UNIVERSIDADE DE SÃO PAULO INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

PAULO MARQUES PIERRY

Transcritômica comparativa de cepas de Xylella fastidiosa

Versão original da Tese defendida

São Paulo

27/03/2017

PAULO MARQUES PIERRY

Transcritômica comparativa de cepas de Xylella fastidiosa

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Títulode Doutor em Ciências (Bioquímica)

Orientadora: Profa. Dra. Aline Maria da Silva

São Paulo 2017

DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO DO CONJUNTO DAS QUÍMICAS

Pierry, Paulo Marques

Transcritômica comparativa de cepas de *Xylella fastidiosa /* Paulo Marques Pierry. - São Paulo, 2017.

Orientadora: Aline Maria da Silva.

196 f.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Química, Programa de Pós-Graduação em Ciências Biológicas. Área de Concentração: Bioquímica.

Descritores: 1. Transcritômica. 2. Clorose Variegada de Citrus. 3. *Xylella fastidiosa*. I. da Silva, Aline Maria II. Universidade de São Paulo. Instituto de Química. Programa de Pós-Graduação em Ciências Biológicas (Bioquímica). III. Doutorado.

Dedico...

Aos meus pais e irmão, à Erika e a todos que me ajudaram.

AGRADECIMENTOS

À Profa. Dra. Aline Maria da Silva por todos os ensinamentos, atenção e suporte durante o Doutorado.

Ao Profs. Drs. Steve Lindow e Michelle Igo, da Universidade da Califórnia, EUA, pelas discussões e informações sobre o cultivo de *Xylella fastidiosa* no meio mínimo PIM-6.

Aos Profs. Drs. Sergio Verjovski-Almeida, Eduardo Reis, Walter Terra, Clélia Ferreira, Suely Gomes, Regina Baldini e Glaucia Souza por permitirem a utilização de seus equipamentos e por eventuais empréstimos de reagentes e materiais.

A todos os professores que tive contato em disciplinas e seminários em todos os Institutos da USP que frequentei.

À Dra. Layla Farage e à Claudia Teixeira, responsáveis pelo Centro Avançado de Tecnologias em Genômica (CATG) do IQ-USP, pela ajuda e discussões sobre os procedimentos de sequenciamento dos transcritomas.

Aos técnicos dos laboratórios do Departamento de Bioquímica do IQ-USP pela valiosa ajuda em diversos momentos, e principalmente ao Alexandre Sanchez, técnico do nosso laboratório e grande amigo, à Dóris Araújo e à Bianca Dazzani.

Aos colegas de laboratório Oséias Feitosa, Ana Paula de Souza, Luciana Antunes, Joaquim Martins, Deibs Barbosa, Paulo Zaini, Roberta Verciano e Raquel Riyuzo pela amizade, conversas, viagens, discussões, ajuda em experimentos e companhia durante o Doutorado.

Aos amigos e colegas bioinformatas, Deibs Barbosa, Joaquim Martins e Rodrigo Guarischi pela ajuda e discussões sobre os métodos de análises computacionais de dados de RNA-Seq.

Aos demais colegas da pós-graduação pela troca de experiências e disposição em colaborar.

Aos funcionários do IQ-USP por proporcionarem um ambiente agradável de trabalho.

Aos meus amigos de fora da USP pelo apoio incondicional durante todo meu Doutorado.

À minha namorada, Erika, por todo o amor, carinho, apoio, e maravilhosa companhia em todos os anos que estamos juntos.

Aos meus pais e família por me proporcionarem todo o amor, suporte e apoio para que eu pudesse me dedicar inteiramente ao Doutorado.

Ao CNPq, CAPES, FAPESP e NUBIC/USP pelo apoio financeiro e à FAPESP pela bolsa de Doutorado.

RESUMO

Pierry, Paulo Marques. **Transcritômica comparativa de cepas de** *Xylella fastidiosa.* 2017. 196p. Tese de Doutorado - Programa de Pós-Graduação Ciências Biológicas (Bioquímica). Instituto de Química, Universidade de São Paulo, São Paulo.

O fitopatógeno Xylella fastidiosa coloniza o lúmen dos vasos do xilema de seus hospedeiros e o aparelho bucal do inseto-vetor. É responsável por doenças de extrema gravidade em videira, laranjeira, e oliveira, entre outras plantas de relevância econômica. Há evidência de especificidade entre cepas de X. fastidiosa e as diferentes espécies de plantas que colonizam, mas as bases moleculares desta interação são desconhecidas. O objetivo central deste trabalho foi elucidar o repertório completo de genes expressos e/ou diferencialmente expressos por diferentes cepas X. *fastidiosa* em meios e tempos de cultivo distintos e relacionar as respostas transcricionais a mecanismos de virulência, patogenicidade e especificidade ao hospedeiro. Foram sequenciados, analisados e comparados os transcritomas de cepas de laranjeiras (9a5c, J1a12, U24d e Fb7), de cafeeiro (3124), de hibisco (Hib4), ameixeira (Pr8x) e de videira (Temecula1), no início e fim da fase a exponencial de crescimento populacional em meio rico PWG e em meio mínimo PIM6, que mimetiza a seiva do xilema. Foi observado que a maioria dos genes de X. fastidiosa é expressa, ainda que, dependendo da cepa e da condição experimental, 40-80% dos transcritos sejam pouco abundantes. Por outro lado, foi verificado um conjunto de transcritos muito abundantes, uma parte deles comuns a todas as cepas, e que incluem os ncRNAs 6S e RNAseP, além de transcritos de microcinas, proteases, lipases, proteínas de resposta a estresse e proteínas de função desconhecida. Além da definição de perfis transcricionais, foram descritas as regiões 5' e 3' não-traduzidas dos transcritos. As estruturas de 545 e 386 operons expressos, respectivamente pelas cepas 9a5c e Temecula1, também foram mapeadas, e pela primeira vez foi obtido o perfil de sRNAs expressos por X. fastidiosa. As análises de expressão diferencial entre transcritomas das duas fases de crescimento no mesmo meio indicam que o estresse gerado pela limitação nutricional do meio PIM6 exigiu mudanças mais drásticas na expressão gênica do que no meio PWG. Foi também observado que diferentes cepas respondem de maneiras distintas a uma mesma condição, indicando que genes ortólogos são regulados de formas diferentes. Além disso, a transcritômica comparativa revelou diferenças relevantes na regulação gênica de cepas de hospedeiros vegetais distintos que podem estar relacionadas a especificidade ao hospedeiro. Por fim, as análises dos transcritomas evidenciaram vários genes candidatos que poderão ser futuramente investigados quanto ao seu papel na biologia e na virulência de X. fastidiosa.

Palavras chave: fitopatógeno; *Xylella fastidiosa*; Clorose Variegada dos Citros; Doença de Pierce; transcritoma; RNA-Seq

ABSTRACT

Pierry, Paulo Marques. **Comparative transcriptomic of** *Xylella fastidiosa* strains. 2017. 196p. PhD Thesis - Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

The phytopathogen Xylella fastidiosa colonizes the lumen of xylem vessels from its hosts and the mouth apparatus of the insect-vector. It is responsible for severe diseases in grapevine, orange and olive trees, among other plants of economic relevance. There is evidence for specificity between X. fastidiosa strains and the different plant species they colonize, but the molecular bases of this interaction are unknown. The main objective of this work was elucidate the complete repertoire of expressed and/or differentially expressed genes by different X. fastidiosa strains in distinct media and growth times and associate transcriptional responses to virulence mechanisms, pathogenicity and host specificity. Transcriptomes of orange strains (9a5c, J1a12, U24d and Fb7), coffee (3124), hibiscus (Hib4), plum (Pr8x) and grapevine (Temecula1), were sequenced, analyzed and compared from cells at the beginning and end stages of exponential growth phase in rich medium PWG and in minimum medium PIM6, which mimics xylem sap. It was observed that the majority of X. fastidiosa genes is expressed, although, depending of the strain and experimental condition, 40-80% of transcripts are less abundant. On the other hand, it was verified a set of more abundant transcripts, some of them shared by all strains, including 6S and RNAseP ncRNAs as well as transcripts for microcins, proteases, lipases, stress response proteins and proteins of unknown function. Besides the definition of transcriptional profiles, 5' and 3' untranslated regions of transcripts were described. The structure of 545 and 386 expressed operons, respectively for 9a5c and Temecula1 strains, were also mapped, and for the first time the expressed profile of sRNAs in X. fastidiosa was obtained. The differential expression analyzes between transcriptomes of two growth phases in the same medium indicate that the stress generated by nutritional limitation of PIM6 medium required more drastic changes in gene expression than PWG medium. It was also observed that different strains respond in distinct manners to a same condition, indicating that orthologous genes are regulated in different ways. Moreover, comparative transcriptomics revealed relevant differences in gene regulation of strain of distinct plant hosts that can be related to host specificity. Lastly, transcriptomic analyzes pointed to several gene candidates that could be further investigated for their roles in X. fastidiosa biology and virulence.

Keywords: phytopathogen; *Xylella fastidiosa*; Citrus Variegated Chlorosis; Pierce´s Disease, transcriptome; RNA-Seq.

LISTA DE ABREVIATURAS E SIGLAS

ALS: Escaldadura da folha de amendoeira asRNA: RNA anti-senso ATP: adenosina trifosfato BLAST: Basic Local Alignment Search Tool CAGe: Cap Analysis of Gene Expression CATG: Centro Avançado de Tecnologias em Genômica, IQ-USP c-di-GMP: diguanilato cíclico ou di-GMP cíclico cDNA: DNA complementar CDS: sequência codificadora CEFAP: Centro de Facilidades de Apoio à Pesquisa, ICB-USP. CLS: Escaldadura da folha de cafeeiro CVC: Clorose Variegada dos Citros CWDE: enzima degradadora de parede celular DEPC: dicarbonato de dietila DNAse: desoxirribonuclease dNTP: desoxirribonucleotídeo 5' trifosfato DO: densidade óptica DRS: Direct RNA Sequencing DSF: Diffusible Signaling Factor EDTA: ácido etilenodiaminotetracético EFSA: European Food Safety Authority **ESTs:** Expressed Sequence Tags FPKM: Fragments per kilobase transcript per million reads FUNDECITRUS: Fundo de Defesa da Citricultura GO: Gene Ontology IMG/ER: Integrated Microbial Genomes-Expert Review KEGG: Kyoto Encyclopedia of Genes and Genomes LPS: lipopolissacarídeo MIAME: Minimum Information about a Microarray Experiment MIRA: Mimicking Intelligent Read Assembly mRNA: RNA mensageiro NCBI: National Center for Biotechnology Information ncRNA: non coding RNA ou RNA não codificador OLS: Escaldadura da folha de oleandro ou espirradeira OMV: vesícula de membrana externa OQDS: Síndrome do Rápido Declínio das Oliveiras **ORESTES:** Open Reading Frame Expressed Sequence Tags **ORF**: *open read frame* pb: pares de base PCR: reação em cadeia da polimerase PD: Doença de Pierce PE: Paired-End

PET: PE Tags

PLS-peach: Escaldadura da folha de pessegueiro PLS-plum: Escaldadura da folha de ameixeira qPCR: PCR quantitativo RBS: *ribosome binding site* RLK: quinases do tipo receptor RLP: proteínas do tipo receptor **RNAP: RNA polimerase** RNase: ribonuclease RPKM: Reads per kilobase per million mapped reads rpm: rotações por minuto rRNA: RNA ribossomal RT-PCR: PCR precedido por transcrição reversa RT-qPCR: qPCR precedido de transcrição reversa SNP: polimorfismo de um único nucleotídeo sRNA: small RNA ou pequeno RNA TE: Tampão Tris contendo EDTA TF: transcription factor ou fator de transcrição Tris: tris-(hidroximetil)-aminometano tRNA: RNA transportador **TSS:** Transcription Start Sites UTR: untranslated region ou região não traduzida UV: ultravioleta **VT**: Virtual Terminator

SUMÁRIO

1. Introdução	12
1.1. Regulação gênica em bactérias	12
1.1.1. Regulação do início da transcrição	13
1.1.1.1. Fatores de transcrição	14
1.1.1.2. Fatores Sigma (σ)	16
1.1.1.3. Regulação epigenética	20
1.1.2. Regulação do término da transcrição e pós-transcricional	21
1.1.2.1. Antiterminação e atenuação	22
1.1.2.2. Riboswitches	23
1.1.2.3. RNAs não codificadores	26
1.1.2.4. RNAs anti-senso	28
1.1.2.5. Regulação da estabilidade do transcrito	30
1.1.3. Mecanismos de sinalização celular em bactérias	33
1.2. Métodos de análise de transcritomas	36
1.2.1. Microarranios de DNA	37
1.2.2. RNA-Seq	40
1.3. Aspectos da biologia de Xvlella fastidiosa	47
1.3.1. Doencas causadas por X. <i>fastidiosa</i> , formas de transmissão e controle	248
1.3.2 As epidemias de X fastidiosa	51
1.3.3. Variabilidade entre cenas e genomas de <i>Xylella</i>	
134 Fatores de virulência de X <i>fastidiosa</i>	59
135 Análises transcritômicas em X fastidiosa	66
2. Objetivos	72
3. Procedimentos experimentais	
3.1. Manutenção e cultivo de Xvlella fastidiosa.	
3.2. Extração de RNA total	74
3.3. Purificação de RNA total com DNase	75
3.4. Verificação da eficiência do tratamento com DNase	76
3.5. Preparação de bibliotecas de cDNA para seguenciamento (RNA-Seg)	77
3.5.1. Ouantificação de RNA por método fluorimétrico	77
3.5.2. Depleção de rRNAs	77
3.5.3. Preparação de bibliotecas de cDNA	
3.5.4 Preparação de bibliotecas de pequenos RNAs (sRNAs)	78
3.5.5 Ouantificação de bibliotecas de cDNA por aPCR absoluto	79
3.6 Sequenciamento de DNA no equipamento MiSeq (Illumina)	80
3.7 Análises bioinformáticas e estatísticas	81
4. Resultados e Discussão	83
4.1. Curva de crescimento das cepas 9a5c e Temecula1 em meio PWG e PIM6	85
4.2 Obtenção dos transcritomas por RNA-Seq	88
4.3 Análise da qualidade das sequências dos transcritomas	91
4.4 Mapeamento das sequências dos transcritomas nos genomas de referência	93
4.5 Análise de correlação entre réplicas biológicas e entre diferentes transcritom	as .95
4.6. Identificação de regiões não-traduzidas (UTR_s) e operons	101
4.7 Análise dos transcritomas baseada nos valores de FPKM	104
4.7.1. Panorama da expressão gênica das cepas 9a5c e Temecula1	104
4.7.2. Descrição dos genes mais expressos	107
4.7.3. Análise de rotas metabólicas expressas	
4.8. Análises de expressão gênica diferencial	
4.8.1. Comparações entre os transcritomas	115
1 5	

4.8.2. Categorização funcional dos genes diferencialmente expressos	2
4.8.2.1. Análise por ontologia gênica12	2
4.8.2.2. Análise por vias KEGG12	7
4.8.2.3. Análise de genes relacionados à virulência e patogenicidade12	9
4.9. Identificação de pequenos RNAs (sRNAs) expressos pela cepa 9a5c13	4
4.10. Curva de crescimento populacional de seis cepas de X. fastidiosa	8
4.11. Obtenção dos transcritomas por RNA-Seq14	0
4.12. Análises da qualidade das sequências dos transcritomas	3
4.13. Mapeamento das sequências dos transcritomas nos genomas de referência14	4
4.14. Análise de correlação das réplicas biológicas e entre diferentes transcritomas 14.	5
4.15. Análise dos transcritomas baseada nos valores de FPKM	1
4.15.1. Panorama da expressão gênica das seis cepas	1
4.15.2. Descrição dos genes mais expressos	5
4.15.3. Análise de rotas metabólicas expressas	1
4.16. Análises de expressão gênica diferencial16	2
4.16.1. Análise de genes relacionados à virulência e patogenicidade	7
5. Considerações finais	4
6. Bibliografia17	6
Súmula curricular	7

1. Introdução

1.1. Regulação gênica em bactérias

Microrganismos estão sempre sujeitos a variadas mudanças nas condições ambientais, tais como flutuações na disponibilidade de nutrientes ou na presença de agentes tóxicos, que podem afetar sua densidade celular e/ou causar estresses diversos em suas células. Para enfrentar tais adversidades e garantir sua sobrevivência, as células devem ser capazes de responder prontamente, alterando o conjunto de genes expressos de modo a manter o funcionamento adequado de suas atividades celulares. A rápida adaptação de microrganismos, principalmente de bactérias, pode ser explicada pela plasticidade de suas redes de regulação transcricional (Balleza *et al.*, 2009).

A expressão do repertório gênico é finamente controlada por meio de diversos mecanismos reguladores, promovendo a eficiência metabólica e energética dos processos celulares, sendo que nenhum produto celular deverá ser sintetizado se não for necessário naquele momento. A mudança da expressão de genes na célula é desencadeada pela percepção de sinais externos, que podem ou não ser internalizados, embora a maioria seja detectado por sensores proteicos localizados nas membranas celulares. Tais sinais desencadeiam vias complexas de sinalização celular que resultam na ativação ou repressão seletiva de genes (Marques, 2012; Snyder *et al.*, 2013).

Em cromossomos bacterianos, os genes estão frequentemente organizados em operons, sendo, portanto, transcritos como mRNAs policistrônicos. Dessa forma, uma unidade de transcrição pode ser definida por uma região promotora regulatória, um sítio de início da transcrição, uma ou mais regiões codificantes (ORF ou CDS) e um sítio de terminação da transcrição (Browning e Busby, 2004). Com isso, bactérias podem responder de forma rápida e eficiente aos sinais que recebem, uma vez que o mRNA policistrônico codifica ORFs que atuam, geralmente, em uma mesma rota metabólica (Keene, 2007). O modelo de operon foi proposto como mecanismo de regulação coordenada da expressão de genes relacionados com o transporte e catabolismo de lactose (Jacob e Monod, 1961). Nesse modelo, a proteína repressora (repressor *lac*) quando ligada ao indutor, no caso alolactose, sofre mudança conformacional e perde afinidade pelo operador (elemento regulatório ao qual o repressor se liga e que controla

a expressão dos genes do operon), resultando na desrepressão do operon (Lewis, 2005). Em operons, unidades transcricionais se sobrepõem; sempre há um promotor que controla a transcrição do conjunto completo de genes, mas podem existir promotores internos regulando determinados genes do operon, ou mesmo regiões de terminação da transcrição prematuras, fazendo com que apenas alguns genes do operon sejam parte do mRNA policistrônico (Browning e Busby, 2004).

A regulação da expressão gênica em bactérias pode ocorrer em diversos momentos desde ativação da transcrição até a estabilidade do produto gênico final (Silva-Rocha e de Lorenzo, 2010). A transcrição é o principal ponto de regulação da expressão gênica (Browning e Busby, 2004), porém mecanismos pós-transcricionais, tais como a regulação da tradução ou da estabilidade de transcritos e de proteínas são também importantes em bactérias. Os diversos pontos e mecanismos de regulação gênica em bactérias serão detalhados adiante.

Ademais, vale destacar que a regulação gênica bacteriana pode ser mediada, em seus diferentes níveis, tanto pela ação de proteínas como de RNAs regulatórios. Como também detalharemos adiante, proteínas regulatórias estão representadas, por exemplo, pelos fatores σ da RNA polimerase (RNAP) e fatores de transcrição (TF) (Balleza *et al.*, 2009) enquanto que os RNAs regulatórios estão representados pelos *riboswitches* (Winkler e Breaker, 2003), RNAs não codificadores atuantes em *trans (non coding RNAs*, ncRNAs) (Mandin *et al.*, 2013) e RNAs anti-senso atuantes em *cis* (asRNAs) (Sesto *et al.*, 2013).

1.1.1. Regulação do início da transcrição

O início da transcrição depende do reconhecimento da sequência de DNA específica do gene a ser transcrito pela maquinaria de transcrição, além da formação de um complexo aberto (bolha de transcrição) no DNA, que irá preceder o processo de elongação da transcrição. Controlar o início da transcrição é uma das principais formas de regulação gênica e ocorre por meio de diversos mecanismos, que em sua maioria envolvem a ligação de moléculas a sequências específicas no DNA localizadas principalmente na região promotora do gene a ser regulado.

1.1.1.1. Fatores de transcrição

Fatores de transcrição são proteínas regulatórias que reconhecem sequências específicas no DNA e promovem a ativação ou inibição da transcrição de seus genes alvos. Aqueles que estimulam a transcrição são chamados de ativadores e agem estabilizando a ligação da RNAP ao promotor. Seus sítios de ligação normalmente estão localizados a montante da posição -35 do promotor, mas também podem ser encontrados entre os sítios -35 e -10. Enquanto isso, aqueles que impedem o acesso da RNAP ao promotor são chamados de repressores e seus sítios de ligação são chamados de operadores, sendo normalmente encontrados à jusante da posição -35 do promotor. Geralmente, o número de genes codificando fatores de transcrição aumenta com o número total de genes do organismo. Tais fatores são difusíveis pela célula e, portanto, são considerados elementos *trans*-atuantes que se ligam aos elementos *cis*-regulatórios no DNA localizados nas regiões promotoras dos genes regulados (Browning e Busby, 2004; Balleza *et al.*, 2009).

Em alguns casos, os fatores de transcrição não são capazes de, isoladamente, se ligar aos seus sítios de ligação. Moléculas como íons, açúcares, aminoácidos e outros metabólitos podem atuar como indutores, caso participem na ligação de um fator ativador ou na remoção de um repressor, ou co-repressores, caso participem da ligação de um fator repressor ou remoção de um ativador. A grande maioria dessas moléculas atua na regulação de genes que codificam vias metabólicas relacionadas à sua síntese, em um processo de regulação por *feedback* (Browning e Busby, 2004; Marques, 2012).

Fatores de transcrição e proteínas regulatórias são classificados em diversas famílias, nas quais compartilham domínios funcionais conservados (Snyder *et al.*, 2013). Pelo menos dois domínios são necessários: um domínio sensor de sinal, que pode diretamente se associar a um ligante ou interagir com outras proteínas; e outro domínio que interage diretamente com o seu respectivo sítio de ligação na sequência do DNA, sendo o motivo estrutural hélice-volta-hélice o mais encontrado nos reguladores bacterianos. Apesar de membros da mesma família ter em comum determinado domínio, não necessariamente eles respondem ao mesmo estímulo ou se ligam às mesmas sequências no DNA (Balleza *et al.*, 2009).

Geralmente, reguladores negativos se ligam aos promotores e interferem diretamente na RNAP, enquanto que reguladores positivos se ligam a sequências específicas a montante do promotor e auxiliam no início da transcrição, embora existam fatores de transcrição que apresentam papel regulatório duplo (Madan Babu e Teichmann, 2003; Balleza *et al.*, 2009). Além disso, mais de um fator de transcrição podem atuar juntos em uma mesma região regulatória. Há que se considerar também a ocorrência de uma degeneração da interação entre os fatores e seus sítios de ligação, ou seja, existem diferentes sítios capazes de recrutar o mesmo fator e diferentes fatores podem reconhecer sítios similares (Balleza *et al.*, 2009).

