotas de 50 µL, 100 µL e 350 µL foram semeadas em placas de meio LB contendo carbenicilina 100 µg/mL. As placas foram incubadas a 37°C por uma noite. Algumas das colônias obtidas (geralmente 6 a 10 colônias) foram inoculadas individualmente em 3 mL de meio LB contendo os antibióticos necessários. As culturas líquidas foram incubadas a 37°C por uma noite a 200 rpm e utilizadas para preparação de plasmídeos ou como pré-inóculo em volumes maiores de cultivo.

3.12.3. Preparação de plasmídeos e digestão com enzimas de restrição

As preparações de plasmídeos foram realizadas utilizando-se o kit *Pure Link HiPure Plasmid DNA purification* (Invitrogen). Se necessário a quantidade de DNA foi estimada pela medida da absorbância a 260nm, onde $Abs_{260nm} = 1$ equivale a 50µg de DNA fita dupla/mL.

As preparações plasmidiais foram submetidas a reações de digestão com enzimas de restrição apropriadas (1U/µg DNA) a 37°C por 2 h, seguindo-se análise em eletroforese em gel de agarose 1%.

3.12.4. Indução da expressão de proteínas recombinantes

As construções pAE-*pilQC*, pAE-*pilQN* e pAE-*pilY1N* foram utilizadas na transformação de bactérias BL21(DE3) Gold e os clones selecionados foram inoculados em 3 mL de meio LB contendo carbenicilina 100 µg/mL e crescidos a 37°C, por uma noite a 200 rpm. Uma alíquota 500 µL desta cultura foi transferida para 10 mL de LB contendo antibiótico na mesma concentração, seguindo-se a incubação a 37°C e 200 rpm. Quando a cultura atingiu uma $DO_{600nm} = 0,6$ uma alíquota de 1mL foi retirada (cultura não induzida), o IPTG foi adicionado para concentração final de 1mM e alíquotas da suspensão bacteriana (cultura induzida) foram retiradas após 4 horas de incubação a 37°C. Amostras das culturas induzidas e não induzidas foram ressuspendidas, respectivamente, com 200 µL e 100 µL de tampão de amostra para análise por SDS-PAGE. O restante da cultura foi submetido à centrifugação 4.000 xg por 20 min a 4°C e armazenado a -80°C.

A determinação do conteúdo proteico nas amostras obtidas a partir da expressão em bactéria foi realizada segundo o método descrito por Smith, *et al.* (1985) (Smith e Krohn, 1985) tendo como padrão BSA (albumina de soro bovino).

3.12.5. Purificação das proteínas recombinantes

O sedimento correspondente a ~50 mL de culturas bacterianas induzidas com IPTG foi suspenso em 1,5 mL de tampão de lise (Tris-HCl 10 mM, pH 7,0, NaCl 100 mM, imidazol 10 mM, PMSF 1 mM e glicerol 10%), e a suspensão foi submetida a sonicação com 3 pulsos de 15 seg cada, posição 3 (Branson Sonifier 450) com intervalos de incubação de 1 minuto em gelo entre cada pulso. O lisado foi submetido à centrifugação a 9.880 xg por 30 min a 4°C. O sobrenadante foi recolhido e o sedimento suspenso em 1 mL do tampão de lise.

Aproximadamente 1 mL do sobrenadante (fração solúvel) foi então incubado com 100 µL da resina de Ni²⁺-NTA agarose (Qiagen) previamente equilibrada com 2,5 mL de tampão de lise a 4°C por 1 h em homogeneizador (*Heavy dutyrotator*, Cole Parmer) em baixa velocidade. Foram realizadas 2 lavagens com 4 mL de tampão de lise por centrifugação a 16.000 xg por 1 minuto. A proteína foi eluída com 150 µL de tampão de lise acrescido de 25, 50, 100, 150 e 300 mM de imidazol, seguindo-se incubação por 15 min em gelo, centrifugação nas condições anteriores e coleta do sobrenadante. Alíquotas das frações eluídas foram analisadas por eletroforese em SDS-PAGE e fração contendo a proteína recombinante pura foi armazenada a -80°C em glicerol 10%.

Após purificação as proteínas recombinantes foram separadas em SDS-PAGE e as bandas correspondentes (~25 kDa para pAE-PilY1N e ~20kDa para pAE-PilQC) foram recortadas do gel e homogeneizadas em água Milli-Q com um homogeneizador do tipo Potter-Elvehjem, antes da imunização dos coelhos.

A concentração das proteínas recombinantes purificadas foi estimada pelo método de Bradford tendo como padrão albumina de soro bovino (Bradford, 1976).

3.12.6. Produção de anticorpos policlonais

Para produção de anticorpos policlonais 1000 µg de PilY1N recombinante purificada e 500 µg de PilQC recombinante purificadas em um volume final de 800µL foi misturadas a 800 µL de adjuvante de Freund completo (Sigma) seguindo-se imunização de 2 coelhos machos (um coelho para cada proteína). Antes da imunização, um volume total de 10 mL de sangue foi retirado do coelho para a obtenção do soro pré-imune. Uma nova imunização foi repetida trinta dias após a primeira nas mesmas condições, exceto pelo uso do adjuvante de Freund incompleto (Sigma). Após dez dias, o sangue do coelho foi coletado para a obtenção do soro imune. Estes procedimentos foram realizados no Biotério de Produção e

Experimentação da Faculdade de Ciências Farmacêuticas e do Instituto de Química da USP, seguindo-se rigorosamente suas normas para experimentação em animais.

O anticorpo foi semipurificado de acordo com procedimento já estabelecido (Jordão *et al.*, 1996). O sangue coletado (pré-imune e imune) foi incubado por 1 h a 37°C, seguindo-se incubação a 4°C por uma noite. Esse material foi submetido à centrifugação a 3.000xg por 10 min a 4°C. Ao sobrenadante (soro) foi adicionada uma solução de sulfato de amônio saturada pH 6,8 numa proporção 1:1. Essa etapa foi seguida de incubação por uma noite a temperatura ambiente e centrifugação 5.000xg por 15 min a 4°C. O precipitado obtido foi suspenso em solução 50% saturada de sulfato de amônio pH 6,8 e submetido à centrifugação nas mesmas condições anteriores. O precipitado final foi suspenso em 2 mL de NaCl 0,1M e submetido à diálise por 20 horas contra 1.000 volumes de 0,1M de NaCl a 4°C com uma troca. Por fim, o material obtido foi distribuído em microtubos e mantido a -80°C.

3.13. Eletroforese unidimensional em gel de poliacrilamida contendo SDS (SDS-PAGE)

A eletroforese de proteínas foi realizada em gel de separação de 12% de poliacrilamida e gel de empilhamento de 4,0% (Laemmli, 1970) em placas (sistema mini-Protean®, BioRad). As alíquotas foram suspensas em tampão de amostra para eletroforese 1X [Tris 60mM, pH 6,8; β-mercaptoetanol 5% (v/v); glicerol 10% (v/v); SDS 2% (p/v) e azul de bromofenol 0,05% (p/v)], fervidas por 3 min e aplicadas no gel. A corrida foi realizada a 200V até que o azul de bromofenol atingisse o limite inferior do gel. As proteínas foram visualizadas por coloração com azul de Coomassie R 0,2% (p/v), preparado em metanol 50% (v/v) e ácido acético 10% (v/v), seguindo-se descoloração com ácido acético 7% (v/v) em metanol 30% (v/v).

3.14. Transferência eletroforética para membrana de nitrocelulose e detecção imunológica de proteínas

Após SDS-PAGE, os polipeptídeos foram transferidos para membrana de nitrocelulose (Bio-Rad) (Harlow, 1988). Após eletroforese o gel foi incubado por 10 min em tampão de transferência (Tris base 48 mM, glicina 39 mM, SDS 0,037% e metanol 20%) e transferido para uma montagem na seguinte ordem: 3 folhas de papel Whatman, membrana de nitrocelulose previamente tratada com tampão de transferência, gel de poliacrilamida e 3 folhas de papel Whatman. A transferência foi realizada por 1 h a 12 V e 0.3 A.

Após a transferência, a membrana de nitrocelulose foi incubada com PBS (Tris.HCl 10mM pH7,4 e NaCl 150mM) pH 7.4 por 10 min. Em seguida, a membrana foi incubada em solução de bloqueio (*Odyssey Blocking Buffer*, LI-COR Biosciences) por 1 h. A solução de bloqueio contendo anticorpo policlonal (1:250 para anti-PilQC e 1:500 para anti-PilY1N) foi adicionada, seguindo-se incubação por 12 horas à 5°C. A membrana de nitrocelulose foi lavada com PBS (Tris.HCl 10mM pH7,4 e NaCl 150mM) contendo Tween 20 0,1% por 4 vezes em períodos alternados de 5 min antes da adição do anticorpo secundário *IRDye 680RD Goat anti-Rabbit IgG* (1:1500), seguindo-se incubação por 1 h. Após 1 h de incubação a membrana de nitrocelulose foi lavada outra vez com PBS (Tris.HCl 10mM pH7,4 e NaCl 150mM) contendo Tween 20 0,1% 4 vezes em períodos alternados de 5 min. A captura da imagem foi feita no aparelho *Odyssey Infrared Imaging System* (LI-COR Biosciences) no comprimento de onda 700 nm e intensidade 3.5 com uma resolução de 169 µm.

3.15. PCR quantitativa precedida de transcrição reversa (RT-qPCR)

3.15.1. Purificação de RNA total

Para purificação de RNA total de *X. fastidiosa* foi utilizado o *Pure Link RNA Mini Kit* (Invitrogen) segundo o protocolo descrito pelo fabricante com algumas modificações. Os precipitados de células obtidos a partir de 1 mL das culturas foram suspensos com 1 mL do reagente Trizol e incubados à temperatura ambiente por 5 min, seguindo-se adição de 200 µL de cloroformio, com forte inversão do tubo por 15 seg, e incubação por mais 3 min a temperatura ambiente. As suspensões foram então centrifugadas a 12.000 xg por 15 min a 4°C. A fase aquosa foi recolhida em tubos limpos, adicionando-se em seguida 400 µL de etanol (70%). Após agitação vigorosa do tubo, as suspensões foram transferidas, 700 µL de cada vez, para filtros -coletores acoplados a tubos coletores, seguindo-se centrifugação a 12.000xg por 15 seg em temperatura ambiente. Os filtrados foram descartados e os filtros-coletores foram reinseridos nos mesmos tubos coletores. Esta etapa foi repetida até as amostras originais serem processadas por completo. Os filtros-coletores foram lavados com 700 µL de tampão de lavagem I por centrifugação a 12.000 xg por 15 seg em temperatura ambiente e, em seguida reinseridos em tubos limpos. Os filtros-coletores foram então lavados com 500 µL de tampão de lavagem II contendo etanol por centrifugação a 12.000 xg por 15 seg em temperatura ambiente. O filtrado foi descartado, seguindo-se nova lavagem com tampão de lavagem II e centrifugação por mais 1 min a 12.000 xg para secagem da membrana com o RNA adsorvido. Os filtros-coletores inseridos em

microtubos e o RNA foi eluído pela adição de 50 µL de água livre de RNase (água pré-tratada com DEPC (dietilpirocarbonato) 0,1 % (Ausubel *et al.*, 1995). Após a incubação em temperatura ambiente por 1 min os tubos foram submetidos a centrifugação por 2 min a 12.000 xg em temperatura ambiente e a solução contendo RNA total purificado foi transferida para novo tubo.

A quantificação da concentração de RNA foi feita através da determinação das absorbâncias (Abs) a 260 nm e 280 nm em espectrofotômetro ND-1000 (NanoDrop Technologies), sendo que $Abs_{260\text{ nm}} = 1$ corresponde a 40 µg/mL de RNA quando a razão $Abs_{260\text{ nm}}/Abs_{280\text{ nm}} \cong 1,9$. A integridade do RNA total foi avaliada no equipamento BioAnalyzer (Agilent), o qual permite a avaliação de amostras com concentrações de RNA entre 0,5–50 ng/µL. A matriz do gel aplicada nos capilares do *chip* utilizado no BioAnalyzer contém uma mistura de fluoróforos e marcadores de tamanho. A ligação dos fluoróforos ao marcador de tamanho e ao RNA resulta na emissão de fluorescência que é quantificada, permitindo a separação do rRNA 16S e 23S e a verificação da integridade do RNA total a partir da atribuição de um número de integridade do RNA (RIN) que varia de 1 a 10. As preparações de RNA utilizadas apresentaram RIN entre 7,0 e 8,2.

Finalmente, o RNA total foi tratado com RQ1 DNase (Promega) para a remoção de eventual DNA genômico contaminante. Para cada 4 µg de RNA foram adicionadas 4 U de RQ1 DNase, 1 µL de Tampão de DNase (10x), o volume de 13 µL completado com água DEPC, e a reação era incubada a 37 °C por 30 min. Após esse período a RQ1 DNase foi inativada utilizando 1 µL de *stop solution*, a quantificação foi feita através da determinação das absorbâncias a 260 nm e 280 nm em espectrofotômetro ND-1000 (NanoDrop Technologies) e as amostras foram armazenadas a -20 °C até que se procedessem os ensaios de RT-qPCR.

3.15.2. RT-qPCR precedida por transcrição reversa

Os ensaios de RT-qPCR para CDS selecionadas foram realizados com oligonucleotídeos (iniciadores) específicos planejados com o auxílio do programa PRIMER3 em conjunto com o software *Bioedit* (este último com o intuito de obter sequências de oligonucleotídeos adequadas a todas as cepas analisadas), visando a geração de *amplicons* com tamanho de aproximadamente 150 pb. As sequências dos pares de oligonucleotídeos utilizados estão apresentadas na Tabela 5.

A síntese do cDNA foi realizada utilizando-se 4 µg de RNA total, pré-tratado com 1 U de DNase RQ1 (Promega) para cada 1 µg de RNA. Para a síntese do cDNA, foi acrescentado 1 µL (aproximadamente 500 ng) de iniciadores randomicos, 1 µL da

mistura de dNTP a 10 mM e água DEPC para um volume final de 13 µL. Essa reação foi incubada a 65 °C durante 5 min e em seguida, transferida para o gelo por no mínimo 1 min. A este volume foram adicionados 4 µL de tampão 5x para a síntese de primeira fita do cDNA, 1 µL de DTT (Ditiotreitol) a 0,1 M, 1 µL de RNase OUT (20 U) e 1 µL (200 U) da enzima transcriptase reversa Superscript III (Invitrogen). A reação foi então incubada a 25 °C por 5 min e em seguida, a 50 °C por 60 min. A inativação da reação foi realizada através de incubação a 70 °C durante 15 min. Em seguida, o cDNA recém-sintetizado foi tratado com RNase H (*Fermentas Life Science*), diluído com água livre de nuclease para 125 µL e armazenado a -20 °C até o uso.

Tabela 5. Pares de iniciadores utilizados nas reações de RT-qPCR

CDS	Iniciadores diretos	Iniciadores reversos
XF0986	GTGGATTTGGTGTGAAAGG-3	CGCAGCGATCAGAGAAAGAT
XF0889	CTGGATCAAGACAAGGTC-3	AGGTCGTCTTGCGAGGCTG
XF1113	GAGGAAGAGGTGCGACTTA	CTCCTCGTTCAACGTGCTC
XF1540	TTTGTATGCGATAGGCATGC	ATGTAAGGTGCGCAAGATCC
XF0478	TCAATTCCTGTTGTCGCAAC	GAATTCTGGCGAGCATTGAT
XF0080	GTCACCTTCAACCTGCGAAG	GATGGCAAATCGTGTTTCAG
XF2466	GAATAATCTGCCAATGACGC	GAATAATCTGCCAATGACGC

Para cada reação de RT-qPCR, foram utilizados aproximadamente 15 ng de cDNA, 3 µL da mistura de iniciadores (direto e reverso) específicos para cada CDS, o que corresponde a uma concentração de 600 nM, 0,175 µL de mix contendo o fluoróforo (*SYBR Green* – Applied Biosystems). Para todos os lotes de reações foram incluídos controles negativos tais como reação com RNA total (sem ter passado pela síntese de cDNA) para assegurar a ausência de DNA genômico residual.

As reações de amplificação foram realizadas em um termociclador em tempo real (Applied Biosystems, modelo Gene Amp 5700 ou modelo 7300), utilizando-se o seguinte programa: 2 min a 50 °C, 2 min a 95 °C e 50 ciclos de 15 segundos a 95 °C e 30 s a 60 °C. O *software* do equipamento gera curvas de amplificação e dissociação dos iniciadores para cada amostra analisada e oferece o cálculo automático do CT (*Cycle Threshold*) que corresponde ao número de ciclos de reação realizados até que a amplificação seja detectada a partir da emissão de fluorescência. A quantidade de cDNA das amostras foi normalizada de acordo com a expressão do gene controle (*dnaQ*; XF2157) como descrito por (Zaini *et al.*, 2008). A relação entre a expressão do gene de interesse e a expressão do gene controle foi calculada pelo método $2^{-\Delta\Delta CT}$, conforme proposto por (Livak e Schmittgen, 2001). Os ensaios foram realizados em triplicata, a partir de duas réplicas biológicas.

4. Resultados e Discussão

4.1. Descrição das cepas de *X. fastidiosa* selecionadas para o sequenciamento do genoma

Neste trabalho selecionamos as cepas U24d, Fb7 e 3124 de *Xylella fastidiosa* para o sequenciamento genômico completo.

A cepa **U24d** é parte de uma coleção de 360 isolados obtidos a partir de laranjeiras doce com sintomas de CVC de distintas regiões no Estado de São Paulo (Neves Paulista, Gavião Peixoto, Santa Rita do Passa, Paulinea e Bauru) em diferentes épocas do ano de 2002 (Coletta-Filho e Machado, 2003). As plantas do pomar em que foi isolada a cepa U24d apresentavam sintomas severos de CVC (Figura 8). Estes isolados foram caracterizados através de marcadores moleculares RAPD (*Random Amplification of Polymorphic DNA*) e SSR (*Single Sequence Repeats*), sendo que os haplótipos *multilocus* característicos da cepa U24d foram encontrados em todas as regiões citrícolas de São Paulo (Coletta-Filho e Machado, 2003). Assim, a cepa U24d pode ser considerada um representante do haplótipo prevalente nos pomares paulistas na época em que foi isolada. Inoculações desta cepa em plantas de laranja doce mantidas em casa de vegetação resultaram em sintomas muito severos, muitas vezes superando os normalmente observados em plantas inoculadas com a 9a5c ou outras cepas (H. Della Colleta-Filho, comunicação pessoal).

A cepa **Fb7** foi isolada na região de Corrientes, na Argentina (da Silva *et al.*, 2007). A CVC apareceu na Argentina nos anos de 1980, onde foi chamada de *Pecosita* ou *falsa mancha grisienta*. Há relatos que essa doença seria contemporânea ao primeiro relato da CVC no Brasil (de Coll *et al.*, 2000). A sintomatologia de plantas com *Pecosita* assemelha-se a sintomatologia de plantas infectadas com cepas virulentas de citros no Brasil (Figura 9). Análises moleculares mostraram que a cepa Fb7 é mais relacionada com a cepa 9a5c do que outras obtidas de citros que atualmente tem presença abrangente no parque citrícola paulista. Estas diferenças são marcantes principalmente em relação aos elementos geneticamente móveis (da Silva *et al.*, 2007). É intrigante esta relação entre estirpes isoladas em épocas e regiões geograficamente distintas, porém de hospedeiro de mesma espécie.

Figura 8. Laranjeira doce com sintomas de CVC nível 2. A. Folhas com sintomas avançados de CVC apresentando clorose e necrose disseminada por toda planta. **B.** Frutos acometidos pela doença no estágio avançado. As imagens são representativas do pomar de onde foi isolada a cepa U24d. Fotos cedidas pelo Dr. Helvécio Della Coletta-Filho (Centro Apta Citros Sylvio Moreira, Cordeirópolis, SP).

Figura 9. Laranjeira doce com sintomas de *pecocita*. A. Planta com sintomas nível 2 da doença. A seta aponta galhos da planta com sintomas avançados de *pecocita*. B. Planta com nível 1 da doença. Os círculos evidenciam as regiões das folhas com necrose. As imagens são representativas do pomar de onde foi isolada a cepa Fb7. Fotos cedidas pela Dra. Norma Costa (*Estación Experimental Agropecuaria Concordia, Argentina*).

A cepa **3124** foi isolada de cafeeiro com sintomas de escaldadura foliar (CLS) em Matão, SP em 1997. Esta é uma doença amplamente generalizada em culturas de café por todas as áreas produtivas em São Paulo e outros Estados brasileiros, sendo que muitas vez sua ocorrência se dá adjacente a pomares de citros afetados com sintomas de CVC. A CLS foi relatada pela primeira vez em São José do Rio Preto, São Paulo, Brasil, em 1987. Todavia, a *X. fastidiosa* só foi detectada em cafeeiros através de testes sorológicos em 1995 (Chang *et al.*, 1993; Hartung *et al.*, 1994; Rosato *et al.*, 1998; da Costa *et al.*, 2000). Os sintomas da doença se caracterizam pela diminuição dos entrenós no caule, perda prematura de folhas velhas e escaldadura nas regiões apicais e marginais das folhas. Embora sua significância econômica atual não esteja

quantificada, há registro de que há mais de dez anos esta doença tenha sido responsável pela redução de 30% na produção de café em algumas plantações na região de São José do Rio Preto, SP (Li *et al.*, 2001).

Em ensaios de virulência e patogenicidade, a cepa 3124 demonstrou capacidade de colonização (Li *et al.*, 2001). Em média, seis meses depois da inoculação em cafeeiro, foram verificados os primeiros sintomas, que se agravaram após oito meses (Figura 10). Além disso, ensaios de infecção cruzada mostraram que a cepa 2105, isolada de citros, conseguiu infectar cafeeiro e apresentar sintomas três semanas após a inoculação, corroborando a hipótese de que regiões que apresentam as culturas de citros e café próximas têm o potencial de transmissão de *X. fastidiosa* de citros para cafeeiro (Li *et al.*, 2001). Entretanto, estudos posteriores mostraram que a cepa 3124, embora colonize plantas de café, não resultou em sintomas típicos de escaldadura da folha do cafeeiro após 11 meses da inoculação, possivelmente esses sintomas devam aparecer anos após a infecção no laboratório ou no campo (Almeida *et al.*, 2008). Este mesmo estudo também mostrou que a cepa 3124 não coloniza citros e que sua estrutura genética é distinta da cepa 9a5c segundo de estudos de MLST (Almeida *et al.*, 2008).

Figura 10. Cafeeiro como sintomas de escaldadura foliar. A. A planta à esquerda apresenta escaldadura da folha do cafeeiro, com redução dos entrenós no caule, perda prematura de folhas velhas, deformação das folhas e a presença da doença na região apical e marginal das folhas. À direita, planta controle saudável. **B.** À direita e à esquerda folhas com sintomas da escaldadura na região apical e marginal. No centro, folha controle sem apresentar a doença. Imagens retiradas de Li *et al.*, 2001.

4.2. Avaliação do fenótipo das cepas no cultivo *in vitro*

O cultivo das cepas U24d, Fb7 e 3124 foi estabelecido em laboratório, podendo ser realizado em meio PWG (Davis *et al.*, 1981). A figura 11 mostra a cultura das cepas 9a5c, U24d e Fb7 após 7 dias de crescimento em meio PWG. A cepa U24d apresenta um fenótipo muito semelhante à cepa 9a5c no cultivo em meio líquido, tanto em relação à formação de agregados mais compactos, como na formação de biofilme aderido à superfície do frasco de cultivo. Por outro lado, observamos o crescimento predominantemente planctônico da cepa Fb7 e ausência do biofilme aderido à superfície do vidro. A cepa 3124 apresenta um fenótipo menos agregativo em relação ao observado para cepas 9a5c e U24d, e mais próximo ao observado para a cepa J1a12, isolada de citros e não virulenta (dados não mostrados).

Figura 11. Cultivo de cepas de *X. fastidiosa*. A figura mostra culturas das cepas 9a5c (A), U24d (B) e Fb7 (C) em meio PWG após 7 dias de crescimento a 28°C com agitação de 100 rpm.

A morfologia das colônias formadas pelas cepas 9a5c, U24d, Fb7 e 3124 foi avaliada por microscopia ótica após cinco dias de crescimento em PWG. Como mostrado na figura 12 e sumarizado na tabela 6, as colônias destas cepas apresentam características morfológicas peculiares. Entre as diferenças observadas, ressaltamos a ausência da franja e de formação de colônias satélites na cultura da cepa U24d comparativamente a cultura da cepa 9a5c, o que é indicativo da ausência de motilidade (*twitching motility*) (Figura 12A, B e C).

A 3124 diverge muito das características das demais cepas analisadas (U24d, Fb7 e 9a5c). A figura 12D mostra que as colônias desta cepa não exibem definição de seu contorno e tão pouco apresentam a franja. Todavia, a cepa que mais nos

surpreendeu foi a Fb7 (Figura 12E), pois, não forma colônias definidas e a cultura apresenta um aspecto bastante viscoso (Tabela 6).

Figura 12. Morfologia das colônias de *X. fastidiosa*. As colônias das cepas indicadas após cinco dias de crescimento em PWG foram observadas em microscópio óptico invertido com 10X de aumento. **A.** Colônia da cepa 9a5c. O retângulo preto destaca a presença de colônias satélites ao redor de uma colônia. **B.** Colônias da cepa 9a5c. A seta mostra a franja de deslocamento das células da colônia, indicativo de motilidade *twitching*. **C.** Colônia da cepa U24d. Neste caso não foi observado presença de colônias satélites à colônia principal nem a presença da franja de deslocamento celular nas margens desta colônia. **D.** Colônias da cepa 3124. O retângulo branco destaca o aglomerado de colônias sem simetria. **E.** Estrutura formada pela cepa Fb7. A seta mostra a presença da franja de deslocamento.

A variação na morfologia de colônias de diferentes cepas de *X. fastidiosa* tem sido associada a distintos genótipos. Por exemplo, enquanto cepas isoladas de videiras formam colônias bem definidas e rugosas, certos isolados de amendoeira exibem colônias com aspecto menos rugoso (Chen *et al.*, 2007; Galvani *et al.*, 2007). Por outro lado colônias de um isolado de ameixeira (Pr8x) apresentaram colônias definidas com textura aspera, um isolado de hibisco (Hib4) apresentou colônias irregulares com aspecto viscoso (Pierry, 2012). Além disso, já foi verificado que a formação da “franja” também parece depender das condições de cultivo, sendo que a presença de albumina no meio PWG afeta dramaticamente seu aparecimento (Chen *et al.*, 2007; Galvani *et al.*, 2007).

Tabela 6. Características morfológicas das colônias das cepas de *X. fastidiosa*

Cepas	U24d	3124	Fb7	9a5c
Descrição				
Colônias	Definidas	Definidas	Indefinidas	Definidas
Forma	Circular	Irregular	-	Circular
Margem	Ondulada	Indefinida	-	Ondulada
Aparência da superfície	Rugosa	Rugosa	-	Rugosa
Textura	Áspera/seca	Pouco viscosa	Viscosa	Áspera/seca
Elevação	Convexo	Ausente	Ausente	Convexo
“Fanja”	Não	Não	Sim	Sim

As cepas 3124, Fb7 e U24d foram também avaliadas com os pares de *primers* XF0077 R/F e XF1250 R/F. Esses *primers* amplificam as CDS da adesina de fimbrias (XF0077) e da arginina deaminase (XF1250) e têm sido utilizados rotineiramente em nosso laboratório para genotipagem e diferenciação das cepas 9a5c, J1a12, Pr8x e Hib4, respectivamente cepas de virulenta e avirulenta isolada de citros e cepas isoladas de ameixeira e hibisco. A PCR com a combinação destes dois pares de *primers* gera, respectivamente, produtos de 1599 pb e de 1122 pb para a cepa 9a5c, enquanto que para a cepa J1a12 é gerado um padrão de bandas distinto (549 pb para os *primers* referentes a CDS XF0077 e 159 pb para os primers da CDS XF1250), em razão de deleções previamente identificadas nesta região de seu genoma (Koide *et al.*, 2004). A cepa Pr8x exibe produtos de ~3000 pb (primers XF0077 R/F) e de 159 pb

(primers XF1250 R/F) enquanto que a cepa Hib4 apresenta um padrão de bandas idêntico ao observado para a cepa J1a12 (Pierry, 2012).

A figura 13 mostra que os tamanhos dos *amplicons* gerados para cepas U24d e Fb7 correspondem aos tamanhos dos *amplicons* observados para a cepa 9a5c e, portanto não carregam deleções nas CDS XF0077 e XF1250. A cepa 3124 mostra um perfil distinto tanto das cepas U24d e Fb7 como também das demais cepas já analisadas, sugerindo que uma provável deleção na CDS XF0077. A provável ausência dessa adesina deve contribuir para o fenótipo menos agregativo da cepa 3124, como verificado para a cepa J1a12 (Koide *et al.*, 2004; Pierry, 2012).

Assim como verificado para as cepas J1a12, Hib4 e Pr8x (Pierry, 2012) também observamos que o par de primers CVC-1/272-2-int (Pooler e Hartung, 1995), tradicionalmente utilizado para diferenciar cepas de citros das cepas de videiras, resulta na amplificação de um produto de 500pb também para as cepas Fb7, U24d e 3124 (dados não mostrados). Desta forma, confirmamos que este par de *primers* aparentemente, distingue cepas sul-americanas e não apenas cepas isoladas de citros.

Figura 13. Genotipagem das cepas U24d, Fb7 e 3124 de *X. fastidiosa*.

DNA total isolado destas cepas foi submetido a PCR com os pares de primers XF0077 R/F e XF1250 R/Fs e os produtos de amplificação foram separados em gel de agarose 1% corado com brometo de etídio. As setas apontam, de cima para baixo, os *amplicons* de 1599 bp, 1122 pb e 549 pb.

4.4. Pirosequenciamento, montagem e anotação do genoma das cepas U24d, Fb7 e 3124

Para o sequenciamento do genoma das cepas U24d e Fb7 utilizamos a estratégia de *shotgun* (Margulies *et al.*, 2005; Ahmadian *et al.*, 2006) e a metodologia de *Paired End Tags* (PET) (Fullwood *et al.*, 2009) na plataforma 454/Roche. Para o sequenciamento da cepa 3124, utilizamos apenas a estratégia de *shotgun*.