O acesso dos fatores de transcrição para interagir com o DNA também é regulado pelo nível de empacotamento do material genético (Willenbrock e Ussery, 2004). Diversas proteínas (DNA isomerases, DNA chaperonas, proteínas acessórias) atuam nos processos de dobramento e enovelamento do cromossomo bacteriano, o que faz com que algumas regiões fiquem inacessíveis à ligação de proteínas ou da maquinaria transcricional (Balleza *et al.*, 2009). Além do nível de empacotamento, a composição de bases nas regiões regulatórias também constitui um fator relevante para o início da transcrição, uma vez que regiões com um maior conteúdo de A+T do que outras tendem a ser mais facilmente acessadas pela maquinaria transcricional (Dekhtyar *et al.*, 2008).

Um conceito importante na regulação da expressão gênica por fatores de transcrição é o de *regulon* (Balleza *et al.*, 2009), o qual compreende um grupo de genesalvo que sofrem influência de um mesmo conjunto de fatores de transcrição que respondem a determinado estímulo ou condição celular. Por sua vez, um gene também pode ser alvo de mais de um fator de transcrição, pertencendo a mais de um *regulon*. Normalmente, genes presentes no mesmo *regulon* codificam proteínas que participam de processos celulares relacionados, o que justifica o fato deles serem co-regulados. Por outro lado, vários genes podem ser controlados por um determinado estímulo, que acionaria diferentes fatores de transcrição, os quais regulariam seus respectivos alvos, constituindo o chamado *stimulon* (Marques, 2012).

Em *Escherichia coli*, a maioria dos genes de fatores de transcrição estão localizados próximos de seus genes alvos (Janga *et al.*, 2007). Esta proximidade constitui uma vantagem para a ação direta destes fatores que rapidamente podem encontrar seus sítios regulatórios, uma vez, que muitas vezes, a sua concentração na célula é baixa (Kolesov *et al.*, 2007). Ademais, os *regulons* são evolutivamente flexíveis, pois genes codificadores de fatores de transcrição e seus respectivos sítios de ligação podem ser

adquiridos ou perdidos a partir da transferência horizontal gênica e outros eventos que possam alterar a estrutura do genoma bacteriano (Balleza *et al.*, 2009).

Fatores de transcrição com maior afinidade a determinados sítios tendem a se manter mais tempo ligados ao DNA, sugerindo que tal gene alvo precise ter uma regulação, positiva ou negativa, por um tempo prolongado. No caso de ativadores transcricionais, quando os promotores são fortes e os fatores de transcrição são abundantes, de modo geral a transcrição ocorrerá em uma taxa bem definida. Porém, quando o promotor é fraco e a concentração do fator oscila, a transcrição do gene dependerá de flutuações do acaso para ocorrer (Balleza *et al.*, 2009; Silva-Rocha e de Lorenzo, 2010).

1.1.1.2. Fatores Sigma (σ)

A RNAP bacteriana apresenta cinco subunidades proteicas principais formando seu cerne ou *core*: duas subunidades α (alfa), uma β (beta), uma β ' (beta linha) e uma ω (ômega). O cerne da enzima é capaz de se ligar ao DNA de forma não-específica e iniciar a síntese de RNA a partir de extremidades ou quebras no DNA (Saecker *et al.*, 2011). Entretanto, para iniciar uma transcrição produtiva a partir de sequências específicas precedidas por um promotor, é exigida a presença de uma subunidade proteica dissociável, chamada fator σ (sigma), que se liga ao cerne da RNAP formando a holoenzima (Paget, 2015).

Os fatores σ desempenham diversos papéis-chave na iniciação da transcrição, incluindo o reconhecimento inicial de sequências promotoras, a abertura e estabilização do complexo aberto do DNA no promotor, a interação com outras moléculas ativadoras, além de influenciar na dissociação da RNAP do promotor (o fator σ se desliga da RNAP assim que a fase de elongação é iniciada) (Saecker *et al.*, 2011; Feklistov *et al.*, 2014; Paget, 2015). Enquanto que as funções na iniciação da transcrição nos eucariotos estão divididas entre muitas proteínas, o fator σ procariótico realiza todas essas funções na iniciação (Feklistov *et al.*, 2014).

Existem diferentes fatores σ nas células, sendo que cada um é responsável por regular a transcrição de um determinado conjunto de genes para lidar com as condições desencadeadas por estímulos intra e extracelulares, dirigindo a RNAP à posição correta iniciar a transcrição (Feklistov *et al.*, 2014). Os estímulos podem ser carência

nutricional, início da fase estacionária, mudanças de temperatura, diferentes tipos de estresses (oxidativo, osmótico), entre outros. Os fatores σ são divididos em duas famílias e diversas subfamílias (ou grupos), baseado nas suas origens filogenéticas: a família do σ^{70} (sigma 70) e a família do σ^{54} (sigma 54) (Paget, 2015). Dessa forma, haverá uma competição pelo cerne da RNAP entre os diferentes fatores σ , sendo que a condição atual da célula irá guiar a predominância de uns pelos outros. Uma vez que um gene pode possuir mais de um promotor e, consequentemente, ser regulado por mais de um fator sigma, ele pode ser expresso em mais de uma condição que a célula é exposta.

A família do σ^{70} compreende um maior número de fatores relacionados com as mais variadas funções durante o crescimento. Os membros dessa família foram filogeneticamente e estruturalmente classificados em quatro grupos distintos, os quais são diferenciados pela presença ou ausência de quatro regiões conservadas que correspondem a quatro domínios estruturais (1.1, 2, 3 e 4). Tais domínios estão associados com a ligação ao DNA nas regiões promotoras -10, -10 estendida e -35, características de membros dessa família (Feklistov *et al.*, 2014; Paget, 2015).

O grupo 1 consiste de fatores σ primários e contém todos os domínios estruturais e, consequentemente, podem atingir massas moleculares de ~70 kDa. O fator σ^{70} é o mais abundante e basal na célula sendo responsável pela transcrição de genes relacionados à manutenção do metabolismo e atividade celulares. Um dos pontos de controle de sua atividade é a ligação com uma proteína regulatória, chamada Rsd, a qual o sequestra e impede a ação do fator σ^{70} , atuando como fator anti-sigma. Foi observado que a concentração de Rsd aumenta duas vezes com a entrada de *E. coli* na fase estacionária e que sua ligação ao fator σ^{70} é mais efetiva nesta fase (Paget, 2015).

Os grupos 2 a 4 compreendem os fatores σ alternativos com funções especializadas, e se diferenciam do grupo 1 pela completa ausência do domínio 1.1, a presença variável do domínio 3 e a especificidade de sequências no promotor. Fatores σ do grupo 2 estão envolvidos na adaptação a estresses, incluindo limitação nutricional e outros associados a entrada em fase estacionária. O σ^{S} ou σ^{38} é o mais estudado dos fatores σ do grupo 2 em *E. coli*. Este fator é responsável pela resposta geral a estresse e pela sobrevivência durante a fase estacionária, sendo induzido em resposta a uma variedade de estresses ambientais (Battesti *et al.*, 2011). Membros do grupo 3 são estruturalmente e funcionalmente diversos, mas normalmente contém os domínios 2, 3 e 4 (Feklistov *et al.*, 2014). Eles foram categorizados em quatro subgrupos filogeneticamente distintos e que parcialmente correlacionam-se com suas funções: biossíntese de flagelo (σ^{28} de *E. coli*), resposta a choque de calor (σ^{32} de *E. coli*), estresse geral (σ^{B} de *B. subtilis*) e esporulação (quatro fatores σ em *B. subtilis*) (Paget, 2015). Finalmente, membros do grupo 4 são também conhecidos como grupo com função extracitoplasmática (ECF), dado a sua capacidade de perceber e responder a sinais gerados fora da célula ou na membrana celular (Lonetto *et al.*, 1994). Trata-se do maior e mais diverso grupo de fatores σ em nível de sequência primária, consistindo de, pelo menos, 43 subgrupos filogeneticamente distintos (Staron *et al.*, 2009). Dentre os seus papéis biológicos estão a resposta ao estresse no envelope celular, transporte de ferro, resposta ao estresse oxidativo e estresses em geral, além da regulação de genes de virulência (Kazmierczak *et al.*, 2005; Paget, 2015). Uma vez que fatores σ ECF não possuem os domínios 1.1 e 3, eles são menores, apresentando ~20 kDa (Feklistov *et al.*, 2014; Paget, 2015).

O fator σ^{54} pertence a uma família diferente dos outros fatores σ . Além da separação filogenética, este fator tem a particularidade de depender de uma sequência ativadora no DNA (enhancer), distante ~100 pb à montante do sítio de início da transcrição e que promove a ligação efetiva a suas sequências específicas no promotor. É nesta sequência *enhancer* que ativadores transcricionais irão se ligar, hidrolisar ATP e catalisar a formação de complexos abertos no promotor (Kazmierczak et al., 2005; Zhang e Buck, 2015). O fator σ^{54} reconhece sequências específicas e conservadas nas regiões -12 e -24 dos promotores que regula, sendo mais uma distinção dos membros da família do σ^{70} (Kazmierczak *et al.*, 2005). σ^{54} contém dois domínios conservados (regiões I e III) separados por uma alça de ligação (região II). A região I interage com ATPases, o cerne da RNAP e com a sequência -12 do promotor. Já a região III está principalmente envolvida em se ligar à sequência -24 do promotor. A região II é dispensável, estando ausente no fator σ^{54} de algumas espécies bacterianas. O σ^{54} regula a transcrição de genes relacionados a vários processos como metabolismo de nitrogênio, utilização de fontes alternativas de carbono, quimiotaxia, estresses na membrana e estresses em geral (Zhang e Buck, 2015). Devido à necessidade da ligação de uma proteína ativadora, o processo de transcrição por este fator é melhor regulado, além de poder haver modificações pós-traducionais nas proteínas ativadoras (fosforilação, por exemplo). Outra vantagem deste mecanismo de regulação é que mesmo que dois genes possam ter promotores que se ligam ao fator σ^{54} , eles podem ter proteínas ativadoras

diferentes e, consequentemente, serem regulados em condições distintas (Marques, 2012).

A concentração dos fatores σ pode ser controlada no nível de transcrição, tradução e nível de proteína. Alguns fatores "pró-sigma" são ativados por proteólise, envolvendo a remoção de uma determinada sequência inibitória, geralmente localizada na extremidade N-terminal. Porém, na maioria dos casos a expressão e atividade dos fatores sigma alternativos são controladas pela ação de fatores anti-sigma, os quais são proteínas que se ligam aos fatores σ e impedem que eles se liguem à RNAP. O fator anti-sigma oclui sequências-chave para a ligação, através de interações com os domínios 2 e 4 do fator sigma (Paget, 2015).

Em resposta a algum sinal ou estímulo específico, os fatores anti-sigma desligam-se e liberam os fatores σ , envolvendo diversos mecanismos distintos. Pode ocorrer a percepção direta do sinal pelo anti-sigma, seguida de uma mudança conformacional que liberaria o fator σ . Um sistema de transdução de sinal pode também estar envolvido, no qual uma vez recebido o estímulo extracitoplasmático, clivagens do anti-sigma na membrana liberariam o fator σ no citoplasma. Portanto, fatores anti-sigma são compostos de um domínio de ligação ao fator σ e um domínio sensor/sinalização que responde a um sinal tanto dentro como fora da célula (Paget, 2015).

A atividade dos fatores σ na célula também pode ser regulada por ncRNAs. Um dos exemplos mais conhecidos é do pequeno RNA 6S, que, em *E. coli*, possui 184 nucleotídeos e que se liga à RNAP quando ela está associada ao σ^{70} , impedindo a ligação da holoenzima aos promotores dependentes de σ^{70} . As concentrações do RNA 6S aumentam quando as células entram na fase estacionária, o que resulta no sequestro de RNAP ligadas ao σ^{70} (Wassarman, 2007). Outras moléculas de RNAP ficam então livres para interagir com fatores σ alternativos. Quando a célula sai da fase estacionária e o crescimento é retomado, o aumento na concentração citoplasmática de nucleotídeos trifosfatados é o sinal que habilita a RNAP- σ^{70} a ter atividade de RNA polimerase dependente de RNA e utilizar o próprio 6S como molde para síntese da RNAP de 14–20 nucleotídeos (pRNA). O duplex do 6S com pRNA dissocia-se da RNAP- σ^{70} , é degradado e a holoenzima fica, portanto, livre para associar-se ao promotores dependente de σ^{70} (Wassarman, 2007; Mandin *et al.*, 2013).

1.1.1.3. Regulação epigenética

A participação de mecanismos epigenéticos na regulação gênica tem sido muito estudada na última década. Trata-se de modificações estáveis no DNA com implicações na sua estrutura e acessibilidade da maquinaria de transcrição, mas que não envolvem mutações, rearranjos ou modificações na sequência de bases do DNA. Tais modificações (ex: metilação de bases do DNA; metilação de histonas) são herdáveis pela progênie e correspondem a qualquer informação adicional imposta na sequência de DNA que possa afetar a expressão gênica (Casadesus e Low, 2006; Bird, 2007).

A metilação de adeninas do DNA é o mecanismo epigenético mais conhecido, e ocorre pela transferência do grupo metil da S-adenosilmetionina (SAM) para a sexta posição de resíduos de adenina (N6-metil-adenina ou m6A) por metilases de DNA (Balleza *et al.*, 2009; Su *et al.*, 2016b). O estado metilado frequentemente está associado com repressão da transcrição. A ligação de grupos metil ao DNA ocorre em sítios-alvo específicos, sendo que o padrão de metilação pode variar entre células (Casadesus e Low, 2006). Em bactérias, a remoção da metilação no DNA é normalmente atingida por dois eventos de replicação (Su *et al.*, 2016b).

A metilação do DNA desempenha importantes papéis na biologia bacteriana. Processos como o tempo de replicação do DNA, a divisão dos cromossomos para as células filhas, o reparo do DNA, e o tempo de transposição e conjugação de plasmídeos estão sensíveis aos estados de metilação de regiões específicas no DNA. Todos estes eventos utilizam como sinal o estado de hemimetilação da nova molécula de DNA sintetizada, na qual uma das fitas não está metilada, devido ao processo de replicação semiconservativa (Casadesus e Low, 2006). Esta modificação no DNA pode alterar interações de proteínas regulatórias através de um efeito estérico direto ou por modificação da estrutura do DNA (Casadesus e Low, 2006; Balleza *et al.*, 2009).

O fenômeno chamado de bi-estabilidade (*bistability*) ou variação de fase (*phase variation*) pode acontecer em populações bacterianas nas quais existem dois ou mais estados fenotípicos distintos, controlados por mecanismos epigenéticos com diferentes níveis de complexidade (Casadesus e Low, 2013).

As metilases regulatórias epigenéticas parecem ser derivadas dos conhecidos sistemas de restrição/modificação bacterianos. Trata-se de um sistema de defesa da célula contra sequências de DNA externo ou mesmo viral, os quais são degradados

pelas enzimas de restrição antes que possam ser metilados. De fato, a maioria das metilases presentes em genomas bacterianos são componentes desses sistemas, porém as metilases epigenéticas não possuem uma enzima de restrição cognata, sendo chamadas de "órfãs" (Casadesus e Low, 2006).

Há ainda uma herança epigenética não relacionada ao material genético, realizada através da transmissão para as células-filhas de componentes celulares do citoplasma da célula-mãe em cada ciclo de divisão celular (Balleza *et al.*, 2009). Além disso, algumas espécies bacterianas apresentam complexos eventos de diferenciação celular, os quais alteram propriedades morfológicas e fisiológicas da célula sem que a sequência do DNA seja alterada (Casadesus e Low, 2013).

Em resumo, populações bacterianas podem usar esses padrões de metilação do DNA herdados como uma memória de curto-prazo das condições metabólicas encontradas pela geração anterior, conferindo uma vantagem evolutiva para elas. Dessa forma, se as células bacterianas podem guardar essa informação acerca de experiências passadas, podem usar essa memória para modular seu comportamento na atual condição em que se encontram (Casadesus e Low, 2006; Wolf *et al.*, 2008).

1.1.2. Regulação do término da transcrição e pós-transcricional

Além dos mecanismos regulatórios no início da transcrição, existem também mecanismos que regulam o término da transcrição. A maquinaria transcricional, durante seu processo de elongação, pode encontrar estruturas ou moléculas regulatórias que possam desestabilizá-la, podendo ocasionar uma terminação prematura. Uma vez que, em bactérias, os processos de transcrição e tradução estão frequentemente acoplados, mecanismos de regulação pós-transcricional podem rapidamente atuar no transcrito nascente e exercerem seus efeitos. A regulação pós-transcricional vai depender principalmente de dois fatores: a estabilidade do mRNA, ou seja, sua meia-vida antes que seja degradado, e a capacidade do ribossomo em acessar e se ligar a sequências específicas no mRNA e, consequentemente, iniciar a tradução.

1.1.2.1. Antiterminação e atenuação

Os mecanismos de terminação da transcrição em procariotos podem ser divididos em duas classes: terminação intrínseca, que depende de estruturas em sequências específicas no DNA; terminação dependente da proteína Rho, que independe de segmentos estruturados nas sequências as quais se liga, podendo terminar a transcrição em diferentes sítios no genoma (Nudler e Gottesman, 2002). Neste último mecanismo, a proteína Rho se liga às suas sequências específicas em forma de um hexâmetro e progride na direção 3' do transcrito e, assim que encontrar a RNAP em um sítio de pausa, aborta a transcrição e libera o transcrito (Henkin e Yanofsky, 2002).

Como mencionado anteriormente, os genes de um operon são, na maiorias das vezes, transcritos com um mRNA policistrônico. Porém, em algumas situações, ocorre a transcrição seletiva de genes do operon, dirigida por promotores internos ou alternativos ou pela formação de uma região de terminação da transcrição interna. Neste último caso, dois mecanismos podem ser utilizados pela célula: antiterminação, quando a transcrição acontece sem interrupção; e atenuação, quando transcrição é interrompida (Marques, 2012). A antiterminação acontece quando a molécula regulatória induz a continuação da transcrição, sendo que na ausência do ligante, a terminação da transcrição ocorreria. Já na atenuação da transcrição ocorre o inverso, ou seja, sem a molécula regulatória a transcrição ocorreria normalmente, porém na sua presença a terminação na transcrição é induzida (Gollnick e Babitzke, 2002). Tais moléculas regulatórias atuam em *trans*, ligando-se ao mRNA, alterando sua estrutura secundária e favorecendo ou não a formação da estrutura de terminação da transcrição (Gollnick e Babitzke, 2002).

Na terminação intrínseca bacteriana, um segmento do transcrito sendo sintetizado pela RNAP forma uma estrutura secundária estável, a qual irá afetar o andamento da maquinaria de elongação, podendo gerar pausas. Este mecanismo de terminação também depende de uma sequência rica em nucleotídeos de uridina após a estrutura secundária (Henkin e Yanofsky, 2002; Nudler e Gottesman, 2002). Além dessas características, a reação é estimulada pela presença de proteínas adicionais ligadas ao complexo de elongação que regulam o andamento da RNAP, sendo específicas e se associando somente aos seus respectivos terminadores. No mecanismo de antiterminação, a ação dessas proteínas regulatórias faz com que a transcrição não acabe

no terminador. Um exemplo deste mecanismo ocorre nos operons que codificam os RNA ribossomais em *E. coli*, em que a anti-terminação é mediada pelo complexo de proteínas da família *Nus* (Nudler e Gottesman, 2002).

Alguns terminadores intrínsecos, chamados de atenuadores, estão localizados na sequência líder do primeiro gene de um operon e controlam diretamente a eficiência da transcrição de genes a jusante (Yanofsky, 2000). A formação ou não da estrutura secundária pode determinar se aquela unidade transcricional será transcrita. No processo de atenuação da transcrição, é regulada a capacidade da maquinaria transcricional de passar por este terminador intrínseco localizado no início (Marques, 2012). O primeiro relato de atenuação da transcrição foi com o operon de genes relacionados com a biossíntese do aminoácido triptofano em *E. coli* (Yanofsky, 1981). Além disso, outros operons de biossíntese de aminoácidos também foram descritos como regulados por este mecanismo, sendo que a ação deste sítio atenuador irá depender da concentração do aminoácido específico na célula (Gollnick e Babitzke, 2002).

O mecanismo de atenuação da transcrição é uma estratégia regulatória comum e conservada entre genes ortólogos de espécies procarióticas filogeneticamente distantes (Henkin e Yanofsky, 2002; Merino e Yanofsky, 2005). Além disso, a conservação se mantém independente da organização dos respectivos operons. Tal fato sugere que possa ter acontecido uma seleção positiva para este mecanismo de atenuação da transcrição, promovendo vantagens evolutivas. Embora o mecanismo sensor de um sinal metabólico possa variar entre organismos, a estratégia de resposta foi conservada (Merino e Yanofsky, 2005). Uma das principais vantagens de regular a expressão gênica por estes mecanismos é que pequenas sequências e estruturas no RNA podem mediar decisões regulatórias importantes (Henkin e Yanofsky, 2002).

1.1.2.2. Riboswitches

Uma das estratégias de regulação transcricional e pós-transcricional é a utilização de elementos regulatórios chamados *riboswitches*. Estas estruturas são exclusivamente encontradas na região à montante do códon de início de tradução, na chamada região 5' não-traduzida (5'UTR), a qual não é codificante. A regulação é baseada na percepção de diversos sinais intra e extracelulares, tais como variações em parâmetros físico-químicos ou na concentração de determinada molécula. Ao contrário dos fatores de

transcrição proteicos e dos ncRNAs regulatórios que atuam regulação em *trans*, os *riboswitches* estão localizados no próprio transcrito onde irão atuar e, portanto, atuam em *cis* (Winkler e Breaker, 2005).

Estes elementos podem modular a expressão da sequência codificadora, uma vez que podem interferir na elongação da transcrição, na estabilidade do RNA e na eficácia da tradução do mRNA (Breaker, 2011). O controle mediado pelo *riboswitch* pode operar negativamente na formação de um terminador transcricional, inibindo a síntese completa do transcrito, ou inibir a tradução a partir da geração de estruturas secundárias em sequências importantes, como no *ribosome binding site* (RBS) (sequência *Shine-Dalgarno*) e no códon de início da tradução. Em outros casos, sua ação pode ser positiva, liberando as sequências-chave para que a tradução possa ocorrer (Mandin *et al.*, 2013).

Os *riboswitches* sensíveis a parâmetros físico-químicos podem ser classificados em responsivos a mudanças na temperatura, no pH e na força iônica. Os primeiros são chamados de elementos termossensores ou RNA termômetros e a estrutura de sua região 5'UTR é afetada por mudanças de temperatura, devido a enovelamentos alternativos que a molécula pode assumir. Dentre os RNAs termômetros estão os mRNAs dos genes *rpoH e cspA*, os quais respondem ao choque por calor e por frio, respectivamente, sendo encontrados em várias espécies bacterianas. Exemplos de outros mRNAs que apresentam tais estruturas codificam proteínas com atividade de chaperonas ou de proteases, que atuam sobre proteínas desnaturadas pelo estresse térmico (Mandin *et al.*, 2013; Sherwood e Henkin, 2016).

O primeiro exemplo de *riboswitch* sensível a variações no pH foi o do mRNA do gene *alx* de *E. coli* (Nechooshtan *et al.*, 2009) que codifica para uma proteína com provável função de transportador. Em condições alcalinas, além do aumento na expressão do gene *alx*, o *riboswitch* sofre alteração conformacional e a tradução da proteína Alx é aumentada. Já um exemplo de *riboswitch* que é modulado por variações na concentração de íons é a M-box, a qual responde a Mg⁺² e é encontrada em genes codificadores de proteínas transportadoras deste íon. Em baixa concentração do gene à jusante. Com a ligação do íon ao *riboswitch*, a M-box sofre uma mudança de conformação e sequestra a porção da alça antiterminadora, além de formar um terminador, resultando na inibição da expressão destes genes (Ramesh e Winkler, 2010).

Os riboswitches que interagem com ligantes podem ser classificados em três categorias de acordo com o tipo do ligante: proteína, RNA ou metabólitos (Winkler e Breaker, 2005; Mandin et al., 2013). Os elementos mais simples em sequência e estrutura são aqueles que se ligam a fatores proteicos enquanto que os riboswitches que ligam metabólitos apresentam características mais complexas de sequência e estrutura, sendo também mais diversos (Winkler e Breaker, 2005). Geralmente, os metabólitos se ligam a riboswitches de mRNAs que codificam proteínas de vias de sua própria biossíntese e/ou transporte, ou vias nas quais eles são utilizados. Dentre os riboswitches mais comuns desta classe estão os que ligam purinas, aminoácidos, c-di-GMP, vitaminas e/ou cofatores (adenosilcobalamina, tiamina pirofosfato, flavina mononucleotideo ou FMN e S-adenosilmetionina). Além disso, já foi descrita a presença de mais de um riboswitch no mesmo mRNA organizados in tandem e responsivos a diferentes metabólitos (Breaker, 2011; Mandin et al., 2013; Sherwood e Henkin, 2016).

Outra importante classe de riboswitch é daquelas que ligam tRNAs, as chamadas Tboxes (Gutierrez-Preciado et al., 2009). Estes elementos interagem com tRNAs e apresentam conformações alternativas dependendo se o tRNA está ligado, ou não, ao seu respectivo resíduo aminoacil. A interação entre a T-box e seu tRNA envolve uma sequência específica dentro da *T-box* e o anticódon do tRNA. Quando há uma interação com um tRNA carregado, ocorre a formação de uma hélice terminadora e consequente parada da elongação da transcrição ou oclusão do RBS; caso o tRNA não esteja carregado, ocorre a interação desta hélice com a extremidade 3' livre do tRNA, desestabilizando-a e fazendo com que a transcrição ou tradução possam continuar (Gutierrez-Preciado et al., 2009; Mandin et al., 2013). Mais de 90% destes elementos precedem sequências codificando para aminoacil-tRNA-sintetases e proteínas relacionadas ao metabolismo de aminoácidos (Gutierrez-Preciado et al., 2009). As Tboxes atuam principalmente na regulação da elongação da transcrição, mas também podem, em alguns casos, atuar na regulação da tradução do gene-alvo. T-boxes que regulam a terminação da transcrição são mais comuns em bactérias Gram-positivas com baixo conteúdo G+C, enquanto que T-boxes que controlam o início da tradução predominam em Gram-positivas com alto conteúdo G+C e em bactérias Gram-negativas (nas quais a regulação por *T-box* é menos comum) (Vitreschak *et al.*, 2008).

Introdução

Um *riboswitch* é composto de dois domínios distintos. O primeiro corresponde a um aptâmero evolutivamente conservado quanto a sua sequência e estrutura, o qual corresponde ao sítio de ligação específico para determinado metabólito. Seu comprimento pode variar de cerca de 70 a 200 nucleotídeos. O segundo é um domínio funcional conhecido por plataforma de expressão, localizado logo após o aptâmero. Suas sequências, estrutura e tamanho são mais diversas entre as diferentes classes de *riboswitches*. Em resumo, quando o ligante se liga ao aptâmero, mudanças conformacionais ocorrem na plataforma de expressão, exercendo, assim, a regulação da expressão dos genes presentes no mRNA (Winkler e Breaker, 2003; Breaker, 2011; Sherwood e Henkin, 2016).

Outra correlação evolutiva entre tipos de *riboswitches* está na observação que bactérias Gram-positivas tendem a apresentar *riboswitches* atuantes na atenuação da transcrição, enquanto que bactérias Gram-negativas teriam maior ocorrência de *riboswitches* que controlam o início da tradução. Entretanto, tal afirmação pode ser complicada pela ocorrência de eventos de transferência horizontal de genes e operons ou de duplicação no mesmo genoma. Uma vez que a regulação baseada em *riboswitch* não exige outros fatores para atuar, sua funcionalidade após transferência horizontal é completamente possível (Vitreschak *et al.*, 2004).

1.1.2.3. RNAs não codificadores

Genomas bacterianos são bastante compactos quanto ao seu tamanho e seu número de genes, apresentando alta densidade gênica e regiões intergênicas mais curtas e menos frequentes do que genomas eucarióticos. Mesmo assim, é crescente a descrição de novos RNAs não-codificadores (ncRNAs) bacterianos, que são normalmente transcritos a partir de regiões intergênicas Em *E. coli*, os ncRNAs são geralmente monocistrônicos, possuindo seu próprio promotor e terminador de transcrição intrínseco. Os ncRNAs atuam como reguladores da transcrição, da tradução ou da estabilidade de RNAs-alvo, além de ligarem-se a proteínas regulatórias (Storz *et al.*, 2011; Mandin *et al.*, 2013).

Os primeiros RNAs regulatórios descobertos em *E. coli* foram o 4.5S, 6S, tmRNA, RNaseP e Spot42, sendo todos relativamente muito abundantes (Mandin *et al.*, 2013). O primeiro relato de determinação da função de um ncRNA foi a descoberta de um RNA em *E. coli* complementar à região 5'UTR do mRNA de *ompF*, resultando na inibição de sua tradução (Mizuno *et al.*, 1984).

Os ncRNAs bacterianos são relativamente pequenos variando entre 50-500 nucleotídeos (Mandin *et al.*, 2013; Ternan, 2013). São divididos em duas classes: aqueles que atuam em mRNAs e aqueles que se ligam a proteínas. Os ncRNAs formam duplex com seus mRNAs-alvo a partir de um pareamento de bases imperfeito, sendo também importante a sua estrutura secundária. Além disso, atuam em mRNAs codificados em loci distantes no genoma, caracterizando uma atuação em *trans* (Mandin *et al.*, 2013). O número de ncRNAs varia com o tamanho do genoma, sendo que genomas maiores exigem maior complexidade na regulação gênica. O genoma de *E. coli* possui cerca de 4.000 genes e 50 a 100 ncRNAs preditos (Gottesman, 2004).

Alguns ncRNAs necessitam de uma proteína chaperona acessória para serem capazes de desempenhar suas funções. A mais estudada e amplamente conhecida é a proteína Hfq, que parece ser exigida para a ação da maioria dos ncRNAs em seus mRNAs-alvo, principalmente para estabilizar o pareamento de bases imperfeito entre eles (Vogel e Luisi, 2011). Hfq tem alta afinidade por RNA fita simples rico em adeninas e uracilas, e geralmente se liga em regiões no ncRNA que não possuem estrutura secundária. Além disso, foi demonstrado que a maioria dos ncRNAs perdem estabilidade em células mutantes para o gene codificando Hfq, sugerindo que ela os protege da degradação por RNases (De Lay *et al.*, 2013). Entretanto, nem todos os genomas bacterianos contém um gene homólogo de Hfq. Mesmo quando ele está presente, já foi visto sendo dispensável para a atuação de alguns ncRNAs. Uma hipótese possível é que Hfq não seria exigida quando a interação ncRNA e seus mRNAs-alvo fosse muito forte (Jousselin *et al.*, 2009).

Assim como os outros elementos regulatórios, os ncRNAs podem exercer um efeito negativo ou positivo sobre o seu alvo. A maioria deles pareia com o mRNA em uma região sobrepondo ou próximo ao RBS, sendo que o duplex impede o acesso do ribossomo e, portanto, inibindo a tradução (Wagner e Romby, 2015). O mRNA fica então desprotegido da ação de RNases, as quais devem ser recrutadas ativamente. Acredita-se que a proteína Hfq possa ter um papel nesse recrutamento (De Lay *et al.*, 2013; Mandin *et al.*, 2013).

Por outro lado, poucos exemplos de regulação positiva mediada por ncRNA têm sido descritos (Wagner e Romby, 2015). Estes RNAs também podem atuar na ativação

da tradução e/ou aumentando a estabilidade de seus mRNA alvos. A ação na ativação se dá pelo deslocamento de um elemento inibitório interno contido na região 5'UTR do mRNA-alvo. Muitas vezes, esta inibição é resultado da formação de uma estrutura secundária no mRNA, devido à complementaridade de bases entre duas regiões. Na ausência do ncRNA, a estrutura previne o acesso do ribossomo ao RBS, inibindo a tradução. Quando ocorre a ligação do ncRNA à esta estrutura, o ribossomo é então capaz de acessar as sequências de ligação ao DNA e dar início à tradução. Com a tradução em andamento, a presença de ribossomos protege o mRNA da ação de RNases, aumentando indiretamente a sua estabilidade (Mandin *et al.*, 2013). Em resumo, ncRNAs podem exercer efeitos globais na expressão gênica, muitas vezes equiparandose à capacidade regulatória de fatores de transcrição proteicos, sedimentando sua importância na fisiologia bacteriana (Ternan, 2013; Wagner e Romby, 2015).

Além de validações experimentais e predições *in silico* da presença de ncRNAs, diversos bancos de dados e ferramentas computacionais têm sido criados contendo informações que podem ser rapidamente acessadas e analisadas. Um dos mais antigos é o Rfam, o qual contém famílias de ncRNAs representadas por alinhamentos múltiplos e predições de estrutura secundária (Griffiths-Jones *et al.*, 2003; Burge *et al.*, 2013). Diversos outros bancos de dados de ncRNAs foram criados, tais como fRNAdb (Kin *et al.*, 2007), sRNAMap (Huang *et al.*, 2009), sRNAdb (Pischimarov *et al.*, 2012), entre outros. Li e colaboradores desenvolveram um banco de dado chamado BSRD, o qual afirma conter nove vezes mais sRNAs (*small RNAs*) validados experimentalmente do que os outros. Além disso, foi integrada ao banco uma plataforma de análise de dados de RNA-Seq, chamada sRNADeep, o que permite caracterizar sRNAs a partir de dados de transcritomas obtidos por sequenciamento em larga escala (Li *et al.*, 2013).

1.1.2.4. RNAs anti-senso

Além dos ncRNAs transcritos a partir de regiões intergênicas, diversos estudos têm observado a presença de eventos de transcrição de ncRNAs em regiões anti-senso a regiões codificadoras já descritas. Estes RNAs anti-senso (asRNA) exercem sua ação regulatória nos seus respectivos genes da fita oposta à região de onde foram transcritos, apresentando perfeita complementaridade aos seus alvos, gerando interações mais estáveis (Georg e Hess, 2011; Mandin *et al.*, 2013; Sesto *et al.*, 2013). Tal regulação é

considerada em *cis*, uma vez que o asRNA atua próximo de onde é sintetizado e, consequentemente, fornece uma resposta de regulação muito mais rápida. Alguns asRNAs respondem a mudanças ambientais, podendo sofrer alterações na sua expressão sob certas condições. Também há relatos sobre efeitos regulatórios mediados por asRNAs relevantes para a virulência e patogenicidade bacteriana (Georg e Hess, 2011; Mandin *et al.*, 2013).

Com o avanço das tecnologias de sequenciamento de RNA, um grande número de asRNAs tem sido descrito em uma ampla variedade de espécies microbianas. Os asRNAs podem ser classificados de acordo com a sua localização: sobrepondo à região 5', sobrepondo à região 3' ou localizados internamente na unidade transcricional senso. Seus tamanhos podem variar de pequenos, 100 a 300 nucleotídeos, a muito mais longos, de 700 a 3.500 nucleotídeos (Georg e Hess, 2011; Sesto *et al.*, 2013).

Os asRNAs podem exercer sua ação na estabilidade, tradução e transcrição de suas moléculas-alvo. A interação de um asRNA com seu mRNA alvo pode alterar a estrutura secundária de ambas as moléculas ao formar um RNA dupla-fita. Estas mudanças alteram a estabilidade e meia-vida dos RNAs e podem resultar na degradação completa de ambos por RNases específicas (Georg e Hess, 2011; Sesto *et al.*, 2013). Entretanto, em outros casos, asRNAs também são capazes de proteger seus mRNAs-alvos da degradação, a partir da sua interação (Sesto *et al.*, 2013). O pareamento do asRNA com seu mRNA alvo também pode ocluir os sítios de ligação ao ribossomo, inibindo a tradução (Georg e Hess, 2011).

Os asRNAs podem ainda interferir com a transcrição do RNA senso, em razão de um evento de colisão entre dois complexos de elongação da RNA polimerase que sejam convergentes e em fitas opostas, resultando na terminação prematura de um ou ambos os eventos de transcrição (Sesto *et al.*, 2013). Outros mecanismos possíveis para esta interferência seriam por oclusão do promotor, quando ocorre uma repressão exercida por um complexo de início de transcrição na formação de outro complexo a montante e em direção oposta ou pela dissociação de um complexo aberto pela colisão com outro complexo de elongação vindo na direção oposta, antes que a primeira polimerase possa proceder na elongação (Georg e Hess, 2011; Mandin *et al.*, 2013). Além da interferência na transcrição, asRNAs podem atuar na terminação da transcrição, sendo que os mecanismos de interferência e atenuação podem ocorrer simultaneamente (Sesto *et al.*, 2013).

2013). Além de regular genes codificadores de proteínas, asRNAs também podem sobrepor-se e afetar a expressão de pequenos ncRNAs (Sesto *et al.*, 2013).

Apesar de todos os trabalhos descreverem a ação em *cis* de asRNAs, baseado na sua perfeita complementaridade com seus alvos, um estudo realizado com *Staphylococcus aureus* descreveu um asRNA atuando em *trans* na inibição da tradução de seu mRNA-alvo (Sayed *et al.*, 2011).

1.1.2.5. Regulação da estabilidade do transcrito

A instabilidade metabólica é uma das principais características de moléculas de mRNA (Mackie, 2013). A degradação é um grande componente do metabolismo geral de RNAs e desempenha um importante papel em regular os seus níveis intracelulares (Deutscher, 2006). De modo geral, a meia-vida de um mRNA bacteriano é de poucos minutos, desde a sua transcrição até a degradação.

A estabilidade de cada mRNA é única, dependendo, principalmente, da função da proteína que codifica, sendo determinada pela combinação de características de sua sequência, de possíveis estruturas secundárias, da eficiência na tradução e de sua possível ligação a moléculas regulatórias (Mackie, 2013). Como já mencionado, a ligação de proteínas ou pequenos RNAs regulatórios em seus respectivos sítios pode estimular ou inibir a degradação do mRNA-alvo. A presença de ribossomos também interfere na estabilidade de mRNAs, uma vez que podem ocluir sítios de reconhecimento de RNases. Entretanto, sabe-se que o acúmulo de mRNAs pode não significar uma super-produção das proteínas codificadas, uma vez que, mesmo quimicamente estabilizados, os mRNAs podem estar funcionalmente inativados por algum repressor traducional, por um asRNA, ou por motivos estruturais que obstruem os sítios de início da tradução (Regnier e Arraiano, 2000).

A degradação de RNAs é catalisada por RNases, sendo que existem diversas RNases bacterianas com funções específicas e eficiência variável, permitindo uma degradação seletiva de RNAs (Mackie, 2013). Geralmente, os mRNAs são degradados, inicialmente, por clivagem endonucleolítica, seguida de clivagens adicionais do mRNA em fragmentos menores, culminando com a remoção de nucleotídeos terminais por exonucleases $3' \rightarrow 5'$ (Regnier e Arraiano, 2000; Deutscher, 2006; Condon, 2007). Dentre as diversas RNases presentes em *E. coli*, a endonuclease RNase E é a principal responsável pela clivagem da maioria dos RNAs da célula, atuando através de um complexo multienzimático chamado degradossomo (De Lay *et al.*, 2013; Mackie, 2013). A RNase E tem como preferência moléculas de RNA-alvo que apresentem uma extremidade 5' monofosforilada acessível, além de exigir, no mínimo, quatro nucleotídeos em fita-simples na extremidade 5' para se ligar eficientemente (Condon, 2007; Mackie, 2013). Esta preferência impede que esta enzima degrade RNAs intactos, os quais apresentam um trifosfato na extremidade 5'. Sua ação também cliva ligações fosfodiéster em regiões ricas em adeninas e uracilas, gerando extremidades 3'-OH e 5'-monofosfato nas moléculas-alvo (Misra e Apirion, 1979; Deutscher, 2006). Entretanto, outra via menos comum, independente da extremidade 5' monofosforilada, também pode ser usada na degradação de mRNAs e no processamento de tRNAs, e neste caso, outra ribonuclease realiza o corte no interior do mRNA, permitindo em seguida a ação da RNase E (Mackie, 2013).

A concentração e atividade de RNase E está sujeita a uma complexa regulação. A autorregularão combina a disponibilidade de seu próprio mRNA à concentração de seus substratos na célula (Sousa et al., 2001). Quando a atividade de RNase E excede a demanda para processamento e degradação de RNA, a síntese de seu mRNA é interrompida e ele se torna o próprio alvo preferencial para a degradação, regulando a concentração de RNase E. Além da autorregulação, a modulação da atividade de RNase E pode ser realizada por proteínas regulatórias, as quais respondem a mudanças nas condições de crescimento de forma a modificar sua atividade, tanto globalmente como contra determinado grupo de mRNAs. Assim que a condição normal é reestabelecida, a autorregulação pode voltar a operar (Mackie, 2013). Além de reguladores proteicos, a atividade de RNase E também pode ser regulada através da ação de ncRNAs (Morita et al., 2005; Storz et al., 2011; Vogel e Luisi, 2011). A função inicial dos ncRNAs é a inibição do início da tradução, porém a degradação do alvo é consequência da inatividade do transcrito (Condon, 2007). A chaperona de RNA Hfq atua no processo de regulação de duas maneiras distintas: ela pode se ligar à RNase E, em seguida a um ncRNA e direcionar o complexo para o mRNA-alvo; ou ela primeiramente facilita a interação e formação do duplex entre o ncRNA e mRNA-alvo, sendo então posteriormente reconhecido pela RNase E (De Lay et al., 2013; Mackie, 2013).

Outras RNases também desempenham funções específicas com diferentes mecanismos de ação, embora com atuações menores, como as endonucleases RNase G,

RNase III, RNase P, RNase I e oligoribonucleases (Condon e Putzer, 2002; Condon, 2007). A RNase G apresenta a porção catalítica homóloga a da RNase E, apresentando substratos e propriedades enzimáticas muito similares a dela (Condon e Putzer, 2002). A RNase III cliva RNAs que contenham estruturas secundárias e também autorregula sua expressão pela indução da degradação de seus próprios mRNAs, estando ausente em arqueias (Regnier e Arraiano, 2000; Condon e Putzer, 2002). A RNase P é uma ribozima responsável por gerar uma extremidade 5' madura em todos os tRNAs e de catalisar clivagens de alguns poucos mRNAs (Condon e Putzer, 2002; Esakova e Krasilnikov, 2010). A RNase I é uma nuclease que degrada RNAs livres no ambiente, sendo encontrada exclusivamente no espaço periplasmático de alfa- e gama-proteobactérias (Condon e Putzer, 2002). E as oligorribonucleases são responsáveis por converterem produtos de oligoribonucleotídeos da última fase da degradação de mRNA em monoribonucleotídeos (Mackie, 2013).

A extremidade 3' de mRNAs bacterianos é protegida do ataque de exonucleases devido a estruturas secundárias, sendo na maioria dos casos um terminador de transcrição independente de Rho. Após a clivagem pelas endonucleases, novas extremidades 3' desprotegidas são geradas e podem então ser degradadas por exorribonucleases. As principais exorribonucleases atuando em mRNAs de E. coli são RNase II, PNPase e RNase R (Condon, 2007). A RNase II degrada apenas RNAs fitasimples, sendo completamente inibida por estruturas secundárias. As outras duas degradam RNAs com estruturas secundárias, embora seja dependente da presenca de uma cauda poli(A) na extremidade 3', a qual foi sintetizada pela poli(A) polimerase (PAP1) com a ajuda da Hfq (Condon, 2007). Juntas, as atividades hidrolítica da RNase II e fosforolítica da PNPase geram oligorribonucleotídeos di e monofosfatados, respectivamente (Regnier e Arraiano, 2000). PNPases são mais um exemplo de autorregulação da expressão de seus próprios mRNAs, induzindo sua própria degradação em casos de excesso (Regnier e Arraiano, 2000). Os oligoribonucleotídeos gerados durante a degradação de RNAs são degradados por oligoribonucleases (Ghosh e Deutscher, 1999). Nenhuma exonuclease com habilidade de degradar extremidades 5' de mRNAs foi identificada em bactérias. Na ausência de tal enzima, sugere-se que a degradação no sentido 5' \rightarrow 3' seria devido a várias clivagens endonucleolíticas 5' \rightarrow 3', as quais geram extremidades 3' livres que são então degradadas por exonucleases $3' \rightarrow 5'$ (Regnier e Arraiano, 2000).

As RNases também são capazes de realizar a maturação ou processamento de precursores de rRNAs e tRNAs (Mackie, 2013). Em células em crescimento, a estabilidade dos rRNAs parece ser uma consequência de sua incorporação em ribossomos e proteção por proteínas ribossomais (Deutscher, 2003). No caso de tRNAs, a presença de estruturas secundárias e terciárias lhes confere resistência à ação de RNases, além da associação com aminoacil-tRNA sintetases, fatores de elongação e ribossomos (Deutscher, 2003). Tais RNAs estáveis não são geralmente degradados durante o crescimento exponencial, embora em certas condições (estresses, tratamento com agentes tóxicos), a degradação dessas moléculas ocorrerá. Em condições de escassez de nutrientes ou em fase estacionária, estes RNAs podem ser degradados para serem utilizados como nutrientes para a célula (Deutscher, 2003). A degradação de RNAs estáveis também envolve clivagens endonucleolíticas iniciais seguida por digestão exonucleolítica dos fragmentos resultantes, similar ao que acontece com os mRNAs, sendo que as mesmas RNases participam em ambos os processos (Regnier e Arraiano, 2000; Deutscher, 2006).

1.1.3. Mecanismos de sinalização celular em bactérias

A capacidade de bactérias em sobreviver e se adaptar a um ambiente em constante mudança e sob os mais variados tipos de estresses lhes confere grande vantagem evolutiva, sendo que desenvolveram complexos sistemas de sinalização celular para gerar respostas adaptativas ao seu ambiente (Parkinson, 1993). Para tal, as bactérias são capazes de perceber a condição atual do ambiente, transmitir a informação para o interior da célula e desencadear uma resposta, geralmente refletida em mudanças na expressão gênica.