Um pré-requisito para o bom resultado do pirosequenciamento é que a qualidade e a quantidade de DNA sejam seguramente determinadas. A tabela 7 mostra as razões das absorbâncias medidas a 230nm, 260nm e 280nm de preparações de DNA total extraído a partir de 25 mL de cultura das 3 cepas após 7 dias de cultivo em meio PWG. A integridade do DNA total foi verificada por eletroforese em gel de agarose 1% (Figura 14). Estas análises indicaram que o DNA purificado das três cepas atendia os requisitos para o pirosequenciamento.

Tabela 7. Análise do DNA genômico purificado das cepas de *X. fastidiosa*

Cepa	Razão A260nm/280nm	Razão A260nm/230nm	Concentração de DNA genômico em ng/mL
U24d	1,80	1,81	185
3124	1,93	1,84	200
Fb7	1,85	1,86	190

Para o sequenciamento *shotgun* foram utilizados ~500ng de DNA total isolado das três cepas, o qual foi rompido mecanicamente por nebulização em fragmentos de 500 a 800 pb. No caso dos genomas de U24d e Fb7 foram realizadas corridas independentes, sendo cada genoma sequenciado em metade da placa. A outra metade foi utilizada no sequenciamento de amostras não relacionadas. Para o genoma da cepa 3124 foi realizada uma corrida independente utilizando-se a totalidade da placa (PTP). No protocolo *Paired End Tags* foram utilizados ~15µg de DNA total rompido mecanicamente em fragmentos ~3 kpb, sendo realizada uma corrida e cada genoma (U24d e Fb7) carregado em ¼ da PTP. Infelizmente não foi possível realizar a estratégia PET para a cepa 3124. As etapas de ambos os protocolos de pirosequenciamento estão descritas no item 3.9.

O resumo do desempenho do pirosequenciamento desses 3 genomas está apresentado na tabela 8. As variações no desempenho refletem diferenças na qualidade e quantidade do DNA total e/ou no protocolo de pirosequenciamento, visto que os sequenciamentos *shotgun* foram realizados em épocas distintas no CATG do

IQUSP. É importante relatar que duas corridas de sequenciamento *shotgun* dessas cepas apresentaram resultados insatisfatórios devido a problemas técnicos apresentados pelo equipamento 454/Roche instalado no CATG do IQUSP, e foram, portanto, descartadas (dados não mostrados).

Figura 14. Análise eletroforética do DNA genômico das cepas U24d, 3124 e Fb7 de *X. fastidiosa*. A foto ilustra a integridade das amostras de DNA total utilizadas para o pirosequenciamento, separadas em gel de agarose 1%. Os marcadores de tamanho (DNA 1Kbp Ladder) estão na canaleta M.

Tabela 8. Desempenho do pirosequenciamento dos genomas das cepas U24d, Fb7 e 3124.

Cepa	Total de sequências <i>shotgun</i>	Tamanho médio das sequências <i>shotgun</i>	Bases sequenciadas (<i>shotgun</i>)	Bases sequenciadas (PET) ^a	Cobertura estimada ^b
U24d	523.884	398	$\sim 20 \times 10^7$	$19,1 \times 10^6$	$\sim 81 \times$
Fb7	346.699	245	$\sim 8,5 \times 10^7$	$60,6 \times 10^6$	$\sim 25 \times$
3124	1.651.875	438	$\sim 72,3 \times 10^7$	-----	$\sim 267 \times$

^aBibliotecas para estratégia PET foram construídas com fragmentos de 3 Kbp

^bRelativa ao genoma da cepa 9a5c ($\sim 2,7 \times 10^6$ pb)

As sequências obtidas com o pirosequenciamento *Paired Ends Tags* foram reunidas com as sequências obtidas no pirosequenciamento por *shotgun* e submetidas aos programas *Newbler* (454 Life Sciences/Roche), *MIRA* (http://chevreux.org/projects_mira.html) e *Phrap* (<http://www.phrap.org/phredphrap/general.html>), sendo que os melhores resultados foram obtidos com o programa *Newbler* (J. P. Kitajima comunicação pessoal). A montagem completa *ab initio* desses três genomas foi possível graças a cuidadosas intervenções manuais realizadas pelo Dr. J.P. Kitajima, gerando a primeira versão destes genomas com cromossomos principais de 2.681.334 pb, 2.733.974 pb e 2.748.594 pb, respectivamente para as cepas U24d, Fb7 e 3124 (Tabela 9). A montagem das sequências da cepa U24d revelou também a existência de um plasmídeo de 51.156 pb, enquanto que as cepas Fb7 e 3124, aparentemente não possuem plasmídeos.

As sequências completas desses genomas foram submetidas para anotação automática na plataforma IMG/ER (*Integrated Microbial Genomes-Expert Review*, <http://img.jgi.doe.gov/>). As características gerais desses genomas estão resumidas na tabela 9, que inclui também os genomas das cepas Temecula1 e 9a5c como referências. Vale destacar que os genomas dessas duas cepas foram submetidos a reanotação na plataforma IMG/ER, para uniformizar o critério de reconhecimento e anotação de CDS, para efeito das análises comparativas descritas neste trabalho. A reanotação apresentou algumas discrepâncias com a anotação original e publicamente disponível dos genomas de Temecula1 e 9a5c, principalmente quanto ao número de genes codificadores de proteínas (Tabela 9) e quanto à definição dos inícios de tradução, frequentemente mais à montante nas anotações originais (dados não mostrados).

Tabela 9. Características gerais dos genomas das cepas Temecula1, 9a5c, U24d, Fb7 e 3124

Características	Temecula1	9a5c	U24d	Fb7	3124
Tamanho do genoma (pb)	2.519.802	2.679.305	2.681.334	2.733.974	2.748.594
Conteúdo de G+C	51.78%	52.67%	52.68%	52.63%	52.63%
Genes codificadores de proteínas	2.417 (2041) ^a	2.578 (2.769) ^a	2.618	2.715	2.669
Genes codificadores de proteínas com função predita	16.24 (65.59%)	1.750 (66.31%)	1.764 (67%)	1.784 (66%)	1.786 (65.45%)
Genes codificadores de proteínas sem predição de função	793 (32.0%)	828 (31.3%)	854 (33%)	931 (34%)	883 (32.3%)
Operons de rRNA	2	2	2	2	2
rRNA 5S	2	2	2	2	2
rRNA 16S	2	2	2	2	2
rRNA 23S	2	2	2	2	2
Genes de tRNA	49	50	50	50	50
Plasmídios	1 1346 pb	2 51.158pb 1.285 pb	1 51.156 pb	-	-

^aNúmero de genes na anotação original publicamente disponível

4.5. Análises comparativas dos genomas das cepas U24d, Fb7 e 3124 com genomas das cepas 9a5c e Temecula1

Como apresentado na tabela 9, a média do conteúdo G+C dos genomas das cepas que sequenciamos está mais próxima ao da cepa 9a5c (~52%). O tamanho do cromossomo principal, o número predito de genes codificadores de proteínas e a média de genes sem predição de função diferem ligeiramente, sendo que entre esses cinco genomas, o menor genoma é o da cepa Temecula1 e o maior é o da cepa 3124.

Os genomas que sequenciamos apresentaram regiões de desvio de conteúdo GC (*GC skew*) esperadas para a espécie *X. fastidiosa* (Figura 15), os quais apresentam amplitudes superiores quando comparados às espécies de integrantes do gênero *Xanthomonas* (Pieretti *et al.*, 2009). Entretanto, a cepa Fb7 mostra alteração deste padrão de *GC skew*, provavelmente devido a eventos de rearranjo genômico relacionados a elementos genéticos móveis, como regiões de profagos e ilhas genômicas, como será melhor discutido mais adiante.

Figura 15. Representação circular dos genomas de *X. fastidiosa*. Os atlas dos genomas foram gerados com o GenomaViz. De fora para dentro, círculo verde CDSs da fita *forward*; círculo azul CDSs da fita *reverse*; Média de GC% em verde; GC Skew em vermelho para baixo e azul para cima.

Como já mencionado, dentre as cepas que sequenciamos, somente a U24d apresentou um plasmídeo de 51.156 pb. Esse plasmídeo é idêntico e colinear ao pXF51 da cepa 9a5c (Marques *et al.*, 2001), como esquematizado na Figura 16.

Figura 16. Alinhamento dos plasmídeos pXF51 das cepas 9a5c e U24d. O diagrama de *Dotplot* foi gerado a partir da comparação das sequências de nucleotídeos dos plasmídeos pXF51 das cepas 9a5c e U24d com o programa *MUMmer* disponível na plataforma IMG/ER. A colinearidade é perfeita como indica a reta vermelha.

A comparação entre as sequências de nucleotídeos do cromossomo principal das cepas U24d, Fb7 e 3124 com os genomas de referência, cepas 9a5c (sul-americana) e Temecula1 (norte-americana) foram feitas usando os programas *MUMmer* (Delcher *et al.*, 2003), disponível na plataforma IMG/ER, e *Mauve* (Darling *et al.*, 2010). Como mostra a figura 17, o genoma da cepa U24d é praticamente colinear com a cepa 9a5c. Todavia, o alinhamento com a cepa Temecula1, mostrou 14 regiões principais de quebra, indicando a ocorrência de rearranjos genômicos. Interessantemente, o genoma da cepa Fb7 mostra uma maior colinearidade com a cepa norte-americana (Temecula1) do que com a cepa sul-americana (9a5c), sendo 9 regiões principais de quebra no primeiro caso e 12 regiões principais de quebra no segundo caso (Figuras 17 e 18). O mesmo foi observado para o genoma da cepa 3124, que também apresentou maior colinearidade em relação à cepa Temecula1 do que com a cepa 9a5c, com 13 e 19 regiões de quebra, respectivamente (Figuras 17 e 18). Curiosamente, todos os cinco genomas apresentam um *cluster* de sintenia, tanto da posição 1 a 273.834 pb, quanto da posição 2.618.073 pb até o final dos genomas analisados (Figura 18). Essas regiões estão próximas à origem de replicação e

abrigam genes essenciais à sobrevivência da *X. fastidiosa*, como genes relacionados com o metabolismo de carbono, aminoácidos, lipídios, ácido nucléico, replicação, transcrição, mecanismos de transdução de sinal e defesa.

Figura 17. Alinhamento das seqüências de nucleotídeos dos cromossomos principais das cepas 3124, U24d e Fb7 com as cepas Temecula1 e 9a5c. Os diagramas de “Dotplot” foram gerados a partir da comparação das seqüências de nucleotídeos dos genomas das cepas indicadas com o programa *MUMmer* disponível na plataforma IMG/ER. As regiões de similaridade encontradas na mesma orientação estão indicadas em azul, exceto a U24d que está indicada em vermelho, e em vermelho, estão indicadas regiões de similaridade em orientação oposta (inversão). Os alinhamentos à esquerda representam a sintenia das cepas U24d, Fb7 e 3124 em relação à cepa Sul-Americana, 9a5c e os alinhamentos à direita representam a sintenia destas cepas em relação à cepa Norte-Americana, Temecula1.

Figura 18. Alinhamento das seqüências de nucleotídeos dos genomas de *X. fastidiosa*. Os blocos em diferentes cores são as regiões de identidades entre os genomas. Os blocos invertidos representam as inversões existentes nos genomas. Alinhamento foi realizado no programa Mauve 2.0.

Os genomas anotados das cepas U24d, Fb7 e 3124 e os genomas reanotados das cepas 9a5c e Temecula1 foram comparados para identificação de genes ortólogos e exclusivos. Estas análises foram realizadas através da plataforma IMG/ER, com parâmetros de mínimo de 80% de identidade e E-value $10e^{-10}$, associando-se análises adicionais utilizando BLASTp. Como mostrado na tabela 10, o número de genes ortólogos é semelhante entre as cinco cepas comparadas (1555 a 1608 genes codificadores de proteínas). Em contra partida, o número de genes exclusivos variou consideravelmente entre as cepas. Alguns desses genes exclusivos codificam toxinas, integrases de profagos, endonucleases e recombinases sítio-específicas. O menor número de genes exclusivos para as cepas 9a5c e U24d reflete o fato de que essas cepas têm genomas muito semelhantes entre si.

Tabela 10. Número de ortólogos e genes exclusivos nas cinco cepas de *X. fastidiosa*

Cepa	Ortólogos com função predita	Ortólogos sem predição de função ^a	Genes exclusivos função predita	Genes exclusivos sem predição de função ^a
9a5c	1.582	317	0	15
3124	1.608	318	33	84
U24d	1.601	313	0	25
Fb7	1.602	376	32	83
Temecula1	1.555	326	27	262

^a Sequências codificantes sem *hits* para qualquer categoria COG.

Entre os genes exclusivos detectados na comparação destes cinco genomas, destacamos um que codifica uma metaloprotease dependente de zinco (XFFB7_01080) no genoma da cepa Fb7 (Figura 19). Análises adicionais revelaram que um ortólogo deste gene (100% idêntico) está presente em duas cópias no genoma da cepa Hib4 (isolada de hibisco). Em algumas espécies de *Xanthomonas*, uma enzima desta família está relacionada à degradação da matriz extracelular associada à glicoproteína rica em prolina do pecíolo (Dow *et al.*, 1998), a qual tem papel na proteção contra o ataque microbiano (Bowles, 1990). Em *Xanthomonas campestris* pv. *campestris* sugere-se que uma zinco metaloprotease pode ter importância no desenvolvimento da doença da podridão negra (Dow *et al.*, 1998). A parede celular primária do xilema é constituída de matriz extracelular associada à glicoproteína rica em prolina (Sherrier e Vandenbosch, 1994). Desta forma, é razoável especular que esta zinco metaloprotease, se secretada, esteja relacionada à degradação da parede celular no xilema, facilitando assim, a colonização na planta hospedeira pela cepa Fb7.

A análise completa das alterações (polimorfismos, inserções e deleções) nos genes codificadores de proteínas preditos nos genomas das cepas U24d, Fb7 e 3124 ainda está em andamento. A seguir detalharemos uma alteração que detectamos no genoma da cepa U24d e outra detectada em Fb7. A primeira corresponde a inserção de nucleotídeos no gene que codifica a proteína PilQ na cepa U24d. A segunda corresponde a uma inserção de um nucleotídeo no gene que codifica a proteína RpfC na cepa Fb7. Em ambos os casos as alterações supostamente resultam em perda de função destas proteínas, com reflexo no fenótipo destas cepas.


```

MP_Fb7      -----MNYSVIPPYIFSKIIDTGSSDQSRARSTLAHVQMLMTETKT----- 42
ZMP1_HIB4  -----MNYSVIPPYIFSKIIDTGSSDQSRARSTLAHVQMLMTETKT----- 42
ZMP2_HIB4  -----MNYSVIPPYIFSKIIDTGSSDQSRARSTLAHVQMLMTETKT----- 42
ZMP_XPE     ---MSAFVSARRAGILPPYLLEHVAQAAPAHARQCALLSRQITACLRLQQRALGLLARAD 56
ZMP_XCC     MEIAMSFAFVAPRRAGILPPYLLDHVAQAAPERARHCAQLTRHITAQLRQRRAGQLLAR-- 58
              . . : : * * : : : : . . . * * : . * .

ZMP_Fb7      --KPEGLHITTTGKQLREIYDAQQTQNLPGTLVRMEKSREND DVTVNEAYDYLGVTYDFY 100
ZMP1_HIB4    --KPEGLHITTTGKQLREIYDAQQTQNLPGTLVRMEKSREND DVTVNEAYDYLGVTYDFY 100
ZMP2_HIB4    --KPEGLHITTTGKQLREIYDAQQTQNLPGTLVRMEKSREND DVTVNEAYDYLGVTYDFY 100
ZMP_XPE      APAADAKAQQTAPAVQRRRIYDAQQTALPGTLVRDEGAAATQDVAVTEAYDYLGATHDFF 116
ZMP_XCC      DSADDAAPTATTD TAVRRHLYDARQGTALPGVLVREEGAPPTDDVAVTEAYDYLGATHAFF 118
              : . * : : * : : * : * : * : : : : * : : * : : * : : * : : * : :

ZMP_Fb7      KEAYQRNSLDNKGLPLKATVHYGKDYQNAFWNGHQMVFGDGDGEI FNRFTSAIDVVAHEL 160
ZMP1_HIB4    KEAYQRNSLDNKGLPLKATVHYGKDYQNAFWNGHQMVFGDGDGEI FNRFTSAIDVVAHEL 160
ZMP2_HIB4    KEAYQRNSLDNKGLPLKATVHYGKDYQNAFWNGHQMVFGDGDGEI FNRFTSAIDVVAHEL 160
ZMP_XPE      QTVYGRNSIDAAGMPLIGSVHYERGYDNAFWNGEQMVFGDGDGEV FNRFTIALD VVGHEL 176
ZMP_XCC      QQVYARNSIDDAGMPLLGTVHYERNYDNAFWTGEQMVFGDGDGEI FTRFTIALD VVAHEL 178
              : . * * : * * : : * : : * : : * : * : * : * : * : * : * : * : * : * : * :

ZMP_Fb7      SHGMTSEADLIYFQQAGALNESISDVFGSLVKQFNKKQTADQADWI IGEGLLMKGINGK 220
ZMP1_HIB4    SHGMTSEADLIYFQQAGALNESISDVFGSLVKQFNKKQTADQADWI IGEGLLMKGINGK 220
ZMP2_HIB4    SHGMTSEADLIYFQQAGALNESISDVFGSLVKQFNKKQTADQADWI IGEGLLMKGINGK 220
ZMP_XPE      THGVTERTANLIYQQSGALNESISDVFGVLKQYTLRQSADQADWI IAGLLMPGIKGV 236
ZMP_XCC      THGVIER TANLIYQQSGALNESVSDVFGVLKQYALRQDAAQADWL VGAGMFLPGVQGV 238
              : * : * * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

ZMP_Fb7      GIRSMKPGSAYDDPLL GKDPQPAHMNDYVQTRDDNGGVHLNSGIPNHAFYLA AIALGGY 280
ZMP1_HIB4    GIRSMKPGSAYDDPLL GKDPQPAHMNDYVQTRDDNGGVHLNSGIPNHAFYLA AIALGGY 280
ZMP2_HIB4    GIRSMKPGSAYDDPLL GKDPQPAHMNDYVQTRDDNGGVHLNSGIPNHAFYLA AIALGGY 280
ZMP_XPE      GLRSMQAPGSAYDDPALGKDPQPATMAGYVDTQEDDGGVHYNSGIPNHAFYRAA VAIGGA 296
ZMP_XCC      ALRSMQAPGTAYDDPALGKDPQPAHMDAYVDTQEDDGGVHYNSGIPNRAFQRAA VAIGGY 298
              : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

ZMP_Fb7      AWKIAGYAWYDTVCDKHL PQDADFKTFAEFTIKHGKTRFNKSVG--HTIETAWQQVGVLP 338
ZMP1_HIB4    AWKIAGYAWYDTVCDKHL PQDADFKTFAEFTIKHGKTRFNKSVG--HTIETAWQQVGVLP 338
ZMP2_HIB4    AWKIAGYAWYDTVCDKHL PQDADFKTFAEFTIKHGKTRFNKSVG--HTIETAWQQVGVLP 338
ZMP_XPE      AWEKTGRIWYRALTG GELAGADFATFAALTVEVASAEYGANSSETAAVQQAWRDVGVLA 356
ZMP_XCC      AWEKAGRIWYRALTG GALSASADFATFAALTVRVASTDYGAGSAEASAVEQAWRDVGVLA 358
              * : : * * : : . * . * * * * : * : . . : : . . . : : * : * : * : * :

```

Figura 19. Comparação da sequência de aminoácidos da metaloprotease dependente de zinco. Alinhamento da sequência de aminoácidos da sequência de aminoácidos da metaloprotease dependente de zinco (ZMP) predita nos genomas das cepas Fb7 e Hib4 com os ortólogos em *Xanthomonas perforans* (XPE) e *Xanthomonas campestris* pv *campestris* (XCC).

4.6. Análise de PilQ na cepa U24d

Embora os genomas das cepas U24d e 9a5c sejam 99% idênticos e colineares (Figuras 17 e 18), detectamos que o gene que codifica a proteína PilQ (CDS XF0373 em 9a5c) apresenta uma inserção de 10 nucleotídeos na cepa U24d (Figura 20 e Figura S1). Esta alteração foi confirmada por PCR com um par de *primers*, em que um dos oligonucleotídeos alinha na inserção (Tabela 3, Figura S1) originando um *amplicon* de 347bp (Figura 20). Essa alteração foi também confirmada por sequenciamento desta região do gene *pilQ* da cepa U24d pelo método de Sanger (dados não mostrado). Essa inserção possibilitou também a caracterização de uma marca genética para genotipagem e diferenciação desta cepa das demais cepas existentes no laboratório.

A inserção de 10 nucleotídeos em *pilQ* no genoma de U24d resulta em uma alteração da fase de leitura e aparecimento de um códon de terminação prematuro (Figura S1). A proteína truncada não possui o domínio *secretin* (Figura 21), o qual é responsável para que PilQ secrete as pilinas que vão formar a fibra do *pilus* tipo IV (Mattick, 2002).

Figura 20. Inserção de 10 nucleotídeos no gene *pilQ* da cepa U24d. O trecho de alinhamento mostrado foi extraído do alinhamento completo das sequências de *pilQ* de 9a5c e U24d apresentado na figura suplementar S1 (Apêndice). A inserção foi confirmada por PCR com o par de *primers* PilQ-InsF e PilQ-InsR (Tabela 3) com a amplificação do segmento de DNA 347bp apenas para o DNA genômico de U24d. A foto mostra o gel de agarose 1% contendo brometo de etídeo. Os marcadores de tamanho em Kpb estão indicados à esquerda e o *amplicon* à direita.

A.
B.

PilQ_9a5c	MIDLDSMMTLFEVLGLRFFRRIASVRVCAIWVCMFAWCCIVGTSAENKQAIIPVPKVA	60
PilQ_U24d	MIDLDSMMTLFEVLGLRFFRRIASVRVCAIWVCMFAWCCIVGTSAENKQAIIPVPKVA	60
PilQ_9a5c	NAPLSVSKIDFKRGDDGSGRLILKFDGQGATPDLRTQGGTVLVDLGTAVLPTVLQRQINV	120
PilQ_U24d	NAPLSVSKIDFKRGDDGSGRLILKFDGQGATPDLRTQGGTVLVDLGTAVLPTVLQRQINV	120
PilQ_9a5c	VDFATPVQRIDAKPMKGGAQLVLSTKGAFESLAYQTGDEYVVEIVPRKGEAAMGGAITPE	180
PilQ_U24d	VDFATPVQRIDAKPMKGGAQLVLSTKGAFESLAYQTGDEYVVEIVPRKGEAAMGGAITPE	180
PilQ_9a5c	SVKSAAKISAQGYSGRPVTFNFQDVPVRTVQLIADESNIIVASDTVQGNVTLRLINV	240
PilQ_U24d	SVKSAAKISAQGYSGRPVTFNFQDVPVRTVQLIADESNIIVASDTVQGNVTLRLINV	240
PilQ_9a5c	PWDQALDIVLRAKGLDKRRDEKVIWVAPQQLAKYEQEKEDARIAIENREGLITDYVQIN	300
PilQ_U24d	PWDQALDIVLRAKGLDKRRDEKVIWVAPQQLAKYEQEKEDARIAIENREGLITDYVQIN	300
PilQ_9a5c	YHSATVIFKALTEAKMGCGGCTPNNNSDDAFLSPRGRLVADERTNTLMISDIPKKVAC	360
PilQ_U24d	YHSATVIFKALTEAKMGCGGCTPNNNSDDAFLSPRGRLVADERTNTLMISDIPKKVAC	360
PilQ_9a5c	MRTLIQHIDRPVDQVLIIEGRIVATDSFARDLGAKFGVGAASRFSNTATIGSHVTADS	420
PilQ_U24d	MRTLIQHIDRPVDQVLIIEGRIVATDSFARDLGAKFGVGAASRFSNTATIGSHV----	415
PilQ_9a5c	STSATRGLNVDLGGPTNTTASVLPSTAYTLLGPKFNLDLELSALQQESRGEVISNPRIVT	480
PilQ_U24d	-----	-----
PilQ_9a5c	ANQREGYIKQGKEIGYVTITGGVAGGQATPNVQFKEAVLELRVTPTITNDNRVFLNMTVK	540
PilQ_U24d	-----	-----
PilQ_9a5c	KDEIDQMITIANFGTVPTLNKREINTAVLVDDGQTVVIGGVYEFSDRNSVSKVPFLGDVP	600
PilQ_U24d	-----	-----
PilQ_9a5c	FLGNLFKKRGKDKQKAELLIFVTPKVLRVANATAAVH	637
PilQ_U24d	-----	-----

Figura 21. Comparação da sequência de aminoácidos proteínas *pilQ* de U24d e 9a5c. A. Esquema de domínios de PilQ da cepa 9a5c pelo PFAM, indicando em verde o domínio N- terminal de secretinas, em vermelho o domínio pequeno também de secreção e em azul o domínio secretina. **B.** Alinhamento da sequência de aminoácidos de PilQ de 9a5c e da forma truncada predita na cepa U24d e que não contém o domínio secretina.

O nocaute de *pilQ* na cepa Temecula1 de *X. fastidiosa*, resulta na ausência de *pilus* tipo IV e de motilidade *twitching*. Em videiras infectadas com essa cepa mutante, não foi observado colonização basipetal ao longo do xilema. Assim, a cepa nocaute para o gene *pilQ* perde a capacidade de colonizar e consequentemente tem sua virulência afetada (Meng *et al.*, 2005).

A mutação em *pilQ* detectada no genoma da cepa U24d parece explicar a ausência da “franja” em sua colônias, como mostrado na Figura 12C. Confirmamos através de microscopia eletrônica de varredura a completa ausência de *pilus* do tipo IV (*pilus* longos) na cepa U24D (Figura 22). Desta forma, concluímos que a cepa U24d não é capaz de realizar motilidade *twitching*, sendo, portanto um mutante natural deste fenótipo.

Contudo, ainda resta a questão se a proteína *pilQ* truncada é expressa na cepa U24d e qual seria o destino da proteína PilY1 (adesina localizada na extremidade das fibras do *pilus* tipo IV). Para abordar estas questões planejamos a obtenção de anticorpos policlonais contra estas proteínas como será descrito a seguir.

Figura 22. Microscopia eletrônica de varredura das cepas 9a5c e U24d. A. Células da cepa 9a5c com *pilus* longos (*pilus* do tipo IV) e *pilus* curtos em um de seus pólos. B. Células da cepa U24d apresentando apenas o *pilus* curto em um dos pólos. Barras de escala correspondem a 2,5 μm.

4.6.1. Obtenção de anticorpos policlonais anti-rPilQC e anti-rPilY1N e detecção de PilQ e PilY1 nas cepas U24d e 9a5c

Com intuito de produzir anticorpos policlonais dirigidos para PilQ e PilY1, planejamos a obtenção de proteínas recombinantes correspondentes à região N-terminal e C-terminal de PilQ e à região N-terminal de PilY1, ambas da cepa 9a5c. Os segmentos de DNA codificando estas regiões foram amplificados por PCR, clonados no vetor pGEM-T e posteriormente sub-clonados no vetor de expressão pAE (Ramos *et al.*, 2004), como apresentado na figura 23.

As construções pAE-*pilQN*, pAE-*pilQC* e pAE-*pilY1N* foram utilizadas para transformar bactérias competentes BL21(DE 3)*Gold* seguindo-se a indução da expressão das respectivas proteínas recombinantes com IPTG por 4 h a 37°C. A indução das proteínas recombinantes rPilQC e rPilY1N foi confirmada através de *Western blot* com anticorpo monoclonal anti-His-tag (Figura 24), sendo as proteínas detectadas na fração insolúvel. Os níveis de indução das duas proteínas recombinantes ficaram abaixo do desejado, uma vez que não detectamos a indução das duas proteínas recombinantes em SDS-PAGE corado com *Coomassie Blue*. Nas condições testadas não detectamos a expressão da proteína rPilQN (dados não apresentados).

As proteínas rPilQC e rPilY1N foram solubilizadas a partir dos corpos de inclusão com 0,3 % de sarcosil (Marshak *et al.*, 1996) e submetidas a cromatografia de afinidade em resina Ni²⁺-NTA agarose e eluição com concentrações crescentes de imidazol. A proteína recombinante rPilQC foi eluída com 50mM e 100mM de imidazol, todavia, não tivemos sucesso na purificação de rPilY1N através da cromatográfica de afinidade (dados não apresentados). Portanto a proteína recombinante rPilY1N foi obtida em maior escala e excisada diretamente do gel após separação em SDS-PAGE.

Figura 23. Clonagem das regiões N e C terminal de PilQ e N-terminal de PilY1. **A.** Os segmentos de DNA codificando estas regiões foram amplificados por PCR e clonados no vetor pGEM-T. O gel mostra as respectivas construções obtidas digeridas com *XhoI* e *EcoRI*: pGEM-*pilQC*, (canaleta 2), pGEM-*pilQN* (canaleta 4) e pGEM-*pilY1* (canaleta 6). Os plasmídios não digeridos das respectivas construções foram aplicados nas canaletas 3, 5 e 7. **B.** Os insertos excisados das construções em pGEM-T foram subclonados no vetor pAE. Para confirmar o sucesso da clonagem, as construções pAE-*pilQC* (canaleta 2); pAE-*pilQN* (canaleta 4) e pAE-*pilY1* (canaleta 6) foram digeridas com as enzimas de restrição *XhoI* e *EcoRI*. Os plasmídios não digeridos das respectivas construções foram aplicados nas canaletas 3, 5 e 7. As setas apontam os fragmentos de ~504pb, 616pb e 583pb correspondentes aos insertos liberados das respectivas construções. As amostras foram analisadas em gel de agarose 1,0% corado com brometo de etídeo. O marcador de tamanho 1Kb DNA ladder foi aplicado na canaleta 1.

Figura 24. Expressão das proteínas recombinantes rPilQC e rPilY1 em BL21(DE3) Gold. A cultura de bactérias BL21(DE3) Gold transformada com a construção pAE-PilQC e pAE-PilY1, crescida a 37°C, por uma noite, foi diluída para $DO_{600nm} = 0,1$ seguindo-se incubação a 37°C. Após DO_{600nm} atingir 0,6, foi adicionado IPTG 1mM, seguindo-se incubação por 3 horas a 37°C. As bactérias foram lisadas e o extrato bruto foi centrifugado. O sobrenadante (fração solúvel) foi removido e o sedimento (corpos de inclusão) suspenso no mesmo tampão utilizado na lise das bactérias. A fração solúvel (S) e o sedimento (P) foram separados em SDS-PAGE analisados por *Western blot* com anticorpo monoclonal anti-His-tag (*His-probe (H-3) HRP mouse monoclonal IgG1, Santa Cruz Biotechnology*). As imagens foram obtidas a partir da detecção da fluorescência do anticorpo secundário fluorescente (1:15000) no *aparelho Odyssey Infrared Imaging System*. As setas mostram a massa molecular esperada para as proteínas recombinantes.

Na obtenção de anticorpos anti-PilQC e anti-PilY1 foram utilizados dois coelhos imunizados com ~500ug de cada proteína recombinante obtidas como descrito acima. Um dos coelhos foi imunizado com uma dose de proteína recombinante purificada em coluna de níquel, rPilQC, (Figura 24) e o outro submetido a duas imunizações com proteína recombinante rPilY1N parcialmente purificada. Os antisoros obtidos reconheceram as proteínas recombinantes rPilQC e rPilY1 porém apresentaram reação cruzada com outras proteínas do extrato bacteriano, sendo que aparentemente anti-rPilQC mostrou-se mais específico do que anti-PilY1 (Figura 25).

Estes anti-soros foram avaliados contra o extrato celular total das cepas 9a5c e U24d (Figura 26). Como esperado, a cepa U24d não expressa a proteína PilQC, ao contrário da cepa 9a5c, em que o anticorpo detectou um polipetídeo com a massa molecular esperada para esta proteína (~69kDa).

Por outro lado, verificamos que o antisoro anti-rPilY1N, que se mostrou inespecífico contra o extrato celular da cepa BL21(DE3) (Figura 25), reconheceu um polipeptídeo de ~50kDa, aparentemente mais abundante na cepa U24D (Figura 26).

Mais ensaios são necessários para confirmarmos se este polipeptídeo é relacionado à proteína PilY1, cuja massa molecular predita é de ~158kDa. Uma possibilidade é que PilY1 sofra degradação na cepa U24d que não dispõe da proteína PilQ funcional e portanto não localiza a adesina na extremidade do pilus do tipo IV.

Infelizmente, como não obtivemos o anticorpo contra a região amino-terminal de PilQ, não foi possível avaliar se a proteína truncada é expressa na cepa U24d.

O conjunto desses resultados indica que a cepa U24d é incapaz de montar pilus tipo IV e, assim, realizar o *twitching motility* por não expressar PilQ com domínio *secretin*. Estudos anteriores usando a cepa norte-americana Temecula1 mostraram que o *twitching motility* aumenta a colonização do xilema e é importante para estruturar os estágios iniciais da formação do biofilme (De La Fuente *et al.*, 2007). Além disso, como já mencionado, o nocaute de *pilQ* resultou na ausência de *pilus* tipo IV e perda de *twitching motility* e franja de colônia, sendo que este mutante não foi capaz de colonizar sistemicamente videiras e só foi detectado próximo do ponto de inoculação (Meng *et al.*, 2005).

Figura 25. Western blot com anticorpos anti-rPilQC e anti-rPilY1N contra as respectivas proteínas recombinantes. Foram usados 1ug do extrato bacteriano em cada canaleta. Da esquerda para direita encontra-se a amostra do extrato da bactéria BL 21 (De3) Gold sem o plasmídio (Ex. B); Amostra da cultura expressando rPilQC em (A) e rPilY1N em (B) não induzida (NI) e induzida (I) por 5 horas com 1mM de IPTG a 37°C. Os anticorpos anti-rPilQC e anti-rPilY1 foram utilizados, respectivamente na diluição 1:250 (A) e 1:500 (B). As setas indicam a massa molecular referente a cada proteína recombinante. Foi usado o anticorpo secundário fluorescente na diluição de 1:15000. As imagens foram capturadas no aparelho *Odyssey Infrared Imaging System*.

Figura 26. Western Blot com anticorpos anti-rPilQC e anti-rPilY1N contra extratos das cepas U24d e 9a5c. O extrato total das cepas U24d e 9a5c (~9ug de proteína total) foi separado em SDS-PAGE, seguindo-se transferência para membrana de nitrocelulose e incubação com os anti-soros anti-rPilQC e anti-rPilY1N, respectivamente na diluição de 1:250 e 1:500. Foi usado o anticorpo secundário fluorescente na diluição de 1:15.000. As imagens foram capturadas no aparelho *Odyssey Infrared Imaging System*. As setas apontam os polipeptídeos reconhecidos pelos anticorpos indicados.

4.6.2. Avaliação da virulência e eficiência de colonização da cepa U24d

Para avaliar o efeito da possível ausência de motilidade *twitching* na cepa U24d *in planta*, realizamos um ensaio de deslocamento de *X. fastidiosa* em laranjeiras variedade pêra, hospedeiro natural das cepas U24d e 9a5c. Os ensaios foram feitos em laranjeiras com dois meses de idade e as inoculações foram realizadas em 24 laranjeiras, sendo 12 com U24d e 12 com 9a5c. Das 12 laranjeiras inoculadas com estas cepas, seis plantas foram inoculadas na região superior e seis na região inferior do caule. Após 7 meses de inoculação, coletamos as folhas ao longo do caule (em distâncias determinadas a partir do ponto de inóculo) para extração de DNA dos pecíolos e realização de ensaios de PCR com os primers CVC-1/272-1 como diagnóstico da presença de *X. fastidiosa*. Os resultados estão apresentados nas tabelas 11 e 12. Das 12 plantas inoculadas com a cepa 9a5c, todas foram colonizadas. A colonização mostrou-se distribuída ao longo do xilema, com o deslocamento máximo de 30 cm do ponto de inoculação. Para a cepa U24d só foi observada reação de PCR positiva em uma planta, a 10 cm de distância do ponto de inoculação nas plantas inoculadas na região superior. É plausível supor que a

ausência desta cepa nos dez primeiros centímetros na maioria das plantas inoculadas deve-se a falta de motilidade *twitching* da cepa U24d. Possivelmente, as células desta cepa formam um biofilme no ponto de inoculação e acabam morrendo pela limitada condição de nutriente neste local. A literatura descreve que o mutante de *X. fastidiosa* deficiente na motilidade *twitching* exibe aumento na formação de biofilme (Meng *et al.*, 2005).

Paralelamente, realizamos ensaios de infecção em citros para comparar a sintomatologia das plantas infectadas com a cepa U24d com a cepa de referência 9a5c. Nesse experimento, os primeiros sinais de CVC foram observados oito meses depois da inoculação em duas plantas das 12 infectadas com a cepa U24d (Figura 27). No entanto, os sintomas foram somente observados próximo ao ponto de inoculação, confirmando os efeitos da ausência de motilidade *twitching* nessa cepa. Em contra partida, a planta inoculada com a cepa 9a5c apresentou os sintomas de CVC generalizados após 18 meses em duas plantas das 12 inoculadas (Figura 27). Ainda que a colonização não tenha sido efetiva em todas as plantas inoculadas, os resultados sugerem que a cepa U24d resulte em sintomas precoces, porém localizados, enquanto que a cepa 9a5c causa sintomas por toda a planta, porém em um prazo de tempo maior do que a cepa U24d.

Tabela 11. Dados de deslocamento de *X. fastidiosa* inoculada na região superior do caule

Cepas inoculadas	Distância em cm					
	-40	-30	-20	-10	+10	+20
9a5c ^a	0/6	5/6	4/6	5/6	4/6	3/6
U24d ^a	0/6	0/6	0/6	1/6	0/6	0/6

^aForam inoculadas 6 plantas com cada cepa

Tabela 12. Dados de deslocamento de *X. fastidiosa* inoculada na região inferior do caule

Cepas inoculadas	Distância em cm					
	-40	-30	-20	-10	+10	+20
9a5c ^a	0/6	5/6	4/6	5/6	4/6	3/6
U24d ^a	0/6	0/6	0/6	0/6	0/6	0/6

^aForam inoculadas 6 plantas com cada cepa

Figura 27. Sintomatologia de CVC em plantas de inoculadas com as cepas U24d e 9a5c. A. Sintomas de CVC em laranjeira (*Citrus sinensis* L. Osbeck) após 18 meses de inoculação com a cepa 9a5c. Os sintomas se apresentaram generalizados na planta. **B.** Sintomas de CVC em laranjeira (*Citrus sinensis* L. Osbeck) após seis meses da inoculação com a cepa U24d. Os sintomas apareceram pontuais e próximos ao ponto de inoculação. As setas em vermelho indicam a região superior da planta aonde foi realizada a inoculação. Em detalhe, estão destacadas as manchas cloróticas nas folhas.

4.7. Análise de *rpfC* da cepa Fb7

Na análise comparativa dos genomas das cepas Fb7 e 9a5c detectamos que o gene que codifica a proteína RpfC (CDS XF1114 em 9a5c) apresenta a inserção de um nucleotídeo de guanina na posição 587 do gene ortólogo na cepa Fb7 (Figura 28 e Figura S2) que resulta no surgimento de um códon de parada prematuro. Ainda que a validação desta alteração através de sequenciamento desta região pelo método de Sanger não tenha sido realizada, aspectos do fenótipo exibido pela cepa Fb7 são sugestivos da veracidade desta mutação.

Como esquematizado na figura 28, a proteína RpfC eventualmente expressa na cepa Fb7 estaria truncada, apresentando apenas o domínio sensor que é transmembranar. Por outro lado, se eventualmente a proteína RpfC for traduzida a partir de um códon de iniciação em fase após a inserção, a proteína seria solúvel e

conteria os domínios necessários para sua atividade de histidina quinase. Em qualquer das duas possibilidades, consideramos para análise dos fenótipos da cepa Fb7, que esta cepa é um mutante natural de *rpfC* e que não expressa a proteína RpfC funcional.

Nas avaliações rotineiras de formação de biofilme e agregação celular, observamos que a deposição de biofilme na parede do frasco ar/líquido é insignificante na cepa Fb7, bem como a formação de agregados celulares no cultivo em meio líquido, características essas contrárias às observadas nas cepas U24d e 9a5c (Figura 11).

A partir dessa constatação da ausência de um biofilme maduro na cepa Fb7 aliada à observação de que esta cepa porta uma mutação em *rpfC*, postulamos que a expressão de genes relacionados à formação do biofilme e agregados celulares poderia estar alterada nesta cepa. Estes genes tem expressão modulada positivamente por DSF que atua através da sinalização mediada por RpfC e RpfG (Chatterjee *et al.*, 2008a; Chatterjee *et al.*, 2008b; Ryan e Dow, 2008; Guo *et al.*, 2012)

O sistema de dois componentes RpfC (histidina-quinase sensora) e RpfG (regulador da resposta) medeia a sinalização celular em resposta a presença de DSF (*Diffusible signal factor*) no meio extracelular, permitindo assim, regular positivamente fatores de virulências tais como estrutura do biofilme, enzimas extracelulares e motilidade (Chatterjee *et al.*, 2008a; Chatterjee *et al.*, 2008b; Ryan e Dow, 2008; Guo *et al.*, 2012). Em *Xanthomonas* e em *X. fastidiosa*, postula-se que RpfC liga-se ao RpfF na ausência de DSF, na presença do mesmo, a enzima RpfF é liberada tornando-se ativa para participar da síntese de DSF (Figura 7). O RpfC descomplexado da interação proteína-proteína com RpfF, pode se autofosforilar e na sequência fosforilar RpfG (fosfodiesterase), tornando-a ativa para degradar c-di-GMP a GMP (Dow *et al.*, 2006; Dow, 2008; Ryan *et al.*, 2010). A mudança nos níveis de c-di-GMP afeta os níveis da expressão do regulador global Clp, o qual diretamente e indiretamente regula a expressão de vários genes (Zhang, 2010).

A.	rpfc_9a5c	GGTCTGATTGCCATCCCTTGGTATTTTCAGTCGCTGCTGAAGGCATTGATCCAAGCCCTT	540
	rpfc_Fb7	GGTCTGATTGCCATCCCTTGGTATTTTCAGTCGCTGCTGAAGGCATTGATCCAAGCCCTT	540

	rpfc_9a5c	AAGGATGCACGTCATGCTAACGAGGCTAAGAGTCGCTTCTGGCCAA-TATGAGCCATGA	599
	rpfc_Fb7	AAGGATGCACGTCATGCTAACGAGGCTAAGAGTCGCTTCTGGCCAAAGTATGAGCCATGA	600

B.	Rpfc_9a5c	MLKKAKWLC SRLSQRPDSEHVQNLRLVIVSLFIIYLNACWDGETDLNTRSRIIWSVLLSD	60
	Rpfc_Fb7	MLKKAKWLC SRLSQRPDSEHVQNLRLVIVSLFIIYLNACWDGETDLNTRSRIIWSVLLSD	60

	Rpfc_9a5c	LGVSLALMLAIVMHPQISHVRCIGIIADYTSLTVMMLMGEAGSPLYVLCWVTIGNGL	120
	Rpfc_Fb7	LGVSLALMLAIVMHPQISHVRCIGIIADYTSLTVMMLMGEAGSPLYVLCWVTIGNGL	120

	Rpfc_9a5c	RYGCAYLLVTTLSALSFLMVILVSAYWKS NPF LAWGLLIGLIAIPWYFQSLKALIQAL	180
	Rpfc_Fb7	RYGCAYLLVTTLSALSFLMVILVSAYWKS NPF LAWGLLIGLIAIPWYFQSLKALIQAL	180

	Rpfc_9a5c	KDARHANEAKSRFLANMSHEFRTPLNGLSGITEVLATRLDAEQRECLKAIQASVHSVLS	240
	Rpfc_Fb7	KDARHANEAKSRFLAKYEP-----	199

	Rpfc_Fb7.2	MLTRLRVASWPSMSHEFRTPLNGLSGITEVLATRLDAEQRECLKAIQASVHSVLS	56

	Rpfc_9a5c	LIEEVLDISRIEAGKIRILHDTFSLQEVIGSVGLILQPQIRNKVLEYRVDVGLDVPWELF	300
	Rpfc_Fb7.2	LIEEVLDISRIEAGKIRILHDTFSLQEVIGSVGLILQPQIRNKVLEYRVDVGLDVPWELF	116

	Rpfc_9a5c	GDAGYVRQVLLNIVGNAVKFTERGGITLCVSVLNRSRERSVVLRFVADTGIGVPLGMRA	360
	Rpfc_Fb7.2	GDAGYVRQVLLNIVGNAVKFTERGGITLCVSVLNRSRERSVVLRFVADTGIGVPLGMRA	176

	Rpfc_9a5c	RLFDMFEQADVGLDRRYQGSGLGTAIAKGLIELMGSGIFEEENVPCGSLFWFELPFELGK	420
	Rpfc_Fb7.2	RLFDMFEQADVGLDRRYQGSGLGTAIAKGLIELMGSGIFEEENVPCGSLFWFELPFELGK	236

	Rpfc_9a5c	EPVVAQELPEASQCTEASNVIAFSNPFRLRHRARVKSFLVLVADDHETNRMVLSRILDKAG	480
	Rpfc_Fb7.2	EPVVAQELPEASQCTEASNVIAFSNPFRLRHRARVKSFLVLVADDHETNRMVLSRILDKAG	296

	Rpfc_9a5c	HKLLCVDGAEAVLDCLISMEFDVVIIDLHMPGMSGLEMLKQLRVMQSAGLPYTPVLVLSA	540
	Rpfc_Fb7.2	HKLLCVDGAEAVLDCLISMEFDVVIIDLHMPGMSGLEMLKQLRVMQSAGLPYTPVLVLSA	356

	Rpfc_9a5c	DAASDSIRCCKEAGARAFLSKPVLATKLLDLAEIAEGDGGALPYAALSQSSVQQHQVFD	600
	Rpfc_Fb7.2	DAASDSIRCCKEAGARAFLSKPVLATKLLDLAEIAEGDGGALPYAALSQSSVQQHQVFD	416

	Rpfc_9a5c	PTGLDDLAE LMDDDGFEERFIVQSLRDAEHCQQTMLSA AETNQWQEVREQAHA LRGVLGH	660
	Rpfc_Fb7.2	PTGLDDLAE LMDDDGFEERFIVQSLRDAEHCQQTMLSA AETNQWQEVREQAHA LRGVLGH	476

	Rpfc_9a5c	LGLMRASSLAE LMRIPDWQLQSEWCARVQVLYEELRSGRDALEARRQERRRSGILQ	717
	Rpfc_Fb7.2	LGLMRASSLAE LMRIPDWQLQSEWCARVQVLYEELRSGRDALEARRQERRRSGILQ	533

Figura 28. Inserção de 1 nucleotídeo no gene *rpfF* da cepa Fb7. A. Trecho de alinhamento extraído do alinhamento completo das sequências de *rpfF* de 9a5c e Fb7 apresentado na figura suplementar S2 (Apêndice). **B.** Alinhamento da sequência de aminoácidos de RpfC de 9a5c e da forma truncada predita na cepa Fb7 correspondendo a região aminoterminal da proteína. O alinhamento inclui o polipeptídeo que seria eventualmente gerado caso ocorra início da tradução após o ponto da inserção do nucleotídeo em *rpfC* de FB7 (RpfC_Fb7.2). Os domínios funcionais de histidina quinase estão coloridos em verde e nas caixas em vermelho, azul e amarelo. O peptídeo sinal predito está sombreado.

Como pode ser observado na tabela 13, o gene *pspA*, que codifica uma das principais adesinas afimbriais (similares à hemaglutinina) de membrana em *X. fastidiosa* (Voegel *et al.*, 2010), tem seus níveis de transcritos diminuído aproximadamente 10 vezes na cepa Fb7 em relação a cepa 9a5c. Essa CDS na cepa Temecula1 (Norte-Americana) está envolvida na aderência e formação de biofilme na planta e no inseto vetor, sendo este um dos principais mecanismos de virulência de *X. fastidiosa* (Voegel *et al.*, 2010).

Verificamos que MrkD, uma adesina do *pilus* curto (Li *et al.*, 2007), também apresenta níveis de transcritos muito baixos ou inexistentes na cepa Fb7. O nocaute de *fimA* na cepa Temecula1 não possui *pili* curtos e exibe um fenótipo menos agregativo na superfície do vidro entre o ar/meio e apresenta uma maior franja da colônia em relação ao selvagem (De La Fuente *et al.*, 2007). Coincidente com estas observações, a cepa Fb7 é menos agregativa e apresenta uma franja maior (Figuras 11 e 12) em relação à 9a5c. Diante desses dados, é provável que o fenótipo apresentado pela cepa Fb7, como ausência de um biofilme maduro e agregados celulares esteja relacionado diretamente com a diminuição da expressão desses genes.

Tabela 13. Quantificação relativa da expressão gênica por RT-qPCR ^a

ORFs ^b	Genes	Nível de expressão ^c
XF0889	<i>pspA</i>	0.09 ±0.05
XF2466	<i>pglA</i>	0.45 ±0.03
XF2398	<i>hemolysin</i>	0.40 ±0.02
XF0986	<i>cgsA</i>	0.13 ±0.16
XF2367	<i>gumD</i>	6.27 ±0.19
XF1540	<i>clp</i>	0.54 ±0.25
XF0080	<i>mrkD</i>	0.0005 ±0.11
XF1113	<i>rpfG</i>	0.23 ±0.20
XF0478	<i>pilY1</i>	0.11 ±0.25

^a Quantidade relativa de RNA nas células da cepas 9a5c é igual a 1. O *dnaQ* foi usado como normalizador.

^b Conforme <http://aeg.lbi.ic.unicamp.br/xf/>.

^c Razão de expressão da cepa Fb7 em relação a cepa 9a5c.

Observamos, entretanto, que o gene *gumD*, que faz parte do *operon gum*, e está envolvido na produção de exopolissacarídeo (EPS) (Chatterjee *et al.*, 2008a), tem seus níveis de transcritos aumentados em Fb7 em relação 9a5c. O aumento da expressão do *operon gum* provavelmente justifica seu aspecto viscoso. Entretanto o conteúdo de EPS produzido por Fb7 relativamente à cepa 9a5c ainda não foi avaliado.

O mutante de *rpfC* na cepa Temecula1 apresenta nível de virulência aumentado, com aumento de aderência e formação de biofilme na parede de vidro do

frasco na interface ar/líquido (Chatterjee *et al.*, 2008a). Em *Xanthomonas campestris campestris* (XCC) a mutação em *rpfC* resulta em menor nível de aderência e o biofilme diminui consideravelmente (Torres *et al.*, 2007). Na cepa Fb7, que tem RpfC provavelmente não funcional, apresenta um fenótipo que se assemelha ao de XCC, como a diminuição da formação de biofilme e da aderência ao vidro (Figura 11).

Além disso, o mutante de *clp* em XCC apresenta produção de EPS e aderência célula-célula. Na cepa Fb7, ocorre uma diminuição relativa dos níveis de RNA de *mrkD* e *pspA* e aumento no *gumD* (Tabela 13). Interessantemente, análises *in silico* mostraram que os promotores das CDSs *mrkD*, *pspA* e *pilY1* possuem sítios de ligação regulador Clp (dados não publicados), o que reforça a hipótese que as três CDSs, que estão relacionadas diretamente com a virulência e patogenicidade, estão sobre controle transcricional de Clp (Figura 7).

Na cepa Temecula, a mutação na CDS *cgsA* (diguamilato ciclase), responsável pela produção de c-di-GMP, causa uma diminuição na formação de biofilme, aderência, virulência e transmissão de células pelo inseto vetor. No entanto, não há mudança significativa nos níveis transcricionais dos genes *mrkD* e *pspA* em relação ao selvagem (Chatterjee *et al.*, 2010). Entretanto, diminuição dos níveis de c-di-GMP em XCC aumenta os fatores de virulência e motilidade (Ryan *et al.*, 2007). Na cepa Fb7 os níveis transcricionais de *cgsA* também são menores em relação a 9a5c (Tabela 13).

O conjunto destas observações quanto à expressão deste conjunto de genes na cepa Fb7 aliado às informações disponíveis na literatura, propomos um modelo de sinalização mediada por DSF da cepa Fb7, o qual vai de encontro em alguns aspectos ao modelo já proposto para *X. fastidiosa* e corrobora em outros com o modelo para XCC. Neste modelo, *rpfC* mutado não produz proteína RpfC ativa, o qual não interage com a proteína RpfF. A falta de interação entre essas duas proteínas deixa RpfF livre para participar da via de síntese de DSF. Desta forma, haverá um aumento nos níveis de DSF pela cepa Fb7, onde parte deste será externalizado. Com o *rpfC* mutado não ocorrerá fosforilação da proteína RpfG, impedindo sua atividade de fosfodiesterase. Como consequência os níveis de c-di-GMP são aumentados, ainda que a expressão do gene *cgsA* esteja diminuída. Com o aumento dos níveis de c-di-GMP, ocorre sua interação com a proteína Clp e evitando que este ligue a promotores de genes relacionados com virulência, como *mrkD* e *pspA*, que em nossos ensaios de RT-qPCR tem expressão diminuída na cepa Fb7 (Tabela 13). Outro nível de regulação pelo aumento dos níveis de c-di-GMP poderia ser pela redução da expressão de *clp* (por *feedback*). Desta forma, na cepa Fb7, a sinalização desencadeada a partir do sensor/regulador do sistema de dois componentes RpfC, passando pelo segundo

mensageiro, c-di-GMP está comprometida e a expressão das CDSs relacionadas com a formação de um biofilme maduro e aderência ao substrato e agregação célula-a-célula não ocorre (Figura 29).

Figura 29. Proposta de modelo de sinalização célula-a-célula mediada por DSF na cepa Fb7 de *X. fastidiosa*. A CDS *rpfC* que codifica o sensor do sistema de dois componentes não produziria RpfC funcional. Desta forma, RpfF estaria livre para participar da via de biossíntese de DSF. Como RpfC mutado não é capaz de transduzir o sinal de DSF no meio, nem tão pouco se complexa com RpfF, DSF seria produzido constitutivamente. Essa mutação em RpfC não ativaria o RpfG através da fosforilação, que por sua vez não degradaria c-di-GMP, passando esse a se acumular intracelularmente. Esse acúmulo inibiria por *feedback* a expressão da CDS *Clp* ou se ligaria alostericamente a proteína *Clp*, não permitindo o mesmo de se ligar ao promotor de genes relacionados à virulência, como *mrkD*, *pspA* e *pglA*, reduzindo consequentemente a expressão destas CDSs. Essa redução nos níveis de transcritos desses genes relacionados ao biofilme explicaria a ausência de biofilme na cepa Fb7.

4.8. A heurística da dinâmica evolutiva dos genomas de *X. fastidiosa*

4.8.1. Pangenoma de *X. fastidiosa*

Vários estudos têm demonstrado que cepas bacterianas de uma mesma espécie diferem fenotipicamente e genotipicamente, sendo que ao sequenciar vários genomas da mesma espécie, novos genes poderão ser descobertos (Akopyants *et al.*, 1998; Lawrence e Hendrickson, 2005; Dagan e Martin, 2007; Koonin e Wolf, 2008; Lapierre e Gogarten, 2009). Estas observações suportam a *Hipótese do Estado da Distribuição do Genoma (DGH)*, na qual um repertório de genes de uma determinada espécie de bactéria está distribuído entre diferentes cepas (Garth *et al.*, 2008). Assim, uma dada espécie bacteriana possui e mantém um genoma principal (*core-genome*) e um repertório de genes dispensáveis distribuídos entre as diferentes cepas daquela espécie. Este padrão pode ser explicado assumindo que os novos genes são trazidos para uma população, por exemplo, por transferência horizontal (HGT), propiciando então, uma vantagem evolutiva para as bactérias (Dagan e Martin, 2007).

Análises comparativas entre múltiplos genomas têm revelado extensiva diversidade genômica intraespécie (Pallen e Wren, 2007). No entanto, dada à facilidade de gerar sequências de genomas inteiros, torna-se cada vez mais interessante saber quantos genomas deveriam ser sequenciados para representar exatamente o repertório de genes do genoma de uma dada espécie. Uma maneira de resolver este problema é perguntar quantos novos genes são identificados cada vez que um novo genoma de uma espécie de interesse é sequenciado (Tettelin *et al.*, 2008).

Com esta abordagem, trabalhos pioneiros determinaram *core genome* de *Streptococcus agalactiae* e de *Haemophilus influenzae* (Tettelin *et al.*, 2005; Hogg *et al.*, 2007). A característica marcante desses estudos foi a constatação de que uma percentagem significativa de cada sequência genômica é específica para cada cepa e, por conseguinte, cada novo genoma sequenciado fornecia um número de genes novos não caracterizados. Assim, o repertório de genes dessas espécies é significativamente maior do que de uma única cepa. Isso levou os autores ao conceito de pangenoma ou supragenoma bacteriano, que nada mais é do que a soma do core genome (genes comuns a todas as cepas), genoma dispensável (genes presente em algumas, mas não todas as cepas estudadas) e genes novos (genes encontrados especificamente em cada cepa adicionada ao estudo) (Tettelin *et al.*, 2005; Hogg *et al.*, 2007). Vale salientar que o genoma dispensável não só contribui para a diversidade da espécie e,

provavelmente, fornece funções que não são essenciais para o seu nicho ecológico, mas conferem vantagens seletivas incluindo adaptação ao nicho, resistência a antibióticos e a capacidade de colonizar novos hospedeiros (Tettelin *et al.*, 2005).

Como listado na tabela 2, onze genomas de *X. fastidiosa* já estão sequenciados, destes, oito estão publicamente disponíveis (Simpson *et al.*, 2000b; Bhattacharyya *et al.*, 2002a; Van Sluys *et al.*, 2003; Zhang *et al.*, 2011) e três foram recentemente sequenciados em nosso grupo (Pierry, 2012). A este conjunto de genomas, somam-se os três genomas descritos neste trabalho. Desta forma, dispomos do total de 14 genomas sequenciados de *X. fastidiosa*, o que nos permite realizar um estudo da distribuição de genes nos genomas da espécie *X. fastidiosa*, através da genômica comparativa, visando a definição do pangenoma desta espécie.

O pangenoma dos 14 genomas *X. fastidiosa* foi dividido em três conjuntos de genes (*clusters*): *clusters* de ortólogos compartilhados entre todas as cepas, que defini o *core genome*; *clusters* de ortólogos compartilhados entre duas ou mais cepas (*clusters* de genes acessórios ou dispensáveis ou de contingência) e *clusters* de genes únicos de cada cepa (*clusters* de genes novos).

Como esperado e de acordo com a *Hipótese do Estado da Distribuição do Genoma (DGH)* (Garth *et al.*, 2008), o número total de *clusters* de ortólogos de *Xylella*, representando o pangenoma das 14 cepas foi 3.644, isto é, um número maior do que qualquer genoma individual das cepas estudadas. Dentro desse pangenoma, os *clusters* de genes do *core genome* representam 45%, enquanto os *clusters* de genes acessórios, quais estão em duas ou mais combinações, mas não em todas as cepas, representam 20% e os *clusters* de genes únicos (*clusters* de genes novos) de cada cepa representam 34,6%, (Tabela 14). Estas mesmas análises foram feitas com os 14 genomas excluindo as regiões de profagos, profagos remanescentes e ilhas genômicas, não sendo observadas mudanças significativas nas representatividades destes *clusters*. Contudo, o número de *clusters* de genes novos diminuiu de 1.263 (considerando os EGMs) para 1079 (desconsiderando os EGMs) mostrando assim, que os EGMs contribuem para o perfil do pangenoma (Tabela 14).

Tabela 14. Número de *clusters* de ortólogos do pangenoma das 14 cepas de *X. fastidiosa* calculado com e sem EGMS

Categorias dos clusters dos pangenomas	Pangenoma com EGMS		Pangenoma sem EGMS	
	Número de clusters de genes	Representatividade no pangenoma	Número de clusters de genes	Representatividade no pangenoma
Core genome	1.641	45%	1317	41%
Genes acessórios	740	20%	794	24,8%
Genes novos	1.263	34,6%	1079	33%
Total de clusters	3.644	100%	3190	100%

Para determinar o comportamento do número de clusters de genes compartilhados entre todas as cepas (*core genome*), sobre adição sequencial de novas sequências de genomas, foi realizado um decaimento exponencial em função dos dados. O resultado de todas as combinações possíveis ordenada para cada adição de uma nova cepa está mostrado na figura 30. Como esperado, o número de genes compartilhados inicialmente decresce com adição de uma nova sequência de genoma. Entretanto, o número de *clusters* de genes do *core genome* apresenta uma tendência para uma assíntota horizontal perto de 1.641 *clusters* de genes. Como esperado, esses *clusters* possuem genes relacionados com os processos metabólicos centrais de *X. fastidiosa* e na constituição de profagos. Em estudos do pangenoma de outras bactérias os genes que compõem o *core genome* também estão relacionados com os processos metabólicos centrais (Tettelin *et al.*, 2005; Hogg *et al.*, 2007; Rasko *et al.*, 2008; Conlan *et al.*, 2012).