O mecanismo inicia-se com a percepção do estímulo ambiental, seguida do processamento do sinal. O estímulo pode ser um fenômeno físico-químico, como temperatura e osmolaridade, ou a disponibilidade de diferentes moléculas no ambiente, como nutrientes ou agentes tóxicos, além da percepção da densidade populacional (Waters e Bassler, 2005; Beier e Gross, 2006). A percepção do estímulo pode ser sentida diretamente no citoplasma, ou por ser um sinal físico-químico ou por conseguir atravessar membranas celulares. Outros são percebidos indiretamente por meio de

proteínas localizadas na membrana que recebem o sinal e transferem a mensagem para o interior da célula. Muitas vezes, nem sempre é o estímulo em si que desencadeia a maquinaria de sinalização, mas o efeito que ele causa (Parkinson, 1993; Stock *et al.*, 2000).

Um dos sistemas de transdução de sinal mais encontrado e utilizado em bactérias é o sistema de dois componentes (Parkinson, 1993; Hoch, 2000; Beier e Gross, 2006). Nesse sistema, a proteína sensora apresenta um domínio funcional com atividade de proteína-quinase que, ao perceber o estímulo/sinal, se autofosforila em um resíduo de histidina, pela transferência do fosfato de uma molécula de ATP (Hoch, 2000; Stock et al., 2000). A transmissão da informação continua com a transferência do grupo fosfato da proteína sensora para um resíduo de aspartato no regulador de resposta. A fosforilação do regulador de resposta altera sua atividade, um estado que é mantido até sua desfosforilação, que pode ser autocatalisada ou por outras proteínas com atividade de fosfatase (Parkinson, 1993; Hoch, 2000; Stock et al., 2000; Beier e Gross, 2006). Em alguns sistemas de dois componentes, a transferência do grupo fosfato da proteína sensora para o regulador de resposta pode envolver proteínas intermediárias. Além disso, a maioria dos reguladores de resposta são fatores de transcrição com domínios de ligação ao DNA, desencadeando mudanças na expressão gênica de seus genes-alvo (Parkinson, 1993; Stock et al., 2000), ou são proteínas com atividade enzimática regulada pela fosforilação (Galperin, 2004; Beier e Gross, 2006).

Sistemas de dois componentes podem sofrer sobreposição, uma vez que mais de uma histidina-quinase pode interagir com o mesmo regulador de resposta e vice-versa, gerando competição e/ou diálogo (*cross-talk*) entre eles. Porém, a afinidade da quinase pelo seu regulador cognato é sempre maior, sendo geralmente codificados em um mesmo operon. Há correlação positiva entre o número de sistemas de dois componentes e o tamanho do genoma. Da mesma forma, bactérias de vida livre também codificam mais sistemas de dois componentes do que bactérias que habitam ambientes mais controlados (por exemplo, patógenos que habitam o interior de seus hospedeiros), uma vez que necessitam de uma maior versatilidade metabólica (Beier e Gross, 2006).

O sistema de transdução também pode variar na forma como transmite a mensagem. Alguns sistemas transmitem a informação utilizando pequenas moléculas intermediárias, chamadas segundo-mensageiros, que somente gerarão a resposta quando atingirem determinada concentração. Este tipo de transmissão de informação pode

atingir vários sistemas simultaneamente, uma vez que a molécula sinalizadora pode ser identificada por várias proteínas. Dentre os principais segundo-mensageiros bacterianos estão três derivados de nucleotídeos, cAMP ou cGMP, ppGpp e c-diGMP (Gorke e Stulke, 2008; Hauryliuk *et al.*, 2015; Hengge *et al.*, 2016).

O mecanismo de regulação por repressão catabólica, no qual há a participação de cAMP, merece ser detalhado. O termo foi pela primeira vez cunhado no início da década de 1960 (Magasanik, 1961). Trata-se de um sistema no qual a célula opta por utilizar primeiramente uma fonte de carbono mais facilmente metabolizada do que outra, sendo a glicose o monossacarídeo preferencial na maioria dos casos. Um exemplo de genes sujeitos a tal regulação é o operon *lac* (Jacob e Monod, 1961). Enquanto houver glicose no meio, ela será primariamente utilizada e os genes para a utilização de lactose não serão plenamente transcritos, mesmo na presença de lactose. Quando toda a glicose for consumida, o operon passa a ser plenamente transcrito e as enzimas para utilização da lactose serão produzidas. A papel do cAMP é o de se ligar ao ativador da transcrição do operon lac. A repressão catabólica também age da mesma forma sobre outros operons que utilizam outras fontes de carbono, como aqueles relacionados ao uso de galactose, arabinose e maltose, todos também regulados por cAMP (Gorke e Stulke, 2008). Além disso, a repressão catabólica tem papel importante na expressão de genes de virulência, uma vez que tais genes podem proporcionar a bactéria o acesso a novas fontes de nutrientes. Entretanto, algumas bactérias patogênicas não possuem tal mecanismo, enquanto outras realizam a chamada repressão catabólica reversa, na qual a glicose não é a fonte de carbono preferencial (Gorke e Stulke, 2008). Um exemplo é a utilização preferencial de acetato e outros intermediários do ciclo do ácido tricarboxílico como fonte de carbono e energia em *Pseudomonas aeruginosa* (Collier *et al.*, 1996).

Mecanismos de sinalização e sistemas de dois componentes também são usados no controle da densidade da população bacteriana. O processo chamado de sinalização por percepção de quórum (*quorum sensing*) é baseado na percepção da concentração de pequenas moléculas secretadas para o meio extracelular por células da população (Waters e Bassler, 2005), sendo a concentração destas moléculas diretamente proporcional ao aumento do número de células. Os genes que são alvo de mecanismos de percepção de quórum são distintos nas diferentes espécies bacterianas, podendo envolver genes relacionados à virulência (toxinas, mobilidade, produção de antibióticos, biofilme, exopolissacarídeos, etc.) e a outros aspectos da fisiologia da célula
(morfologia de colônias, produção de pigmentos, competência, esporulação, etc) (Hoch, 2000).

1.2. Métodos de análise de transcritomas

Diversos métodos foram desenvolvidos para análise parcial ou completa de transcritomas, os quais apresentam características próprias quanto ao grau de dificuldade para realização, acurácia, e rendimento (*throughput*) quanto a quantidade de genes que podem ser analisados simultaneamente. Enquanto alguns métodos possibilitam a análise de um número reduzido de genes por vez, outros podem descrever o perfil transcricional global de um organismo, isto é, descrever o repertório completo de moléculas de RNA presentes na célula em determinada fase de crescimento ou em uma condição ambiental particular. As análises de transcritomas completos também são utilizadas para definir os genes que fazem parte de *regulons* e *stimulons*.

Até o advento da técnica de PCR, o estudo de um número limitado de genes de sequências previamente conhecidas era realizado com métodos baseados em hibridização de ácidos nucleicos, como Northern Blot e ensaios de proteção contra RNases (VanGuilder et al., 2008), que requerem quantidades significativas de RNA total (>10 µg). O surgimento da PCR, tornou possível a quantificação da abundância de transcritos de interesse previamente convertidos em cDNA com o método denominado RT-PCR (PCR precedido por transcrição reversa), que embora tenha boa sensibilidade, permitindo a detecção de mRNAs pouco abundantes, e utilize menores quantidades de RNA de total (~1µg), é um método semi-quantitativo (Sellner e Turbett, 1998). O aprimoramento desta metodologia veio com o surgimento do qPCR (PCR quantitativo, em tempo real), que uma vez precedido de transcrição reversa (RT-qPCR) possibilita a quantificação relativa ou absoluta de transcritos, com alta sensibilidade e especificidade (Schmittgen et al., 2000; VanGuilder et al., 2008). Frequentemente, o método de RTqPCR é utilizado na validação de análises de expressão gênica globais, para confirmar variações na expressão de genes mais importantes ou mais interessantes (Canales et al., 2006).

A primeira abordagem de maior rendimento para análise de transcritomas foi o sequenciamento das extremidades 3' e 5' de cDNAs, gerando-se sequências de 100 a

800 pb denominadas de ESTs (Expressed Sequence Tags). Este recurso foi amplamente utilizado na análise de transcritomas eucarióticos pelo sequenciamento de clones de bibliotecas de cDNA geradas pela transcrição reversa de mRNAs poliadenilados, ainda que limitado pela ausências de sequências de porções internas dos transcritos (Adams et al., 1991; Nagaraj et al., 2006). Esta limitação foi contornada por um protocolo alternativo chamado ORESTES (Open Reading Frame Expressed Sequence Tags), o qual acopla uma etapa de PCR às preparações de cDNA, com o objetivo de enriquecer sequências de regiões codificadoras dos cDNAs (Dias Neto et al., 2000). em Estratégias de normalização, envolvendo hibridização subtrativa de bibliotecas de cDNA, são empregadas para facilitar a detecção de transcritos raros (Bonaldo et al., 1996). O perfil de expressão gênica de células ou tecidos pode ser comparado a partir da análise abundância relativa de ESTs, se as bibliotecas de cDNA não forem normalizadas (Gruber, 2007), ainda que tais análises sejam limitadas quanto a sua precisão quando comparadas a outras metodologias para análise de transcritomas como os microarranjos de DNA e o RNA-Seq, que serão apresentados a seguir.

1.2.1 Microarranjos de DNA

Os microarranjos de DNA permitem a quantificação da abundância relativa de dezenas de milhares de RNAs transcritos em diferentes células ou tecidos submetidos ou não a condições ou tratamentos de interesse. Como o nome sugere, os microarranjos contêm fragmentos de DNA de sequências conhecidas (sondas), imobilizados em uma superfície sólida (lâminas de vidro ou quartzo) em um arranjo de milhares de pontos ordenados espacialmente (Schena *et al.*, 1995; Dharmadi e Gonzalez, 2004; Slonim e Yanai, 2009). Assim como a técnica de *Northern blot*, a técnica dos microarranjos baseia-se na hibridização de moléculas de ácidos nucleicos complementares (Dharmadi e Gonzalez, 2004). Assim, os RNAs são purificados de células e utilizados como moldes para síntese de cDNAs fluorescentes (marcados com fluoróforos acoplados aos dNTPs) que são, então, utilizados para hibridização do microarranjo. Por fim, um escaneamento é realizado e o sinal de intensidade da luz emitida pelo fluoróforo em cada ponto do microarranjo é registrado, e valor obtido é utilizado na determinação da

abundância de cada alvo (transcrito) na amostra de RNA (Dharmadi e Gonzalez, 2004; Clarke e Zhu, 2006; Malone e Oliver, 2011).

O conjunto de sondas é desenvolvido com base na sequência genômica ou de transcritos (cDNAs) do organismo em estudo. Inicialmente as sondas eram geradas por PCR gerando fragmentos de 200 a 500 pares de bases que eram depositados em lâminas de vidro (Schena *et al.*, 1995). Atualmente, as sondas são oligonucleotídeos sintéticos de 50-80 bases (Dharmadi e Gonzalez, 2004) constituindo os chamados *oligoarrays*, que podem representar sequências genômicas completas como nos *tilling arrays* possibilitando a verificação do nível de expressão de qualquer região do genoma, independentemente de conter ou não um gene predito (Filiatrault, 2011).

As amostras que terão seus níveis de expressão comparados podem ser hibridizadas separadamente, cada uma em um microarranjo (*single-color* ou *single-channel*), ou juntas em um mesmo (*two-color* ou *two-channel*), em função da marcação do cDNA com um ou dois fluoróforos distintos (Moreau *et al.*, 2003; Slonim e Yanai, 2009). No primeiro caso, a intensidade do sinal de hibridização é usada para determinar a concentração das moléculas-alvo (quantificação absoluta) (Moreau *et al.*, 2003). No segundo caso, irá ocorrer uma hibridização competitiva das duas amostras marcadas, eliminando a variabilidade técnica que aconteceria em duas hibridizações distintas e permitindo uma comparação direta entre as amostras em uma única hibridização (Moreau *et al.*, 2003; Altman e Hua, 2006).

Diversos fatores podem influenciar na reprodutibilidade de experimentos de microarranjos, entre eles o desenho experimental inadequado, diferenças no tipo de microarranjo utilizado e variações na sensibilidade dos *scanners* a laser (Irizarry *et al.*, 2005). Além disso, a técnica de microarranjos apresenta outras limitações, incluindo baixa acurácia das medições de expressão, especialmente para transcritos pouco abundantes, em virtude da fluorescência de fundo (*background*) que pode ocorrer, dependendo das condições de hibridização empregadas. Recomenda-se que sejam observadas as especificações do MIAME (*Minimum Information about a Microarray Experiment*), que fornece uma guia para anotação de microarranjos, incluindo o desenho experimental, desenho do microarranjo, detalhes das amostras e tratamentos, condições de hibridização, medidas e controles de normalização, de modo a garantir que os dados obtidos possam ser reproduzidos e/ou comparados (Brazma *et al.*, 2001; Moreau *et al.*, 2003).

Introdução

A comparação de intensidades de fluorescência de um dado ponto do microarranjo resultante da hibridização com duas amostras distintas resulta na obtenção da "razão de expressão" (ou *fold change*), o que reflete diretamente a abundância relativa do respectivo transcrito nessas amostras. A determinação final das intensidades do sinal de fluorescência leva em conta o sinal específico produzido pelo transcrito-alvo marcado, o sinal de hibridização-cruzada e o sinal de fundo não-específico (Zhao *et al.*, 2014). Os dois últimos sinais interferentes dificultam a detecção de genes pouco expressos. Por outro lado, transcritos muito abundantes tendem a saturar o sinal, complicando a detecção de variações nos níveis destes mRNAs entre amostras distintas (Dharmadi e Gonzalez, 2004).

A padronização das análises estatísticas é de extrema importância para comparar dados de microarranjos. O uso de testes t e ANOVAs, cortes de *p*-valor, correções Bonferroni, normalizações e cortes de *fold change*, devem ser escolhidos cuidadosamente. Mudanças nos parâmetros de níveis de significância dos genes diferencialmente expressos e do corte de *fold change* podem levar a diferentes interpretações de dados de microarranjos (Dalman *et al.*, 2012). Normalizações dos dados brutos são utilizadas para controlar ou eliminar a variação técnica dentro de um estudo (Slonim e Yanai, 2009). Além disso, um experimento de microarranjo deve incluir pelo menos três réplicas biológicas e técnicas, de modo a estimar, com adequado poder estatístico, a significância das mudanças observadas (Lee *et al.*, 2000; Clarke e Zhu, 2006) bem como avaliar a probabilidade de variações ao acaso ou devidas a artefatos da técnica (Dharmadi e Gonzalez, 2004).

Embora se recomende que resultados de experimentos de microarranjos devam ser validados por outros métodos distintos, como *Northern blot* ou RT-qPCR (Dharmadi e Gonzalez, 2004; Clarke e Zhu, 2006), deve-se atentar que microarranjos podem subestimar em até uma ordem de grandeza os valores de razão de expressão determinados por RT-qPCR (Conway e Schoolnik, 2003; Moreau *et al.*, 2003). Além disso, em comparações entre experimentos de microarranjos e de proteômica, a abundância de transcritos nem sempre reflete os níveis da proteína correspondente, especialmente no caso de eventos pós-transcricionais. Resultados de análises proteômicas com géis bi-dimensionais podem subestimar o número total de proteínas/genes induzidos em 2 a 4 vezes comparativamente aos microarranjos.

Entretanto, na maioria dos casos, os níveis de transcrição e de proteínas correlacionamse (Conway e Schoolnik, 2003).

A técnica microarranjos de DNA foi muito utilizada em estudos de transcritômica em bactérias. Como exemplos, citamos os estudos de Oshima e colaboradores (Oshima *et al.*, 2002) e de Han e colaboradores (Han *et al.*, 2004) que empregaram microarrays construídos com *amplicons* de todas as ORFs anotadas nos genomas. No primeiro caso, os microarranjos de DNA foram empregados para avaliar o perfil transcricional de 36 mutantes de sistemas de dois-componentes em *E. coli* sob uma única condição de cultivo, possibilitando identificação de redes de interação e regulação cruzada entre diferentes sistemas de dois componentes. O segundo estudo, analisou as mudanças de expressão do patógeno *Yersinia pestis* em resposta a variação de 26°C para 37°C, que são temperaturas que a bactéria enfrenta no seu ciclo de vida em seus hospedeiros naturais, evidenciando um aumento da expressão de genes relacionados a fatores de virulência, reguladores transcricionais, prófagos, entre outros. Mais recentemente, estudos de transcritomas utilizando *tiling arrays* e, principalmente, RNA-Seq tem contribuído substancialmente para compreensão da complexidade, plasticidade e da regulação da expressão gênica em procariotos (Sorek e Cossart, 2010).

1.2.2. RNA-Seq

A tecnologia denominada RNA-Seq (Sequenciamento de RNA) consiste no sequenciamento de todos os transcritos de uma célula utilizando abordagens de sequenciamento de DNA de alto-desempenho (Nagalakshmi *et al.*, 2010; Malone e Oliver, 2011; Wolf, 2013; Zhao *et al.*, 2014), coletivamente denominadas plataformas de sequenciamento de última geração (NGS, *Next-Generation Sequencing*) (Mardis, 2008).

O RNA-Seq substitui plenamente a técnica de microarranjos na descrição de transcritomas completos, apresentando vantagens como menor custo e menor complexidade experimental (Mardis, 2008; Mortazavi *et al.*, 2008; Morozova *et al.*, 2009). Além disso, RNA-Seq não depende do conhecimento prévio da sequência do genoma, o que é relevante na confecção dos microarranjos (Malone e Oliver, 2011; Mutz *et al.*, 2013; Zhao *et al.*, 2014). Além disso, alguns problemas e vieses

encontrados na técnica de microarranjos são evitados, como diferenças na eficiência das sondas e hibridização-cruzada e/ou não-específica (Balleza *et al.*, 2009; Nagalakshmi *et al.*, 2010; Wolf, 2013; Zhao *et al.*, 2014). Entretanto, assim como qualquer outra técnica, a análise de transcritomas por RNA-Seq também é vulnerável a erros e vieses derivados de uma eventual baixa cobertura do sequenciamento ou da ambiguidade no processo de mapeamento das sequências ao genoma de referência.

A cobertura do sequenciamento, ou seja, quantas sequências devem ser obtidas por amostra para efetivamente contemplar todo o transcritoma, incluindo genes com baixa expressão, é uma das questões importantes (Wang *et al.*, 2009; Malone e Oliver, 2011; Wolf, 2013). Em geral, quanto maior o genoma, mais complexo é seu transcritoma e uma maior profundidade de sequenciamento é necessária para uma cobertura adequada (Wang *et al.*, 2009). Além disso, é possível observar uma heterogeneidade das sequências através de uma região expressa, ou seja, a profundidade de sequenciamento ao longo do comprimento do transcrito pode variar. Cobertura e heterogeneidade não são problemas em microarranjos, uma vez que as sondas usadas são conhecidas (Malone e Oliver, 2011). Outro problema que pode ser encontrado na análise de dados de RNA-Seq é a ambiguidade de mapeamento das sequências em múltiplos locais do genoma, sendo principalmente devido à presença de genes parálogos e regiões repetitivas (Wang *et al.*, 2009; Wolf, 2013). Sequenciamento com a estratégia de *paired-end* ajuda a mapear cada sequência no seu correto local no genoma (Wang *et al.*, 2009).

Alguns estudos realizaram a comparação entre os resultados obtidos com o microarranjos de DNA e RNA-Seq. Zhao e colaboradores (Zhao *et al.*, 2014) encontraram alta sobreposição entre os resultados de ambas as tecnologias em experimentos com células T humanas, embora também tenham observado genes únicos em cada uma delas. Um número maior de genes foi encontrado utilizando a estratégia de RNA-Seq, sendo mais sensível em detectar genes com expressão muito baixa. Além disso, foi observada uma maior acurácia na medida da expressão de genes muito abundantes, enquanto que em microarranjos pode ocorrer saturação de hibridização nestes casos, impossibilitando uma quantificação confiável (Marioni *et al.*, 2008; Zhao *et al.*, 2014). O mesmo foi encontrado em um estudo com *Drosophila pseudoobscura*, no qual foi encontrada alta congruência nos resultados de expressão de microarranjos e de RNA-Seq, com o mesmo viés em transcritos com baixa expressão (Malone e Oliver,

2011). Nookaew e colaboradores realizaram tal comparação entre tecnologias com a levedura *Saccharomyces cerevisiae* e também encontraram boa concordância em relação à quantificação relativa da expressão gênica (Nookaew *et al.*, 2012).

Além da aplicação principal que é quantificação da abundância relativa de transcritos, novos genes podem ser descobertos com a tecnologia de RNA-Seq, incluindo pequenos ncRNAs (Morozova *et al.*, 2009; Sorek e Cossart, 2010; Malone e Oliver, 2011; Nookaew *et al.*, 2012; Mutz *et al.*, 2013; Wolf, 2013). Melhorar e corrigir a anotação de genomas e identificar as sequências das regiões não-traduzidas (UTRs) nas extremidades 5' e 3' dos transcritos também são aplicações importantes desta tecnologia (Wang *et al.*, 2009; Nagalakshmi *et al.*, 2010; Sorek e Cossart, 2010; Nookaew *et al.*, 2012; Mutz *et al.*, 2013). Em organismos eucarióticos, dados gerados por RNA-Seq podem ser utilizados para definir a expressão alelo-específica, mapear fronteiras éxon/íntron e determinar variantes de *splicing* (Wang *et al.*, 2009; Nookaew *et al.*, 2012; Mutz *et al.*, 2014).

Outra aplicação do RNA-Seq é a capacidade de definir os sítios de início de transcrição (*Transcription Start Sites*, TSS) (Wang *et al.*, 2009; Filiatrault, 2011; Ozsolak e Milos, 2011). O mapeamento das TSSs aprimora a definição dos transcritos e permite a caracterização de regiões promotoras que regulam a expressão de cada um deles (Sorek e Cossart, 2010; Ozsolak e Milos, 2011). Um dos primeiros métodos de mapeamento de TSSs em larga escala foi o *Cap Analysis of Gene Expression* (CAGe) (Shiraki *et al.*, 2003). Porém, esta tecnologia exige elevada quantidade de RNA e gera apenas sequências curtas para cada TSS (Ozsolak e Milos, 2011).

Dados de RNA-Seq também contribuem para compreender a estrutura de operons bacterianos e a regulação da expressão dos genes constituintes (Sorek e Cossart, 2010; Filiatrault, 2011). A predição derivada da anotação de genomas é geralmente baseada na ocorrência de uma curta distância entre genes consecutivos, conservação da ordem dos genes em organismos relacionados e na identificação de associação funcional entre os genes da vizinhança. A detecção da expressão de transcritos policistrônicos, que incluem ou não sequências de determinados genes do operon, são informações valiosas para o entendimento da regulação e da plasticidade de transcritomas procarióticos (Sorek e Cossart, 2010).