O número de *clusters* de genes novos identificados nos 14 genomas analisados apresentou também um decaimento exponencial (Figura 31). Nesse, ~90 *clusters* de genes novos são adicionados a cada inclusão de uma nova sequência de genoma de *X. fastidiosa*, sendo que a maioria dos genes dentro desses *clusters* são hipotéticos ou genes relacionados com EGMS. Entretanto, no gráfico que representa o pangenoma, é observada uma curva em crescimento contínuo, que aumenta a novas adições de sequência de genomas, mostrando que novos genes são adicionados a cada adição de genoma (Figura 30). Desta forma, os resultados do pangenoma e dos genes novos, faz do genoma da espécie *X. fastidiosa* um genoma aberto, matematicamente infinito. Esse modelo também ocorre em *E. coli*, *Streptococcus agalactiae*, *Streptococcus epidermidis* e *Haemophilus influenzae* (Tettelin *et al.*, 2005; Hogg *et al.*, 2007; Tettelin *et al.*, 2008; Conlan *et al.*, 2012). Um caso contrário a esse, é o genoma da *Buchnera aphidicola*, que apresenta um genoma fechado, não

apresentando novos genes quando novas adições de sequência são inseridas, obtendo assim, um gráfico com crescimento exponencial em que após a soma de um número determinado de genomas atinge o platô (Medini *et al.*, 2005; Mira *et al.*, 2010).

Apesar de *X. fastidiosa* encontrar-se basicamente em dois ambientes (cibário do inseto vetor ou no xilema da planta hospedeira) (Chatterjee *et al.*, 2008a), verifica-se uma alta diversidade genética deste fitopatógeno, uma vez que, mesmo sequenciando 14 genomas desta espécie não foi possível cobrir o repertório total de genes da bactéria, como também já observado para outras espécies bacterianas (Hiller *et al.*, 2007). Essa característica do genoma da *X. fastidiosa* ser matematicamente infinito deve-se provavelmente ao fato da presença de grande quantidade de EGMs como será discutido adiante. Assim, é provável que *X. fastidiosa* tenha tido um ancestral com genoma fechado, porém no decorrer da evolução deste fitopatógeno, eventos de redução do genoma (Pieretti *et al.*, 2009) propiciaram a perda do sistema de imunidade CRISPR/cas, qual está envolvido diretamente na barreira contra entrada de EGMs em um genoma (Karginov e Hannon, 2010; Barrangou e Horvath, 2012). Essa perda provavelmente tornou *X. fastidiosa* mais permissiva a aquisição de muitos EGMs, passando assim, a bactéria a ter um genoma aberto. Essa perda de CRISPR/cas além de tornar o genoma mais suscetível a EGMs e tornar o genoma aberto, seria responsável, provavelmente, pela especiação dos genomas das cepas de *X. fastidiosa* e conseqüentemente possa ter refletido na especiação das cepas dos diferentes hospedeiros.

Figura 30. Pan-genoma e core genome de *X. fastidiosa*. O número de *clusters* de genes foram plotados em função do número de genomas sequencialmente adicionados. Os círculos são os valores obtidos para diferentes combinações de cepas dada por $n!/(n-i)!.i!$. Os pequenos círculos são as medianas dos valores obtidos nas combinações. A curva contínua em vermelho é a linha de tendência para uma assíntota horizontal perto de 1.641 no gráfico do core genoma.

Figura 31. Descoberta de novos genes em *X. fastidiosa*. O número de *clusters* de genes foram plotados em função do número de genomas sequencialmente adicionados. Os círculos são os valores obtidos para diferentes combinações de cepas, dada por $n!/(n-i)!.i!$. Os pequenos círculos são as médias dos valores obtidos das combinações. A curva contínua em vermelho é a linha de tendência. A curva tem uma tendência para uma assíntota vertical próximo de 80.

4.8.2. Identificação de regiões de profagos e ilhas genômicas nos 14 genomas de *X. fastidiosa*

As integrases são um dos componentes principais relacionados com Elementos Genéticos Móveis (EGM), e sua identificação quase sempre está associada a regiões de profagos. Sua principal função é mediar o acesso do EGM ao genoma hospedeiro (Groth e Calos, 2004). Desta forma, para encontrar regiões de profagos e ilhas genômicas, os genomas das cepas de *X. fastidiosa* foram varridos em busca de integrases, segundo metodologia já usada anteriormente para genomas de *X. fastidiosa* (de Mello Varani *et al.*, 2008). A classificação das integrases foi feita de acordo com as seguintes características: (a) forma completa e potencialmente funcional, apresentando 325 resíduos de aminoácidos (segundo o banco de dados PFAM); (b) forma completa apresentando FS/SCF (*frameshift / stop codon in frame*); (c) forma incompleta (ou fragmento), isto é, apresenta tamanho menor do que 280 resíduos de aminoácidos (de Mello Varani *et al.*, 2008). A tabela 15 mostra todas as integrases mapeadas nos 14 genomas de *X. fastidiosa*. Vale salientar que essas integrases não representam todas as regiões de profagos e profagos remanescentes das cepas de *X. fastidiosa*.

Um fato interessante quanto ao número de integrases de *X. fastidiosa* é que as cepas da América do Sul tem quase o dobro de integrases envolvidas diretamente com EGMs em relação as cepas da América do Norte é notável também que as cepas sul-americanas têm número maior de integrases completas e um menor número de integrases incompletas com *frameshift* em relação às cepas norte-americanas (Tabela 15) (Tabela suplementar S1). Essas evidências refletem diretamente no número e no tipo de rearranjo cromossômico das cepas estudadas. É provável que as cepas da América do Norte tenham passado por um processo mais acentuado de deterioração (Canchaya *et al.*, 2003) dos EGMs do que as da América do Sul.

Tabela 15. Número e características das integrases das cepas Norte e Sul-Americanas.

Cepa/integrases	Número de integrases	Integrases completas	Integrases completas (<i>frameshift</i>)	Integrases incompletas
9a5c	11	8	1	2
U24d	11	8	1	2
J1a12	14	12	0	2
Fb7	13	8	1	4
3124	12	10	0	2
HIB4	12	8	0	4
PR8X	12	9	0	3
Temecula	9	4	3	1
GB514	8	5	1	2
M23	10	5	0	5
EB-92.1	5	1	0	4
Dixon	12	7	1	4
M12	9	5	0	4
Ann1	10	8	0	2

Os bacteriófagos desempenham um papel importante na evolução dos genomas bacterianos por dois fatores principais: sua abundância global mundial (Wommack e Colwell, 2000) e sua capacidade de transferência lateral entre bactérias (Ochman *et al.*, 2000; Rodriguez-Valera *et al.*, 2009), permitindo assim, alterar a biologia da célula bacteriana, através de fatores de virulência como, toxinas extracelulares, proteínas efetoras que modulam a invasão a célula hospedeira e fatores de adesão ao substrato (Boyd e Brussow, 2002). Desta forma, os EGMs são significativas forças evolutivas entre os procariotos por acrescentar mecanismos de adaptação ecológica aos genomas das bactérias (Beiko *et al.*, 2005; Aminov, 2011).

Tendo por base tais informações foram varridos os genomas das cepas de *X. fastidiosa* da América do Norte e do Sul em busca de EGMs. Para tanto, foi realizado inicialmente uma busca através da plataforma PHAST (PHAge Search Tool) (Zhou *et*

al., 2011). No entanto, essa ferramenta não foi capaz de identificar todos os elementos móveis, sendo necessária então, uma inspeção manual cuidadosa na plataforma IMG/ER para encontrar outras regiões de profagos, profagos remanescentes e IGs. Essa inspeção manual teve como ponto de partida a busca por integrases, método esse já usado na identificação de EGMs em *X. fastidiosa* (de Mello Varani et al., 2008).

A média do número total de elementos móveis encontrados nas cepas de *X. fastidiosa* da América do Sul foram 13,7, já os encontrados nas cepas da América do Norte foram 12. Esse resultado também se repete para as regiões de profagos/IG, sendo em média 10,8 EGMs para as cepas da América do Sul e 9,7 EGMs para as cepas da América do Norte. Todavia, o tamanho médio desses EGMs é em torno de 225,8 Kb nas cepas norte-americanas e 381,0 Kb nas cepas sul-americanas. Já os fragmentos de profagos/IG estão distribuídos equivalentemente entre todas as cepas. Entretanto, as regiões de profagos/IG exclusivas são superiores nas cepas da América do Norte em relação às cepas de *X. fastidiosa* da América do Sul (Tabela 16).

A espécie *X. fastidiosa* tem uma representatividade muito grande de regiões de profagos/IG por genoma. Em média, essa representatividade ultrapassa 12%. Na família *Xanthomonadaceae* essa espécie é a que tem provavelmente a maior representatividade de EGMs por genoma (dados não mostrados). Interessantemente, grande parte dos EGMs das cepas da América do Sul é compartilhada. No entanto, os EGMs das cepas da América do Norte são poucos compartilhados. Menos ainda são observados os EGMs comuns entre as cepas das Américas do Norte e do Sul (Tabela 16).

Estudos anteriores identificaram EGMs de *X. fastidiosa* das cepas 9a5c, Temecula1 e Dixon (Simpson et al., 2000b; de Mello Varani et al., 2008). Todavia, neste trabalho identificamos os EGMs das cepas então sequenciadas da América do Sul (U24d, Fb7, 3124, J1a12, HIB4 e PR8X); os EGMs das cepas da América do Norte, não analisadas antes (GB514, EB-92.1, M23, M12 e Ann1) e acrescentamos novas identificações de regiões nas cepas já estudadas, como: a cepa Temecula1 - três novas regiões de profagos e cepa Dixon, uma nova região de profago. Nesses estudos, vimos também que esses EGMs identificados são encontrados em outros gêneros de organismos, como, *Aggregatibacter*, *Stenotrophomonas* e *Microcystis* (Tabela suplementar S2).

Tabela 16. Descrições do número e tamanho das regiões de profagos/IG das cepas Sul e Norte-Americanas

Profagos/IG / cepa	9a5c	U24d	Fb7	3124	J1a12	HIB4	PR8X	Temecula1	GB514	M23	EB-9.2	Dixon	M12	Ann1
Total de regiões de profagos/IG	11	11	16	13	16	16	13	13	11	14	18	17	19	13
Número de Regiões de profagos/IG	9	9	14	14	11	12	10	11	9	12	4	15	6	11
Número de regiões de fragmentos de profagos/IG	2	2	2	2	5	3	1	2	2	2	4	2	3	2
Regiões de profagos/IG exclusivas	0	0	5	4	7	5	2	2	2	1	6	11	4	10
Soma de regiões de profagos /IG (Kb)	324,4	96,8	398,6	447,57	501,97	393,17	305,17	233,1	134,2	339,1	67,7	375,4	227	304,4
Representatividade das regiões de profagos/IG nos respectivos genomas (%)	12,10	11,60	14,57	16,28	17,99	13,95	11,44	9,25	5,38	9,40	14,27	4,27	9,17	11,62
Maior região de profago (Kb)	49,3	49,3	73,5	95,4	147,4	56,6	65	42,9	40,1	42,9	3,7	60,8	60,7	62,6
Regiões de profagos/IG compartilhadas entre as cepas Sul ou Norte-americanas	11	11	11	9	9	10	11	12	9	13	1	6	5	3
Regiões de profagos/IG compartilhados entre as cepas Sul e Norte-Americanas	5 (xf1; xfp2; xfp4; xfp10 e cvc-r1)/82 (6.1%)													

Um fato notável nos 14 genomas das cepas de *X. fastidiosa* é a distribuição das regiões de profagos, profagos remanescente e ilhas genômicas. Nessa, os EGMs não se encontram distribuídos uniformemente ao longo do cromossomo, mas ao contrário, estão concentrados na região central dos genomas, tanto nas cepas da América do Norte quando nas cepas da América do Sul (Figura 32). A exceção é o genoma da GB514, que tem a maior concentração de EGMs na região inicial e final do genoma. Em uma análise mais detalhada desse genoma é possível verificar que essas regiões correspondem a regiões centrais nos outros genomas cuja posição foi alterada por translocações com inversões. Outro fato que merece atenção é a concentração de alguns EGMs no final do genoma da cepa Dixon. Todavia, não causam mudanças posicionais no genoma.

Os EGMs identificados e mapeados segundo sua posição no genoma assumem posições diferentes, segundo as cepas correlacionadas (Figura 32). As cepas 9a5c e U24d, em especial, possuem os mesmos EGMs nas mesmas posições em relação aos genomas analisados. Todavia, o mesmo evento não é visto para as outras cepas, sendo possível um mesmo EGM se encontrar em praticamente todas as cepas, no entanto, em posições totalmente diferentes. Essas variações nas posições dos EGMs observadas nas 14 cepas de *X. fastidiosa* estão relacionadas com eventos de rearranjo cromossômico, sendo em alguns casos, translocações, translocações com inversões ou somente inversões (Figura 32). Em trabalhos anteriores com algumas cepas de *X. fastidiosa* (9a5c, Temecula1 e Dixon) já foi constatado a existência de rearranjo cromossômico em regiões flanqueadas por integrases (Bhattacharyya *et al.*, 2002a; Van Sluys *et al.*, 2003; de Mello Varani *et al.*, 2008). Todavia, nossas observações mostram pela primeira vez tanto as correlações dos EGMs nas diferentes cepas de *X. fastidiosa*, quanto o posicionamento preferencial dos elementos móveis em determinada região do genoma (Figura 32).

As 14 cepas de *X. fastidiosa* somam 181 EGMs, com um total de 5.730 genes presentes (com redundância). Todavia, aproximadamente 52% desse total são genes hipotéticos. Entretanto, dos genes que fazem parte dos EGMs, ~48% tem função definida segundo COG, e destes, boa parte são genes relacionados à codificação de proteínas de fagos. Contudo, vale ressaltar a presença de outros genes que estão provavelmente relacionados à adaptabilidade da *X. fastidiosa* ao seu nicho. Um desses genes é a metaloprotease discutida no item 4.5. A tabela suplementar S1 apresenta uma lista desses genes cuja contribuição à adaptabilidade patógeno/hospedeiro poderia ser estudada em maior detalhe.

Figura 32. Distribuição de EGMs nos genomas das cepas sul e norte-Americanas. Os retângulos de diferentes cores representam as regiões de progafos/IGs nas várias regiões dos genomas. As linhas conectando os retângulos representam os EGMs em comum nas diferentes cepas. Os retângulos pretos conectados por linhas representam as IGs. Já os retângulos pretos isolados são EGMs exclusivos de cada cepa. Figura construída a partir do programa Mauve 2.0 .

Nos estudos de pangenoma, verificamos que 5.730 (com redundância) genes dos EGMs ou 1.062 *clusters* (não redundante) estão contidos no pangenoma de *X. fastidiosa*. Destes, 328 (30,8%) *clusters* enquadram-se na divisão de *clusters* do *core genome*; 478 (45%) estão presentes na divisão dos *clusters* de genes acessórios (presente em, pelo menos, 2 cepas) e 256 (24,1%) *clusters* foram identificados na divisão dos *clusters* de genes novos (presentes em uma cepa) (Tabela suplementar S1). Interessantemente, a maior parte dos *clusters* das regiões de fagos está presente na divisão dos genes acessórios do pangenoma. Região essa, que os estudos de pangenoma consideram de suma importância para evolução do organismo, uma vez

que estão presentes genes que conferem vantagem ao organismo, como: a) genes de adaptação ao nicho; b) resistência a antibiótico e c) capacidade de colonizar novos hospedeiros (Medini *et al.*, 2005; Tettelin *et al.*, 2005; Rasko *et al.*, 2008; Tettelin *et al.*, 2008)

É provável que o grande número de EGMs encontrados nas cepas de *X. fastidiosa* possam estar relacionados tanto com a divergência destas cepas estreitamente relacionadas, quanto com a aquisição de novos fatores de virulência, patogenicidade e adaptabilidade para com os diferentes hospedeiros. Esse tipo de especiação genômica de cepas proximamente relacionadas tem sido sugerido em outros organismos, a exemplo o gênero *Streptococcus*, para o qual foram revelados fatores de virulência putativos codificados por EGMs que provavelmente contribuem na adaptabilidade da espécie ao seu habitat (Banks *et al.*, 2002; Boyd e Brussow, 2002; Canchaya *et al.*, 2003; Brussow *et al.*, 2004; Frost *et al.*, 2005).

4.8.3. Sistema de imunidade CRISPR-cas em *X. fastidiosa*

CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) são pequenas sequências de DNA processadas por proteínas *cas* que são inseridas no genoma do hospedeiro. Essas sequências medeiam uma forma de aquisição de imunidade contra patógenos virais específicos, principalmente bacteriófagos (Deveau *et al.*, 2010; Karginov e Hannon, 2010; Barrangou e Horvath, 2012; Szczepankowska, 2012). O sistema funciona resumidamente da seguinte maneira: (a) a bactéria incorpora material genético de fagos dentro dos espaços CRISPRs; (b) uma vez adquiridos esses espaços, a bactéria torna-se resistente a tipos de fagos que contém esses tipos de sequências. Parte integral do mecanismo de defesa anti-fago deve-se a proteínas associadas ao CRISPR, proteínas *cas*. As sequências que codificam essas proteínas são usualmente encontradas adjacentes e *upstream* as sequências de CRISPR, e mostram uma alta similaridade a endonucleases (Vale e Little, 2010).

Tendo por base tais observações, realizamos análise nos 14 genomas de *X. fastidiosa* no servidor *CRISPRfinder* online (<http://crispr.u-psud.fr/Server/>) com o objetivo de identificar sequências CRISPR e genes que codificam proteína *cas*. A tabela 17 mostra os possíveis candidatos CRISPR, bem como o número e a posição nos respectivos genomas.

Nas cepas sul-americanas foram identificados candidatos CRISPR somente nas cepas Fb7 e HIB4, já nas cepas norte-americanas foram identificados em todas as cepas (Temecula1, GB514, M23, EB-92.1, Dixon, M12 e Ann1). Todavia, não detectamos genes que codificassem proteínas *cas* em nenhum dos genomas das

cepas analisadas. Essas proteínas são necessárias para a funcionalidade do sistema de imunidade CRISPR/cas (Deveau *et al.*, 2010). Desta forma, apesar de termos encontrado candidatos CRISPRs nos genomas estudados, o sistema de imunidade não seria funcional, uma vez que as proteínas cas estão ausentes. É possível que no passado as cepas de *X. fastidiosa* possuísem o sistema CRISPR/cas completo, mas, por erosão do genoma (Pieretti *et al.*, 2009) os genes que codificavam as proteínas cas foram perdidas do sistema tornando-o não funcional. Desta maneira, sem esse sistema de imunidade, as cepas de *X. fastidiosa* tornaram-se consideravelmente mais permissivas a aquisição de EGMs.

Vários estudos recentes informam que a ausência do Sistema de Imunidade CRISPR-cas torna a bactéria mais suscetível a aquisição de EGMs (Plagens *et al.*, 2012; Richter *et al.*, 2012; Stephen, 2012). Um desses exemplos é o caso da *Pseudomonas aeruginosa*, que demonstrou que a presença do sistema CRISPR/cas não só evita a aquisição de material genético de fagos, mas também broqueia a replicação dos mesmos (Cady *et al.*, 2012). Outros estudos realizados com *Streptococcus agalactiae* mostraram que o sistema CRISPR/cas controla a diversidade de EGMs na população (Lopez-Sanchez *et al.*, 2012). Assim, o sistema CRISPR/cas é uma barreira natural contra EGMs que possam ser introduzidos nos genomas de bactérias hospedeiras.

Tabela 17. Candidatos CRISPR nos genomas das cepas Sul e Norte-Americanas.

Cepa	Candidatos Questionáveis	Posição de CRISPR no genoma	Tamanho total do CRISPR	Tamanho da Sequência consenso das DR^a	Número de sequência consenso das DR^b	Número de espaços
FB7	Candidato 1	1225857..1226005	148	26	3	2
	Candidato 2	2613784..1957153	148	26	3	2
	Candidato 3	2613784..2613869	85	26	2	1
	Candidato 4	2723623..2723771	85	26	3	2
HIB4	Candidato1	506111.. 506211	100	23	2	1
Temecula1	Candidato 1	1851617..1851692	75	24	2	1
GB514	Candidato 1	458857..458932	75	24	2	1
M23	Candidato 1	1867432..1867507	75	24	2	1
M12	Candidato 1	974946..975034	88	32	2	1
	Candidato 2	2258103..2258180	77	26	2	1
Dixon	Candidato 1	329094..329182	88	32	2	1
	Candidato 2	215322..215399	77	26	2	1
	Candidato 3	13416..13516	100	23	2	1
Ann-1	Candidato 1	1083287..1083398	111	31	2	1
EB-92.1	Candidato 1	99474.. 99549	75	24	2	1

A e B (DR) – repetições direta

4.8.4. Filogenia das cepas Sul e Norte-Americanas de *X. fastidiosa*

A Análise de Sequência Multilocus (MLSA) é um método de reconstrução filogenética adotado para a obtenção de uma classificação taxonômica mais robusta de cepas proximamente relacionadas e com alto nível de recombinação genética. O princípio básico deste método consiste essencialmente na concatenação de várias sequências de ortólogos para o mesmo conjunto de organismos, resultando em uma "supermatriz", que é utilizada para inferir a filogenia (Maiden *et al.*, 1998; Hanage *et al.*, 2005). Dentre as vantagens deste método, destacam-se as análises filogenéticas de cepas ou espécies estreitamente relacionadas, bem como a inferência de transferência lateral de genes e recombinação homologa. O MLSA tem um grande potencial de resolução taxonômica devido ao número de caracteres analisados, e a maior sensibilidade ao impacto dos sinais conflitantes (incongruência filogenética, por exemplo), resultantes de eventos de transferência horizontal de genes (Papke *et al.*, 2007).

Desta maneira, adotamos o método de MLSA e o modelo sugerido pelo *jModelTeste 0.1.1* (Posada, 2008) para reconstrução filogenética das cepas *X. fastidiosa*. Como esperado, a filogenia gerada separou as cepas das Américas do Norte e do Sul em dois maiores clados com suporte estatístico (*bootstrap*) aceitável >70% (Figura 33). Entretanto, o clado da América do Norte tem seus ramos internos e externos mais definidos do que os da América do Sul. Essa maior definição dos ramos onde se encontram as cepas norte-americanas deve-se supostamente ao fato destas cepas terem menor número de recombinação genética, possivelmente por enfrentarem uma maior restrição geográfica, terem vetores diferentes e uma gama de hospedeiros que limita eventos de recombinação (Almeida *et al.*, 2008). Em contra partida, para as cepas da América do Sul, os ramos internos e externos praticamente se sobrepõem, provavelmente por ocorrer com maior frequência eventos de recombinação genética entre essas cepas (Nunney *et al.*, 2012). Em estudos evolutivos entre cepas de café e de citros, sugere-se a ocorrência acentuada de recombinação genética entre ambos grupos, pelo fato de compartilharem o mesmo vetor (Almeida *et al.*, 2008). Interessantemente, o número de EGMs compartilhados entre as cepas da América do Norte é menor do que o número de EGMs compartilhados entre as cepas da América do Sul (Tabela 16).

Na filogenia mostrada na Figura 33, a cepa 3124 (isolada de cafeeiro) se encontra na base do clado das cepas de citros, e, portanto provavelmente teve um ancestral comum a estas cepas. Isso é corroborado por estudos de introgressão

genética que sugerem que as cepas de citros e café foram originadas de uma subespécie que infecta ameixa, presente na região desde 1935 (possivelmente *X. fastidiosa* subsp. *multiplex*.) (Nunney *et al.*, 2012). Todavia, a cepa Argentina de citros Fb7, provavelmente compartilhou um ancestral comum com as cepas de citros 9a5c e, portanto, é provável que seja a cepa de citros mais antiga de *X. fastidiosa* já estudada. Quanto à cepa U24d, é um isolado geneticamente e morfológicamente muito parecido com a cepa 9a5c, haja vista os resultados apresentados no item 4.5., e corroborados na filogenia, em que as duas cepas apresentam-se em um mesmo clado.

Desta forma, de acordo com a inferência filogenética, somada as informações a respeito da restrição geográfica que as cepas norte-americanas foram submetidas, sugerimos que provavelmente tenha ocorrido uma divergência alopátrica nas cepas de *X. fastidiosa* naquela região. Todavia, a falta dessa restrição geográfica para com as cepas sul-americanas, exceto a cepa Fb7, provavelmente submeteu essas cepas a uma divergência simpátrica. Diversos estudos têm demonstrado a importância do isolamento geográfico na divergência de isolados de bactérias, em que permitiriam uma pressão seletiva diferenciada, contribuindo assim na evolução de cepas de bactérias (Whitaker *et al.*, 2003; Whitaker, 2006).

Figura 33. Árvore filogenética construída a partir de um grupo de seqüências de 7 genes *housekeeping*. O método filogenético utilizado foi o de Máxima Verossimilhança, com modelo de substituição GTR + G. A árvore foi construída com as seqüências dos genes *leuA* (XF1818), *petC* (XF0910), *malF* (XF2447), *cysG* (XF0832), *hoIC* (XF0136), *nuoL* (XF0316) e *glfT* (XF0656), sendo a identidade das CDS em parenteses aquelas da cepa 9a5c.. As análises foram realizadas utilizando o software PAUP4b10 com 1000 réplicas (*bootstrap*). As árvores foram construídas com *Xanthomonas albilineans* como grupo externo.

4.9. Um novo isolado de *X. fastidiosa* em Cruz das Almas – Bahia

O estado da Bahia é o segundo maior produtor de citros no Brasil, com cerca de 50.000 ha. A Clorose Variegada de Citros (CVC) foi detectada pela primeira vez no estado em 1997. Desde então, vem se disseminando lentamente. Hoje, a CVC ocupa uma grande área no norte da Bahia onde causa grande prejuízo aos citricultores. O Recôncavo Baiano é uma das principais regiões citricultoras do Litoral Norte. Até o momento a CVC só havia sido constatada no Litoral Norte, mas a partir de outubro de 2009 novos focos foram detectados no Recôncavo e Sul do estado (Laranjeira *et al.*, 2008). Em contato com o consultor dos citricultores da região de Cruz das Almas-BA, engenheiro agrônomo Domingos Magalhães, fomos informados que todas as fazendas com pomares de citros, na região de Rio Real, BA, possuem focos de CVC, sendo esta uma doença que mais causa prejuízos econômicos aos citricultores na Bahia (Figura 34).

Diante deste quadro, decidimos em colaboração com o Dr. Abelmon da Silva Gesteira (EMBRAPA Cruz das Almas, Bahia), isolar *X. fastidiosa* nas fazendas mais atingidas por CVC. Para tal realizamos visita a esta região em dezembro de 2011 para coletar material vegetal com o intuito de isolar *X. fastidiosa* de plantas sintomáticas. Além dos sintomas típicos de CVC, observamos que em algumas plantas as folhas apresentavam-se enroladas em torno da nervura principal e com o fruto no final do caule com os sintomas típicos da doença. Tivemos sucesso na obtenção de um isolado de *X. fastidiosa*, o qual aparentemente forma menos biofilme do que a cepa 9a5c. Este novo isolado, que denominamos XRB1 (*Xylella* – Rio Real – Bahia 1), deverá ter seu genoma sequenciado em breve.

Figura 34. Imagens de um pomar com sintomas de CVC na fazenda de citros Valter e Bina, Rio Real, BA. As imagens A, B, C e D ilustram folhas com sintomas de CVC. A figura D apresenta um sintoma na folha de citros peculiar, o enrolamento da folha em torno da nervura principal. As imagens E, F, G e H ilustram as características dos frutos com o citros infectados com *X. fastidiosa*.

5. Conclusão

Realizamos neste trabalho o sequenciamento, a montagem e anotação do genoma de duas cepas de citros, U24d e Fb7, virulentas isoladas de laranjeira com sintomas de CVC, predominantes nos pomares citrícolas, respectivamente, do Estado de São Paulo e em Bella Vista, Corrientes na Argentina, e da cepa 3124, isolada de cafeeiro com sintomas de requeima do cafeeiro, em São José do Rio Preto, São Paulo. Somente a cepa U24d apresentou um plasmídeo idêntico ao plasmídeo pXF51 da cepa 9a5c.

As análises comparativas destes três novos genomas com os genomas das cepas de referência 9a5c (cepas sul-americana isolada de laranjeira) e Temecula1 (cepa norte-americana isolada de videira) permitiram constatar que a cepa 3124 possui um dos maiores genomas até então sequenciado. O genoma da cepa U24d é praticamente idêntico ao da cepa 9a5c. Todavia, os genomas das cepas Fb7 e 3124 apresentam maior colinearidade com a cepa Temecula1 do que com as cepas Sul-Americanas. Os novos genomas sequenciados exibem conteúdo C+G típico da espécie *X. fastidiosa*, com regiões atípicas relacionadas diretamente com a presença de Elementos Genéticos Móveis (EGMs). As análises de genes ortólogos demonstraram um número relativamente parecido entre as cepas sequenciadas e as cepas de referência. Já os genes exclusivos entre essas cepas apresentam disparidade, o que mostra significativa diversidade biológica entre essas cepas da mesma subespécie *pauca*. Destaca-se nesta análise também o gene de uma metaloprotease dependente de zinco que é exclusivo das cepas Fb7 e Hib4 (cepa de hibisco). Este gene pode estar relacionado com a degradação dos pontos de membranas encontrados no xilema da planta hospedeira, facilitando assim, a colonização desta cepa no xilema.

Demonstramos também que os fenótipos das cepas U24d, Fb7 e 3124 apresentam diferenças importantes em relação à cepa de referência 9a5c que podem ser correlacionadas a diferenças genotípicas. Destacamos a ausência de motilidade *twitching* da cepa U24d *in vitro* e *in planta*. Nas análises comparativas do genoma da cepa U24d verificamos uma inserção de 10 nucleotídeos no gene *pilQ* (gene relacionado com a externalização das pilinas do *pilus* tipo IV), a qual gera um códon de parada pré-maturo. A falta da proteína PilQ foi confirmada por *Western blot* e a ausência da produção do *pilus* tipo IV confirmada por microscopia eletrônica de varredura. Apesar desta alteração, a cepa U24d mostrou patogênica, embora sua disseminação sistêmica esteja comprometida.