Qualquer plataforma de sequenciamento de alto-desempenho pode ser utilizada na em experimentos de RNA-Seq, ainda que o método de preparo de bibliotecas e o

Introdução

processo de sequenciamento variem entre elas (Mardis, 2008). A preparação de bibliotecas para sequenciamento inicia-se com a extração da amostra de RNA total sem qualquer contaminante e mantendo-se a integridade dessas moléculas (Nagalakshmi *et al.*, 2010; Malone e Oliver, 2011; Wolf, 2013). Uma vez a fração de mRNAs pode corresponder a apenas 1–5% do RNA total em células procarióticas, é fundamental que seja realizada o enriquecimento desta fração minoritária de RNAs, através da depleção dos abundantes rRNAs (Sorek e Cossart, 2010; Malone e Oliver, 2011; Wolf, 2013). O enriquecimento pode ser feito por captura dos rRNAs ou hibridização subtrativa com utilização de sondas que correspondem a posições conservadas nos rRNAs 16S e 23S, degradação seletiva de RNAs monofofosrilados na extremidade 5' por exonucleases, poliadenilação seletiva de mRNAs e captura de RNAs que estão associados com determinada proteína (ex: chaperona de RNA Hfq) com emprego de anticorpo específico (Sorek e Cossart, 2010; Filiatrault, 2011).

Após a etapa de enriquecimento, o RNA é fragmentado, convertido em cDNA dupla-fita por transcrição reversa e adaptadores são incorporados, por PCR, às extremidades dos cDNAs (Malone e Oliver, 2011; Wolf, 2013). Após quantificação e normalização das bibliotecas de fragmentos, procede-se o sequenciamento, que na maioria dos estudos atuais, é realizado na plataforma Illumina, a partir de uma única extremidade (estratégia *single-end*) ou a partir de ambas (estratégia *paired-end*) (Fullwood *et al.*, 2009; Malone e Oliver, 2011; Ozsolak e Milos, 2011; Wolf, 2013).

O sequenciamento por *paired-end tag* (PET) é uma das estratégias para melhorar a eficiência do sequenciamento de DNA e auxiliar na montagem. Neste tipo de protocolo, sequências curtas e pareadas das duas extremidades de fragmentos de DNA são extraídas e covalentemente ligadas formando construções *ditag* para efeito do sequenciamento de alto-desempenho e mapeamento no genoma de referência. A distância entre as duas extremidades pode ser usada para relacionar *contigs* em montagens de genomas, incluindo regiões contendo repetições (Fullwood *et al.*, 2009).

Após o sequenciamento, os dados brutos de sinais de fluorescência, registrados em diversas imagens, devem ser convertidos em sequências de bases em um processo realizado pelas próprias plataformas de sequenciamento utilizando algoritmos de identificação de base (*base calling*) do fabricante. Um escore de qualidade para cada base é calculado, indicando a confiabilidade de cada base identificada. Todos os dados de saída das plataformas são no formato padrão FASTQ, embora o cálculo do escore de

qualidade difira entre elas (Mutz *et al.*, 2013). Tal processamento dos dados é importante para reduzir os erros nas análises de imagem e identificação de base, podendo remover sequências de baixa-qualidade (Wang *et al.*, 2009).

Após a obtenção das sequências (*reads*), seguem-se análises bioinformáticas e estatísticas, que incluem verificação da qualidade das sequências com o software FastQC, (Wolf, 2013) e eventual remoção de bases com escores de qualidade abaixo do esperado (*quality trimming*). Também é fundamental avaliar a reprodutibilidade das réplicas biológicas e técnicas dos transcritomas obtidos através de análises de correlação. Experimentos de RNA-Seq têm se mostrado altamente reprodutíveis, evidenciando ser uma técnica robusta e confiável (Wang *et al.*, 2009; Nagalakshmi *et al.*, 2010).

As sequências obtidas com os dados de RNA-Seq podem ser tanto mapeadas contra um genoma ou transcritos de referência ou ser utilizadas em uma montagem de novo para reconstrução dos transcritos (Wang et al., 2009; Sorek e Cossart, 2010; Grabherr et al., 2011; Malone e Oliver, 2011; Nookaew et al., 2012; Mutz et al., 2013; Wolf, 2013). Entretanto, problemas na montagem podem existir dificultando a reconstrução de todos os transcritos a partir de sequências curtas, devido a erros de sequenciamento, diferenças na cobertura de cada transcrito ao longo de todo o seu comprimento (Grabherr et al., 2011). Em ambas as estratégias de análise, a contagem de sequências mapeadas em um dado transcrito fornece uma medida da sua abundância ou nível de expressão gênica (Mortazavi et al., 2008; Wolf, 2013). Nookaew e colaboradores compararam ambas as estratégias entre si e com a tecnologia de microarranjos. Os resultados das análises estatísticas para obtenção de genes diferencialmente expressos mostraram boa concordância em todas as abordagens utilizadas nas comparações. A mesma concordância foi observada quando os genes diferencialmente expressos foram funcionalmente categorizados, levando a conclusões biológicas similares (Nookaew et al., 2012).

Devido à diferença no comprimento dos transcritos, em uma amostra de biblioteca os transcritos mais longos terão mais fragmentos representando-os nos transcritomas. Dessa forma, foi proposta a medida de densidade da sequência por RPKM (*Reads per kilobase per million mapped reads*) ou FPKM (*Fragments per kilobase of transcript per million fragments mapped*), as quais refletem a concentração molar de um transcrito na amostra inicial a partir da normalização pelo seu comprimento e pelo número total de

sequências no transcritoma (Mortazavi *et al.*, 2008). Esta normalização é utilizada para comparar a expressão de genes dentro do mesmo transcritoma, não sendo necessária em comparações do mesmo transcrito entre diferentes transcritomas. Nas comparações entre duas amostras, outro importante aspecto de normalização é controlar as diferenças na profundidade e qualidade do sequenciamento (Wolf, 2013).

Muitos métodos de análise foram desenvolvidos para identificar genes diferencialmente expressos através de diferentes modelos estatísticos (Nookaew et al., 2012). Para dados de RNA-Seq, o mais adequado é tratar pela distribuição binomial negativa (Wolf, 2013). Vários pacotes de software são utilizados para realizar a análise de expressão diferencial, com diferentes abordagens, modelos e eficiências: edgeR (Robinson et al., 2010), Cuffdiff (Trapnell et al., 2010), DESeq (Anders e Huber, 2010) e DESeq2 (Love et al., 2014), baySeq (Hardcastle e Kelly, 2010), NOISeq (Tarazona et al., 2011), entre outros. Nookaew e colaboradores realizaram uma comparação de todos estes métodos de análise estatística de dados de RNA-Seq, encontrando resultados consistentes entre eles. Foi observado que o edgeR identificou mais genes diferencialmente expressos do que os outros métodos na mesma condição e mesmo cutoff, o que pode significar menos controle de erros do tipo 1, ou seja, maior presença de falsos negativos em relação à hipótese nula (de que as mudanças observadas são ao acaso) (Nookaew et al., 2012). Esse mesmo estudo comparou os genes diferencialmente expressos nos dados de RNA-Seq com os identificados em análises de microarranjos, evidenciando boa concordância entre as duas tecnologias. Com as listas de genes diferencialmente expressos, categorias funcionais devem ser definidas para cada um, sendo a categorização por Gene Ontology (GO) uma das mais utilizadas (Ashburner et al., 2000; Gene Ontology Consortium, 2004).

A tecnologia de RNA-Seq também possibilita obter informação da fita de DNA que foi transcrita, por meio de métodos alternativos de preparação de bibliotecas de fragmentos de cDNA (Wang *et al.*, 2009; Sorek e Cossart, 2010; Filiatrault, 2011; Ozsolak e Milos, 2011). Tais métodos envolvem a ligação de diferentes adaptadores em orientação conhecida relativa às extremidades 5' e 3' dos transcritos ou às moléculas da primeira-fita dos cDNAs (Levin *et al.*, 2010; Ozsolak e Milos, 2011). Outra abordagem depende em marcar uma fita com uma modificação química, tanto no próprio RNA ou durante a síntese da segunda fita de cDNA seguida de degradação da fita não marcada (Parkhomchuk *et al.*, 2009; Levin *et al.*, 2010; Ozsolak e Milos, 2011). Esta estratégia

de sequenciamento com informação de fita permite detectar eventos de transcrição antisenso a algum gene anotado, determinar a fita em que ncRNAs foram transcritos, demarcar corretamente as fronteiras de genes adjacentes transcritos em fitas opostas e corrigir os níveis de expressão de transcritos que se sobrepõem (Wang *et al.*, 2009; Levin *et al.*, 2010; Sorek e Cossart, 2010; Ozsolak e Milos, 2011).

Outras duas importantes aplicações do RNA-Seq são a metatranscritômica e o transcritoma de uma única célula. A metatranscritômica visa a descrição do perfil de expressão de comunidades microbianas complexas (Sorek e Cossart, 2010; Filiatrault, 2011). Já a análise do transcritoma de uma única célula permite um melhor entendimento da heterogeneidade transcricional de células presentes em uma população bacteriana (Sorek e Cossart, 2010; Tang *et al.*, 2011). Tal estratégia pode ajudar a entender como células respondem individualmente a sinais ambientais. A heterogeneidade da expressão gênica pode ser devido a diferenças no status epigenético de seu genoma, ciclo celular, microambiente ou nicho, entre outros fatores. Células individuais podem ser selecionadas manualmente, por microdissecção por laser ou separação por fluorescência (Tang *et al.*, 2011).

Uma estratégia alternativa interessante ao RNA-Seq é o chamado sequenciamento direto de RNA (*Direct RNA Sequencing*, DRS), o qual realiza o sequenciamento de moléculas de RNA diretamente, sem a prévia síntese de cDNA ou a necessidade de passos de ligação e amplificação. A reação de sequenciamento por síntese é realizada usando uma polimerase modificada e análogos de nucleotídeos (chamados de *Virtual Terminator* ou VT), os quais contém um marcador fluorescente e grupos químicos que podem ser clivados e que permitem o sequenciamento (Ozsolak *et al.*, 2009). As vantagens deste método são as quantidades muito pequenas de RNA exigidas e a eliminação de vieses introduzidos no processo de preparação das bibliotecas de fragmentos em experimentos de RNA-Seq convencional (Ozsolak *et al.*, 2009).

A maioria das análises de transcritomas completos visa obter o perfil transcricional momentâneo em condições variadas e/ou avaliar a expressão diferencial em resposta a distintos estímulos. Estudos de transcritoma também estão sendo utilizados para comparar como linhagens ou espécies proximamente relacionadas respondem a determinadas condições ou estímulos, evidenciando respostas transcricionais relevantes para sua adaptação a ambientes distintos e que eventualmente incluem a expressão de potenciais fatores de virulência (Filiatrault, 2011).

1.3. Aspectos da biologia de Xylella fastidiosa

Xylella fastidiosa é um importante fitopatógeno, pertencente ao grupo de bactérias Gram-negativas da classe Gammaproteobacteria, família Xanthomonadaceae (Wells *et al.*, 1987; Hopkins, 1989). As células de *X. fastidiosa* têm formato de bastonete (Almeida e Nunney, 2015) com diâmetro entre 0,25 e 0,50 μm e comprimento que pode variar entre 1,0 a 4,0 μm. Suas colônias são geralmente circulares com bordas lisas ou rugosas. Apresenta crescimento lento no cultivo *in vitro*, com tempo de duplicação que varia entre 9 e 55 horas, dependendo do meio de cultivo. A sua temperatura ótima de crescimento *in vitro* é entre 26°C e 28°C (Davis *et al.*, 1978; Davis *et al.*, 1981; Wells *et al.*, 1987; Hopkins, 1989). Embora classificada como aeróbia obrigatória, foi verificado que, dependendo do meio de cultivo, essa bactéria se comporta com anaeróbia facultativa (Shriner e Andersen, 2014). *X. fastidiosa* coloniza dois ambientes muito distintos: o lúmen dos vasos do xilema de seus hospedeiros vegetais, não sendo capaz de colonizar outros tecidos da planta, e o aparelho bucal de insetos sugadores de seiva do xilema (Chatterjee *et al.*, 2008).

X. fastidiosa é considerada um fitopatógeno generalista em razão da ampla gama de hospedeiros em que já foi encontrada. A lista atualizada da EFSA (*European Food Safety Authority*)¹ inclui 359 espécies de plantas de 240 gêneros e 75 famílias botânicas como hospedeiros naturais e/ou experimentais de *X. fastidiosa* (EFSA, 2016). Contudo, a hipótese é que a gama de hospedeiros dessa bactéria poderia aumentar com o surgimento de novas cepas *X. fastidiosa* em decorrência de eventos de recombinação genética, transferência horizontal de genes e infecção por bacteriófagos (Nunney *et al.*, 2012; Jacques *et al.*, 2016).

Muitos destes hospedeiros vegetais não exibem sintomas quando colonizados por *X. fastidiosa*, e são ditos resistentes à infecção; nesses casos *X. fastidiosa* pode ser considerada um endofítico (Chatterjee *et al.*, 2008). Por outro lado, várias espécies são altamente suscetíveis à infecção por *X. fastidiosa*, nas quais causa doenças de extrema gravidade, como a Clorose Variegada dos Citros (CVC) (Rossetti *et al.*, 1990; Chang *et al.*, 1993), Doença de Pierce das videiras (PD) (Hopkins e Purcell, 2002) e a Síndrome do Rápido Declínio das Oliveiras (OQDS) (Cariddi *et al.*, 2014). Exemplos de outras

¹ https://www.efsa.europa.eu/en/press/news/160209

doenças associadas a infecção por *X. fastidiosa*, são as escaldaduras de folha de ameixeira (*plum*, PLS-*plum*), amendoeira (ALS), cafeeiro (CLS), pessegueiro (PLS-*peach*), oleandro ou espirradeira (OLS), amoreira, cerejeira, carvalho, sabugueiro e hibisco (Hopkins e Purcell, 2002; Chatterjee *et al.*, 2008; Janse e Obradovic, 2010).

A literatura reporta a existência de especificidade entre cepas de *X. fastidiosa* e espécies do hospedeiro vegetal que colonizam (Almeida *et al.*, 2008; Prado *et al.*, 2008; Killiny e Almeida, 2011). Além disso, mesmo espécies relacionadas têm comportamentos distintos em relação à *Xylella*. Por exemplo, entre as espécies de citros, *Citrus sinensis* (laranjeira doce) é mais suscetível a CVC, enquanto que outras espécies (*C. reticulata, C. limonia, C. limon, C. medica* e *C. grandis*) apresentam sintomas moderados ou são resistentes à infecção (Rossetti e De Negri, 1990). Por outro lado, há relatos de uma mesma cepa ter colonizado com sucesso plantas de espécies distintas em infecções artificiais (Li *et al.*, 2001; Lopes *et al.*, 2003; Prado *et al.*, 2008; Lopes *et al.*, 2010; Oliver *et al.*, 2015). As bases moleculares da especificidade de *X. fastidiosa* ao hospedeiro vegetal ainda não são conhecidas, mas certamente não envolvem efetores secretados pelo sistema de secreção tipo 3 (McCann e Guttman, 2007) visto que este sistema está ausente nessa fitobactéria (Killiny e Almeida, 2011).

1.3.1. Doenças causadas por X. fastidiosa, formas de transmissão e controle

Entre as doenças causadas por *X. fastidiosa*, as mais estudadas são a CVC e a PD. Os principais sintomas da CVC são lesões cloróticas intervenais na parte superior de folhas maduras e regiões de necrose na parte inferior correspondentes às áreas de clorose da parte superior. Os frutos apresentam tamanho reduzido, endurecidos e sem sumo e com maturação precoce, inviabilizando o consumo *in natura* e sua utilização na indústria de sucos e derivados (Hopkins, 1989; Janse e Obradovic, 2010). Entretanto, a CVC geralmente não causa morte de indivíduos susceptíveis infectados (Li *et al.*, 2003). Essa doença ocorre, em intensidades diferentes, em quase todas as áreas citrícolas do Brasil e também já foi detectada na Costa Rica, Argentina e Paraguai (Coletta-Filho e Machado, 2003).

A PD tem como principal sintoma a necrose do limbo foliar que se inicia nas bordas e evolui para as regiões centrais. O desenvolvimento dos frutos é severamente afetado, sendo menores e murchos, e, portanto, inadequados para consumo e produção de sucos e vinhos. A PD pode ser considerada mais agressiva do que a CVC, uma vez que em estágios avançados há abscisão das folhas pela parte distal do pecíolo e eventual morte do hospedeiro (Hopkins, 1989; Janse e Obradovic, 2010). Apesar desses sintomas geralmente serem associados ao déficit hídrico gerado pela oclusão do xilema pelo biofilme de *X. fastidiosa*, tem sido sugerido que outros fatores teriam papel importante na geração dos sintomas (Thorne *et al.*, 2006), uma vez que folhas altamente sintomáticas apresentaram baixas concentrações de células do patógeno (Gambetta *et al.*, 2007). A PD é encontrada predominantemente nos EUA, principalmente no estado da Califórnia, mas também há registros dessa doença na América Central e Taiwan (Su *et al.*, 2013; Almeida e Nunney, 2015).

As outras doenças associadas à *X. fastidiosa*, genericamente denominadas de escaldaduras, apresentam sintomas foliares semelhantes às necroses características da PD e também comprometem o desenvolvimento dos frutos (Janse e Obradovic, 2010). Os hospedeiros vegetais ornamentais e florestais são de grande importância na epidemiologia das principais doenças causadas pela *X. fastidiosa*, pois no caso de estarem nas vizinhanças de pomares citrícolas, vinhedos ou olivais, poderiam servir como reservatório da bactéria em razão da relativa inespecificidade para algumas cepas dessa bactéria (Baumgartner e Warren, 2005; Hernandez-Martinez *et al.*, 2007; Chatterjee *et al.*, 2008; Saponari *et al.*, 2013; Almeida e Nunney, 2015).

Sendo uma bactéria restrita ao xilema, *X. fastidiosa* não é encontrada em vida livre ao contrário de outras espécies da família Xanthomonadaceae. A transmissão de *Xylella* entre os hospedeiros vegetais requer a ação de pequenos insetos da família *Cicadellidae (sharpshooters)* e *Cercopidae (spitlebugs)*, popularmente conhecidos como cigarrinhas. Estes insetos têm como hábito sugar o fluido xilemático diretamente dos pecíolos das folhas, e podem eventualmente adquirir células de *X. fastidiosa* durante a sua alimentação. As bactérias então aderem ao forro cuticular das partes bucais do tubo digestório anterior do inseto, o pré-cibário e cibário, que são revestidos com quitina. Em uma próxima alimentação do inseto, as bactérias poderão ser inoculadas diretamente no xilema e, dependendo da titulação de células, colonizar o hospedeiro (Redak *et al.*, 2004; Almeida *et al.*, 2005; Chatterjee *et al.*, 2008; Backus e Morgan, 2011). *X. fastidiosa* é capaz de se multiplicar no aparelho bucal do inseto, sendo o único fitopatógeno transmitido por artrópode que é propagativo, mas que não circula na

hemolinfa (Purcell e Finlay, 1979). Ademais, *X. fastidiosa* produz quitinases, indicando a possível utilização de quitina como nutriente (Killiny *et al.*, 2010; Labroussaa *et al.*, 2017). Diante destas observações, ainda há debate se o inseto é meramente um vetor da bactéria ou pode ser considerado um hospedeiro intermediário, uma vez que a *Xylella* não provoca qualquer dano aparente ao inseto ou prejudica seu desenvolvimento e reprodução.

As cigarrinhas são consideradas generalistas, uma vez que podem se alimentar do fluido xilemático de diversas espécies de plantas e as diversas espécies de cigarrinhas são capazes de transmitir *X. fastidiosa*, embora com diferentes eficiências (Redak *et al.*, 2004). Somando todas essas informações com a relativa inespecificidade de hospedeiro observada em algumas linhagens de *X. fastidiosa*, as probabilidades de transmissão aumentam e dificultam as estratégias de controle da doença. Entretanto, o controle das populações de insetos vetores ainda constitui uma das principais formas de controle das doenças causadas pela *Xylella* (Almeida *et al.*, 2005).

Além da principal forma de transmissão por meio do inseto vetor, a transmissão antrópica também pode disseminar a doença. O comércio de mudas de plantas de interesse econômico e agrícola pode espalhar a bactéria para diversos locais do planeta, caso elas estejam contaminadas com *X. fastidiosa*. Essa bactéria é considerada patógeno de grave risco biológico e seu controle é requerido na inspeção e certificação de mudas que são transportadas de países em que *Xylella* é endêmica (Madden e Wheelis, 2003). Além da disseminação por mudas infectadas, porta-enxertos contaminados também consistiam em fonte de transmissão. Atualmente, são empregados espécies resistentes como porta-enxertos, como o limão-cravo que é usado em mudas de laranjeira doce. O manejo adequado dos pomares, incluindo a poda de galhos contaminados e a erradicação de plantas severamente contaminadas, é também muito importante para controle das doenças juntamente com o controle dos insetos transmissores de *X. fastidiosa* e a utilização de mudas certificadas (Janse e Obradovic, 2010).

Ainda em caráter experimental, outras potenciais formas de controle de doenças como a PD e a CVC envolvem o controle biológico por meio da inoculação de cepas não-virulentas de *X. fastidiosa* como a EB92-1 em videiras (Hopkins, 2005) ou de endofíticos que interferem na crescimento de *X. fastidiosa* (Araujo *et al.*, 2002; Lacava *et al.*, 2004). Outra possibilidade promissora é a fertirrigação com N-acetil-cisteína, a qual se mostrou eficiente no tratamento de plantas com sintomas de CVC (Muranaka *et*

al., 2013). O desenvolvimento de plantas transgênicas resistentes à *X. fastidiosa* também está em consideração (Aguero *et al.*, 2005; Dandekar *et al.*, 2012; Lindow *et al.*, 2014).

1.3.2 As epidemias de X. fastidiosa

X. fastidiosa foi descrita somente no final da década de 1980 (Wells *et al.*, 1987), porém há registros em 1884 de uma grande epidemia em videiras no estado da Califórnia nos EUA, a qual se acredita ser o primeiro relato da Doença de Pierce (PD) (Auger *et al.*, 1974). Seis anos mais tarde, outra epidemia em videiras foi relatada no estado da Flórida (EUA), sendo comparada com a PD (Stoner, 1952; Nunney *et al.*, 2010). Somente em 1892 que a doença recebe oficialmente o nome de PD, uma vez que foi descrita pelo fitopatologista Newton B. Pierce (Hopkins e Purcell, 2002; Janse e Obradovic, 2010; Nunney *et al.*, 2010). Inicialmente, acreditava-se que a PD e doenças relacionadas eram causadas por vírus. Somente em 1978, bactérias foram isoladas de plantas doentes, cultivadas e reinseridas na planta, completando o postulado de Koch, comprovando ser o agente causal da doença. Inicialmente, acreditava-se que essas bactérias forma classificadas como sendo da familia Xanthomonadaceae e, então, nomeadas de *X. fastidiosa* (Wells *et al.*, 1987; Hopkins, 1989).

Os primeiros registros de CVC na América do Sul foram relatados no estado de Minas Gerais em 1987 (Rossetti *et al.*, 1990), seguido por sua ocorrência em São Paulo (Chang *et al.*, 1993). Como já mencionado, a CVC ocorre em quase todas as áreas citrícolas do Brasil e também já foi detectada na Costa Rica, Argentina e Paraguai (Coletta-Filho e Machado, 2003). Já a escaldadura da folha do cafeeiro foi observada pela primeira vez em São Paulo em 1995 (Beretta *et al.*, 1996), em seguida sendo relatada na Costa Rica (Rodríguez *et al.*, 2001).