Outra alteração relevante é a mutação no gene *rpfC* na cepa Fb7 e que possivelmente não produz a proteína sensora RpfC funcional. Interessantemente, esta cepa não forma agregados celulares e biofilme típicos de *X. fastidiosa*, possivelmente pela baixa expressão de genes relacionados com esse fenótipo, por exemplo adesinas fimbrial e afimbrial. A ausência de expressão destes genes parece decorrer da ausência de RpfC funcional. Esta proteína sensora responde ao aumento do DSF extracelular ativando uma via de sinalização mediada por di-cGMP e pelo regulador Clp que controla a expressão de vários genes, entres eles genes relacionados a formação do biofilme. Ao contrário do modelo proposto para cepa Temecula1, na cepa Fb7 a produção de exopolissacarídeo parece não estar co-regulada com a expressão das adesinas, uma vez que o operon *gum* está aparentemente muito expresso na cepa Fb7 apesar da mutação verificada em *rpfC*.

Neste trabalho demonstramos que o pangenoma da *X. fastidiosa* concorda com a Hipótese do Estado da Distribuição do genoma (DGH) e que os EGMs provavelmente contribuíram para tornar o genoma da *Xylella* aberto. Vale ressaltar, que a maioria dos genes de regiões de fagos estão presentes no grupo dos “genes acessórios” do pangenoma. Sendo assim, é provável que o grande número de EGMs encontrados nas cepas de *X. fastidiosa* possam estar relacionados tanto com a divergência destas cepas estreitamente relacionadas, quanto com a aquisição de novos fatores de virulência, patogenicidade e adaptabilidade para com os diferentes hospedeiros.

Verificamos que os EGMs representam ~12% do três novos genomas sequenciados neste trabalho. Além disso, observamos que tanto o número quanto o tamanho desses EGMs são maiores nas cepas sul-americanas do que nas norte-americanas, sendo que se concentram na porção central destes genomas. Interessantemente, as cepas sul-americanas compartilham maior número de EGMs do que as norte-americanas. Uma possível explicação para esse grande número de EGMs nos genomas das cepas de *X. fastidiosa* seria falta do sistema de imunidade CRISPR/cas completo.

Assim, é provável que *X. fastidiosa* tenha tido um ancestral com genoma fechado. Todavia, no decorrer da evolução deste fitopatógeno, eventos de erosão do genoma propiciaram a perda do sistema CRISPR/cas completo, o qual funciona diretamente com uma barreira contra EGMs. Portanto, essa perda provavelmente tornou a *Xylella* mais permissiva a aquisição de muitos EGMs, do tipo profagos e ilhas genômicas, passando assim, a bactéria a ter um genoma aberto.

Finalmente, nossos estudos filogenéticos por MLSA demonstraram, como esperado, uma clara divisão das cepas norte-americanas das cepas sul-americanas.

Observamos que a cepa 3124 posiciona-se na base do clado das cepas de citros, e, portanto provavelmente teve um ancestral comum com estas cepas. Já a cepa Fb7 compartilha um ancestral comum com as cepas de citros 9a5c, e provável que seja a cepa de citros mais antiga de *X. fastidiosa* já estudada. A cepa U24d, como esperado, agrupa com a cepa 9a5c, haja vista as características genômicas e morfológicas compartilhadas por elas. A reconstrução filogenética sugere ainda que as cepas da América do Norte possivelmente tenham sofrido menos recombinação genética do que as cepas da América do Sul, por conta de eventos de divergência alopátricos e simpátricos, respectivamente.

6. Referências

1. **Ahmadian, A., Ehn, M. e Hober, S.** 2006. Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363:83-94.
2. **Akopyants, N. S., Fradkov, A., Diatchenko, L., Hill, J. E., Siebert, P. D., Lukyanov, S. A., Sverdlov, E. D. e Berg, D. E.** 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America*. 95:13108-13113.
3. **Alexeyenko, A., Tamas, I., Liu, G. e Sonnhammer, E. L.** 2006. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* 22:e9-15.
4. **Almeida, R. P., Killiny, N., Newman, K. L., Chatterjee, S., Ionescu, M. e Lindow, S. E.** 2012. Contribution of *rpfB* to cell-to-cell signal synthesis, virulence, and vector transmission of *Xylella fastidiosa*. *Mol Plant Microbe Interact* 25:453-62.
5. **Almeida, R. P., Nascimento, F. E., Chau, J., Prado, S. S., Tsai, C. W., Lopes, S. A. e Lopes, J. R.** 2008. Genetic structure and biology of *Xylella fastidiosa* strains causing disease in citrus and coffee in Brazil. *Appl Environ Microbiol* 74:3690-701.
6. **Almeida, R. P. P., Blua, M. J., Lopes, J. R. S. e Purcell, A. H.** 2005. Vector transmission of *Xylella fastidiosa*: Applying fundamental knowledge to generate disease management strategies. *Annals of the Entomological Society of America* 98:775-786.
7. **Almeida, R. P. P. e Purcell, A. H.** 2006. Patterns of *Xylella fastidiosa* colonization on the precibarium of sharpshooter vectors relative to transmission to plants. *Annals of Entomological Society of America*. 99:84-890.
8. **Altschul, S. F., Gish, W., Miller, W., Myers, E. W. e Lipman, D. J.** 1990. Basic local alignment search tool. *J Mol Biol* 215:403-10.
9. **Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. e Lipman, D. J.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-402.
10. **Alves, E., Leite, B., Marucci, R. C., Pascholati, S. F., Lopes, J. R. e Andersen, P. C.** 2008. Retention sites for *Xylella fastidiosa* in four sharpshooter vectors (Hemiptera: Cicadellidae) analyzed by scanning electron microscopy. *Curr Microbiol* 56:531-8.
11. **Aminov, R. I.** 2011. Horizontal gene exchange in environmental microbiota. *Front Microbiol* 2:158.
12. **Araújo, W. L., Marcon, J., Maccheroni Jr, W., van Elsas, J. D., van Vuurde, J. W. L. e Azevedo, J. L.** 2002. Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. *Applied and Environmental Microbiology*. 68:4906-4914.
13. **Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., Seidman, J. G., Smith, J. A. e Struhl, K.** 1995. *Current Protocols in Molecular Biology*.
14. **Backus, E. A., Andrews, K. B., Shugart, H. J., Carl Greve, L., Labavitch, J. M. e Alhaddad, H.** 2012. Salivary enzymes are injected into xylem by the glassy-winged sharpshooter, a vector of *Xylella fastidiosa*. *J Insect Physiol* 58:949-59.
15. **Backus, E. A. e Morgan, D. J.** 2011. Spatiotemporal colonization of *Xylella fastidiosa* in its vector supports the role of egestion in the inoculation mechanism of foregut-borne plant pathogens. *Phytopathology* 101:912-22.
16. **Bairoch, A.** 2000. The ENZYME database in 2000. *Nucleic Acids Res* 28:304-5.

17. **Banks, D. J., Beres, S. B. e Musser, J. M.** 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends Microbiol* 10:515-21.
18. **Barber, C. E., Tang, J. L., Feng, J. X., Pan, M. Q., Wilson, T. J., Slater, H., Dow, J. M., Williams, P. e Daniels, M. J.** 1997. A novel regulatory system required for pathogenicity of *Xanthomonas campestris* is mediated by a small diffusible signal molecule. *Mol Microbiol* 24:555-66.
19. **Barrangou, R. e Horvath, P.** 2012. CRISPR: new horizons in phage resistance and strain identification. *Annu Rev Food Sci Technol* 3:143-62.
20. **Beiko, R. G., Harlow, T. J. e Ragan, M. A.** 2005. Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A* 102:14332-7.
21. **Bhattacharyya, A., Stilwagen, S., Ivanova, N., D'Souza, M., Bernal, A. et al.** 2002a. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. *Proc Natl Acad Sci U S A* 99:12403-8.
22. **Bhattacharyya, A., Stilwagen, S., Reznik, G., Feil, H., Feil, W. S. et al.** 2002b. Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights. *Genome Res* 12:1556-63.
23. **Binnewies, T. T., Motro, Y., Hallin, P. F., Lund, O., Dunn, D., La, T., Hampson, D. J., Bellgard, M., Wassenaar, T. M. e Ussery, D. W.** 2006. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct Integr Genomics* 6:165-85.
24. **Bowles, D. J.** 1990. Defense-related proteins in higher plants. *Annu Rev Biochem* 59:873-907.
25. **Boyd, E. F. e Brussow, H.** 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol.* 10:560-602.
26. **Bradford, M. M.** 1976. Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248-254.
27. **Brussow, H., Canchaya, C. e Hardt, W. D.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 68:560-602, table of contents.
28. **Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R. e O'Toole, G. A.** 2012. The CRISPR/Cas Adaptive Immune System of *Pseudomonas aeruginosa* Mediates Resistance to Naturally Occurring and Engineered Phages. *J Bacteriol* 194:5728-38.
29. **Canchaya, C., Proux, C., Fournous, G., Bruttin, A. e Brussow, H.** 2003. Prophage genomics. *Microbiol Mol Biol Rev* 67:238-76, table of contents.
30. **Carver, T., Harris, S. R., Berriman, M., Parkhill, J. e McQuillan, J. A.** 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464-9.
31. **Caserta, R., Takita, M. A., Targon, M. L., Rosselli-Murai, L. K., de Souza, A. P., Peroni, L., Stach-Machado, D. R., Andrade, A., Labate, C. A., Kitajima, E. W., Machado, M. A. e de Souza, A. A.** 2010. Expression of *Xylella fastidiosa* fimbrial and afimbrial proteins during biofilm formation. *Appl Environ Microbiol* 76:4250-9.
32. **Chang, C. J., Garnier, M., Zreik, L., Rossetti, V. e Bove, J. M.** 1993. Culture and Serological Detection of the Xylem-Limited Bacterium Causing Citrus Variegated Chlorosis and Its Identification as a Strain of *Xylella fastidiosa*. *Current Microbiology.* 27:137-142.
33. **Chatterjee, S., Almeida, R. P. e Lindow, S.** 2008a. Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annu Rev Phytopathol* 46:243-71.

34. **Chatterjee, S., Killiny, N., Almeida, R. P. e Lindow, S. E.** 2010. Role of cyclic di-GMP in *Xylella fastidiosa* biofilm formation, plant virulence, and insect transmission. *Mol Plant Microbe Interact* 23:1356-63.
35. **Chatterjee, S., Newman, K. L. e Lindow, S. E.** 2008b. Cell-to-cell signaling in *Xylella fastidiosa* suppresses movement and xylem vessel colonization in grape. *Mol Plant Microbe Interact* 21:1309-15.
36. **Chen, J., Civerolo, E. L., Jarret, R. L., Van Sluys, M. A. e de Oliveira, M. C.** 2005. Genetic discovery in *Xylella fastidiosa* through sequence analysis of selected randomly amplified polymorphic DNAs. *Curr Microbiol* 50:78-83.
37. **Chen, J., Hartung, J. S., Chang, C. J. e Vidaver, A. K.** 2002. An evolutionary perspective of Pierce's disease of grapevine, citrus variegated chlorosis, and mulberry leaf scorch diseases. *Curr Microbiol* 45:423-8.
38. **Chen, J., Xie, G., Han, S., Chertkov, O., Sims, D. e Civerolo, E. L.** 2010. Whole genome sequences of two *Xylella fastidiosa* strains (M12 and M23) causing almond leaf scorch disease in California. *J Bacteriol* 192:4534.
39. **Chen, J. C., Groves, R., Zheng, Y. W., Civerolo, E. L., Viveros, M. e Freeman, M.** 2007. Colony morphology of *Xylella fastidiosa* almond leaf scorch strains. *Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie*. 29:225-231.
40. **Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. e Thompson, J. D.** 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497-500.
41. **Chou, F. L., Chou, H. C., Lin, Y. S., Yang, B. Y., Lin, N. T., Weng, S. F. e Tseng, Y. H.** 1997. The *Xanthomonas campestris* gumD gene required for synthesis of xanthan gum is involved in normal pigmentation and virulence in causing black rot. *Biochem Biophys Res Commun* 233:265-9.
42. **Coletta-Filho, H. D., Bittleston, L. S. e Almeida, R. P.** 2011. Spatial genetic structure of a vector-borne generalist pathogen. *Appl Environ Microbiol* 77:2596-601.
43. **Coletta-Filho, H. D. e Machado, M. A.** 2003. Geographical Genetic Structure of *Xylella fastidiosa* from Citrus in Sao Paulo State, Brazil. *Phytopathology* 93:28-34.
44. **Colnaghi Simionato, A. V., da Silva, D. S., Lambais, M. R. e Carrilho, E.** 2007. Characterization of a putative *Xylella fastidiosa* diffusible signal factor by HRGC-EI-MS. *J Mass Spectrom* 42:1375-81.
45. **Conlan, S., Mijares, L. A., Comp Seq Program, N., Becker, J., Blakesley, R. R., Bouffard, G. G., Brooks, S., Coleman, H. L., Gupta, J., Gurson, N., Park, M., Schmidt, B., Thomas, P. J., Young, A., Otto, M., Kong, H. H., Murray, P. R. e Segre, J. A.** 2012. *Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome Biol* 13:R64.
46. **da Costa, P. I., Franco, C. F., Miranda, V. S., Teixeira, D. C. e Hartung, J. S.** 2000. Strains of *Xylella fastidiosa* rapidly distinguished by arbitrarily primed-PCR. *Curr Microbiol* 40:279-82.
47. **da Silva Neto, J. F., Koide, T., Abe, C. M., Gomes, S. L. e Marques, M. V.** 2008. Role of sigma54 in the regulation of genes involved in type I and type IV pili biogenesis in *Xylella fastidiosa*. *Arch Microbiol* 189:249-61.
48. **da Silva, P. I. P.** 2010. Identificação de genes com expressão modulada por estreptomicina e de genes associados à virulência e patogenicidade em *Xylella fastidiosa*., Departamento de Bioquímica, Universidade de São Paulo, São Paulo. 138pp.
49. **da Silva, V. S., Shida, C. S., Rodrigues, F. B., Ribeiro, D. C., de Souza, A. A., Coletta-Filho, H. D., Machado, M. A., Nunes, L. R. e de Oliveira, R. C.** 2007. Comparative genomic characterization of citrus-associated *Xylella fastidiosa* strains. *BMC Genomics* 8:474.

50. **Dagan, T. e Martin, W.** 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci U S A* 104:870-5.
51. **Darling, A. E., Mau, B. e Perna, N. T.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147.
52. **Davis, M. J., French, W. J. e Schaad, N. W.** 1981. Axenic culture of the bacteria associated with phony disease of peach and plum leaf scald. *Current Microbiology*. 6:309-314.
53. **Davis, M. J., Purcell, A. H. e Thomson, S. V.** 1978. Pierce's disease of grapevines: isolation of the causal bacterium. *Science* 199:75-77.
54. **de Coll, O. R., Remes-Lenicove, A. M. M., Agostine, J. P. e Paradell, S.** 2000. Detection of *Xylella fastidiosa* in weeds and sharpshooters in orange groves affected with citrus variegated chlorosis in Misiones., Presented at the 14th International Organization of Citrus Virologist Conference., Argentina.
55. **De La Fuente, L., Burr, T. J. e Hoch, H. C.** 2007. Mutations in type I and type IV pilus biosynthetic genes affect twitching motility rates in *Xylella fastidiosa*. *J Bacteriol* 189:7507-10.
56. **de Mello Varani, A., Souza, R. C., Nakaya, H. I., de Lima, W. C., Paula de Almeida, L. G., Kitajima, E. W., Chen, J., Civerolo, E., Vasconcelos, A. T. e Van Sluys, M. A.** 2008. Origins of the *Xylella fastidiosa* prophage-like regions and their impact in genome differentiation. *PLoS One* 3:e4059.
57. **De Negri, J. D.** 1990. Clorose variegada dos Citros: uma nova anomalia afetando pomares em São Paulo e Minas Gerais . p. 6. *In* Campinas (ed.), Comunicado Técnico nº 82. Extensão Rural, Coord. Ass. Téc. Integral (CATI) ed, Brasil.
58. **de Souza, P. S., de Goes, A., Stuchi, E. S., Jaimes, E. P. G., Wickert, E., da Silva, S. R. e Donadio, L. C.** 2006. Reação de variedades e clones de laranjas a *Xylella fastidiosa*. *Rev. Bras. Frutic.* 28:145-147.
59. **Delcher, A. L., Salzberg, S. L. e Phillippy, A. M.** 2003. Using MUMmer to identify similar regions in large sequence sets., vol. Chapter 10:Unit 10 3.
60. **Deveau, H., Garneau, J. E. e Moineau, S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* 64:475-93.
61. **Dow, J. M., Crossman, L., Findlay, K., He, Y. Q., Feng, J. X. e Tang, J. L.** 2003. Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc Natl Acad Sci U S A* 100:10995-1000.
62. **Dow, J. M. e Daniels, M. J.** 2000. *Xylella* genomics and bacterial pathogenicity to plants. *Yeast* 17:263-71.
63. **Dow, J. M., Davies, H. A. e Daniels, M. J.** 1998. A metalloprotease from *Xanthomonas campestris* that specifically degrades proline/hydroxyproline-rich glycoproteins of the plant extracellular matrix. *Molecular Plant-Microbe Interactions*. 11:1085-1093.
64. **Dow, J. M., Fouhy, Y., Lucey, J. F. e Ryan, R. P.** 2006. The HD-GYP domain, cyclic di-GMP signaling, and bacterial virulence to plants. *Mol Plant Microbe Interact* 19:1378-84.
65. **Dow, M.** 2008. Diversification of the function of cell-to-cell signaling in regulation of virulence within plant pathogenic *Xanthomonas*. *Sci Signal* 1:pe23.
66. **Frost, L. S., Leplae, R., Summers, A. O. e Toussaint, A.** 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 3:722-32.
67. **Fullwood, M. J., Wei, C. L., Liu, E. T. e Ruan, Y.** 2009. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res* 19:521-32.
68. **Galvani, C. D., Li, Y., Burr, T. J. e Hoch, H. C.** 2007. Twitching motility among pathogenic *Xylella fastidiosa* isolates and the influence of bovine serum albumin

- on twitching-dependent colony fringe morphology. *FEMS Microbiol Lett* 268:202-8.
69. **Garth, D. E., N Luisa, H. e Fen, Z. H.** 2008. What makes pathogens pathogenic. *Genome Biology* 9:6-225.
 70. **Ghai, R., Hain, T. e Chakraborty, T.** 2004. GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* 5:198.
 71. **Gmitter Jr, F. G., Chen, C., Machado, M. A., de Souza, A. A., Ollitrault, P., Froehlicher, Y. e Shimizu, T.** 2012. *Tree Genetics & Genomes*, 1^o ed.
 72. **Groth, A. C. e Calos, M. P.** 2004. Phage integrases: biology and applications. *J Mol Biol* 335:667-78.
 73. **Guo, Y., Zhang, Y., Li, J. L. e Wang, N.** 2012. Diffusible signal factor-mediated quorum sensing plays a central role in coordinating gene expression of *Xanthomonas citri* subsp. *citri*. *Mol Plant Microbe Interact* 25:165-79.
 74. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp.:Ser.* 41.
 75. **Hanage, W. P., Fraser, C. e Spratt, B. G.** 2005. Fuzzy species among recombinogenic bacteria. *BMC Biol* 3:3-6.
 76. **Harlow, E. L., D. .** 1988. *Antibodies.*, New York.
 77. **Hartung, J. S., Beretta, J., Brlansky, R. H., Spisso, J. e Lee, R. F.** 1994. Citrus Variegated Chlorosis-Bacterium - Axenic Culture, Pathogenicity, and Serological Relationships with Other Strains of *Xylella fastidiosa*. *Phytopathology*. 84:591-597.
 78. **He, Y. W., Ng, A. Y., Xu, M., Lin, K., Wang, L. H., Dong, Y. H. e Zhang, L. H.** 2007. *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Mol Microbiol* 64:281-92.
 79. **He, Y. W., Wang, C., Zhou, L., Song, H., Dow, J. M. e Zhang, L. H.** 2006. Dual signaling functions of the hybrid sensor kinase RpfC of *Xanthomonas campestris* involve either phosphorelay or receiver domain-protein interaction. *J Biol Chem* 281:33414-21.
 80. **Hill, B. L. e Purcell, A. H.** 1995. Acquisition and retention of *Xylella fastidiosa* by an efficient vector, *Graphocephala atropunctata*. *Phytopathology* 85:2009-2012.
 81. **Hiller, N. L., Janto, B., Hogg, J. S., Boissy, R., Yu, S., Powell, E., Keefe, R., Ehrlich, N. E., Shen, K., Hayes, J., Barbadora, K., Klimke, W., Dernovoy, D., Tatusova, T., Parkhill, J., Bentley, S. D., Post, J. C., Ehrlich, G. D. e Hu, F. Z.** 2007. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. *J Bacteriol* 189:8186-95.
 82. **Hogg, J. S., Hu, F. Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., Post, J. C. e Ehrlich, G. D.** 2007. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol* 8:R103.
 83. **Hopkins, D. L.** 1989. Xylem-Limited Bacterial Pathogen of Plants. *Annual Review of Phytopathology* 27:271-290.
 84. **Hopkins, D. L., Kirkpatrick, B. C., Hill, B. L., Smith, R. e Johnson, D.** 2010. Biological control of Pierce's disease of grapevine with benign strain of *Xylella fastidiosa* subsp. *piercei*.
 85. **Hopkins, D. L. e Purcell, A. H.** 2002. *Xylella fastidiosa*: Cause of Pierce's disease of grapevine and other emergent diseases. *Plant Disease*. 86:1056-1066.
 86. **Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K. et al.** 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* 40:D306-12.

87. **Janse, J. D. e Obradovic, A.** 2010. *Xylella fastidiosa*: its biology, diagnosis, control and risks. *Journal of Plant Pathology* 92:S1.35- S1.48.
88. **Jayaraman, A. e Wood, T. K.** 2008. Bacterial quorum sensing: signals, circuits, and implications for biofilms and disease. *Annu Rev Biomed Eng* 10:145-67.
89. **Jordão, B. P., Terra, W. R., Ribeiro, A. F., Lehane, M. J. e Ferreira, C.** 1996. Trypsin secretion in *Musca domestica* larval midguts: A biochemical and immunocytochemical study. *Insect Biochemistry and Molecular Biology*. 26:337-346.
90. **Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. e Tanabe, M.** 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40:D109-14.
91. **Karginov, F. V. e Hannon, G. J.** 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* 37:7-19.
92. **Killiny, N., Prado, S. S. e Almeida, R. P.** 2010. Chitin utilization by the insect-transmitted bacterium *Xylella fastidiosa*. *Appl Environ Microbiol* 76:6134-40.
93. **Koide, T., Vencio, R. Z. e Gomes, S. L.** 2006. Global gene expression analysis of the heat shock response in the phytopathogen *Xylella fastidiosa*. *J Bacteriol* 188:5821-30.
94. **Koide, T., Zaini, P. A., Moreira, L. M., Vencio, R. Z., Matsukuma, A. Y., Durham, A. M., Teixeira, D. C., El-Dorry, H., Monteiro, P. B., da Silva, A. C., Verjovski-Almeida, S., da Silva, A. M. e Gomes, S. L.** 2004. DNA microarray-based genome comparison of a pathogenic and a nonpathogenic strain of *Xylella fastidiosa* delineates genes important for bacterial virulence. *J Bacteriol* 186:5442-9.
95. **Koonin, E. V. e Wolf, Y. I.** 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* 36:6688-719.
96. **Lacava, P. T., Araújo, W. L., Marcon, J., Maccheroni Jr, W. e Azevedo, J. L.** 2004. Interaction between endophytic bacteria from citrus plants and the phytopathogenic bacteria *Xylella fastidiosa*, causal agent of citrus-variegated chlorosis. *Letters in Applied Microbiology* 39:55-59.
97. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-5.
98. **Lapierre, P. e Gogarten, J. P.** 2009. Estimating the size of the bacterial pan-genome. *Trends Genet* 25:107-10.
99. **Laranjeira, F. F., Silva, L. G., Fonseca, E. L., Silva, S. X. B., Rocha, J. B., Santos-Filho, H. P., Ledo, C. A. S. e Hau, B.** 2008. Prevalence, incidence and distribution of citrus variegated chlorosis in Bahia, Brazil. *Tropical Plant Pathology*. 33:339-447.
100. **Lawrence, J. G. e Hendrickson, H.** 2005. Genome evolution in bacteria: order beneath chaos. *Curr Opin Microbiol* 8:572-8.
101. **Leduc, J. L. e Roberts, G. P.** 2009. Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis* pv. citri. *J Bacteriol* 191:7121-2.
102. **Lee, M. W., Rogers, E. E. e Stenger, D. C.** 2010. Functional characterization of replication and stability factors of an incompatibility group P-1 plasmid from *Xylella fastidiosa*. *Appl Environ Microbiol* 76:7734-40.
103. **Lee, R. F. D. K. S., Beretta, M. J. G., CHAGAS, C. M. e Rossetti, V.** 1991. Citrus variegated chlorosis: a new destructive disease of citrus in Brazil. *Citrus Industry* 72:10-13.
104. **Li, W. B., Pria, W. D., Teixeira, C., Miranda, V. S., Ayres, A. J., Franco, C. F., Costa, M. G., He, C. X., Costa, P. I. e Hartung, J. S.** 2001. Coffee leaf scorch caused by a strain of *Xylella fastidiosa* from citrus. *Plant Disease*. 85:501-505.
105. **Li, W. B., Zreik, L., Fernandes, N. G., Miranda, V. S., Teixeira, D. C., Ayres, A. J., Garnier, M. e Bov, J. M.** 1999. A triply cloned strain of *Xylella fastidiosa*

- multiplies and induces symptoms of citrus variegated chlorosis in sweet orange. *Curr Microbiol* 39:106-8.
106. **Li, Y., Hao, G., Galvani, C. D., Meng, Y., De La Fuente, L., Hoch, H. C. e Burr, T. J.** 2007. Type I and type IV pili of *Xylella fastidiosa* affect twitching motility, biofilm formation and cell-cell aggregation. *Microbiology* 153:719-26.
 107. **Livak, K. J. e Schmittgen, T. D.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25:402-8.
 108. **Lopes, F. F., Neves, M. F., Kalaki, R. B. e Trombin, V. G.** 2010. O Retrato da Citricultura Brasileira, vol. 1º. 2010, Brasil.
 109. **Lopez-Sanchez, M. J., Sauvage, E., Da Cunha, V., Clermont, D., Ratsima Hariniaina, E., Gonzalez-Zorn, B., Poyart, C., Rosinski-Chupin, I. e Glaser, P.** 2012. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol Microbiol* 85:1057-71.
 110. **Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M. e Spratt, B. G.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95:3140-5.
 111. **Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S. et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-80.
 112. **Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N. N. e Kyrpides, N. C.** 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40:D115-22.
 113. **Marques, M. V., da Silva, A. M. e Gomes, S. L.** 2001. Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*. *Plasmid* 45:184-99.
 114. **Marshak, S., Totary, H., Cerasi, E. e Melloul, D.** 1996. Purification of the beta-cell glucose-sensitive factor that transactivates the insulin gene differentially in normal and transformed islet cells. *Proceedings of the National Academy of Sciences of the United States of America*. 93:15057-15062.
 115. **Mattick, J. S.** 2002. Type IV pili and twitching motility. *Annu Rev Microbiol* 56:289-314.
 116. **Medini, D., Donati, C., Tettelin, H., Massignani, V. e Rappuoli, R.** 2005. The microbial pan-genome. *Curr Opin Genet Dev* 15:589-94.
 117. **Meng, Y., Li, Y., Galvani, C. D., Hao, G., Turner, J. N., Burr, T. J. e Hoch, H. C.** 2005. Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. *J Bacteriol* 187:5560-7.
 118. **Meyer, M. M.** 2010. The effects of cold temperature on *Xylella fastidiosa*-infected *Vitis vinifera*. Plant Pathology, University of California, Davis.
 119. **Mira, A., Martin-Cuadrado, A. B., D'Auria, G. e Rodriguez-Valera, F.** 2010. The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol* 13:45-57.
 120. **Moreira, L. M., De Souza, R. F., Digiampietri, L. A., Da Silva, A. C. e Setubal, J. C.** 2005. Comparative analyses of *Xanthomonas* and *Xylella* complete genomes. *OMICS* 9:43-76.
 121. **Newman, K. L., Almeida, R. P., Purcell, A. H. e Lindow, S. E.** 2004. Cell-cell signaling controls *Xylella fastidiosa* interactions with both insects and plants. *Proc Natl Acad Sci U S A* 101:1737-42.
 122. **Newman, K. L., Chatterjee, S., Ho, K. A. e Lindow, S. E.** 2008. Virulence of plant pathogenic bacteria attenuated by degradation of fatty acid cell-to-cell signaling factors. *Mol Plant Microbe Interact* 21:326-34.