Até meados dos anos 90, havia registros deste fitopatógeno exclusivamente no continente americano. Entretanto, a ocorrência de *Xylella* foi relatada nos continentes europeu e asiático: pereiras e videiras em Taiwan (Leu e Su, 1993; Su *et al.*, 2013); videiras em Kosovo (Berisha *et al.*, 1998); amendoeiras na Turquia (Güldumlaut~r *et al.*, 2005); videiras e amendoeiras no Irã (Amanifar *et al.*, 2014); oleandro,

amendoeiras, oliveiras, cerejeiras, além de duas plantas arbustivas (*Polygala myrtifolia* e *Westringia fruticosa*) na Itália (Saponari *et al.*, 2013; Cariddi *et al.*, 2014; Saponari *et al.*, 2014). Acredita-se que a *Xylella* tenha sido introduzida nesses locais por meio de mudas infectadas oriundas de países em que essa bactéria é endêmica, principalmente das Américas. Vale relatar, que em dois episódios recentes, as severas políticas de controle e manejo atualmente vigentes evitaram introduções de mudas infectadas. No primeiro, mudas de cafeeiros com sintomas de escaldadura da folha do cafeeiro (CLS) oriundas da Costa Rica e Honduras foram encontradas nos Países Baixos, sendo diagnosticadas e destruídas (Bergsma-Vlami *et al.*, 2015). No segundo episódio, quatro mudas de cafeeiro infectadas, oriundas do Equador e México, foram encontradas na França e também foram destruídas (Jacques *et al.*, 2016).

Com as recentes ocorrências de X. fastidiosa na região mediterrânea, diversos países como Portugal (Pereira, 2015), Espanha, Grécia, Marrocos, Tunísia, Egito, Líbia, entre outros, estão em alerta e tomando as providências necessárias para evitar sua disseminação. Apesar desses cuidados, a bactéria foi recentemente encontrada no arquipélago das Ilhas Baleares, território pertencente à Espanha e localizado no Mar Mediterrâneo. Oliveiras, amendoeiras, cerejeiras, entre outras espécies, foram diagnosticadas com X. fastidiosa em outubro de 2016, nas duas das maiores ilhas do arquipélago, Mallorca e Ibiza². Um recente estudo em que foi realizada uma análise da topologia da rede de olivais indica que o Sul da Itália está se tornando um reservatório para X. fastidiosa, com poucas chances de erradicação, sugerindo que o foco deve ser direcionado para estratégias de manejo e controle do vetor de modo a evitar espalhamento do patógeno a outros países e pomares (Strona et al., 2017). Interessantemente, em seu livro escrito em 1789, o cientista e professor de Medicina e Filosofia Cosimo Moschettini descreve os sintomas de uma doença em oliveiras similares aos observados na Síndrome do Rápido Declínio das Oliveiras (OQDS) causada por X. fastidiosa e que tem aterrorizado os agricultores italianos nos últimos anos (Saponari et al., 2013; Cariddi et al., 2014; Saponari et al., 2014). A Brusca, como ele denominou a doença, foi observada próxima da região onde ocorreu o surto de OODS de 2013^3 .

² https://phys.org/news/2017-02-spain-balearic-islands-deadly-olive.html#jCp

³ http://www.olivenews.gr/en/article/7668

1.3.3. Variabilidade entre cepas e genomas de Xylella

Todos os isolados de *X. fastidiosa* tem sido classificados, até o momento, como uma única espécie, sendo que apresentam suficiente variabilidade genética para serem divididos em linhagens ou cepas (Hopkins, 1989; Mehta e Rosato, 2001). Há, entretanto, uma exceção, a cepa PLS229, isolada de uma árvore de pera-asiática (*Pyrus pyrifoliae*) no distrito de Houli, Taiwan, que foi proposta como uma nova espécie, *Xylella taiwanensis* (Su *et al.*, 2014; Su *et al.*, 2016a).

Dezenas de linhagens de X. fastidiosa isoladas de variados hospedeiros já foram descritas e comparadas com base em sequências de genes marcadores filogenéticos, o que permitiu a proposição de seis subespécies: fastidiosa, multiplex, sandyi, pauca, tashke e morus (Schaad et al., 2004; Schuenzel et al., 2005; Randall et al., 2009; Nunney et al., 2014b; Almeida e Nunney, 2015; Jacques et al., 2016). Entre as cepas de X. fastidiosa subsp. fastidiosa (anteriormente chamada de subsp. piercei) estão as que causam PD e doenças em amendoeira, alfafa, carvalho. Acredita-se que a subespécie fastidiosa tenha sido introduzida nos EUA, oriunda da América Central (Costa Rica), e que tenha recombinado com um ancestral nativo da subespécie *multiplex*, o deve ter lhe conferido a capacidade de infectar videiras, gerando o surto de PD ocorrido em 1884 (Nunney et al., 2010; Nunney et al., 2013). Cepas da subespécie multiplex estão associadas com uma grande variedade de hospedeiros incluindo doenças de escaldadura em amendoeiras, ameixeira, pessegueiro, carvalho, entre outros. Acreditava-se que esta subespécie era nativa de climas temperados na América do Norte, porém uma cepa causadora da escaldadura da folha de ameixeira (PLS) foi encontrada em 1935 na Argentina, Brasil e Paraguai (French e Kitajima, 1978), possivelmente após introdução proveniente dos EUA (Nunes et al., 2003).

X. fastidiosa subsp. *pauca* compreende as cepas causadoras de CVC, CLS e outras doenças causadas por esta bactéria na América do Sul (Almeida *et al.*, 2008; Nunney *et al.*, 2014a). Dados filogenéticos indicam que a *Xylella* seria nativa da América do Sul, estando presente de forma endofítica em algum hospedeiro ainda desconhecido. Mesmo com a introdução de espécies de citros, originárias do continente asiático (principalmente Índia e China), e de espécies de cafeeiros, originárias do continente africano (Etiópia), sendo estabelecidas como grandes monoculturas por

vários anos, não havia registros de doenças causadas por *X. fastidiosa* nestes hospedeiros. Nunney e colaboradores sugerem que a *X. fastidiosa* subsp. *pauca*, como a conhecemos hoje, tenha se originado via recombinação intersubespecífica entre a cepa de *X. fastidiosa* subsp. *pauca* nativa da América do Sul e a cepa de *X. fastidiosa* subsp. *multiplex* que infecta ameixeira, tornando-se patogênica para citros e café (Nunney *et al.*, 2012).

Os isolados de oliveiras infectadas por *X. fastidiosa* na Argentina e na Itália foram também classificados como sendo da subespécie *pauca* (Elbeaino *et al.*, 2014; Haelterman *et al.*, 2015). Em um estudo filogenético com sequências de sete genes *housekeeping*, a cepa italiana CoDiRO (isolada de oliveira com OQDS) agrupou-se no clado contendo cepas da subespécie *pauca* distantes das cepas brasileiras de citros e cafeeiro, mas próximas de cepas de cafeeiros do Equador (Jacques *et al.*, 2016). Tal fato corrobora a alta similaridade genética desta cepa com um isolado de *Xylella fastidiosa* subsp. *pauca* da América Central, geograficamente mais próximo do Equador do que do Brasil (Giampetruzzi *et al.*, 2015a). Entretanto, outras seis cepas de *X. fastidiosa* isoladas de oliveiras assintomáticas na Califórnia foram classificadas como *X. fastidiosa* subsp. *multiplex*. Inoculações em videiras e amendoeiras resultaram em sintomas característicos de cepas desta subespécie. Portanto, as oliveiras na Califórnia estavam apenas servindo como reservatório de cepas de *Xylella* patogênicas a outras espécies de hospedeiros (Krugner *et al.*, 2014). Tal fato é mais uma evidencia de especificidade entre cepas e seus respectivos hospedeiros vegetais.

Schuenzel e colaboradores estimaram que o clado contendo as subespécies *fastidiosa* e *sandyi* se dividiram da subespécie *multiplex* há cerca de 30.000 anos (Schuenzel *et al.*, 2005). Análises com a subespécie *pauca*, um grupo externo em relação as subespécies *fastidiosa* e *sandyi*, indicaram que ela divergiu há 60.000 anos (Nunney *et al.*, 2012).

As outras três subespécies envolvem cepas que infectam apenas uma espécie de hospedeiro cada. Cepas de *X. fastidiosa* subsp *sandyi* são responsáveis por causar doenças em oleandro ou espirradeira (*Nerium oleander*). É provável que esse genótipo tenha sido introduzido nos EUA 100 anos após a epidemia em videiras na Califórnia (Purcell *et al.*, 1999; Yuan *et al.*, 2010). A subespécie *X. fastidiosa* subsp *tashke* é representada por cepas isoladas de plantas sintomáticas da espécie ornamental *Chitalpa tashkentensis* (Randall *et al.*, 2009). Por fim, *X. fastidiosa* subsp *morus* foi proposta

para cepas que causam doença em amoreira, as quais também podem ter surgido de eventos de recombinação intersubespecífica entre linhagens das subespécies *fastidiosa* e *multiplex* (Nunney *et al.*, 2014b).

Até o momento, 38 cepas de *Xylella*, com diversas origens e hospedeiros, tiveram seus genomas sequenciados (Tabela 1), permitindo estudos mais aprofundados de suas diferenças fenotípicas e genômicas. A cepa 9a5c, isolada de uma planta de citros com sintomas de CVC no estado de São Paulo (Li *et al.*, 1999), foi o primeiro genoma de um fitopatógeno sequenciado no mundo (Simpson *et al.*, 2000) e se tornou referência para os outros genomas que vieram a seguir. Como pode ser observado na Tabela 1, genomas de cepas de citros são predominantes, tendo sido isoladas de diferentes cidades do estado de São Paulo, diferentes estados do Brasil, além de uma cepa da Argentina. Enquanto isso, genomas de cepas norte-americanas mostram uma maior diversidade e um equilíbrio entre os hospedeiros, embora a maioria ainda tenha sido isolada de videiras. As primeiras sequências genômicas de cepas norte-americanas obtidas foram de Ann-1 (oleandro) e Dixon (amendoeira), embora incompletas. A cepa Temecula1 foi o primeiro genoma completo de cepas norte-americanas a ser publicado (Van Sluys *et al.*, 2003), servindo como referência para os genomas de cepas norte-americanas.

Os primeiros genomas de *Xylella* sequenciados fora do continente americano foram o da cepa PLS229, isolada de pereira em Taiwan (Su *et al.*, 2014), e o da cepa CoDiRO, isolada de oliveira na Itália (Giampetruzzi *et al.*, 2015a). Recentemente, foi proposto que a cepa PLS229 apresenta um genoma diferente o suficiente para ser classificada como uma nova espécie do gênero *Xylella*, a *Xylella taiwanensis*. Os valores médios de identidade de nucleotídeo entre esta cepa e outras de *X. fastidiosa* foram de 83,4 a 83,9%, significativamente menores do que os 95% encontrados entre bactérias da mesma espécie. Entretanto, a similaridade de sequência entre os genes dos rRNAs 16S, comumente empregados para caracterizar espécies, foi maior do que 98%, acima dos 97% exigidos para justificar que duas cepas sejam da mesma espécie. Porém, outras diferenças foram encontradas: apenas 87% de similaridade entre as sequências do espaçador 16S-23S, SNPs em genes *housekeeping*, perfis de ácidos graxos distintos e diferentes fenótipos. Além disso, uma árvore filogenética com sequências de genes de rRNA 16S posicionaram PLS229 entre os membros dos gêneros *Xylella* e *Xanthomonas*. Tais dados justificaram a proposição de nova espécie (Su *et al.*, 2016a).

Introdução

Uma vez que os genomas sequenciados são originários de diferentes cepas de uma mesma espécie, a maior parte da estrutura e organização do genoma bem como o conteúdo gênico é concordante entre eles. Mesmo assim, diferenças significativas em suas sequências ainda foram encontradas, refletindo nos fenótipos de suas cepas e nos diferentes níveis de virulência. Estudos de genômica comparativa revelaram a ocorrência de grandes rearranjos de blocos de sequências nos cromossomos (translocações e inversões), *indels* (inserções e deleções) de nucleotídeos em genes importantes e polimorfismos de um único nucleotídeo (SNPs). Diferenças também foram encontradas no conteúdo flexível do genoma, que incluem sequências de elementos transponíveis, como prófagos e plasmídeos, e outros genes que podem estar relacionados com a especificidade entre cepa e hospedeiro (Nunes *et al.*, 2003; Van Sluys *et al.*, 2003; Koide *et al.*, 2004; da Silva *et al.*, 2007; Doddapaneni *et al.*, 2007; de Mello Varani *et al.*, 2008; Pierry, 2012; Santana, 2012; Barbosa *et al.*, 2015).

Em algumas cepas não foram encontrados plasmídeos, enquanto em outras um ou mais deles estão presentes (Simpson et al., 2000; Marques et al., 2001; Lee et al., 2010; Stenger et al., 2010; Rogers e Stenger, 2012), os quais eventualmente contem genes que podem conferir alguma vantagem à cepa, como por exemplo, a toxina PemK que atua como endoribonuclease (Lee et al., 2012). X. fastidiosa tem a maior razão entre o número de genes de bacteriófagos e o tamanho do genoma, sendo que o conteúdo de sequências relacionadas a prófagos nos genomas é muito diverso, com cada cepa apresentando um perfil único de tipos e quantidades (Varani et al., 2013). Postula-se que uma vez ativos, os prófagos podem ser responsáveis por rearranjos cromossômicos em larga escala no genoma, além de carregarem genes potencialmente importantes para a sobrevivência, virulência e para a capacidade em colonizar diversas espécies de hospedeiros (Nunes et al., 2003; Chen et al., 2005; de Mello Varani et al., 2008). O alto número de sequências oriundas de fagos nos genomas de X. fastidiosa pode ser devido à ausência do sistema CRISPR-Cas completo (Santana, 2012), o qual tem importante papel de imunidade e proteção contra moléculas de ácidos nucleicos invasoras (Sorek et al., 2008; Deveau et al., 2010).

Introdução

Сера	Hospedeiro	Origem [∆]	Tamanho (Mb)	%GC	Scaffolds	CDS ⁺	Referências
9a5c	Citros	Brasil (SP)	2,73	52,64	3	2310	(Simpson et al., 2000)
J1A12	Citros	Brasil (SP)	2,87	52,81	3	2378	(Pierry, 2012)
U24D	Citros	Brasil (SP)	2,73	52,62	2	2246	(Santana, 2012)
XRB	Citros	Brasil (BA)	2,71	52,24	74	2288	Não publicado
B111	Citros	Brasil (SP)	2,68	52,36	77	2274	Não publicado
11399	Citros	Brasil (SE)	2,74	52,64	36	2281	(Niza et al., 2016)
CVC0251	Citros	Brasil (SP)	2,74	52,51	130	2305	Não publicado
CVC0256	Citros	Brasil (SP)	2,70	52,51	128	2280	Não publicado
3124	Cafeeiro	Brasil (SP)	2,75	52,63	1	2261	(Santana, 2012)
COF0324	Cafeeiro	Brasil (MG)	2,77	52,43	143	2352	Não publicado
6с	Cafeeiro	Brasil (SP)	2,61	52,36	46	2192	(Alencar et al., 2014)
32	Cafeeiro	Brasil (SP)	2,61	52,50	56	2155	(Alencar <i>et al.</i> , 2014)
HIB4	Hibisco	Brasil (SP)	2,88	52,69	2	2388	(Pierry, 2012)
PR8X	Ameixeira	Brasil (SP)	2,71	52,59	2	2229	(Pierry, 2012)
FB7	Citros	Argentina (Corrientes)	2,70	52,51	2	2150	(Santana, 2012)
Temecula1	Videira	EUA (CA)	2,52	51,80	2	2157	(Van Sluys <i>et al.</i> , 2003)
GB514	Videira	EUA (TX)	2,52	51,77	2	1992	(Schreiber IV et al., 2010)
ATCC 35879	Videira	EUA (FL)	2,52	51,80	16	2116	Não publicado
Stag's Leap	Videira	EUA (CA)	2,51	51,70	15	2087	(Chen et al., 2016)
Ann-1	Oleandro	EUA (CA)	2,78	52,07	2	2375	(Bhattacharyya et al., 2002)
Dixon	Amendoeira	EUA (CA)	2,62	52,00	32	2255	(Bhattacharyya et al., 2002)
M12	Amendoeira	EUA (CA)	2,48	51,90	1	2054	(Chen et al., 2010)
M23	Amendoeira	EUA (CA)	2,57	51,76	2	2193	(Chen et al., 2010)
EB92.1	Sabugueiro	EUA (FL)	2,48	51,50	168	2070	(Zhang <i>et al.</i> , 2011)
MUL0034	Amoreira	EUA (CA)	2,67	51,97	2	2234	Não publicado
Mul-MD	Amoreira	EUA (MD)	2,52	51,60	101	2077	(Guan et al., 2014b)
Sy-VA	Sicômoro	EUA (VA)	2,48	51,60	128	2042	(Guan <i>et al.</i> , 2014a)
Griffin-1	Carvalho vermelho	EUA (GA)	2,39	51,70	84	1855	(Chen et al., 2013)

Tabela 1. Lista de cepas de Xylella fastidiosa que tiveram seus genomas sequenciados, e suas principais características.

Introdução								
ATCC 35871	Ameixeira	EUA (GA)	2,42	51,70	58	1966	Não publicado	
BB01	Mirtilo	EUA (GA)	2,52	51,80	84	2943	(Van Horn <i>et al.</i> , 2017)	
CO33	Cafeeiro	Itália (Costa Rica)*	2,68	51,70	96	2224	(Giampetruzzi <i>et al.</i> , 2015b)	
COF0407	Cafeeiro	Costa Rica (San José)	2,54	51,84	172	2081	Não publicado	
OLS0478	Oleandro	Costa Rica (San José)	2,56	51,92	48	2115	Não publicado	
OLS0479	Oleandro	Costa Rica (San José)	2,54	51,93	183	2070	Não publicado	
CoDiRO	Oliveira	Itália (Apulia)	2,54	51,97	12	2076	(Giampetruzzi <i>et al.</i> , 2015a)	
CFBP8072	Cafeeiro	França (Equador)*	2,50	51,90	278	1987	(Jacques et al., 2016)	
CFBP8073	Cafeeiro	França (México)*	2,58	51,60	328	2141	(Jacques et al., 2016)	
PLS229 (X. taiwanensis)	Pereira	Taiwan (Houli)	2,73	53,10	1	2152	(Su et al., 2014)	

^Aentre parêntesis o estado, província ou distrito do país de origem da cepa; ⁺genes codificadores de proteínas; *país de origem das mudas infectadas com *X. fastidiosa*.

1.3.4. Fatores de virulência de X. fastidiosa

O sequenciamento do primeiro genoma de *X. fastidiosa* (cepa 9a5c, agente causal da CVC e isolada de citros no Brasil) forneceu informações importantes e essenciais para o entendimento dos principais fatores de virulência e patogenicidade dessa bactéria. Este conhecimento tem sido constantemente enriquecido com o sequenciamento de genomas de cepas de outras localidades e de diversos hospedeiros, aliado a estudos bioquímicos, genéticos e fitopatológicos realizados, principalmente, com as cepas de referência 9a5c e Temecula1, que é o agente causal da PD.

A bactéria X. fastidiosa tem como principal característica ser muito adesiva, possuindo vários genes codificadores de adesinas fimbriais e afimbriais (Simpson et al., 2000; Chatterjee et al., 2008; Zaini et al., 2015). Os pili de Xylella estão sempre localizados em apenas um dos polos da célula e podem ser de dois tipos, dependendo de seus tamanhos e funções. O pilus curto (as células de X. fastidiosa possuem vários pili curtos), também chamado de chaperoneusher, possui funções relacionadas à adesão a substratos e superfícies, além da formação de biofilme e agregação (Simpson et al., 2000; Li et al., 2007; Caserta et al., 2010). Uma vez que a Xylella não possui flagelo, ela possui um pilus longo, também chamado de pilus do tipo IV, o qual é utilizado pela célula para realizar a motilidade do tipo *twitching*, além de participar na formação de biofilme (Meng et al., 2005; Li et al., 2007). Esse mecanismo funciona com a polimerização do pilus para o exterior da célula, adesão ao substrato por uma proteína adesina localizada na sua extremidade, seguido da despolimerização do pilus, puxando a célula para onde a adesina está aderida. Como a bactéria habita um ambiente com fluxo constante, ambas as adesinas são importantes para que ela possa se manter presa no substrato (De La Fuente et al., 2007). Foi demonstrado que a motilidade por twitching por pilus longo ocorre predominantemente contra uma corrente de fluxo em câmara de microfluídica, sugerindo que X. fastidiosa é capaz de se mover na direção oposta do fluxo do fluido xilemático (Meng et al., 2005). Tal motilidade bacteriana permite que a bactéria possa atingir outros vasos do xilema e colonizar a planta de forma sistêmica. Esta movimentação parece envolver quimiotaxia, pois foram identificados genes ortólogos a sistemas quimiossensoriais no genoma deste fitopatógeno (Cursino et al., 2011).

Além das adesinas fimbriais, as adesinas afimbriais também são de extrema importância para a sobrevivência da população bacteriana. Entre elas estão genes que codificam para hemaglutininas e adesinas autotransportadoras triméricas (Xad). Todas contribuem em diferentes intensidades para a adesão ao substrato, mas sua função principal parece ser adesão célula-célula, uma vez que elas estão distribuídas na membrana externa e não limitadas a um dos polos da célula, como é o caso das adesinas fimbriais. Este tipo de adesão tem papel fundamental para a formação de um biofilme denso (Guilhabert e Kirkpatrick, 2005; Caserta *et al.*, 2010; Voegel *et al.*, 2010). Por outro lado, já foi mostrado que uma das adesinas não fimbriais (Xad1) é componente de vesículas de membrana externa (OMV) as quais desempenham um papel modulador da adesão de células de *X. fastidiosa* ao substrato tanto *in vitro* como *in planta* (Ionescu *et al.*, 2014). Além disso, a aquisição de *X. fastidiosa* pelo inseto-vetor durante sua alimentação em plantas infectadas é mais eficiente se as células estão mais aderentes (Almeida *et al.*, 2005).

Além das proteínas de adesão para a formação do biofilme de Xylella, os genomas desta bactéria também possuem o operon gum que contem todos os genes necessários para a produção e secreção de exopolissacarídeos que forma a goma fastidiana (Simpson et al., 2000; Chatterjee et al., 2008). A secreção deste material auxilia na estruturação da matriz extracelular do biofilme que, juntamente com o aglomerado de células e outras moléculas como proteínas e ácidos nucleicos, formam estruturas com densos agregados. O biofilme passa por cinco fases no seu ciclo, incluindo o início com suas células fundadoras, passando pela maturação, até a sua liberação e dispersão para colonizar novos vasos e locais (de Souza et al., 2004; Caserta et al., 2010). Além de proteger contra adversidade no ambiente como a ação de outras bactérias e ataques de mecanismos de defesa do hospedeiro, esta estrutura auxilia na captação de nutrientes do meio, além de manter presas células mortas que também servem para a sua nutrição (Parsek e Singh, 2003; Hall-Stoodley et al., 2004; Danhorn e Fuqua, 2007). A formação destes aglomerados de biofilmes densos resulta na oclusão parcial ou total de vasos do xilema, dificultando ou impedindo que o fluido xilemático possa atingir as partes superiores do hospedeiro. A carência hídrica e nutricional parece ser uma das principais causas dos sintomas das diversas doenças causadas por Xylella (Chatterjee et al., 2008). Entretanto, os sintomas também podem estar associados com a resposta sistêmica do hospedeiro frente à presença do patógeno (Gambetta et al., 2007).