123. **Nunney, L., Elfekih, S. e Stouthamer, R.** 2012. The importance of multilocus sequence typing: cautionary tales from the bacterium *Xylella fastidiosa*. *Phytopathology* 102:456-60.
124. **Ochman, H., Lawrence, J. G. e Groisman, E. A.** 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299-304.
125. **Pallen, M. J. e Wren, B. W.** 2007. Bacterial pathogenomics. *Nature* 449:835-42.
126. **Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D. e Doolittle, W. F.** 2007. Searching for species in *haloarchaea*. *Proc Natl Acad Sci U S A* 104:14092-7.
127. **Perez-Donoso, A. G., Sun, Q., Roper, M. C., Greve, L. C., Kirkpatrick, B. e Labavitch, J. M.** 2010. Cell wall-degrading enzymes enlarge the pore size of intervessel pit membranes in healthy and *Xylella fastidiosa*-infected grapevines. *Plant Physiol* 152:1748-59.
128. **Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., Couloux, A., Darrasse, A., Gouzy, J., Jacques, M. A., Lauber, E., Manceau, C., Mangenot, S., Poussier, S., Segurens, B., Szurek, B., Verdier, V., Arlat, M. e Rott, P.** 2009. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics* 10:616.
129. **Pierry, P. M.** 2012. Pirosequenciamento e análise comparativa de genomas do fitopatógeno *Xylella fastidiosa*., Biblioteca Digital de Teses e Dissertações da USP Brasil. 1-148.
130. **Pinel, N., Davidson, S. K. e Stahl, D. A.** 2008. *Verminephrobacter eiseniae* gen. nov., sp. nov., a nephridial symbiont of the earthworm *Eisenia foetida* (Savigny). *Int J Syst Evol Microbiol* 58:2147-57.
131. **Plagens, A., Tjaden, B., Hagemann, A., Randau, L. e Hensel, R.** 2012. Characterization of the CRISPR/Cas subtype I-A system of the *hyperthermophilic crenarchaeon Thermoproteus tenax*. *J Bacteriol* 194:2491-500.
132. **Pollumaa, L., Alamae, T. e Mae, A.** 2012. Quorum sensing and expression of virulence in pectobacteria. *Sensors (Basel)* 12:3327-49.
133. **Pooler, M. R. e Hartung, J. S.** 1995. Specific PCR detection and identification of *Xylella fastidiosa* strains causing citrus variegated chlorosis. *Curr Microbiol* 31:377-81.
134. **Posada, D.** 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25:1253-6.
135. **Punta, M., Coghill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A. e Finn, R. D.** 2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290-301.
136. **Purcell, A. H. e Hopkins, D. L.** 1996. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* 34:131-151.
137. **Ramos, C. R., Abreu, P. A., Nascimento, A. L. e Ho, P. L.** 2004. A high-copy T7 *Escherichia coli* expression vector for the production of recombinant proteins with a minimal N-terminal His-tagged fusion peptide. *Braz J Med Biol Res* 37:1103-9.
138. **Randall, J. J., Goldberg, N. P., Kemp, J. D., Radionenko, M., French, J. M., Olsen, M. W. e Hanson, S. F.** 2009. Genetic analysis of a novel *Xylella fastidiosa* subspecies found in the southwestern United States. *Appl Environ Microbiol* 75:5631-8.
139. **Rasko, D. A., Rosovitz, M. J., Myers, G. S., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R., Chaudhuri, R., Henderson, I. R., Sperandio, V. e Ravel, J.** 2008. The pangenome structure of

- Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol 190:6881-93.
140. **Redak, R. A., Purcell, A. H., Lopes, J. R., Blua, M. J., Mizell, R. F., 3rd e Andersen, P. C.** 2004. The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology. Annu Rev Entomol 49:243-70.
 141. **Reddy, V. S., Kumar, Y. N., Raghavendra, A., Sowjanya, G., Kumar, S., Ramyasree, G. e Reddy, G. R.** 2012. In silico model of DSF synthase RpfF protein from *Xanthomonas oryzae* pv. Oryzae: a novel target for bacterial blight of rice disease. Bioinformation 8:504-7.
 142. **Remm, M., Storm, C. E. e Sonnhammer, E. L.** 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314:1041-52.
 143. **Richter, H., Zoepfel, J., Schermuly, J., Maticzka, D., Backofen, R. e Randau, L.** 2012. Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic Acids Res 40:9887-96.
 144. **Rodriguez-Valera, F., Martin-Cuadrado, A. B., Rodriguez-Brito, B., Pasic, L., Thingstad, T. F., Rohwer, F. e Mira, A.** 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7:828-36.
 145. **Roper, M. C., Greve, L. C., Warren, J. G., Labavitch, J. M. e Kirkpatrick, B. C.** 2007. *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. Mol Plant Microbe Interact 20:411-9.
 146. **Rosato, Y. B., Neto, J. R., Miranda, V. S., Carlos, E. F. e Manfio, G. P.** 1998. Diversity of a *Xylella fastidiosa* population isolated from *Citrus sinensis* affected by Citrus variegated chlorosis in Brazil. Systematic and Applied Microbiology. 21:593-598.
 147. **Rossetti, V. e De Negri, J. D.** 1990. Clorose variegada dos citros. Laranja 11:1-14.
 148. **Rossetti, V., Garnier, M., Beretta, M. J. G., Teixeira, A. R. R., Quaggio, J. A., Battaglia, O. C., Gomes, M. P. e Negri, J. D.** 1990. Resultados preliminares de estudos sobre uma nova anormalidade dos citros observada nos Estados de São Paulo e Minas Gerais. Summa Phytopathologica 16:1-13.
 149. **Rozen, S. e Skaletsky, H.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-86.
 150. **Ryan, R. P. e Dow, J. M.** 2008. Diffusible signals and interspecies communication in bacteria. Microbiology 154:1845-58.
 151. **Ryan, R. P., Fouhy, Y., Lucey, J. F., Crossman, L. C., Spiro, S., He, Y. W., Zhang, L. H., Heeb, S., Camara, M., Williams, P. e Dow, J. M.** 2006. Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. Proc Natl Acad Sci U S A 103:6712-7.
 152. **Ryan, R. P., Fouhy, Y., Lucey, J. F., Jiang, B. L., He, Y. Q., Feng, J. X., Tang, J. L. e Dow, J. M.** 2007. Cyclic di-GMP signalling in the virulence and environmental adaptation of *Xanthomonas campestris*. Mol Microbiol 63:429-42.
 153. **Ryan, R. P., McCarthy, Y., Andrade, M., Farah, C. S., Armitage, J. P. e Dow, J. M.** 2010. Cell-cell signal-dependent dynamic interactions between HD-GYP and GGDEF domain proteins mediate virulence in *Xanthomonas campestris*. Proc Natl Acad Sci U S A 107:5989-94.
 154. **Ryan, R. P., Vorholter, F. J., Potnis, N., Jones, J. B., Van Sluys, M. A., Bogdanove, A. J. e Dow, J. M.** 2011. Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. Nat Rev Microbiol 9:344-55.
 155. **Scarpari, L. M., Lambais, M. R., Silva, D. S., Carraro, D. M. e Carrer, H.** 2003. Expression of putative pathogenicity-related genes in *Xylella fastidiosa* grown at low and high cell density conditions in vitro. FEMS Microbiol Lett 222:83-92.

156. **Schreiber, H. L. I., Koirala, M., Lara, A., Ojeda, M., Dowd, S. E., Bextine, B. e Morano, L.** 2010. Unraveling the first *Xylella fastidiosa* subsp. *fastidiosa* genome from Texas. *Southwestern Entomologist* 35:479-483.
157. **Schuenzel, E. L., Scally, M., Stouthamer, R. e Nunney, L.** 2005. A multigene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen *Xylella fastidiosa*. *Appl Environ Microbiol* 71:3832-9.
158. **Selengut, J. D., Haft, D. H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W. C., Richter, A. R. e White, O.** 2007. TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Research* 35:D260-4.
159. **Sherrier, D. J. e Vandenbosch, K. A.** 1994. Localization of Repetitive Proline-Rich Proteins in the Extracellular-Matrix of Pea Root-Nodules. *Protoplasma*. 183:148-161.
160. **Silva, M. S., De Souza, A. A., Takita, M. A., Labate, C. A. e Machado, M. A.** 2011. Analysis of the biofilm proteome of *Xylella fastidiosa*. *Proteome Sci* 9:58.
161. **Simpson, A. J., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M. et al.** 2000a. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406:151-7.
162. **Simpson, A. J., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M. et al.** 2000b. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature* 406:151-9.
163. **Slater, H., Alvarez-Morales, A., Barber, C. E., Daniels, M. J. e Dow, J. M.** 2000. A two-component system involving an HD-GYP domain protein links cell-cell signalling to pathogenicity gene expression in *Xanthomonas campestris*. *Mol Microbiol* 38:986-1003.
164. **Smith, P. K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H. et al.** 1985. Measurement of protein using bicinchoninic acid. *Anal Biochem* 150:76-85.
165. **Sondermann, H., Shikuma, N. J. e Yildiz, F. H.** 2012. You've come a long way: c-di-GMP signaling. *Curr Opin Microbiol* 15:140-6.
166. **Stenger, D. C., Lee, M. W., Rogers, E. E. e Chen, J.** 2010. Plasmids of *Xylella fastidiosa* mulberry-infecting strains share extensive sequence identity and gene complement with pVEIS01 from the earthworm symbiont *Verminephrobacter eiseniae*. *Physiological and Molecular Plant Pathology* 74:238-245.
167. **Stephen, T. A.** 2012. Bacterial 'immunity' against bacteriophages. *Bacteriophage* 2:50–54.
168. **Studholme, D. J., Downie, J. A. e Preston, G. M.** 2005. Protein domains and architectural innovation in plant-associated Proteobacteria. *BMC Genomics* 6:17.
169. **Sun, Q., Greve, L. C. e Labavitch, J. M.** 2011. Polysaccharide compositions of intervessel pit membranes contribute to Pierce's disease resistance of grapevines. *Plant Physiol* 155:1976-87.
170. **Szczepankowska, A.** 2012. Role of CRISPR/cas system in the development of bacteriophage resistance. *Adv Virus Res* 82:289-338.
171. **Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J. e Natale, D. A.** 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41.
172. **Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D. et al.** 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102:13950-5.

173. **Tettelin, H., Riley, D., Cattuto, C. e Medini, D.** 2008. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* 11:472-7.
174. **Thompson, J. D., Gibson, T. J. e Higgins, D. G.** 2002. Multiple sequence alignment using ClustalW and ClustalX., vol. Chapter 2:Unit 2 3.
175. **Torres, P. S., Malamud, F., Rigano, L. A., Russo, D. M., Marano, M. R., Castagnaro, A. P., Zorreguieta, A., Bouarab, K., Dow, J. M. e Vojnov, A. A.** 2007. Controlled synthesis of the DSF cell-cell signal is required for biofilm formation and virulence in *Xanthomonas campestris*. *Environ Microbiol* 9:2101-9.
176. **Vale, P. F. e Little, T. J.** 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. *Proc Biol Sci* 277:2097-103.
177. **Van Sluys, M. A., de Oliveira, M. C., Monteiro-Vitorello, C. B., Miyaki, C. Y., Furlan, L. R. et al.** 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *J Bacteriol* 185:1018-26.
178. **Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E., Menck, C. F., Da Silva, A. C., Ferro, J. A., Oliveira, M. C., Setubal, J. C., Kitajima, J. P. e Simpson, A. J.** 2002. Comparative genomic analysis of plant-associated bacteria. *Annu Rev Phytopathol* 40:169-89.
179. **Voegel, T. M., Warren, J. G., Matsumoto, A., Igo, M. M. e Kirkpatrick, B. C.** 2010. Localization and characterization of *Xylella fastidiosa* haemagglutinin adhesins. *Microbiology* 156:2172-9.
180. **Wang, L. H., He, Y., Gao, Y., Wu, J. E., Dong, Y. H., He, C., Wang, S. X., Weng, L. X., Xu, J. L., Tay, L., Fang, R. X. e Zhang, L. H.** 2004. A bacterial cell-cell communication signal with cross-kingdom structural analogues. *Mol Microbiol* 51:903-12.
181. **Wells, J. M., Raju, B. C., Hung, H.-Y., Weisburg, W. G., Mandelco-Paul, L. e Brenner, D. J.** 1987. *Xylella fastidiosa* gen. nov., sp. nov: Gram-Negative, Xylem-Limited, Fastidious Plant Bacteria Related to *Xanthomonas* spp. . *International Journal of Systematic Bacteriology* 37:36-143.
182. **Whitaker, R. J.** 2006. Allopatric origins of microbial species. *Philos Trans R Soc Lond B Biol Sci* 361:1975-84.
183. **Whitaker, R. J., Grogan, D. W. e Taylor, J. W.** 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301:976-8.
184. **Wilgenbusch, J. C. e Swofford, D.** 2003. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics Chapter 6:Unit 6 4*.
185. **Wommack, K. E. e Colwell, R. R.** 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* 64:69-114.
186. **Yang, L., Hong, L., Takahashi, Y., Chen, F., Walker, M. A. e Civerolo, E. L.** 2011. Proteomic analysis of grapevine stem in response to *Xylella fastidiosa* inoculation. *Physiological and Molecular Plant Pathology* 75:90-99.
187. **Zaini, P. A., De La Fuente, L., Hoch, H. C. e Burr, T. J.** 2009. Grapevine xylem sap enhances biofilm development by *Xylella fastidiosa*. *FEMS Microbiol Lett* 295:129-34.
188. **Zaini, P. A., Fogaca, A. C., Lupo, F. G., Nakaya, H. I., Vencio, R. Z. e da Silva, A. M.** 2008. The iron stimulon of *Xylella fastidiosa* includes genes for type IV pilus and colicin V-like bacteriocins. *J Bacteriol* 190:2368-78.
189. **Zhang, L. H.** 2010. A novel C-di-GMP effector linking intracellular virulence regulon to quorum sensing and hypoxia sensing. *Virulence* 1:391-4.
190. **Zhang, S., Flores-Cruz, Z., Kumar, D., Chakrabarty, P., Hopkins, D. L. e Gabriel, D. W.** 2011. The *Xylella fastidiosa* biocontrol strain EB92-1 genome is very similar and syntenic to Pierce's disease strains. *J Bacteriol* 193:5576-7.
191. **Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. e Wishart, D. S.** 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39:W347-52.

7. Apêndice

Figura S1. Alinhamento das sequências de nucleotídeos da CDS pilQ das cepas 9a5c e U24d de *Xylalla fastidiosa*. A inserção que resulta em uma mutação de *fremeshift* está indicada por um retângulo em vermelho. (*) indica nucleotídeos idênticos.

```

XF_9a5c      ATGATTGATTTAAGGGATAGTATGATGACTTTATTTGAAGTCTTAGGCTTGCCTTTTTTT 60
XF_U24d      ATGATTGATTTAAGGGATAGTATGATGACTTTATTTGAAGTCTTAGGCTTGCCTTTTTTT 60
*****

XF_9a5c      AGGCGTATAGCCTCGGTCCGCGTATGTGCAATTTGGGTATGCATGGCTTTCGCTTGGTGT 120
XF_U24d      AGGCGTATAGCCTCGGTCCGCGTATGTGCAATTTGGGTATGCATGGCTTTCGCTTGGTGT 120
*****

XF_9a5c      TGCATCGTAGGAACGGTGTCTGGCTGAAAACAAGCAAGCGATTCCAGTGCCTAAAGTTGCC 180
XF_U24d      TGCATCGTAGGAACGGTGTCTGGCTGAAAACAAGCAAGCGATTCCAGTGCCTAAAGTTGCC 180
*****

XF_9a5c      AATGCGCCACTAAGCGTTTCCAAGATCGATTTCAAACGTGGTGATGATGGCTCTGGCCGT 240
XF_U24d      AATGCGCCACTAAGCGTTTCCAAGATCGATTTCAAACGTGGTGATGATGGCTCTGGCCGT 240
*****

XF_9a5c      CTGATTTTGAAGTTTGATGGGCAGGGAGCCACTCCTGACCTGCGTACGCAGGGTGGCACC 300
XF_U24d      CTGATTTTGAAGTTTGATGGGCAGGGAGCCACTCCTGACCTGCGTACGCAGGGTGGCACC 300
*****

XF_9a5c      GTGCTGGTTGACCTGGGGACAGCTGTGTTGCCGACTGTCTTGCAACGCCAGATCAACGTT 360
XF_U24d      GTGCTGGTTGACCTGGGGACAGCTGTGTTGCCGACTGTCTTGCAACGCCAGATCAACGTT 360
*****

XF_9a5c      GTTGACTTTTGCTACCCAGTGCAGCGTATCGACGCTAAGCCTATGGGTAAGGGTGACACAA 420
XF_U24d      GTTGACTTTTGCTACCCAGTGCAGCGTATCGACGCTAAGCCTATGGGTAAGGGTGACACAA 420
*****

XF_9a5c      CTTGTGTTGAGCACAAAGGGAGCGTTTGAGTCGCTTGCTTACCAGACAGGGGATGAATAC 480
XF_U24d      CTTGTGTTGAGCACAAAGGGAGCGTTTGAGTCGCTTGCTTACCAGACAGGGGATGAATAC 480
*****

XF_9a5c      GTGGTTGAAATTGTGCCACGCAAGGGAGAGGCTGCCATGGGTGGGGCGATCACCCAGAA 540
XF_U24d      GTGGTTGAAATTGTGCCACGCAAGGGAGAGGCTGCCATGGGTGGGGCGATCACCCAGAA 540
*****

XF_9a5c      TCGGTCAGTAAGTCCGCAGCGAAGATTTCCGCACAAGGGTACAGTGGTCGTCCAGTAACG 600
XF_U24d      TCGGTCAGTAAGTCCGCAGCGAAGATTTCCGCACAAGGGTACAGTGGTCGTCCAGTAACG 600
*****

XF_9a5c      TTTAACTTCCAAGACGTTCCAGTACGAACGGTTTTCAGTTGATCGCTGATGAGTCCAAT 660
XF_U24d      TTTAACTTCCAAGACGTTCCAGTACGAACGGTTTTCAGTTGATCGCTGATGAGTCCAAT 660

```



```

*****

XF_9a5c      CTCAATATTGTTGCTTCCGACACTGTGCAAGGCAATGTCACTCTACGTTTGATTAATGTG  720
XF_U24d      CTCAATATTGTTGCTTCCGACACTGTGCAAGGCAATGTCACTCTACGTTTGATTAATGTG  720
*****

XF_9a5c      CCATGGGATCAAGCGCTCGACATCGTTTTGCGTGCTAAAGGATTGGATAAGCGCCGTGAT  780
XF_U24d      CCATGGGATCAAGCGCTCGACATCGTTTTGCGTGCTAAAGGATTGGATAAGCGCCGTGAT  780
*****

XF_9a5c      GAGAAAGTAATTTGGGTTGCGCCGCAACAGGAGCTTGCTAAATACGAGCAAGAAAAAGAG  840
XF_U24d      GAGAAAGTAATTTGGGTTGCGCCGCAACAGGAGCTTGCTAAATACGAGCAAGAAAAAGAG  840
*****

XF_9a5c      GACGCGCGGATCGCGATTGAAAATCGTGAGGGGTTAATTACCGATTATGTCCAAATAAAT  900
XF_U24d      GACGCGCGGATCGCGATTGAAAATCGTGAGGGGTTAATTACCGATTATGTCCAAATAAAT  900
*****

XF_9a5c      TATCATTGCGCCACGGTCATTTTTAAGGCACTGACAGAAGCCAAAGGTATGGGTGGGGGG  960
XF_U24d      TATCATTGCGCCACGGTCATTTTTAAGGCACTGACAGAAGCCAAAGGTATGGGTGGGGGG  960
*****

XF_9a5c      GGTAACGGTACACCTAATAATAATAGTGATGATGCTTTCTTATCGCCACGTGGTCGGCTT  1020
XF_U24d      GGTAACGGTACACCTAATAATAATAGTGATGATGCTTTCTTATCGCCACGTGGTCGGCTT  1020
*****

XF_9a5c      GTAGCTGATGAGCGCACTAATACGTTGATGATTAGCGATATTCCGAAGAAAGTCGCGCAA  1080
XF_U24d      GTAGCTGATGAGCGCACTAATACGTTGATGATTAGCGATATTCCGAAGAAAGTCGCGCAA  1080
*****

XF_9a5c      ATGCGTACTCTCATCCAGCATATTGATCGTCCTGTGGACCAAGTGTGATTGAGGGCCGC  1140
XF_U24d      ATGCGTACTCTCATCCAGCATATTGATCGTCCTGTGGACCAAGTGTGATTGAGGGCCGC  1140
*****

XF_9a5c      ATCGTGATCGCTACCGATAGCTTCGCGCGTGATTGCGGGCCAAGTTTGGTGTTGGTGCT  1200
XF_U24d      ATCGTGATCGCTACCGATAGCTTCGCGCGTGATTGCGGGCCAAGTTTGGTGTTGGTGCT  1200
*****

XF_9a5c      GCTAGCCGTTTCAGCGATAAACTGCAACGATCG-----GTAGTCACGTCACTAC  1250
XF_U24d      GCTAGCCGTTTCAGCGATAAACTGCAACGATCGGTAGTCACGTGTAGTCACGTCACTAC  1260
*****

XF_9a5c      TGCTGATAGCTCTACATCGGCTACGCGGGGGCTTAATGTTGACTTGGGCGGTCCGACCAA  1310
XF_U24d      TGCTGATAGCTCTACATCGGCTACGCGGGGGCTTAATGTTGACTTGGGCGGTCCGACCAA  1320
*****

XF_9a5c      CACCACTGCTTCAGTTCTCCCGAGTTTAGCCTACACTTTGCTTGACCTAAATTTAATCT  1370
XF_U24d      CACCACTGCTTCAGTTCTCCCGAGTTTAGCCTACACTTTGCTTGACCTAAATTTAATCT  1380
*****

XF_9a5c      AGATCTTGAGCTTTCGGCATTCGACGAGGAATCCCGTGGCGAGGTGATTCAAACCCCCG  1430

```

XF_U24d	AGATCTTGAGCTTTCGGCATTCGAGCAGGAATCCCGTGGCGAGGTGATTCAAACCCCCG	1440

XF_9a5c	TATCGTCACTGCAAACCAGCGTGAGGGATATATCAAACAAGGTAAGGAAATTGGGTATGT	1490
XF_U24d	TATCGTCACTGCAAACCAGCGTGAGGGATATATCAAACAAGGTAAGGAAATTGGGTATGT	1500

XF_9a5c	CACGATCACTGGAGGAGTCGCTGGTGGTCAAGCCACTCCGAATGTCCAATTTAAAGAGGC	1550
XF_U24d	CACGATCACTGGAGGAGTCGCTGGTGGTCAAGCCACTCCGAATGTCCAATTTAAAGAGGC	1560

XF_9a5c	TGTTTTGGAGTTGCGTGTGACCCCAACGATCACCAATGACAACCGTGTGTTTCTCAACAT	1610
XF_U24d	TGTTTTGGAGTTGCGTGTGACCCCAACGATCACCAATGACAACCGTGTGTTTCTCAACAT	1620

XF_9a5c	GACTGTCAAAAAGGACGAGATTGATCAAATGATTACGATCGCCAATTTTCGGTACCGTGCC	1670
XF_U24d	GACTGTCAAAAAGGACGAGATTGATCAAATGATTACGATCGCCAATTTTCGGTACCGTGCC	1680

XF_9a5c	GACCTTGAACAAGCGTGAGATCAACACTGCCGTATTGGTTGATGATGGTCAAACGTGTGT	1730
XF_U24d	GACCTTGAACAAGCGTGAGATCAACACTGCCGTATTGGTTGATGATGGTCAAACGTGTGT	1740

Figura S2. Alinhamento das sequências de nucleotídeos da CDS rpfC das cepas 9a5c e Fb7 de *Xylalla fastidiosa*. A inserção que resulta em uma mutação de *fremeshift* está indicada por uma seta preta. (*) indica nucleotídeos idênticos.

```

rpfC_9a5c      ATGCTCAAGAAAGCTAAATGGCTATGCAGTCGGCTATCGCAGCGCCGGATAGTGAGCAT  60
rpfC_Fb7      ATGCTCAAGAAAGCTAAATGGCTATGCAGTCGGCTATCGCAGCGCCGGATAGTGAGCAT  60
                *****

rpfC_9a5c      GTCCAAAACCTGATCAGACTTGTGATCGTCTCGCTTTTCATCATTTACCTAAACGCATGT  120
rpfC_Fb7      GTCCAAAACCTGATCAGACTTGTGATCGTCTCGCTTTTCATCATTTACCTAAACGCATGT  120
                *****

rpfC_9a5c      TGGGATGGTGAGACTGACCTCAATACATCACGGATCATCTGGTCGGTGCTGCTCTCCGAT  180
rpfC_Fb7      TGGGATGGTGAGACTGACCTCAATACATCACGGATCATCTGGTCGGTGCTGCTCTCCGAT  180
                *****

rpfC_9a5c      CTTGGTGTGTCGCTGGCGCTGATGCTTGCGATTGTCATGCATCCGCAGATCTCGCACGTG  240
rpfC_Fb7      CTTGGTGTGTCGCTGGCGCTGATGCTTGCGATTGTCATGCATCCGCAGATCTCGCACGTG  240
                *****

rpfC_9a5c      CGGCGTTGTATTGGCATCATTTGCCGACTACACATCACTGACGGTGCTGATGTTGCTGATG  300
rpfC_Fb7      CGGCGTTGTATTGGCATCATTTGCCGACTACACATCACTGACGGTGCTGATGTTGCTGATG  300
                *****

rpfC_9a5c      GGCAGAGCGGGTTCGCCGCTGTATGTCTATGTTTGTGGGTAACCATCGGTAATGGCTTA  360
rpfC_Fb7      GGCAGAGCGGGTTCGCCGCTGTATGTCTATGTTTGTGGGTAACCATCGGTAATGGCTTA  360
                *****

rpfC_9a5c      CGTTATGGCTGTGCGTATTTGCTGGTCACTACGACCCTTAGTGCGTTGAGTTTCCTGATG  420
rpfC_Fb7      CGTTATGGCTGTGCGTATTTGCTGGTCACTACGACCCTTAGTGCGTTGAGTTTCCTGATG  420
                *****

rpfC_9a5c      GTGATCCTCGTTTCAGCGTACTGGAAGTCGAATCCGTTTCTGGCTTGGGGATTACTGATT  480
rpfC_Fb7      GTGATCCTCGTTTCAGCGTACTGGAAGTCGAATCCGTTTCTGGCTTGGGGATTACTGATT  480
                *****

rpfC_9a5c      GGTCTGATTGCCATCCCTTGGTATTTTCAGTCGCTGCTGAAGGCATTGATCCAAGCCCTT  540
rpfC_Fb7      GGTCTGATTGCCATCCCTTGGTATTTTCAGTCGCTGCTGAAGGCATTGATCCAAGCCCTT  540
                *****

rpfC_9a5c      AAGGATGCACGTCATGCTAACGAGGCTAAGAGTCGCTTCCTGGCCAA-TATGAGCCATGA  599
rpfC_Fb7      AAGGATGCACGTCATGCTAACGAGGCTAAGAGTCGCTTCCTGGCCAAAGTATGAGCCATGA  600
                *****↑*****

rpfC_9a5c      ATTCGCACGCCGCTGAATGGTTTGTGCGGTATCACGGAGGTGCTGGCCACAACGCGTTT  659
rpfC_Fb7      ATTCGCACGCCGCTGAATGGTTTGTGCGGTATCACGGAGGTGCTGGCCACAACGCGTTT  660
                *****

```

rpfc_9a5c	GGATGCCGAGCAAAGGGAATGCTTGAAAGCGATCCAGGCATCGGTACACAGTGTGTTGTC	719
rpfc_Fb7	GGATGCCGAGCAAAGGGAATGCTTGAAAGCGATCCAGGCATCGGTACACAGTGTGTTGTC	720

rpfc_9a5c	CTTGATTGAGGAAGTACTCGACATCTCAAGAATTGAGGCAGGTAAGATCCGTATCCTCCA	779
rpfc_Fb7	CTTGATTGAGGAAGTACTCGACATCTCAAGAATTGAGGCAGGTAAGATCCGTATCCTCCA	780

rpfc_9a5c	CGATACTTTCTCCTTGCAGGAGGTGATTGGCAGTGTGGTCTGATCCTGCAACCACAGAT	839
rpfc_Fb7	CGATACTTTCTCCTTGCAGGAGGTGATTGGCAGTGTGGTCTGATCCTGCAACCACAGAT	840

rpfc_9a5c	CCGCAATAAAGTGTGGAATATCGTGTCGATGTGGGGTTGGATGTTCCCGAATGGTTGTT	899
rpfc_Fb7	CCGCAATAAAGTGTGGAATATCGTGTCGATGTGGGGTTGGATGTTCCCGAATGGTTGTT	900

rpfc_9a5c	CGGCGATGCTGGTTATGTCCGTCAGGTGTTATTGAACATTGTCGGTAATGCCGTCAGTT	959
rpfc_Fb7	CGGCGATGCTGGTTATGTCCGTCAGGTGTTATTGAACATTGTCGGTAATGCCGTCAGTT	960

rpfc_9a5c	TACCGAGCGTGGTGGGATTACATTATGTGTCAGCGTGTGAACCGTTCTAGAGAGCGCAG	1019
rpfc_Fb7	TACCGAGCGTGGTGGGATTACATTATGTGTCAGCGTGTGAACCGTTCTAGAGAGCGCAG	1020

rpfc_9a5c	TGTGGTGTGCGTTTGAAGTTGCAGATACTGGCATTGGCGTGCCATTGGGAATGCGGGC	1079
rpfc_Fb7	TGTGGTGTGCGTTTGAAGTTGCAGATACTGGCATTGGCGTGCCATTGGGAATGCGGGC	1080

rpfc_9a5c	TCGGCTGTTTCGACATGTTTCGAGCAGGCTGATGTTGGTTTGGATCGTCGTTATCAAGGAAG	1139
rpfc_Fb7	TCGGCTGTTTCGACATGTTTCGAGCAGGCTGATGTTGGTTTGGATCGTCGTTATCAAGGAAG	1140

rpfc_9a5c	CGGACTGGGTACCGCGATTGCTAAGGGTTTGATTGAATTGATGGGTGGAAGCATCGGGTT	1199
rpfc_Fb7	CGGACTGGGTACCGCGATTGCTAAGGGTTTGATTGAATTGATGGGTGGAAGCATCGGGTT	1200

rpfc_9a5c	CGAGGAGAACGTCCCTTGTGGGAGCTTGTTTTGGTTCGAATTGCCGTTTGAAC TTGGAAA	1259
rpfc_Fb7	CGAGGAGAACGTCCCTTGTGGGAGCTTGTTTTGGTTCGAATTGCCGTTTGAAC TTGGAAA	1260

rpfc_9a5c	GGAACCGGTGGTCGCGCAAGAGCTACCCGAAGCATCGCAGTGTACGGAAGCGAGTAATGT	1319
rpfc_Fb7	GGAACCGGTGGTCGCGCAAGAGCTACCCGAAGCATCGCAGTGTACGGAAGCGAGTAATGT	1320

rpfc_9a5c	CATTGCGTTTTTCCAATCCATTCTACGGCATCGTGCCCGTGTCAAGAGCCTGCGCGTGCT	1379
rpfc_Fb7	CATTGCGTTTTTCCAATCCATTCTACGGCATCGTGCCCGTGTCAAGAGCCTGCGCGTGCT	1380

rpfc_9a5c	GGTCGCTGATGACCATGAAACTAACCGCATGGTGTATCCAGAATCCTTGATAAGGCGGG	1439
rpfc_Fb7	GGTCGCTGATGACCATGAAACTAACCGCATGGTGTATCCAGAATCCTTGATAAGGCGGG	1440

```

*****

rpfc_9a5c      ACACAAGCTGTTATGCGTGGACGGTGC GGAAGCAGTGTGGATTGTTTGATCAGTATGGA 1499
rpfc_Fb7       ACACAAGCTGTTATGCGTGGACGGTGC GGAAGCAGTGTGGATTGTTTGATCAGTATGGA 1500
*****

rpfc_9a5c      ATTCGATGTTGTGATCATCGATCTGCACATGCCTGGAATGAGCGGTTTGGAGATGCTCAA 1559
rpfc_Fb7       ATTCGATGTTGTGATCATCGATCTGCACATGCCTGGAATGAGCGGTTTGGAGATGCTCAA 1560
*****

rpfc_9a5c      ACAGTTGCGGGTCATGCAAAGTGTGGTCTGCCTTATACACCGGTGCTGGTGCTGAGTGC 1619
rpfc_Fb7       ACAGTTGCGGGTCATGCAAAGTGTGGTCTGCCTTATACACCGGTGCTGGTGCTGAGTGC 1620
*****

rpfc_9a5c      CGATGCGGCGTCCGATTTCGATCCGCTGTTGTAAAGAGGCAGGGGCGCGTGCCTTCCTGTC 1679
rpfc_Fb7       CGATGCGGCGTCCGATTTCGATCCGCTGTTGTAAAGAGGCAGGGGCGCGTGCCTTCCTGTC 1680
*****

rpfc_9a5c      CAAACCCGTGCTTGCTACCAAGCTGCTGGATGTACTTGCCGAAATTGCTGAAGGGGATGG 1739
rpfc_Fb7       CAAACCCGTGCTTGCTACCAAGCTGCTGGATGTACTTGCCGAAATTGCTGAAGGGGATGG 1740
*****

rpfc_9a5c      TGGTGCATTACCTTACGCTGCATTATCGCAATCCTCAGTTCAGCAGCATCAGGTGTTCTGA 1799
rpfc_Fb7       TGGTGCATTACCTTACGCTGCATTATCGCAATCCTCAGTTCAGCAGCATCAGGTGTTCTGA 1800
*****

rpfc_9a5c      TCCGACGGGTTTGGATGATCTGGCTGAGTTGGATATGGACGATGGTTTCGAAGAGCGGTT 1859
rpfc_Fb7       TCCGACGGGTTTGGATGATCTGGCTGAGTTGGATATGGACGATGGTTTCGAAGAGCGGTT 1860
*****

rpfc_9a5c      CATAGTGCAGTCTTTGCGAGATGCCGAACACTGCCAGCAGACAATGTTGTCTGCCGCTGA 1919
rpfc_Fb7       CATAGTGCAGTCTTTGCGAGATGCCGAACACTGCCAGCAGACAATGTTGTCTGCCGCTGA 1920
*****

rpfc_9a5c      AACTAACCAGTGGCAGGAGGTGCGGGAACAAGCGCATGCACTGCGTGGTGTGCTCGGTCA 1979
rpfc_Fb7       AACTAACCAGTGGCAGGAGGTGCGGGAACAAGCGCATGCACTGCGTGGTGTGCTCGGTCA 1980
*****

rpfc_9a5c      TCTTGGTTTAATGCGCGCTTCGTGCTGGCGGCAGAGTTGATGCGTATCCCAGATTGGCA 2039
rpfc_Fb7       TCTTGGTTTAATGCGCGCTTCGTGCTGGCGGCAGAGTTGATGCGTATCCCAGATTGGCA 2040
*****

rpfc_9a5c      GTTGCAATCGGAGTGGTGTGCTCGTGTCCAAGTGTGTATGAAGAACTCCGGTCTGGTCG 2099
rpfc_Fb7       GTTGCAATCGGAGTGGTGTGCTCGTGTCCAAGTGTGTATGAAGAACTCCGGTCTGGTCG 2100
*****

rpfc_9a5c      TGATGCGCTTGAAGCGCGCCGGCAGGAGCGCCGGCGCAGTGGCATTTTACAATGA 2154
rpfc_Fb7       TGATGCGCTTGAAGCGCGCCGGCAGGAGCGCCGGCGCAGTGGCATTTTACAATGA 2155
*****

```

Tabela S1. CDSs de integrase que se encontram nas regiões de profagos/IG que não estão envolvidas com a produção de fagos.