Vale ressaltar, que a estrutura de biofilme formada na parte anterior do tubo digestivo do inseto é diferente do que é observado nos vasos do xilema. As células se posicionam lado a lado e se mantém em posição vertical, aderida ao substrato pelo polo que contém seus pili curtos (Newman *et al.*, 2004; Almeida e Purcell, 2006). Dessa forma, o tubo digestório do inseto não é ocluído e consequentemente não impede o fluxo de nutrientes. Durante a

alimentação, algumas células de *X. fastidiosa* podem se desprender e atingir as porções posteriores do tubo digestório e servirem como alimento ao inseto. No momento de uma nova alimentação de um vetor carregando *Xylella* em outro hospedeiro, foi proposto um mecanismo de inoculação por egestão/salivação de células do patógeno no interior de vasos do novo hospedeiro (Backus e Morgan, 2011; Backus *et al.*, 2015).

A produção de lipopolissacarídeos (LPS) também é conhecida como um importante fator de virulência em Xylella, atuando como uma barreira seletiva para a entrada de certas substâncias na célula (Clifford et al., 2013). LPS é uma macromolécula predominante na membrana externa de bactérias Gram-negativas e, por estar localizada na superfície celular, media as interações entre a célula bacteriana e o ambiente ao redor. Por este motivo, também é descrito no papel de desencadear de respostas de defesa basais dos hospedeiros. Sua molécula é composta de um componente conservado de oligossacarídeo com lipídeo A e uma porção variável de antígeno O. Tal antígeno mostrou-se importante para a adesão em superfícies, agregação célula-célula e maturação do biofilme, etapas essenciais para a virulência e para uma infecção bem-sucedida. Ensaios de infecção com mutantes para este antígeno mostraram uma diminuição da capacidade de colonização no hospedeiro vegetal (Clifford et al., 2013). Uma vez que o LPS atua na capacidade de adesão aos vasos do xilema, hipotetizou-se que também teria importância na adesão de Xylella ao tubo digestório do inseto vetor. Um mutante para o antígeno O foi utilizado, sendo confirmada a diminuição da adesão da bactéria à parede do tubo digestório (Rapicavoli et al., 2015). Dessa forma, foi confirmado o importante papel do LPS na virulência de X. fastidiosa em todas as etapas do seu ciclo de vida. Além disso, um estudo recente mostrou diferenças na sequência de um gene que codifica para uma das proteínas que compõe as moléculas de LPS em cepas de cafeeiro, sugerindo que os diferentes perfis de LPS de cepas distintas podem influenciar na interação com seu respectivo hospedeiro e vetor (Alencar et al., 2017).

Quando o biofilme atinge a sua fase final de desenvolvimento, ele passa a liberar células que irão se dispersar para eventualmente colonizar outros locais do hospedeiro. Para tal, as células precisam atravessar a parede do vaso do xilema em que se encontra para então acessar e colonizar o vaso adjacente. Na anatomia do xilema existem algumas regiões de comunicação entre os vasos, locais onde a composição de carboidratos é diferente do restante da parede do vaso, apresentando poros por onde passam o fluido xilemático e seus nutrientes (Perez-Donoso *et al.*, 2010). Entretanto, o diâmetro desses poros é inferior ao diâmetro da célula bacteriana, impedindo que ela atravesse livremente por eles. Como forma de contornar

Introdução

esta restrição, uma possibilidade é a degradação da parede do vaso. Interessantemente, o genoma de X. fastidiosa possui diversos genes codificadores de enzimas degradadoras de parede celular (Cell wall degrading enzymes, CWDE). Existe uma variedade delas, com especificidades e eficiências distintas, desde celulases, hemicelulases, endoglucanases, glicohidrolases, poligalacturonases, entre outras (Simpson et al., 2000). Todas as CWDEs são hipoteticamente secretadas para o meio extracelular pelo sistema de secreção de tipo II (Chatterjee et al., 2008). Além de serem capazes de degradar os carboidratos das regiões de comunicação entre vasos, algumas dessas proteínas são responsáveis por degradar os subprodutos das primeiras reações. Dessa forma, a bactéria poderia utilizar tais compostos como nutrientes (Killiny e Almeida, 2009). Acredita-se que este ponto também seja determinante para a especificidade de hospedeiro, uma vez que a composição de carboidratos dos pontos de comunicação de diferentes hospedeiros vegetais pode variar, além da variação do repertório de enzimas ativas das diferentes cepas. Isto poderia influenciar na capacidade e eficiência na colonização de uma cepa em um determinado hospedeiro. Diferentemente de todos os demais fitopatógenos já estudados, os genomas de X. fastidiosa não possuem genes para o sistema de secreção do tipo III, o qual é usado para injetar proteínas efetoras diretamente no interior das células dos hospedeiros (Simpson et al., 2000; Van Sluys et al., 2003; Chatterjee et al., 2008).

O hospedeiro vegetal, ao perceber ataque às suas estruturas internas, pode passar a expressar mecanismos de defesa contra o patógeno. Dentre eles estão a secreção de tilóis e géis que, embora ocluam aquele vaso já danificado e colonizado, impede que a bactéria se espalhe (Chatterjee *et al.*, 2008). Outra importante estratégia é a produção e liberação de espécies reativas de oxigênio, como superóxido e peróxido de hidrogênio, os quais são tóxicos para membranas celulares. Entretanto, a bactéria também é capaz de se defender contra estes ataques, uma vez que possui genes codificadores para catalases, superóxido dismutases, glutationa peroxidase, glutationa S-transferase, que atuam nas espécies reativas de oxigênio, degradando-as e impedindo seu efeito. Os reguladores de transcrição *Ohr* (Cussiol *et al.*, 2003) e *oxyR* (Toledo *et al.*, 2011) são ativados na presença destes compostos e modulam a expressão destes genes de resposta.

Além de ser capaz de se defender de ataques do hospedeiro, a *Xylella* apresenta vários genes codificando para um arsenal de toxinas. Dentre elas estão hemolisinas pertencentes à família RTX e diversas microcinas, como as colicinas V (Simpson *et al.*, 2000). As primeiras são teoricamente capazes de formar poros na membrana plasmática de células do hospedeiro

(Simpson *et al.*, 2000). As microcinas por sua vez poderiam ter função de combater possíveis bactérias que estejam competindo por espaço e nutrientes no xilema. Já foram relatadas espécies dos gêneros *Bacillus, Curtobacterium, Methylobacterium, Enterobacter* e *Pseudomonas* habitando os vasos do xilema de hospedeiros de *Xylella* (Araujo *et al.*, 2002; Lacava *et al.*, 2006). Além disso, foi realizado um estudo de metagenômica da parte anterior do tubo digestório de cigarrinhas, que revelou a presença dos gêneros *Wolbachia, Methylobacterium, Sphingomonas, Agrobacterium, Ralstonia, Caulobacter*, entre outros (Rogers e Backus, 2014). Tal microbiota poderia potencialmente interagir com *Xylella* no inseto vetor. A *Xylella* possui genes para o transporte das toxinas para o meio extracelular, além de proteínas de imunidade ou anti-toxinas (Merfa *et al.*, 2016; Santiago *et al.*, 2016). A hipótese é de que tais toxinas desempenham papel na virulência e patogenicidade da bactéria, e que o repertório de toxinas pode variar entre cepas.

Outros dois conjuntos de enzimas consideradas fatores de virulência e importantes para a colonização de Xylella são as proteases e lipases. Diversas proteases foram preditas pela anotação dos genomas desta bactéria, incluindo serina-proteases, metalo-proteases (algumas dependentes de zinco) e uma cisteína-protease (Simpson et al., 2000; Fedatto et al., 2006; Nogaroto et al., 2006). Esta última, denominada xilellaína, foi encontrada expressa em uma cepa patogênica de citros, mas não foi detectada em uma cepa não patogênica, sugerindo sua importância na patogênese (Nogaroto et al., 2006). Outras duas proteases de uma cepa isolada de videira foram caracterizadas (PD0218 e PD0956) (Chen et al., 2008; Gouran et al., 2016). Dois estudos definiram o proteoma extracelular de isolados de citros de X. fastidiosa e ambos encontraram proteases sendo secretadas (Smolka et al., 2003; Mendes et al., 2016). Tais proteases teriam papel em auxiliar as CWDEs na degradação das membranas de comunicação entre vasos, promovendo a movimentação e disseminação da bactéria na planta, além de fornecer subprodutos de degradação que podem ser usados para a nutrição do patógeno e influência na virulência in planta (Smolka et al., 2003; Fedatto et al., 2006; Gouran et al., 2016). As lipases também foram implicadas na virulência de Xylella em dois trabalhos recentes com o gene PD1703, o qual codifica para a lipase LesA. Primeiramente, o sequenciamento do genoma de uma cepa não-virulenta de Xylella, a EB92-1, evidenciou a ausência de LesA juntamente com outros genes de virulência (Zhang et al., 2011). Com a inserção do gene PD1703 em EB92-1, a cepa mostrou um aumento significativo dos sintomas característicos de PD em um ensaio de infecção em videiras (Zhang et al., 2015). Em uma análise do secretoma de uma cepa de videira, LesA foi abundantemente encontrada no

sobrenadante de cultura. Esta lipase foi associada à formação de biofilme, sendo encontrada em vesículas de membrana externa. Além disso, sugeriu-se que sua presença provocaria uma resposta de hipersensibilidade em videiras, e que mutantes deste gene seriam deficientes em virulência. Seu papel na virulência e sua presença em vesículas levou a hipótese de que esta lipase seria levada até as margens de folhas não colonizadas e desencadearia os sintomas iniciais de PD (Nascimento *et al.*, 2016).

O papel das vesículas de membrana externa no transporte de fatores de virulência a longas distâncias vem sendo descrito em trabalhos recentes. Além de carregar toxinas e outras moléculas em seu interior, elas também levam consigo proteínas localizadas na membrana externa de onde se originaram, como adesinas afimbriais (Voegel *et al.*, 2010; Ionescu *et al.*, 2014). A presença destas proteínas na membrana das vesículas lhes confere uma capacidade de adesão, sendo proposto que elas revestiriam as superfícies de contato da bactéria, como as paredes dos vasos do xilema, e modulando a adesão das células às paredes do xilema, e facilitando seu espalhamento sistêmico (Ionescu *et al.*, 2014).

Todos os fatores de virulência e eventos relacionados à sobrevivência e patogenicidade de Xylella estão relacionados direta ou indiretamente com um sistema de sinalização relacionado com a densidade celular da população bacteriana. A bactéria possui um conjunto de genes denominados rpf, os quais têm papel na síntese e secreção de uma pequena molécula lipídica chamada de DSF (Diffusible signaling factor), com função na percepção de quórum (quorum sensing) (Chatterjee et al., 2008; Ryan e Dow, 2011). O gene rpfF codifica uma enoil CoA-hidratase que juntamente com o gene rpfB, que codifica uma ligase de acil-CoA de cadeia longa, estão envolvidos na síntese e processamento de diferentes moléculas de DSF. As células sintetizam DSF e o secretam para o meio extracelular, sendo que a sua concentração é percebida por um sistema de dois componentes compostos pelas proteínas codificadas pelos genes rpfC e rpfG, o sensor e o regulador de resposta, respectivamente. O último possui um domínio HD-GYP com atividade de fosfodiesterase de c-di-GMP, ou seja, de degradação deste segundo mensageiro, o qual participa ativamente na regulação da expressão de diversos genes-alvo deste sistema (Chatterjee et al., 2008). Ao contrário do observado em outras espécies como Xanthomonas, DSF modula negativamente os níveis de c-di-GMP em Xylella (Chatterjee et al., 2008; Seshasayee et al., 2010; Ryan e Dow, 2011). Acompanhando o crescimento populacional, a concentração de DSF no meio irá variar e, consequentemente, gerará diferentes respostas nas células. A proteína sensora na membrana interna percebe as diferenças na concentração de DSF e desencadeia uma cascata

Introdução

de resposta com diferentes membros do sistema, culminando com a mudança da expressão de diversos genes nas células, incluindo aqueles relacionados com virulência. Baseando-se na concentração de DSF, a população bacteriana de Xylella apresentará duas fases distintas no interior dos vasos do xilema. A chamada fase de colonização do hospedeiro vegetal compreende o período de contato inicial, no qual as células foram inoculadas há pouco tempo e, consequentemente, estão em menor número. Ainda não há infecção sistêmica ou mesmo sintomas da doença. Devido às poucas células presentes, a concentração de DSF é baixa, o que implica na regulação positiva de genes relacionados à produção de pilus longo, usados na locomoção dentro dos vasos, e na alta expressão de genes codificando para as CWDE, as quais serão utilizadas para degradar as barreiras físicas existentes entre os vasos e promover a disseminação da bactéria dentro do hospedeiro. A partir do aumento do crescimento populacional bacteriano, o acúmulo de células promove o aumento da concentração de DSF, o qual sinaliza para uma mudança fenotípica, entrando na chamada fase de aquisição pelo inseto. Neste momento, as células passam a expressar genes relacionados com um fenótipo mais séssil, formando grandes aglomerados de células que acabam por ocluir vasos, o que aumenta as chances de bactérias serem adquiridas pelo inseto vetor e serem disseminadas para outros hospedeiros. Devido à oclusão de vasos, sintomas estão presentes nesta fase. Os genes regulados positivamente nesta fase incluem aqueles codificadores de adesinas fimbriais e afimbriais, além da maior produção e secreção de exopolissacarídeo ou goma fastidiana, os quais contribuem para a formação de um biofilme denso (Chatterjee et al., 2008).

Como evidenciado na resposta do sistema de percepção de quórum por DSF, a regulação precisa da expressão gênica é extremamente importante para que a célula possa sobreviver e lidar com os diferentes nichos nos quais se encontra. Ela tem que lidar com as diferenças na resposta à sua infecção nos mais variados hospedeiros e espécies de insetos vetores, os quais podem possuir diferentes e complexas microbiotas em seu interior. Com isso, compreender sistemas de regulação gênica e analisar diferentes respostas a determinadas condições ambientais que a célula encontra é fundamental para buscar o completo entendimento da biologia desse patógeno. Estudos de genômica comparativa entre cepas de *X. fastidiosa* são úteis para a descrição do repertório gênico de cada uma delas e relatar as diferenças encontradas, realizando inferências acerca de seus respectivos fenótipos e níveis de virulência. Entretanto, para que possamos analisar de fato quais genes estão sendo ativamente expressos em determinada condição, estudos de transcritômica são necessários e fornecem informações mais concretas da real participação de cada gene.

1.3.5. Análises transcritômicas em X. fastidiosa

Análises transcritômicas envolvem a descrição de todos os transcritos presentes em determinada condição em que o organismo se encontra. Diversos estudos com *Xylella* foram realizados em resposta a uma variedade de estímulos, condições de cultivo e fases do crescimento e desenvolvimento da população bacteriana.

A grande maioria dos estudos de transcritômica com esta bactéria foi realizado utilizando a tecnologia de microarranjos de DNA. O primeiro estudo analisou diferenças na expressão dos genes em diferentes estados de crescimento da cepa 9a5c, isolada de citros. Foram analisadas populações bacterianas com diferentes números de passagens no cultivo *in vitro*: uma recentemente extraída do hospedeiro, dita como 1^{a.} passagem e outra após 46 passagens. O objetivo seria analisar se o constante crescimento *in vitro* causaria alterações na expressão de genes de virulência. Ensaios de infecção de plantas confirmaram que as bactérias recém-extraídas tiveram uma colonização mais eficiente. A análise de expressão gênica mostrou genes associados com adesão, adaptação ao hospedeiro e virulência, como lipases e colicina V, induzidos na população de 1^{a.} passagem (de Souza *et al.*, 2003). Em outro estudo, foram analisados os genes diferencialmente expressos de células crescendo em biofilme comparado com células planctônicas. Células no biofilme apresentaram um aumento da expressão de genes relacionados com funções metabólicas de manutenção da célula, além de proteínas relacionadas à adesão a superfícies (de Souza *et al.*, 2004).

Nos seus hábitats naturais, como os vasos do xilema de diferentes hospedeiros vegetais e o tubo digestório de diversas espécies de insetos-vetor, a *Xylella* está sujeita uma variedade de fatores de estresse. Dentre eles se enquadram mudanças nas condições ambientais, nas concentrações de nutrientes, além da presença de compostos anti-microbianos e organismos competidores. Koide e colaboradores analisaram a resposta da bactéria ao aumento da temperatura. Além dos genes relacionados a chaperonas e outros genes relacionados a mecanismos de controle de qualidade de proteínas, foi observado um aumento da expressão de genes associados à virulência, como hemaglutininas, hemolisinas, enzimas degradadoras de xilana, entre outros. Genes relacionados à síntese de pilus, respiração aeróbia e síntese proteica foram reprimidos nesta condição de estresse. O estudo ainda sugere um *regulon* de choque térmico (*heat shock*) após analisarem regiões promotoras de genes com expressão diferencial (Koide *et al.*, 2006). Uma condição de carência nutricional foi avaliada

em um estudo com limitadas concentrações de nitrogênio, revelando genes diferencialmente expressos com funções relacionadas ao transporte, assimilação de nitrogênio, biossíntese de aminoácidos e regulação transcricional (da Silva Neto et al., 2010). Outros dois trabalhos avaliaram a resposta da bactéria frente a moléculas com ação antimicrobiana. Fogaça e colaboradores avaliaram o efeito da gomesina, um peptídeo antimicrobiano extraído de aracnídeo, observando um aumento da expressão de genes relacionados com a produção de biofilme, importante mecanismo de defesa bacteriano (Fogaca et al., 2010). As ações do cobre e do antibiótico tetraciclina em Xylella também foram avaliadas em outro estudo com microarranjos. Novamente, genes envolvidos com a proteção das células em biofilme tiveram sua expressão aumentada. Os autores ainda relataram, pela primeira vez em Xylella, a formação de células persistentes no biofilme em resposta a ambos os estresses, sugerindo uma resposta de tolerância a moléculas antimicrobianas (Muranaka et al., 2012). Tais moléculas também podem ser sintetizadas por organismos competidores que possam co-habitar nichos em comum com a Xylella. Dourado e colaboradores analisaram a resposta da Xylella à interação com uma bactéria endofítica de citros, Methylobacterium mesophilicum, em um cocultivo in vitro. Além disso, esta bactéria já foi encontrada sendo transportada pelo mesmo inseto vetor que pode transmitir a Xylella. Os resultados mostraram um aumento da expressão de genes relacionados com produção de energia, estresse, transporte e motilidade, além da repressão de genes associados ao crescimento (Dourado et al., 2015).

A *Xylella* possui uma variedade de genes codificadores de fatores de transcrição e fatores σ^{E} e σ^{54} foram analisados quanto à regulação de seus respectivos genes alvo. No primeiro caso, a condição de choque por calor foi utilizada para definir o *regulon* deste fator sigma, identificando genes codificadores de enzimas envolvidas no enovelamento e degradação de proteínas, transdução de sinal e sistemas de restrição/modificação de DNA (da Silva Neto *et al.*, 2007). O papel do fator σ^{54} foi investigado a partir da comparação com o transcritoma de uma cepa mutante deste gene, evidenciando a regulação de genes envolvidos com a biogênese dos pili curto e longo e na formação de biofilme (da Silva Neto *et al.*, 2008). A mesma estratégia de gerar mutantes para analisar a expressão gênica global regulada por determinado gene em comparação com a cepa selvagem, foi usada para identificar o *regulon* dos fatores de transcrição AlgU e GacA de *Xylella*. Enquanto AlgU regula genes relacionados à agregação celular, adesão, formação de biofilme e virulência (Shi *et al.*, 2007), GacA regula genes que também contribuem para a adesão e formação de biolfime, além de outros com papel na adaptação e tolerância da bactéria a estresses ambientais (Shi *et al.*, 2009).

Sistemas de sinalização de dois componentes e de percepção de quórum também foram investigados por microarranjos. A análise do transcritoma da cepa mutante do gene xhpT (codifica um regulador de reposta) apontou para alteração na adesão a superfícies, agregação célula-célula, produção de exopolissacarídeos e virulência em videiras (Voegel *et al.*, 2013). O sistema de percepção de quórum já descrito, baseado nos genes rpf e na produção de DSF, foi analisado a partir da construção de uma cepa mutante para o gene rpfF, principal atuante na síntese da molécula sinalizadora DSF. O *regulon* dependente de rpfF envolve genes relacionados a adesão e formação de biofilme, regulando positivamente a expressão de hemaglutininas e hemolisinas, e negativamente a maioria dos genes de síntese dos pili longo e curto, de goma fastidiana e de colicinas V (Wang *et al.*, 2012).

Interessantemente, além da resposta da bactéria sob diversas condições, estudos têm focado na resposta do hospedeiro vegetal a infecção por *Xylella*. A análise do transcritoma, por meio de microarranjos, de *Arabidopsis thaliana* inoculada com *X. fastidiosa*, revelou aumento da expressão de genes relacionados em limitar o dano por estresse oxidativo enquanto que outros genes normalmente responsivos a infecções por patógenos ou estresse biótico não tiveram expressão alterada. Os autores sugerem que tal resposta da planta estaria mais relacionada ao déficit hídrico induzido pelo bloqueio dos vasos do xilema do que à própria infecção bacteriana (Rogers, 2012).

Em um estudo de RNA-Seq, foi analisada a resposta de indivíduos de *C. reticulata* (resistente à colonização por *X. fastidiosa*) um dia após a infecção por *Xylella*. Dentre os transcritos induzidos, foram identificados genes codificadores de proteínas do sistema imune inato do hospedeiro, genes envolvidos com o metabolismo secundário, biossíntese e modificação da parede celular, síntese de ácido abscísico, ácido jasmônico e auxina (Rodrigues *et al.*, 2013).

A resposta de oliveiras à infecção por *X. fastidiosa* também foi investigada utilizando RNA-Seq (Giampetruzzi *et al.*, 2016). O transcritoma de cultivares que exibem sintomas mais brandos, como o cv. *Leccino*, foi comparado ao transcritoma de cultivares suscetíveis a OQDS, como o cv. *Ogliarola salentina*, após infecção por *Xylella*. As análises de genes diferencialmente expressos mostraram que a presença do patógeno aciona uma resposta envolvendo o remodelamento de proteínas da parede celular em ambos os cultivares. A regulação positiva de genes codificando quinases do tipo receptor (RLR) e proteínas do tipo receptor (RLP) foi vista no cv. *Leccino*, mas não no cv. *Ogliarola salentina*. Os autores

concluíram que as diferentes respostas transcricionais observadas podem determinar a menor concentração do patógeno no cv. *Leccino* e sua relativa tolerância à *Xylella*.