Locus_Tag (IMG/ER)	Product
EB92R2_01034	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
EB92R2_01178	nucleotide sugar dehydrogenase
EB92R2_01179	FKBP-type peptidyl-prolyl cis-trans isomerases 1
EB92R2_01180	Glutathione peroxidase
EB92R2_01195	rare lipoprotein A
EB92R2_01196	lytic murein transglycosylase B
EB92R2_01638	prepilin-type N-terminal cleavage/methylation domain
EB92R2_01639	Tfp pilus assembly protein, tip-associated adhesin PilY1
EB92R2_01640	Tfp pilus assembly protein PilX
EB92R2_01641	Tfp pilus assembly protein PilW
EB92R2_01642	type IV pilus modification protein PilV
EB92R2_01643	prepilin-type N-terminal cleavage/methylation domain
XF3124_00550	Tfp pilus assembly protein PilZ
XF3124_00553	dethiobiotin synthase
XF3124_00715	3-deoxy-D-manno-octulosonate cytidyltransferase
XF3124_01159	Fructose-2,6-bisphosphatase
XF3124_01224	Virulence-associated protein and related proteins
XF3124_01454	Beta-lactamase class C and other penicillin binding proteins
XF3124_02210	Peroxiredoxin
XF3124_02218	6,7-dimethyl-8-ribityllumazine synthase
XF3124_02219	3,4-dihydroxy-2-butanone 4-phosphate synthase
XF9a_00438	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XF9a_00439	Tfp pilus assembly protein, major pilin PilA
XF9a_00612	Tfp pilus assembly protein PilZ
XF9a_01475	Glutathione peroxidase
XF9a_01476	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XF9a_01477	nucleotide sugar dehydrogenase
XF9a_01480	L-fucose:H ⁺ symporter permease
XF9a_01481	Sugar kinases, ribokinase family
XF9a_01482	N-acyl-D-glucosamine 2-epimerase
XF9a_01483	D-alanyl-D-alanine carboxypeptidase, serine-type, PBP4 family
XF9a_01485	Methylase involved in ubiquinone/menaquinone biosynthesis
XF9a_01511	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XF9a_01522	Protein involved in cell division
XF9a_01563	Bacteriocin class II with double-glycine leader peptide.
XF9a_01565	SpoVT / AbrB like domain.
XF9a_01577	Microcystin-dependent protein
XF9a_01587	monofunctional biosynthetic peptidoglycan transglycosylase
XF9a_01589	pyridoxal phosphate enzyme, YggS family
XF9a_02181	3-deoxy-D-manno-octulosonate cytidyltransferase
XF9a_02183	glutamate-1-semialdehyde-2,1-aminomutase
XF9a_02186	Lactoylglutathione lyase and related lyases

XF9a_02187	Delta-aminolevulinic acid dehydratase
XF9a_02191	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
XF9a_02336	Archaeal fructose-1,6-bisphosphatase and related enzymes of inositol monophosphatase family
XF9a_02337	dethiobiotin synthase
XF9a_02399	Signal transduction histidine kinase
XF9a_02400	dephospho-CoA kinase
XF9a_02401	Type II secretory pathway, prepilin signal peptidase PulO and related peptidases
XF9a_02402	Type II secretory pathway, component PulF
XF9a_02403	prepilin-type N-terminal cleavage/methylation domain
XF9a_02406	prepilin-type N-terminal cleavage/methylation domain
XFAN_00013	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases
XFAN_00015	Coproporphyrinogen III oxidase
XFAN_00020	phospho-2-dehydro-3-deoxyheptonate aldolase
XFAN_00022	prepilin-type N-terminal cleavage/methylation domain
XFAN_00023	type IV pilus modification protein PilV
XFAN_00024	prepilin-type N-terminal cleavage/methylation domain
XFAN_00025	Tfp pilus assembly protein PilX
XFAN_00026	Tfp pilus assembly protein, tip-associated adhesin PilY1
XFAN_00027	prepilin-type N-terminal cleavage/methylation domain
XFAN_00415	Biopolymer transport proteins
XFAN_00416	Biopolymer transport protein
XFAN_00417	lipid A export permease/ATP-binding protein MsbA
XFAN_00418	tetraacyldisaccharide 4"-kinase
XFAN_00726	glycine dehydrogenase (decarboxylating)
XFAN_00919	Zn-dependent hydrolases, including glyoxylases
XFAN_00923	Nucleoside-diphosphate-sugar epimerases
XFAN_00924	Glycosyltransferases involved in cell wall biogenesis
XFAN_00926	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II
XFAN_00927	Amidases related to nicotinamidase
XFAN_01073	Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis
XFAN_01074	Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis
XFAN_01115	Outer membrane cobalamin receptor protein
XFAN_01119	Outer membrane protein and related peptidoglycan-associated (lipo)proteins
XFAN_01153	Tfp pilus assembly protein, major pilin PilA
XFAN_01327	Periplasmic protease
XFAN_01330	outer membrane autotransporter barrel domain
XFAN_01341	Signal transduction histidine kinase, nitrogen specific
XFAN_01427	Growth inhibitor
XFAN_01935	dethiobiotin synthase
XFAN_02077	Tfp pilus assembly protein, major pilin PilA
XFAN_02141	Methylase involved in ubiquinone/menaquinone biosynthesis
XFAN_02146	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)
XFAN_02148	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)
XFAN_02149	N-acyl-D-glucosamine 2-epimerase
XFAN_02150	Sugar kinases, ribokinase family

XFAN_02151	L-fucose:H ⁺ symporter permease
XFAN_02154	nucleotide sugar dehydrogenase
XFAN_02156	Glutathione peroxidase
XFDX_00177	queuosine biosynthesis protein QueD
XFDX_00182	1-acyl-sn-glycerol-3-phosphate acyltransferase
XFDX_00244	dethiobiotin synthase
XFDX_00245	Archaeal fructose-1,6-bisphosphatase and related enzymes of inositol monophosphatase family
XFDX_00246	RNA methyltransferase, TrmH family, group 1
XFDX_00248	lysine-2,3-aminomutase-related protein
XFDX_00249	translation elongation factor P
XFDX_00250	Cytosine deaminase and related metal-dependent hydrolases
XFDX_00251	ubiquinone biosynthesis O-methyltransferase
XFDX_00252	2-phosphoglycolate phosphatase, prokaryotic
XFDX_00253	Phytoene/squalene synthetase
XFDX_00254	Endopolygalacturonase
XFDX_00255	homoserine O-acetyltransferase
XFDX_00570	Fumarase
XFDX_00603	glucose-inhibited division protein A
XFDX_00609	Aspartyl/asparaginyl beta-hydroxylase and related dioxygenases
XFDX_00849	Delta-aminolevulinic acid dehydratase
XFDX_00850	Lactoylglutathione lyase and related lyases
XFDX_00854	glutamate-1-semialdehyde-2,1-aminomutase
XFDX_00856	3-deoxy-D-manno-octulosonate cytidyltransferase
XFDX_00857	Protein-tyrosine-phosphatase
XFDX_01085	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XFDX_01086	Tfp pilus assembly protein, major pilin PilA
XFDX_01106	NgoMIV restriction enzyme.
XFDX_01276	RecA-family ATPase
XFDX_01284	Restriction endonuclease NotI.
XFDX_01317	Tfp pilus assembly protein PilZ
XFDX_01603	folate-binding protein YgfZ
XFDX_01611	lipoprotein releasing system, transmembrane protein, LolC/E family
XFDX_01612	lipoprotein releasing system, ATP-binding protein
XFDX_01614	Biopolymer transport proteins
XFDX_01615	Biopolymer transport protein
XFDX_01616	lipid A export permease/ATP-binding protein MsbA
XFDX_01617	tetraacyldisaccharide 4"-kinase
XFDX_01672	Signal transduction histidine kinase, nitrogen specific
XFDX_01673	nitrogen regulation protein NR(I)
XFDX_01676	ammonium transporter
XFDX_01677	Nitrogen regulatory protein PII
XFDX_01678	glutamine synthetase, type I
XFDX_01770	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)
XFDX_01771	D-alanyl-D-alanine carboxypeptidase, serine-type, PBP4 family
XFDX_01772	N-acyl-D-glucosamine 2-epimerase

XFDX_01774	L-fucose:H ⁺ symporter permease
XFDX_01777	nucleotide sugar dehydrogenase
XFDX_01778	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XFDX_01823	Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis
XFDX_02032	Cation/multidrug efflux pump
XFDX_02034	Zn-dependent alcohol dehydrogenases
XFDX_02038	Lysozyme M1 (1,4-beta-N-acetylmuramidase)
XFDX_02040	monothiol glutaredoxin, Grx4 family
XFFB7_01013	Fructose-2,6-bisphosphatase
XFFB7_01059	hydro-lyases, Fe-S type, tartrate/fumarate subfamily, alpha region/hydro-lyases
XFFB7_01072	Nitrogen regulatory protein PII
XFFB7_01073	glutamine synthetase, type I
XFFB7_01074	Uncharacterized bacitracin resistance protein
XFFB7_01075	Uncharacterized bacitracin resistance protein
XFFB7_01080	Zinc metalloprotease (elastase)
XFFB7_01154	prepilin-type N-terminal cleavage/methylation domain
XFFB7_01156	dihydrodipicolinate synthase
XFFB7_01157	Glycine cleavage system regulatory protein
XFFB7_01158	Peroxiredoxin
XFFB7_01160	Predicted ATPase related to phosphate starvation-inducible protein PhoH
XFFB7_01327	Protein-tyrosine-phosphatase
XFFB7_01328	3-deoxy-D-manno-octulosonate cytidyltransferase
XFFB7_01329	ATPases involved in chromosome partitioning
XFFB7_01330	glutamate-1-semialdehyde-2,1-aminomutase
XFFB7_01333	Lactoylglutathione lyase and related lyases
XFFB7_01334	Delta-aminolevulinic acid dehydratase
XFFB7_01610	prepilin-type N-terminal cleavage/methylation domain
XFFB7_01611	prepilin-type N-terminal cleavage/methylation domain
XFFB7_01658	Tfp pilus assembly protein, major pilin PilA
XFFB7_01660	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XFFB7_02012	Signal transduction histidine kinase
XFFB7_02013	dephospho-CoA kinase
XFFB7_02014	Type II secretory pathway, prepilin signal peptidase PulO and related peptidases
XFFB7_02015	Type II secretory pathway, component PulF
XFFB7_02016	prepilin-type N-terminal cleavage/methylation domain
XFFB7_02019	prepilin-type N-terminal cleavage/methylation domain
XFFB7_02219	RTX toxins and related Ca ²⁺ -binding proteins
XFFB7_02239	ABC-type multidrug transport system, ATPase component
XFFB7_02240	Glutathione peroxidase
XFFB7_02242	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XFFB7_02243	nucleotide sugar dehydrogenase
XFFB7_02277	monofunctional biosynthetic peptidoglycan transglycosylase
XFFB7_02408	2-polyprenylphenol 6-hydroxylase
XFFB7_02409	Thiol:disulfide interchange protein
XFGB_00158	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase

XFGB_00163	Delta-aminolevulinic acid dehydratase
XFGB_00164	Lactoylglutathione lyase and related lyases
XFGB_00168	glutamate-1-semialdehyde-2,1-aminomutase
XFGB_00170	3-deoxy-D-manno-octulosonate cytidyltransferase
XFGB_00171	Protein-tyrosine-phosphatase
XFGB_00352	Tfp pilus assembly protein PilZ
XFGB_01512	Biopolymer transport protein
XFGB_01513	lipid A export permease/ATP-binding protein MsbA
XFGB_01515	tetraacyldisaccharide 4"-kinase
XFGB_01539	Phosphohistidine phosphatase SixA
XFGB_01541	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthases and related enzymes
XFGB_01548	NAD(P)H:quinone oxidoreductase, type IV
XFGB_01549	nicotinate phosphoribosyltransferase
XFGB_02171	queuosine biosynthesis protein QueC
XFGB_02280	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XFHIB4_00780	3-deoxy-D-manno-octulosonate cytidyltransferase
XFHIB4_01036	biotin biosynthesis protein BioC
XFHIB4_01523	Tfp pilus assembly protein, major pilin PilA
XFHIB4_01527	pantothenate kinase, type III
XFHIB4_01528	birA, biotin-[acetyl-CoA-carboxylase] ligase region
XFHIB4_01529	Uncharacterized enzyme of heme biosynthesis
XFHIB4_01530	Uncharacterized enzyme of heme biosynthesis
XFHIB4_01531	Uroporphyrinogen-III synthase
XFHIB4_01559	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XFHIB4_01605	Zinc metalloprotease (elastase)
XFHIB4_01927	Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain
XFHIB4_01929	DNA-binding ferritin-like protein (oxidative damage protectant)
XFJ1_00004480	Tfp pilus assembly protein, major pilin PilA
XFJ1_00007750	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
XFJ1_00007780	Delta-aminolevulinic acid dehydratase
XFJ1_00007790	Lactoylglutathione lyase and related lyases
XFJ1_00007820	glutamate-1-semialdehyde-2,1-aminomutase
XFJ1_00007830	ATPases involved in chromosome partitioning
XFJ1_00007840	3-deoxy-D-manno-octulosonate cytidyltransferase
XFJ1_00007850	Protein-tyrosine-phosphatase
XFJ1_00009770	diaminopimelate epimerase
XFJ1_00017490	ABC-type multidrug transport system, ATPase component
XFJ1_00017500	Glutathione peroxidase
XFJ1_00017510	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XFJ1_00017520	nucleotide sugar dehydrogenase
XFJ1_00017550	glucose/galactose transporter
XFJ1_00017560	Sugar kinases, ribokinase family
XFJ1_00017570	N-acyl-D-glucosamine 2-epimerase
XFJ1_00017580	D-alanyl-D-alanine carboxypeptidase, serine-type, PBP4 family
XFJ1_00020020	Tfp pilus assembly protein, major pilin PilA

XFJ1_00022740	Glutathione peroxidase
XFJ1_00022750	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1
XFJ1_00022760	thiamine biosynthesis protein ThiC
XFJ1_00022790	Fructose-2,6-bisphosphatase
XFJ1_00023380	Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis
XFJ1_00023410	Kef-type K ⁺ transport system, predicted NAD-binding component
XFJ1_00023420	phosphate ABC transporter, phosphate-binding protein
XFM12_00398	ABC-type sugar transport systems, ATPase components
XFM12_00400	folate-binding protein YgfZ
XFM12_00404	succinate dehydrogenase, flavoprotein subunit, E. coli/mitochondrial subgroup
XFM12_00408	lipoprotein releasing system, transmembrane protein, LolC/E family
XFM12_00410	DNA internalization-related competence protein ComEC/Rec2
XFM12_00414	tetraacyldisaccharide 4''-kinase
XFM12_00458	Phosphatidylserine/phosphatidylglycerophosphate/cardioplin synthases and related enzymes
XFM12_00463	nicotinate phosphoribosyltransferase
XFM12_00653	Tfp pilus assembly protein, major pilin PilA
XFM12_00919	prepilin-type N-terminal cleavage/methylation domain
XFM12_00949	glucose-inhibited division protein A
XFM12_00956	biotin biosynthesis protein BioC
XFM12_01155	Topoisomerase IA
XFM12_01156	Signal transduction histidine kinase, nitrogen specific
XFM12_01157	nitrogen regulation protein NR(I)
XFM12_01161	Nitrogen regulatory protein PII
XFM12_01165	Predicted nucleoside-diphosphate-sugar epimerase
XFM12_01204	Protein involved in cell division
XFM12_01228	2-polyprenylphenol 6-hydroxylase
XFM12_01230	formyltetrahydrofolate deformylase
XFM12_01323	ABC-type multidrug transport system, ATPase component
XFM12_01356	Endodeoxyribonuclease RusA.
XFM23_00405	tetraacyldisaccharide 4''-kinase
XFM23_00431	Phosphatidylserine/phosphatidylglycerophosphate/cardioplin synthases and related enzymes
XFM23_00438	NAD(P)H:quinone oxidoreductase, type IV
XFM23_00439	nicotinate phosphoribosyltransferase
XFM23_01047	Fructose-2,6-bisphosphatase
XFM23_01361	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)
XFM23_01363	D-alanyl-D-alanine carboxypeptidase, serine-type, PBP4 family
XFM23_01364	N-acyl-D-glucosamine 2-epimerase
XFM23_01366	L-fucose:H ⁺ symporter permease
XFM23_01370	nucleotide sugar dehydrogenase
XFM23_01371	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XFM23_01373	ABC-type multidrug transport system, ATPase component
XFM23_01564	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
XFM23_01569	Delta-aminolevulinic acid dehydratase
XFM23_01570	Lactoylglutathione lyase and related lyases
XFM23_01574	glutamate-1-semialdehyde-2,1-aminomutase

XFM23_01576	3-deoxy-D-manno-octulosonate cytidyltransferase
XFM23_01577	Protein-tyrosine-phosphatase
XFM23_01578	DNA modification methylase
XfPR8X_00347	Pyocin activator protein PrtN.
XfPR8X_00353	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XfPR8X_00354	Tfp pilus assembly protein, major pilin PilA
XfPR8X_00377	Protease subunit of ATP-dependent Clp proteases
XfPR8X_00397	DNA adenine methylase (dam)
XfPR8X_00398	NgoMIV restriction enzyme.
XfPR8X_01152	thiamine biosynthesis protein ThiC
XfPR8X_01155	Fructose-2,6-bisphosphatase
XfPR8X_01182	queuosine biosynthesis protein QueC
XfPR8X_01374	glucose/galactose transporter
XfPR8X_02114	Virulence-associated protein and related proteins
XfPR8X_02176	Peroxiredoxin
XFTem_00405	Biopolymer transport proteins
XFTem_00406	Biopolymer transport protein
XFTem_00407	lipid A export permease/ATP-binding protein MsbA
XFTem_00408	tetraacyldisaccharide 4"-kinase
XFTem_00432	Phosphohistidine phosphatase SixA
XFTem_00441	NAD(P)H:quinone oxidoreductase, type IV
XFTem_00442	nicotinate phosphoribosyltransferase
XFTem_01029	Fructose-2,6-bisphosphatase
XFTem_01029	Fructose-2,6-bisphosphatase
XFTem_01078	queuosine biosynthesis protein QueC
XFTem_01170	Protein involved in cell division
XFTem_01188	Signal transduction histidine kinase, nitrogen specific
XFTem_01189	nitrogen regulation protein NR(I)
XFTem_01192	ammonium transporter
XFTem_01193	Nitrogen regulatory protein PII
XFTem_01345	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)
XFTem_01347	D-alanyl-D-alanine carboxypeptidase, serine-type, PBP4 family
XFTem_01348	N-acyl-D-glucosamine 2-epimerase
XFTem_01350	L-fucose:H ⁺ symporter permease
XFTem_01354	nucleotide sugar dehydrogenase
XFTem_01355	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XFTem_01355	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XFTem_01356	Glutathione peroxidase
XFTem_01356	Glutathione peroxidase
XFTem_01389	Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis
XFTem_01545	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
XFTem_01550	Delta-aminolevulinic acid dehydratase
XFTem_01555	glutamate-1-semialdehyde-2,1-aminomutase
XFTem_01555	glutamate-1-semialdehyde-2,1-aminomutase
XFTem_01557	3-deoxy-D-manno-octulosonate cytidyltransferase

XFTem_01558	Protein-tyrosine-phosphatase
XFTem_01730	dethiobiotin synthase
XFTem_01733	Tfp pilus assembly protein PilZ
XFU24D_00442	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XFU24D_00443	Tfp pilus assembly protein, major pilin PilA
XFU24D_00615	Tfp pilus assembly protein PilZ
XFU24D_01495	Glutathione peroxidase
XFU24D_01497	FKBP-type peptidyl-prolyl cis-trans isomerases 1
XFU24D_01498	nucleotide sugar dehydrogenase
XFU24D_01501	L-fucose:H ⁺ symporter permease
XFU24D_01502	Sugar kinases, ribokinase family
XFU24D_01503	N-acyl-D-glucosamine 2-epimerase
XFU24D_01504	D-alanyl-D-alanine carboxypeptidase, serine-type, PBP4 family
XFU24D_01506	Methylase involved in ubiquinone/menaquinone biosynthesis
XFU24D_01532	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
XFU24D_01543	Protein involved in cell division
XFU24D_01585	Bacteriocin class II with double-glycine leader peptide.
XFU24D_01587	SpoVT / AbrB like domain.
XFU24D_01599	Microcystin-dependent protein
XFU24D_01609	monofunctional biosynthetic peptidoglycan transglycosylase
XFU24D_01611	pyridoxal phosphate enzyme, YggS family
XFU24D_02018	glucose-inhibited division protein A
XFU24D_02207	Protein-tyrosine-phosphatase
XFU24D_02208	3-deoxy-D-manno-octulosonate cytidyltransferase
XFU24D_02210	glutamate-1-semialdehyde-2,1-aminomutase
XFU24D_02213	Lactoylglutathione lyase and related lyases
XFU24D_02214	Delta-aminolevulinic acid dehydratase
XFU24D_02371	Archaeal fructose-1,6-bisphosphatase and related enzymes of inositol monophosphatase family
XFU24D_02372	dethiobiotin synthase
XFU24D_02434	Signal transduction histidine kinase
XFU24D_02435	dephospho-CoA kinase
XFU24D_02436	Type II secretory pathway, prepilin signal peptidase PulO and related peptidases
XFU24D_02437	Type II secretory pathway, component PulF
XFU24D_02438	prepilin-type N-terminal cleavage/methylation domain
XFU24D_02441	prepilin-type N-terminal cleavage/methylation domain

Tabela S2. Regiões de progagos e Ilhas genômicas das cepas sul e Norte americanas.

Cepa	Nomenclatura	Integrase	Locus tag (IMG/ER)	Tamanho (pb)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
9a5c	xfp1	phage-related integrase	XF0678	649115-650278	647699-697033	49.3Kb	intacto	Xylella_phage_Xfas5_3	56.02
	xfp2	phage-related integrase	XF2530	2403249-2404268	2358987-2416177	57.1Kb	intacto	Xylella_phage_Xfas5_3	55.2
	xfp3	hypothetical protein	XF1555	1514347-1515201	1514344-1557570	43.2Kb	intacta	Aggregatibacter	54.72
	xfp4	phage-related integrase	XF1642	1585980-1586999	1585873-1640907	55Kb	intacto	Aggregatibacter phage S1249	54.82
	xfp5	hypothetical protein	XF2132	2021692-2022372	2004768-2022372	17.6Kb	-	-	57.2
	xfp6	Hypothetical protein	XF0480	487057-488229	486956-528481	41.5Kb	intacto	Stenotrophomonas_phage_S1	57.89
	xfp7	Site-specific recombinase XerD	XF2132	2021692-2022372	2005406-2024354	18.9Kb	Questionável	Xylella_phage_Xfas5_3	57.07
	cvc-r1	phage-related integrase	XF0631	607242-608483	607242...615973	8.7Kb	-	-	59.2
	cvc-r2	hypothetical protein	XF0968	926826-927659	926649-928116	1.4Kb	-	-	52.9
	cvc-r5 (xfp7.7)	phage-related integrase	XF2288	2177071-2178132	2176140-2198389	22.2Kb	incompleto	Xylella_phage_Xfas5_3	53.80
	giCVC	phage-related integrase	XF1718	1638946-1640907	1631333-1640907	9.5Kb	-	-	53.4
Cepa	Nomenclatura	Integrase	Locus tag (IMG)	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
U24d	xfp1	Site-specific recombinase XerD	XFU24D_0006170	649105-650268	647689-697024	49.3Kb	intacto	Xylella_phage_Xfas5_3	56.02
	xfp2	Site-specific recombinase XerD	XFU24D_00024290	2405283-2406302	2361021-2418210	57.1Kb	intacto	Xylella_phage_Xfas5_3	55.24
	xfp3	Site-specific recombinase XerD	XFU24D_00014450	1514336-1515190	1514333-1557558	43.2Kb	intacto	Aggregatibacter phage S1249	54.72
	xfp4	Site-specific recombinase XerD	XFU24D_00015280	1585967-1586986	1585860-1640893	55Kb	intacto	Aggregatibacter phage S1249	54.82
	xfp5	Site-specific recombinase XerD	XFU24D_00020450	2021668-2022348	2005382-2024330	18.9Kb	intacto	Xylella_phage_Xfas5_3	57.07

	xfp6	Hypothetical protein	XFU24D_0004360	487047-488219	486991-523677	41.5Kb	intacto	Stenotrophomonas_phage_S1	57.89
	xfp7	Site-specific recombinase XerD	XFU24D_00020450	2021668-2022348	2002087-2004817	18.9Kb	Questionável	Xylella_phage_Xfas5_3	59.0
	cvc-r1	Site-specific recombinase XerD	XFU24D_00005720	607232-608473	607232-615963	8.7Kb	-	-	59.2
	cvc-r2	Site-specific recombinase XerD	XFU24D_00009040	926831-927664	926654-928085	1.4Kb	-	-	52.8
	cvc-r5 (xfp7.7)	Phage integrase family.	XFU24D_00021980	2177041-2178102	2177041-2193973	12.2Kb	incompleto	Xylella_phage_Xfas5_3	55.70
	giCVC	Integrase	XFU24D_00016130	1638932-1640893	1631319-1640893	9.5Kb	-	-	53.4
Cepa	Nomenclatura	Integrase	Locus tag (IMG)	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
Fb7	xfp2	Site-specific recombinase XerD	XFFB7_00024140	2410409-2411428	2404395-2452423	48Kb	intacto	Xylella_phage_Xfas5_3	56.23
	xfp6	Hypothetical protein	XFFB7_00016660	1658866-1660038	1611675-1660038	48.3Kb	intacto	Xylella_phage_Xfas5_3	57.13
	xfp8	Site-specific recombinase XerD	XFFB7_00007560	821424-822251	821393-840330	18.9Kb	questionável	Xylella_phage_Xfas5_3	57.80
	xfp9	Site-specific recombinase XerD	XFFB7_00026960	2679883-2680335	2677901-2713592	35.6Kb	intacto	Xylella_phage_Xfas5_3	57.49
	xfp10	Site-specific recombinase XerD	XFFB7_00004450	487638-488468	487570-496766	9.1Kb	incompleto	Xanthomonas_phage_Cf1c	45.28
	xfp11	Site-specific recombinase XerD	XFFB7_00011510	1155042-1156007	1102754-1165466	62.7Kb	intacto	Aggregatibacter phage S1249	53.88
	Xfp12	Putative phage replication protein RstA	XFFB7_00010360	1073271-1074446	1073002-1097870	24.8Kb	incompleto	Stenotrophomonas_phage_phiSMA9	43.1
	xfp12.2	Site-specific recombinase XerD	XFFB7_00020070	1984555-1985574	1960690-1985574	24.8Kb	incompleto	Xylella_phage_Xfas5_3	53.57
	cvc-r1	Site-specific recombinase XerD	XFFB7_00015370	1541995-1543236	1541537-1548198	6.6Kb	-	-	52.9

	cvc-r2 (xfp9.9)	Site-specific recombinase XerD	XFFB7_00 019080	1923749-1924582	1923328-1996860	73.5Kb	intacto	Aggregatibacter phage S1249	56.16%
	cvc-r5 (xfp7.7)	Phage integrase family	XFFB7_00 013190	1319605-1320666	1319605-1335937	16.3Kb	-	-	53.9
	cvc-r6	hypothetical protein	XFFB7_00 014920	1501901-1502122	1501820-1502287	468pb	-	-	53.2
	Cvc-r7 (xfp8.8)	Sem integrase	-	-	2229789-2249362	19.5Kb	incompleto	Aggregatibacter phage S1249	54.27
	xfp30	Sem integrase	-	-	1073271-1097870	24.5Kb	incompleto	Prophage Xylella fastidiosa Temecula1	47.58
	giCVC	Integrase	XFFB7_00 022810	2295290-2297251	2287678-2297251	9.5Kb	-	-	53.4
	xfp12	Sem integrase	-	-	1059998-1084794	24.8K	-	-	44.5
Cepa	Nomenclatura	Integrase	Locus tag (IMG)	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
3124	xfp1	Site-specific recombinase XerD	XF3124_00 022030	2198687-2199658	2420348-2477917	57.5Kb -	intacto	Xylella_phage_Xfas5_3	55.07
	xfp2	Site-specific recombinase XerD	XF3124_00 024820	2473680-2474699	2432722-2475298	41.9Kb	intacto	Xylella_phage_Xfas5_3	56.81
	Xfp14	Site-specific recombinase XerD	XF3124_00 009260	983979-984998	981997-1000860	18.8Kb	questionável	Xylella_phage_Xfas5_3	56.83
	Xfp15	Site-specific recombinase XerD	XF3124_00 015130	1492467-493486	1447788-1493486	45.6Kb	intacta	Aggregatibacter phage S1249	54.33
	Xfp16	Phage integrase family	XF3124_00 009040	957183-958274	923276-960394	37.1Kb	questionável	Xylella_phage_Xfas5_3	51.43
	Xfp17	Site-specific recombinase XerD	XF3124_00 012840	1288800-1289819	1235213-1289819	54.6Kb	intacto	Aggregatibacter phage S1249	54.12
	Xfp18	Phage integrase family	XF3124_00 015810	1575335-1576171	1566694-1625616	58.9Kb	intacto	Xylella_phage_Xfas5_3	54.48
	Xfp12	Sem integrase	-	-	1208465-1230223	16.3Kb	incompleto	Stenotrophomonas_phage_phiSMA9	43.31
	cvc-r1	Site-specific recombinase XerD	XF3124_00 005060	562823-564064	564522-562478	2.0Kb	--	-	57.3

	cvc-r2 (xfp9.9)	Site-specific recombinase XerD	XF3124_00021040	2121972-2122805	2118607-2214095	95.4Kb	intacto	Xylella_phage_Xfas5_3	56.19
	cvc-r5 (xfp7.7)	Phage integrase family.	XF3124_00007260	787412-788473	770843-788709	17.8Kb	-	-	53.5
	cvc-r6	Site-specific recombinase XerD	XF3124_00005520	605459-605683	604941-606576	468pb	-	-	53.2
	giCVC	Integrase	XF3124_01426	1413310-1415271	1412608-1415835	3.2Kb	-	-	61.6
Cepa	Nomenclatura	Integrase	Locus tag (IMG)	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
J1a12	xfp1	Site-specific recombinase XerD	XFJ1_00008540	865136-866284	809915-957360	147.4Kb	intacto	Xylella_phage_Xfas5_3	55.9
	xfp2	Site-specific recombinase XerD	XFJ1_00026000	2514707-2515726	2473510-2515817	42.3Kb	intacto	Xylella_phage_Xfas5_3	56.7
	xfp6	Hypothetical protein	XFJ1_00004420	484621-485793	484520-525314	40.7Kb	intacto	Stenotrophomonas_phage_S1	58.0
	xfp10	Site-specific recombinase XerD	XFJ1_00012140	1196168-1196998	1189381-1196998	7.8Kb	-	-	46.8
	xfp11	Site-specific recombinase XerD	XFJ1_00009140	908059-909024	908059.. 948764	39.5Kb	questionável	Xylella_phage_Xfas5_3	55.7
	Xfp15	Site-specific recombinase XerD	XFJ1_00017050	1714961-1715980	1714858-1755565	40.7Kb	intacto	Aggregatibacter phage S1249	54.96
	Xfp19	Site-specific recombinase XerD	XFJ1_00023350	2249452-2250471	2205021-2263073	58Kb	intacto	Xylella_phage_Xfas5_3	54.5
	Xfp20	Sem integrase	-	-	1940193-1955573	15.3Kb	incompleto	Salmonella_phage_Fels_1	60.75
	Xfp21	hypothetical protein	XFJ1_00019960	1960468-1961640	1959505-1999443	30.4Kb	completo	Xylella_phage_Xfas5_3	59.72
	Xfp26	Sem integrase	-	-	809169-906646	90.9Kb	completo	Xylella_phage_Xfas5_3	56.75
	cvc-r1	Site-specific recombinase XerD	XFJ1_00005790	603717-604958	603717-612448	8.7Kb	-	-	59.2
	cvc-r6	hypothetical protein	XFJ1_00006240	642563-642784	642563-642532	468pb	-	-	53.2

	Cvc-r8	Site-specific recombinase XerD	XFJ1_0001 4600	1461443-1462414	1464541-1461000	3.5Kb	-	-	51.6
	cvc-r8.8	Phage integrase family.	XFJ1_0001 6300	1626827-1627888	1626827-1634273	7.4Kb	-	-	56.1
	Cvc-r9	Phage integrase family	XFJ1_0001 6380	1633212-1634273	1626827-1633212	6.4Kb	-	-	57.0
	giCVC	Integrase	XFJ1_0001 7910	1793878-1795839	1792529-1795888	3.2Kb	-	-	61.6
Cepa	Nomenclatura	Integrase	Locus tag (IMG)	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
HIB4	xfp1	Sem integrase	-	-	820489- 862701	41.9Kb	questionável	Xylella_phage_Xfas5_3	56.99
	xfp2	Site-specific recombinase XerD	XFHIB4_02 591	2541370-2542389	2489021-2545626	56.6Kb	questionável	Xylella_phage_Xfas5_3	55.29
	xfp4	Site-specific recombinase XerD	XFHIB4_01 672	1611335-1612354	1560690-1612354	51.6Kb	intacto	Aggregatibacter phage S1249	55.10
	xfp6	Hypothetical protein	XFHIB4_00 423	482128-483300	498978-519197	20.2Kb	questionável	Stenotrophomonas_phage_S1	62.2
	xfp10	Site-specific recombinase XerD	XFHIB4_02 066	2004267-2005097	1993336-2005114	11.7Kb	incompleto	Stenotrophomonas_phage_phiSMA9	45.33
	Xfp22	Phage integrase family.	XFHIB4_01 917	1843698-1844759	1842816-1864627	21.8Kb	incompleto	Xylella_phage_Xfas5_3	55.53
	Xfp23	Site-specific recombinase XerD	XFHIB4_01 001	1035220-1036239	1029539-1063587	34Kb	questionável	Xylella_phage_Xfas5_3	55.64
	Xfp24	Phage integrase family.	XFHIB4_01 319	1307775-1308122	1292103-1311560	19.4Kb	incompleto	Stenotrophomonas_phage_phiSMA9	43.91
	Xfp25	Site-specific recombinase XerD	XFHIB4_01 563	1521957-1522976	1491720-1522976	31.2Kb	incompleto	Xylella_phage_Xfas5_3	54.24
	Xfp26	Sem integrase	-	-	810691-862405	44.7Kb	questionável	Xylella_phage_Xfas5_3	56.79
	Xfp27	-	-	-	1650068-1704924	54.8Kb	completo	Aggregatibacter phage S1249	55.01
	cvc-r1 (xfp10.1)	Site-specific recombinase XerD	XFHIB4_00 562	600400-601470	600400-610077	9.8Kb	-	-	58.3

	cvc-r2	Site-specific recombinase XerD	XFHIB4_02 302	2267503-2268336	2265769-2268620	2.8Kb	-	-	52.8
	cvc-r6	Phage integrase family.	XFHIB4_00 607	641903-642124	641903-641872	468pb	-	-	53.2
	giCVC	Integrase	XFHIB4_01 414	1399773-1401725	1398424-1402421	3.9Kb	-	-	60.1
Cepa	Nomenclatura	Integrase	Locus tag (IMG)	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
PR8X	xfp2	Phage integrase family.	XfPR8X_01 491	1498473-1499534	1490398-1514434	24Kb	incompleto	Xylella_phage_Xfas5_3	51.55
	xfp10	Site-specific recombinase XerD	XfPR8X_01 850	1835780-1836610	1821602-1836624	15Kb	incompleto	Stenotrophomonas_phage_phiSMA9	44.92
	xfp11	Site-specific recombinase XerD	XfPR8X_02 169	2146799-2147764	2106966-2147789	40.7Kb	Completo	Aggregatibacter phage S1249	55.88
	Xfp12	Putative phage replication protein RstA	XfPR8X_01 163	1193075-1194250	1185819-1205736	19.9Kb	-	-	45.6
	Xfp15	Site-specific recombinase XerD	XfPR8X_01 425	1418896-1419915	1377720-1430097	52.3Kb	Completo	Aggregatibacter phage S1249	54.50
	Xfp18	Phage integrase family.	XfPR8X_01 550	1541025-1542086	1510665-1556986	46.3Kb	completo	Xylella_phage_Xfas5_3	55.99
	Xfp29	Site-specific recombinase XerD	XfPR8X_00 402	439152-440279	398772-442721	43.9Kb	completo	Stenotrophomonas_phage_S1	58.00
	Xfp29	Site-specific recombinase XerD	XfPR8X_00 941	989432-990451	989432.. 1009114	19.6Kb	-	-	57.8
	cvc-r1	Site-specific recombinase XerD	XfPR8X_00 567	606564-607805	605743- 608391	2.6Kb	-	-	56.9
	cvc-r2 (xfp9.9)	Site-specific recombinase XerD	XfPR8X_02 101	2102566-2103399	2088694-2153779	65Kb	completo	Aggregatibacter phage S1249	54.41
	cvc-r6	hypothetical protein	XfPR8X_00 615	648398-648619	648398-648367	468pb	-	-	53.2
	xfp30	Sem integrase	-	-	1214879-1192710	24.5Kb	incompleto	Prophage Xylella fastidiosa Temecula1	47.58
	giCVC	Integrase	XfPR8X_01 338	1339783-1341750	1339081-1342314	3.2Kb	-	-	61.6

Cepa	Nomenclatura	Integrase	Locus tag	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
Temecula	xpd1 (xfp1)	Site-specific recombinase XerD	XFTem_01319	1330522-1331541	1288547-1331541	42.9Kb	completo	Xylella_phage_Xfas5_3	56.68
	xpd2	Site-specific recombinase XerD	XFTem_01187	1216888-1217907	1170405-1224503	54Kb	completo	Xylella_phage_Xfas5_3	57.01
	xpd3	Site-specific recombinase XerD	XFTem_00430	466592-467611	446956-480065	33.1Kb	questionável	Xylella_phage_Xfas5_3	53.42
	xpd4	Site-specific recombinase XerD	XFTem_01386	1391744-1392763	1362502-1398128	35.6Kb	questionável	Aggregatibacter phage S1249	53.46
	xpd5	Sem integrase	-	-	1117952-1143339	19.2Kb	questionável	Aggregatibacter phage S1249	58.71
	xpd6	hypothetical protein	XFTem_01539	1548922-1549963	1548922-1576573	27.6Kb			52.5
	xpp7	Site-specific recombinase XerD	XFTem_02004	2019801-2020421	2004606-2020502	15.8Kb	incompleto	Xylella_phage_Xfas5_3	55.46
	xpd8	Integrase	XFTem_01859	1857997-1858659	1841691-1861174	19.4Kb	-	-	58.7
	xpd9	Sem integrase	-	-	1121265-1143372	14.8Kb	incompleto	Stenotrophomonas_phage_phiSMA9	44.23
	xpd12	Sem integrase	-	-	1159634-1180707	19.1Kb	questionável	Aggregatibacter phage S1249	58.69
	xps14	Sem integrase	-	-	1142680-1119734	22.8Kb	incompleto	Stenotrophomonas_phage_phiSMA9	44.26
	xpd11	Sem integrase	-	-	1115113-1142680	12.1Kb	incompleto	Stenotrophomonas_phage_phiSMA9	44.83
	pd-r1	Site-specific recombinase XerD	XFTem_00863	946367-947386	946367952972	6.6Kb	-	-	58.5
	pd-r2	Site-specific recombinase XerD	XFTem_01731	1737927-1738253	1737034-1738698	1.6Kb	-	-	54.7
Cepa	Nomenclatura	Integrase	Locus tag	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
GB514	xpd1 (xfp1)	Site-specific recombinase	XFGB_02412	2430784-2431803	2391629-2431803	40.1Kb	intacto	Xylella_phage_Xfas5_3	56.57

		XerD							
	xpd2	Site-specific recombinase XerD	XFGB_022_84	2319977-2320996	2283518-2320996	37.4Kb	completo	Xylella_phage_Xfas5_3	57.37
	xpd3	Site-specific recombinase XerD	XFGB_015_37	1586451-1587470	1566818-1599922	33.1Kb	questionável	Xylella_phage_Xfas5_3	53.42
	xpd4	Sem integrase	-	-	2474061-2491169	17.1Kb	questionável	Aggregatibacter phage S1249	56.99
	xpp7	Site-specific recombinase XerD	XFGB_006_25	627021-627641	611827-629836	18Kb	incompleto	Xylella_phage_Xfas5_3	54.78
	xpd8	Integrase	XFGB_004_78	465235-465897	451116-468307	17.1Kb	-	-	59.1
	xpd10	Phage integrase family.	XFGB_001_52	156188-157228	162612-176369	13.7Kb	-	-	51.8
	xpd11	Sem integrase	-	-	2238606-2246534	7.9Kb	incompleta	Stenotrophomonas_phage_phiSMA9	43.16
	xpd12	Sem integrase	-	-	2264296-2283488	19.1Kb	questionavel	Aggregatibacter phage S1249	58.69
	pd-r1	Site-specific recombinase XerD	XFGB_019_82	2066194-2067213	2066194-2067213	6.6Kb	-	-	58.5
	pd-r2	Site-specific recombinase XerD	XFGB_003_50	345188-345514	345188..345938	1.6Kb	-	-	54.7
Cepa	Nomenclatura	Integrase	Locus tag	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
EB92-1	xpd-r14	Site-specific recombinase XerD	EB92R2_0_0496	831	522067-526321	4.2Kb	-	-	48.2
	pd-r2	Site-specific recombinase XerD	EB92R2_0_1505	327	1558616-1562799	1.6Kb	-	-	53.7
	xpd-4	Site-specific recombinase XerD	EB92R2_0_1829	621	1887580-1890722	3.1Kb	-	-	51.2
	xpd-5	Phage integrase family.	EB92R2_0_0637	1062	689913-692742	2.8Kb	-	-	49.0
	xpd8	Phage integrase family	EB92R2_0_1634	795	1659151-1692935	33.7Kb	questionável	Prophage Xylella fastidiosa Temecula1	55.93

	xpd15	Sem integrase	-	-	1056095-1090329	34.2Kb	intacto	Aggregatibacter phage S1249	55.09
	xpd16	sem integrase	-	-	1159913-1182311	22.3Kb	questionável	Xylella_phage_Xfas5_3	55.70
	xpd17	sem integrase	-	-	1221990-1236627	14.6Kb	incompleto	Microcystis_phage_Ma_LMM01	51.04
Cepa	Nomenclatura	Integrase	Locus tag	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
M23	xpd1	Site-specific recombinase XerD	XFM23_01_335	1340063-1341082	1298087-1341082	42.9Kb	completo	Xylella_phage_Xfas5_3	56.68
	xpd2	Site-specific recombinase XerD	XFM23_01_203	1226428-1227447	1226428-1204296	21.8Kb	questionável	Xylella_phage_Xfas5_3	53.77
	xpd3	Site-specific recombinase XerD	XFM23_00_427	466445-467464	446818-479918	33.1Kb	questionável	Xylella_phage_Xfas5_3	53.43
	xpd4	Site-specific recombinase XerD	XFM23_01_402	1401285-1402304	1371738-1402371	18.9Kb	questionável	Aggregatibacter phage S1249	56.60
	xpd6	hypothetical protein	XFM23_01_558	1563391-1564101	1562088-1589974	27.6Kb	-	-	52.5
	xpp7	Site-specific recombinase XerD	XFM23_02_025	2035612-2036232	2020417-2038427	18Kb	incompleto	Xylella_phage_Xfas5_3	54.81
	xpd8	Integrase	XFM23_01_879	1874474-1873812	1876884-1859689	17.1Kb	-	-	59
	xpd11	Sem integrase	-	-	1125000-1151958	12.1Kb	incompleto	Stenotrophomonas_phage_phiSMA9	44.83
	xpd12	Sem integrase	-	-	1170023-1188768	19.1Kb	questionável	Aggregatibacter phage S1249	58.69
	xpd13 (xfp10)	Site-specific recombinase XerD	XFM23_00_574	617765-618595	617703-627363	9.6Kb	incompleto	Stenotrophomonas_phage_phiSMA9	42.65
	xpd14	Sem integrase	-	-	1129800-1152650	22.8Kb	incompleto	Stenotrophomonas_phage_phiSMA9	44.26
	xpd15	Sem integrase	-	-	1189928-1151841	19.1Kb	questionável	Aggregatibacter phage S1249	58.66
	pd-r1	Site-specific recombinase XerD	XFM23_00_882	955886-956905	955886-962572	6.6Kb	-	-	58.5
	pd-r2	Site-specific	XFM23_01	17531941-	1753194-1753531	1.6Kb	-	-	54.7

		recombinase XerD	749	753520					
Cepa	Nomenclatura	Integrase	Locus tag	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
Dixon	xap1 (xfp1)	Site-specific recombinase XerD	XFDX_001 84	203963-204982	197000-257817	60.8Kb	completo	Xylella_phage_Xfas5 3	55.53
	xap2	Site-specific recombinase XerD	XFDX_005 71	585243-586262	583602-614877	31.2Kb	completo	Xylella_phage_Xfas5 3	53.79
	xap3	Site-specific recombinase XerD	XFDX_008 67	856059-857078	841399-857078	15.8Kb	-	-	52.8
	xap4	Site-specific recombinase XerD	XFDX_011 09	1093112- 1094239	1073420-1095897	22.4Kb	questionável	Stenotrophomonas_ phage_S1	55.35
	xap5	Site-specific recombinase XerD	XFDX_013 19	1300648- 1301259	1299265-1337815	38.5Kb	completo	Xylella_phage_Xfas5 3	56.36
	xap6	Site-specific recombinase XerD	XFDX_016 71	1621068- 1622087	1567627-1632540	64.9Kb	completo	Xylella_phage_Xfas5 3	54.14
	xap7	Site-specific recombinase XerD	XFDX_018 20	1759838- 1760857	1720614-1766233	45.6Kb	completo	Aggregatibacter phage S1249	54.80
	xap8	Site-specific recombinases, DNA invertase Pin homologs	XFDX_020 50	1990783- 1991349	1977760-2020239	42.4Kb	completo	Stenotrophomonas_ phage_S1	57.7
	xap9 (xfp4)	Site-specific recombinase XerD	XFDX_022 29	2128575- 2129540	2128575-2129540	37.4Kb	completo	Aggregatibacter phage S1249	55.20
	xap10 (xfp10)	Site-specific recombinase XerD	XFDX_010 04	1001238- 1002068	1001170-1009635	8.4Kb	incompleto	Stenotrophomonas_ phage_phiSMA9	42.92
	xap12	Site-specific recombinase XerD	XFDX_021 57	2084671- 2085210	2084671-2129565	44.8Kb	completo	Aggregatibacter phage S1249	55.2
	al-r13	Site-specific recombinase XerD	XFDX_007 84	790955-791785	784350-791785	7.4Kb	incompleto	Stenotrophomonas_ phage_phiSMA9	43.49
	xap14	Sem integrase	-	-	811891-838371	26.4Kb	completo	Aggregatibacter phage S1249	58.57

	xap15	Sem integrase	-	-	2598113-2622515	24.4Kb	incompleto	Aggregatibacter phage S1249	56.96
	xap16	Phage integrase family.	XFDX_01963	1909860-1910444	1909860-1920494	10.6Kb	-	-	54.3
	al-r3 (cvc-r1)	Site-specific recombinase XerD	XFDX_01274	1261633-1262865	1261051-1271300	10.2KB	-	-	58.3
	pd-r2	Site-specific recombinase XerD	XFDX_00243	245178-245501	244738-245501	764pb	-	-	54.7
Cepa	Nomenclatura	Integrase	Locus tag	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
M12	xfM1 (xfp1)	Site-specific recombinase XerD	XFM12_01222	1249948-1250967	1197551-1258283	60.7Kb	completo	Xylella_phage_Xfas53	54.49
	xfM2	Site-specific recombinase XerD	XFM12_00454	477765-478784	434006-494454	60.4Kb	completo	Prophage Xylella fastidiosa Temecula1	54.17
	xap2	Site-specific recombinase XerD	XFM12_00917	969527-970546	967875-999472	31.5Kb	completo	Xylella_phage_Xfas53	53.81
	xap13 (xfp10)	Site-specific recombinase XerD	XFM12_01125	1174995-1175825	1168812-1185640	16.8Kb	completo	Stenotrophomonas_phage_phiSMA9	43.40
	xfM16	Predicted site-specific integrase-resolvase	XFM12_00672	708169-708795	692331-714826	22.4Kb	questionável	Stenotrophomonas_phage_S1	54.42
	xfM17	Site-specific recombinase XerD	XFM12_01382	1392425-1393444	1349287-1374651	25.3Kb	completo	Aggregatibacter phage S1249	58.32
	al-r3 (cvc-r1)	Site-specific recombinase XerD	XFM12_01739	1748861-1750093	1748861-1752616	3.7Kb	-	-	-
	pd-r2	Site-specific recombinase XerD	XFM12_01695	1710731-1711057	1709838-1711128	1.3Kb	-	-	55.5
	MI-r5	Site-specific recombinase XerD	XFM12_01949	1974201-1974740	1970054-1974950	4.9Kb	-	-	50.2

Cepa	Nomenclatura	Integrase	Locus tag	Tamanho (bp)	Coordenadas (EGMs)	Tamanho (EGMs)	Integridade	EGMs correspondentes	GC%
nn-1	xfA1	Site-specific recombinase XerD	XFAN_01077	1090820-1091839	1082101-1120567	38.4Kb	completo	Aggregatibacter phage S1249	52.95
	xfM2	Site-specific recombinase XerD	XFAN_01176	1181002-1181937	1166654-1182100	15.4Kb	-	-	52.8
	xfA3	Site-specific recombinase XerD	XFAN_01265	1271680-1272807	13528-31328	17.8Kb	questionável	Xylella_phage_Xfas53	61.35
	xfA4 (xfp10)	Site-specific recombinase XerD	XFAN_01445	1414648-1415439	1398461-1420611	22.1Kb	incompleto	Stenotrophomonas phage phiSHP2	45.00
	xfA5	Site-specific recombinase XerD	XFAN_02117	2040948-2041967	2013976-2048677	34.7Kb	questionável	Xylella_phage_Xfas53	53.61
	xfA6	Site-specific recombinase XerD	XFAN_02225	2128884-2129786	2069343-2131981	62.6Kb	completo	Xylella_phage_Xfas53	54.17
	xfA7	Phage integrase family.	XFAN_00728	765494-766555	757184-798395	41.2Kb	completo	Xylella_phage_Xfas53	56.44
	xfp2	Phage integrase family.	XFAN_00916	931809-932870	885886-944769	58.8Kb	completo	Xylella_phage_Xfas53	55.19
	xfA9	Sem integrase	-	-	442941-467738	24.7Kb	incompleto	Prophage Xylella fastidiosa Temecula1	56.69
	xfA10	Sem integrase	-	-	1319478-1331809	12.3Kb	incompleto	Prophage Xylella fastidiosa Temecula1	51.22
	xfA11	Sem integrase	-	-	1998149-2024224	26Kb	incompleto	Prophage Xylella fastidiosa Temecula1	55.70
	An-r1	Phage integrase family.	XFAN_01744	17076311708692	1707631-1708692	4.1Kb	-	-	54.4
	pd-r2	Site-specific recombinase XerD	XFAN_01936	1896702-1897028	1895941-1897099	1.1Kb	-	-	56.3

Súmula curricular

Wesley Oliveira de Santana

Nascimento: 12/02/1981 – Santo André/SP - Brasil

Formação Acadêmica

- 2004–2008 Graduação em Ciências, habilitado em Biologia.
Universidade do Estado da Bahia, Teixeira de Freitas, BA, Brasil.
- 2009–2012 Curso de Doutorado
Programa de Ciências Biológicas (Bioquímica). Instituto de
Química, Universidade de São Paulo, , São Paulo, Brasil.
Orientadora: Profa. Dra. Aline Maria da Silva.
Bolsista da Coordenação de Aperfeiçoamento de Pessoal de
Nível Superior (CAPES).

Atuação Profissional

- 2004-2009 Escola Cooperativa de Teixeira de Freitas - BA
Professor: Ciências, Biologia e Química
- 2006-2009 Escola Cooperativa de Teixeira de Freitas - BA
Coordenador: Laboratório de Ciências
- 2006-2008 Projeto Universidade Para Todos (UNEB)
Professor: Biologia e Química
- 2006-2008 Cursinho pré-vestibular FASB
Professor: Biologia

Formação complementar

- 2009–2011 Programa de Aperfeiçoamento de Ensino (PAE)
Universidade de São Paulo.
- 2008-2009 Iniciação Científica
Universidade do Estado da Bahia, Teixeira de Freitas, Bahia
Orientador: Prof. Jorge Fortuna

Prêmios

- 2012 Pós-graduação – Oral de melhor trabalho na área de
GENÔMICA, pelo trabalho intitulado “Comparative genomics and
genotype-phenotype relationships of South-American *Xylella*
fastidiosa strains”. Apresentado durante o 58º Congresso
Brasileiro de Genética, em Foz do Iguaçu, PR

Publicações

- de Santana, W. O., Fortuna, J. L. 2012. Microbiota de aparelhos de ar condicionado das áreas críticas de hospitais públicos e particulares e sua relação com as infecções hospitalares. Biociências 18:56-64. (monografia).

- Pierry, P. M., de Santana, W. O., da Silva, P.I.P. Zaini, P. A., Sena M., Beckerdorf, F., Teixeira, D.C., Vasconcelos, A. T., de Almeida, L. G., Verjovski-Almeida, S., Kitajima, J. C., da Silva, A. M. Comparative genomics of a non-pathogenic *Xylella fastidiosa* citrus-isolated strain. *Em preparação*.
- de Santana, W. O., Pierry, P. M., Zaini, P. A., Sena M., Beckerdorf, F., Coletta-Filho, H. D., de Souza, A. A., Machado, M. A., Vasconcelos, A. T., de Almeida, L. G., Verjovski-Almeida, S., Kitajima, J. C., da Silva, A. M. 2012. Comparative genomics and genotype-phenotype relationships of South-American *Xylella fastidiosa* strains. *Em preparação*.

Comunicações em Congressos

de Santana, WO; Pierry, PM; Zaini, PA; Sena M; Beckerdorf, F; Coletta-Filho, HD; de Souza, AA; Machado, MA; Vasconcelos, AT; de Almeida, LG; Verjovski-Almeida, S; Kitajima, JC; da Silva, AM. Comparative genomics and genotype-phenotype relationships of South-American *Xylella fastidiosa* strains. In: 58º Congresso Brasileiro de Genética, 2012, Foz do Iguaçu, PR.

de Santana, WO; Amaral, MS; Beckedorff, F; Coletta-Filho, HD; Souza, AA; Almeida, LGP; Vasconcelos, ATR.; Verjovski-Almeida, S; da Silva, AM. Pirosequenciamento, genômica comparativa e evolução de *Xylella fastidiosa*. 2010. In: 56º Congresso Brasileiro de Genética, 2010, Guarujá, SP.

de Santana, WO; Fortuna, JL. Microbiota de aparelhos de ar condicionado das áreas críticas de hospitais públicos e particulares e sua relação com as infecções hospitalares. In: I Semana de Iniciação Científica, 2008, FASB, BA.

de Santana, WO; Gomes, AA; Brito, J. Produção de caixas entomológicas por alunos da segunda série do ensino médio. In: I encontro de Biologia, 2007, Universidade do Estado da Bahia, BA.

de Santana, WO; Junior, ASL. Taxa fotossintética em feijoeiro *Phaseolus vulgaris* L. a intervalos diferentes de temperatura. In: III seminário de pesquisa e extensão do extremo sul da Bahia, 2006, UNEB, BA.

Outras informações

Co-orientação de Carolina D. Hernandez. Bolsista de iniciação científica PIBIC-CNPq, Instituto de Química, USP. Projeto: Clonagem e expressão recombinante de PilQ de *Xylella fastidiosa*. 2010-2011.

Co-orientação de Oséias Rodrigues Feitosa Junior. Bolsista de iniciação científica FAPESP, Instituto de Química, USP. Projeto: Análise da expressão da poligalacturonase em diferentes cepas de *Xylella fastidiosa*. 2010-2011.