Mudanças nas concentrações e disponibilidades de nutrientes no meio é fator importante que influencia a regulação gênica. Alguns estudos analisaram o panorama global da transcrição na presença de diferentes micro e macronutrientes e em diferentes concentrações. Zaini e colaboradores analisaram a resposta de *Xylella* no crescimento em diferentes níveis de concentração e disponibilidade de ferro. Genes envolvidos com funções regulatórias, patogenicidade e estrutura celular foram regulados nas condições analisadas, incluindo genes para colicinas V e biossíntese de pilus (Zaini *et al.*, 2008). Outro micronutriente analisado foi o cálcio, no primeiro estudo no qual transcritomas de *X. fastidiosa* foram analisados utilizando a tecnologia de RNA-Seq. As células foram cultivadas em meio suplementado com cálcio e células do biofilme foram coletadas para a análise do transcritoma. Os resultados mostraram regulação positiva em resposta ao cálcio de genes relacionados com adesão, motilidade, síntese de exopolissacarídeo, formação de biofilme, síntese de peptidoglicano, funções regulatórias, homeostase de ferro e fagos (Parker *et al.*, 2016).

Diferenças no conteúdo de macronutrientes do ambiente também afetam a expressão gênica global de Xylella, sendo que alguns estudos de transcritômica foram realizados com diferentes fontes de carbono e energia. Pashalidis e colaboradores analisaram o perfil transcricional de Xylella cultivada em meio de cultivo com diferentes concentrações de glicose. Os resultados indicaram uma indução da transcrição de genes de colicinas V e precursores de pilus no meio com alta concentração de glicose (Pashalidis et al., 2005). Outros trabalhos buscaram analisar fontes nutricionais que a bactéria poderia diretamente encontrar no ambiente. Para se locomover no interior dos vasos do xilema de hospedeiros vegetais e entre eles, as já mencionadas CWDE são secretadas e atuam na degradação da parede das membranas de comunicação entre vasos. Dessa forma, subprodutos de degradação são gerados e podem modificar a expressão de genes para a sua utilização. Em um estudo com microarranjos, foram analisados os efeitos da adição de pectina e glucano ao meio de cultivo, os quais provocaram mudanças no fenótipo e no perfil de expressão gênica da população bacteriana. A presença de pectina tornou as células mais adesivas, corroborando o aumento da expressão de genes relacionados a hemaglutininas e produção de goma em células cultivadas nos meios com pectina e glucano, além de outras mudanças de expressão em genes importantes para a patogenicidade. Devido a este fenótipo mais agregativo, a disponibilidade para a aquisição de bactérias pelo inseto vetor foi aumentada, elevando a taxa de transmissão (Killiny e Almeida, 2009). Assim que a *Xylella* é adquirida pelo inseto, ela encontra um ambiente completamente distinto e necessita novamente adequar a expressão de seus genes. O carboidrato predominante neste ambiente é a quitina, que compõe o exoesqueleto que reveste o tubo digestório do vetor, tendo sido evidenciado o seu uso como fonte de nutrientes para *X. fastidiosa*, além de induzir aumento na formação de biofilme bacteriano no inseto vetor (Killiny *et al.*, 2010).

Os estudos in vitro para analisar o efeito pontual de determinado carboidrato são importantes para fornecer dados específicos da regulação da expressão gênica do patógeno. Entretanto, sabe-se que a composição do fluido xilemático é muito mais complexa e que outras moléculas podem influenciar na resposta da bactéria. Zaini e colaboradores mostraram ser possível sustentar uma cultura estável de Xylella em seiva do xilema de videiras, resultando em um aumento da taxa de crescimento e indução da adesão e formação de biofilme (Zaini et al., 2009). Dessa forma, estudos que analisam as mudanças no perfil de expressão gênica em diferentes meios com composições nutricionais distintas, podem fornecer respostas mais próximas do que a bactéria possa encontrar no seu ambiente natural. Um estudo realizou a comparação da expressão gênica de Xylella em cultivo em um meio complexo (BCYE) e um meio definido (XDM₂). Este último foi desenvolvido para melhorar o crescimento de Xylella, tendo sido baseado nas necessidades metabólicas do patógeno reveladas pela análise de seu genoma (Lemos et al., 2003). Foram observados genes diferencialmente expressos com diferentes funções, como energia, metabolismo de proteína, aminoácido e nucleotídeo, transporte, toxinas, entre outros (Travensolo et al., 2009). Um segundo estudo com diferentes meios de cultura analisou o perfil de expressão gênica global de Xylella cultivada em um meio com características nutricionais semelhantes às do fluido xilemático de videira (3G10-R) (Leite et al., 2004), comparando com um meio complexo usado para cultivo rotineiro desta bactéria (PW) (Davis et al., 1981). Os resultados mostraram que os genes regulados positivamente para o meio 3G10R podem estar envolvidos com a colonização do hospedeiro, virulência e competição ambiental, enquanto que nas células cultivadas no meio PW foi observado um aumento da respiração aeróbica e das taxas de crescimento bacteriano (Ciraulo et al., 2010).

Em resumo, muitos trabalhos de análise de transcritomas de *Xylella fastidiosa* já foram realizados abordando os mais variados objetivos. Entretanto, quase todos utilizaram a tecnologia de microarranjos, com exceção de apenas um único estudo (Parker *et al.*, 2016) no

qual a estratégia de RNA-Seq foi utilizada. Como já descrito anteriormente, esta técnica tem diversas vantagens em relação aos microarranjos, possibilitando uma avaliação muito mais aprofundada da expressão de todos os transcritos de uma célula, além de diversas outras análises que podem ser geradas com os dados de sequenciamento. Alguns estudos foram realizados utilizando RNA-Seq como estratégia para análise da expressão gênica global em fitopatógenos, como nas espécies *Xanthomonas campestris* pv. *vesicatoria* (Schmidtke *et al.*, 2011) e *Pseudomonas syringae* pv. *tomato* (Filiatrault *et al.*, 2010). Além disso, trabalhos subsequentes utilizaram os mesmos organismos para outros fins que a técnica de RNA-Seq permite, como a definição dos sítios de início de transcrição (TSS) (Filiatrault *et al.*, 2011) e a identificação de pequenos RNAs não codificantes (Schmidtke *et al.*, 2013).
2. Objetivos

O objetivo central desta tese foi sequenciar, analisar e comparar os transcritomas de diferentes cepas de *Xylella fastidiosa*, no início e fim da fase exponencial de crescimento em meio rico PWG e em PIM6, um meio mínimo que mimetiza a composição da seiva do xilema. A premissa para realização deste trabalho é que a descrição do repertório completo de genes diferencialmente expressos nestas condições experimentais, além de aprofundar o conhecimento das respostas transcricionais de *X. fastidiosa* a variações nas condições de cultivo *in vitro*, poderá evidenciar fatores de virulência e de patogenicidade expressos nestas condições.

Os transcritomas de *X. fastidiosa*, cepas 9a5c (agente causal da CVC) e Temecula1 (agente causal da PD), frente a uma variedade de estímulos e condições de cultivo, foram investigados em trabalhos anteriores (ver item 1.3.5) por meio de microarranjos de DNA. Até o momento, apenas um estudo publicado (Parker *et al.*, 2016) utilizou RNA-Seq para descrever o transcritoma da cepa Temecula1 em meio suplementado com cálcio, revelando o aumento da expressão de genes relacionados a adesão e motilidade que são mecanismos relevantes na virulência e patogenicidade de *X. fastidiosa*. A metodologia de RNA-Seq substitui plenamente a técnica de microarranjos, apresentando vantagens como menor custo, menor complexidade experimental e, principalmente, maior acurácia na identificação e quantificação da abundância de transcritos (Malone e Oliver, 2011; Zhao *et al.*, 2014).

Nesta tese, estão descritos, analisados e comparados os transcritomas de oito cepas de *X. fastidiosa* cultivadas em meio rico e/ou meio mínimo. Esses são transcritomas enriquecidos em sequências de mRNAs obtidos por meio de RNA-Seq. Também está descrito um transcritoma enriquecido em sRNAs da cepa 9a5c.

3. Procedimentos experimentais

3.1. Manutenção e cultivo de Xylella fastidiosa

As cepas utilizadas neste trabalho estão listadas na Tabela 2, bem como seus respectivos locais de origem e plantas hospedeiras de que foram isoladas.

Сера	Hospedeiro	Origem*	Fonte	Referência
9a5c	<i>Citrus sinensis</i> cv. Natal	Macaubal/SP/BRA	Fundecitrus	(Li <i>et al</i> ., 1999)
J1a12	<i>Citrus sinensis</i> cv. Pêra	Jales/SP/BRA	Fundecitrus	(Monteiro <i>et al.</i> , 2001)
U24d	<i>Citrus sinensis</i> cv. Baianinha Piracicaba	Ubarana/SP/BRA	APTA Citros	(Santana, 2012)
Fb7	<i>Citrus sinensis</i> cv. Valencia	Bella Vista/Corrientes/ARG	APTA Citros	(da Silva <i>et al.</i> , 2007)
3124	Coffea arabica	Matão/SP/BRA	APTA Citros	(Almeida <i>et al.</i> , 2008)
Hib4	Hibiscus rosa- sinensis	Campinas/SP/BRA	APTA Citros	(Pierry, 2012)
Pr8x	Prunus salicina	Jarinu/SP/BRA	APTA Citros	(Pierry, 2012)
Temecula1	Vitis vinífera	Temecula/CA/EUA	University of California, Berkeley	(Van Sluys <i>et</i> <i>al.</i> , 2003)

Tabela 2. Cepas de X. fastidiosa utilizadas e seus respectivos hospedeiros de origem.

*As siglas correspondem aos estados e países nos quais as cepas foram isoladas: SP, Estado de São Paulo; BRA, Brasil; ARG, Argentina; CA, Estado da Califórnia; EUA, Estados Unidos da América.

Estoques de células (diversas passagens) foram mantidos em meio PWG contendo 50% de glicerol em freezer a -80°C. A manutenção das culturas foi feita em placas de meio PWG-ágar com repiques semanais e incubação a 28°C, sendo o número de passagens controlado. Após 20 passagens, novo cultivo foi iniciado a partir de estoques congelados a - 80°C.

O meio PW (Davis *et al.*, 1981) contendo glicose (meio PWG) é constituído de fitona peptona 4g/L; peptona de caseína com digestão tríptica 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%. Em alguns experimentos, foi utilizado o meio PIM-6 que é composto de HEPES 10mM pH 6,5; MgSO₄ 1mM; CaCl₂ 3mM; D-glicose 0,1mM; tartarato de sódio 0,1mM, malato de sódio 0,2mM; citrato de sódio 1mM; 0,05 g/L K₂HPO₄; 0,03 g/L KH₂PO₄; L-glutamina 5 mM; 0,025% solução de micronutrientes ATCC 13061; 0,02% *soytone* e 0,04% triptona (Michelle Igo, University of California, Davis, comunicação pessoal).

Para o cultivo em meio líquido, as bactérias, após 7 a 10 dias de cultivo em meio sólido PWG-ágar (PWG contendo ágar 1,5%), foram transferidas para frascos contendo de 50 a 100 mL de meio PWG, os quais foram incubados a 28°C com 170 rpm de agitação por até 15 dias.

Para realização de curvas de crescimento, o cultivo foi realizado em tubos cônicos de 50 mL contendo 10 mL do meio de cultivo PWG ou PIM6, a 28°C com agitação de 170 rpm. O cultivo foi iniciado a partir de um inóculo de bactérias cultivadas por 10 dias em meio PWG (50 mL), previamente submetidas a centrifugação (4000 xg, 10 minutos) e ressuspensas em um volume menor (~10 mL) de meio fresco. O cultivo foi iniciado com DO_{600nm} 0,05 em triplicata para cada dia de medição a qual for realizada após 1, 3, 7, 10, 15 e 21 dias. Em alguns ensaios em meio PIM6 o cultivo foi iniciado com DO_{600nm} 0,25, 0,35 e 0,45 e as medições foram realizadas após 2, 5 e 7 dias. Imediatamente antes das medições da DO_{600nm} , a cultura foi agitada vigorosamente para adequada suspensão de células eventualmente aderidas a parede do tubo (Pashalidis *et al.*, 2005; Fogaca *et al.*, 2010).

3.2. Extração de RNA total

As curvas de crescimento de cada uma das cepas nos meios de cultivo PWG ou PIM6 guiaram os tempos de coleta de células no início e no final da fase exponencial do crescimento bacteriano para extração de RNA total. Os cultivos foram iniciados, a partir de inóculos provenientes do cultivo por 10 dias em PWG-ágar, na DO_{600nm} 0,05, quando realizados em meio PWG, e na DO_{600nm} 0,25 quando em PIM6, sendo que nesse caso as células foram incubadas previamente em PIM6 por 1 dia na DO_{600nm} 0,3 para "aclimatação" ao meio mínimo antes do início do cultivo destinado às extrações de RNA total. Foram obtidas três réplicas biológicas do cultivo das cepas 9a5c e Temecula1 em meio PWG e

Nos tempos de cultivo determinados, as células (50 ou 100 mL de cultura dependendo da cepa e do meio de cultivo) foram coletadas por centrifugação (4000 xg, 10 minutos) e utilizadas imediatamente para extração do RNA total. No caso de células provenientes de cultivo em meio PIM6, o precipitado celular foi incubado por 60 minutos a temperatura ambiente com 2 mL do reagente *LifeGuard (Mobio)*, seguindo-se nova centrifugação para completa remoção do reagente antes do início do protocolo de extração de RNA. Esta etapa se

mostrou essencial para preservar a integridade das moléculas de RNA de células provenientes de meio PIM6, mas sendo dispensável para células de meio PWG.

Para purificação de RNA total de X. fastidiosa foi padronizado um protocolo que combina os protocolos do reagente Trizol (Invitrogen) e o PureLink RNA Mini Kit (Ambion). O reagente Trizol (1 a 3 mL, dependendo da massa celular) é utilizado para a ressuspensão do precipitado celular, e o protocolo do fabricante é realizado até a obtenção da suspensão de RNA após adição de isopropanol e etanol. Nesta etapa do protocolo, RNA em suspensão é então aplicado a coluna de purificação de RNA do kit PureLink. O protocolo segue conforme especificado pelo fabricante, com diversas lavagens com tampões, sendo a eluição final realizada com água-DEPC (água pré-tratada com DEPC, dicarbonato de dietila), adicionada ao centro do filtro. A amostra de RNA purificada foi armazenada em freezer a -80°C. Alíquotas de 1,5 µL foram separadas para quantificação e análise da qualidade em espectrofotômetro NanoDrop e avaliação da integridade por eletroforese capilar no equipamento 2100 BioAnalyzer utilizando-se o RNA 6000 Nano kit (Agilent Technologies). A integridade do RNA é avaliada pelo software do próprio equipamento, o qual calcula os valores do RIN (RNA Integrity Number) (Schroeder et al., 2006) baseado nos tamanhos dos picos observados por todo o perfil eletroforético, incluindo os correspondentes ao rRNAs 16S e 23S. Definimos que amostras com RIN ~ 8,0 eram adequadas para prosseguimento das etapas seguintes.

3.3. Purificação de RNA total com DNase

Para a purificação adicional do RNA total foi utilizada a etapa de tratamento com a DNase do "*Illustra RNASpin Mini RNA isolation Kit*" (*GE Healthcare*) com modificações no protocolo do fabricante. Foram utilizados até 15µg de amostra de RNA total, preparado como descrito no item anterior e diluídos em água-DEPC para um volume final de 100µL, e o dobro do volume da solução de DNase (20μ L) para o tratamento da amostra. A reação foi incubada a temperatura ambiente por 50 minutos e, após diversas lavagens com tampões, a eluição foi realizada com 50 µL de água-DEPC pré-aquecida a 50°C. Para melhorar a eficiência do tratamento, o eluato foi passado novamente pela coluna, e o procedimento de eluição foi repetido. Alíquotas de 1,5 µL foram separadas para análise da integridade por eletroforese capilar no equipamento 2100 BioAnalyzer (Agilent Technologies) como descrito no item 2.2. A eficiência do tratamento com DNase foi verificada por PCR como descrito no item 2.4.

3.4. Verificação da eficiência do tratamento com DNase

A ausência de contaminação com DNA nas preparações de RNA, após o tratamento com DNAse, foi verificada por meio de PCR utilizando-se o par de oligonucleotídeos específico para *X. fastidiosa* (Pooler e Hartung, 1995) indicados abaixo e que geram um amplicon de 500 pb:

CVC-1 5'-AGATGAAAACAATCATGCAAA-3'272-2-int 5'-GCCGCTTCGGAGAGCATTCCT-3'

Amostras de DNA genômico de *X. fastidiosa* foram usadas como controle positivo do ensaio de PCR. O controle negativo da reação de PCR foi realizado substituindo-se a quantidade de RNA ou de DNA por água-DEPC. As reações de PCR (dNTPs 0,2mM; MgCl₂1,5 mM; primer CVC-1 e primer 272-2 0,2 µM; tampão da enzima diluído 1x e Taq DNA Polimerase 2,5U; 100 ng de amostra de RNA ou DNA) foram incubadas em um termociclador *GeneAmp PCR System 9700 (Applied Biosystems)* a 95°C por 5 minutos seguindo-se 40 ciclos de 95°C por 45 segundos, 60°C por 30 segundos e 72°C por 2 minutos. Ao final, as reações foram incubadas a 72°C por 10 minutos e armazenadas a 4°C até análise por eletroforese em gel de agarose.

A eletroforese dos amplicons foi realizada em gel de agarose 1% em tampão TBE (Tris-base 89 mM, ácido bórico 89 mM, EDTA 2 mM pH 8,0) contendo 0,5 mg/mL de brometo de etídio, a 100V até a saída do corante indicador. Aproximadamente 8 µL da PCR foram aplicados misturados com pelo menos 20% do volume total de tampão de amostra (corante azul de bromofenol 0,25% (p/v); sacarose 40% (p/v); Tris-base 10 mM pH 7,5; EDTA 1 mM). Ao final da corrida, o gel foi documentado em um fotodocumentador com luz UV e o tamanho dos amplicons foi estimado com base na comparação à mobilidade de fragmentos de DNA com tamanhos já conhecidos (*GeneRuler 1kb DNA Ladder, Fermentas*).

3.5. Preparação de bibliotecas de cDNA para sequenciamento (RNA-Seq)

3.5.1. Quantificação de RNA por método fluorimétrico

Antes de iniciar a preparação das bibliotecas de cDNA para sequenciamento, foram realizadas quantificações mais precisas das amostras de RNA total purificado utilizando o *Quant-iT RiboGreen RNA Assay kit (ThermoFisher Scientific)* segundo o protocolo do fabricante. As amostras de RNA total purificado foram diluídas para concentrações dentro dos limites de detecção da curva padrão que é gerada a partir de uma amostra de RNA de concentração conhecida fornecida pelo *kit*. A leitura da fluorescência foi realizada em um espectrofluorímetro (excitação em 480 nm e a intensidade de emissão medida a 520 nm). Após a medição, os valores de concentração (η g) das amostras de RNA total purificadas foram extrapolados a partir da curva padrão.

3.5.2. Depleção de rRNAs

As amostras de RNA total quantificadas como descrito acima foram submetidas ao protocolo de depleção de rRNAs, visando enriquecer a preparação em mRNAs. O procedimento está baseado na remoção seletiva de moléculas de rRNAs pela hibridização com uma mistura de oligonucleotídeos de sequências conservadas em rRNAs bacterianos. Para tal foi utilizado o *Ribo-Zero Magnetic Kit (Gram-negative bacteria) (Illumina)*, conforme as instruções do fabricante, mas com algumas modificações na etapa final. O protocolo recomenda a utilização de 1-5 μ g de RNA total purificado, e sempre que possível foi utilizada a quantidade máxima de amostra. Ao final da depleção, 90 μ L do sobrenadante foi transferido para um tubo de 1,5mL e submetido a precipitação com etanol *overnight* a - 20°C. Esta alteração no protocolo do fabricante visou aumentar a eficiência da precipitação do RNA depletado. O precipitado de RNA foi seco e em seguida ressuspendido diretamente na solução do *kit* de preparação da biblioteca de cDNA, permitindo o acoplamento imediato dos dois protocolos.

Um pequeno volume da amostra antes da etapa de precipitação com etanol foi separado para posterior análise da eficiência da depleção de rRNAs por eletroforese capilar no equipamento 2100 BioAnalyzer utilizando-se o RNA 6000 Pico kit (Agilent Technologies).

3.5.3. Preparação de bibliotecas de cDNA

Após a etapa de depleção de rRNAs, foi iniciado imediatamente o protocolo de preparação da biblioteca de cDNA utilizando o TruSeq RNA sample preparation kit v2 (Illumina) com uma modificação para acoplar o protocolo de depleção de rRNA. Para tal, o precipitado de RNA depletado seco foi ressuspendido em 18µL da solução Elute, Prime, Fragment Mix do kit TruSeq RNA. A partir dessa etapa, o protocolo do kit TruSeq RNA foi seguido conforme estabelecido pelo fabricante. Para a síntese da primeira fita de cDNA foi utilizada a transcriptase reversa ImProm II (Promega). Todos os procedimentos foram realizados em placas de 96 poços no termociclador GeneAmp PCR System 9700 (Applied Biosystems). Em diversas etapas, o protocolo requer a purificação da amostra para eliminação de adaptadores, oligonucleotídeos ou de fragmentos de cDNA muito pequenos. Tais purificações foram realizadas com o uso de Agencourt AMPure XP beads (Beckman Coulter Life Sciences). Ao final do protocolo, foram coletados 30 µL da biblioteca final, transferidos para um tubo de 1,5 mL e este armazenado a -20°C. Seguiu-se a análise de distribuição dos fragmentos da biblioteca por eletroforese capilar no equipamento 2100 BioAnalyzer utilizando-se o High Sensitivity DNA kit (Agilent Technologies) e quantificação por qPCR absoluto como descrito no item 3.5.5. Em uma das etapas do protocolo, cada biblioteca de cDNA é identificada pela ligação de adaptadores de sequências conhecidas e distintas (indexes ou barcodes) às duas extremidades dos fragmentos de cDNA, o que possibilita o sequenciamento simultâneo de mais de uma biblioteca de cDNA em uma mesma corrida.

3.5.4. Preparação de bibliotecas de pequenos RNAs (sRNAs)

Para a preparação de bibliotecas de pequenos RNAs (sRNAs) de *X. fastidiosa* foi um adaptado protocolo empregando o *NEBNext Small RNA Library Prep Set for Illumina (New England Biolabs)* que é utilizado na preparação de bibliotecas de microRNAs eucarióticos. Este *kit* exige que as moléculas de RNA estejam monofosforiladas na extremidade 5' e hidroxiladas na extremidade 3'. Os sRNAs bacterianos não processados apresentam extremidade 5' trifosforilada. Dessa forma, foram realizadas duas etapas anteriores ao primeiro passo do *kit*, nas quais foram utilizadas as enzimas *RNA 5' Pyrophosphohydrolase* (RppH) (*New England Biolabs*) e *T4 Polynucleotide Kinase* (T4 PNK) (*New England*

Biolabs). A T4 PNK adicionaria um monofosfato na extremidade 5' de moléculas nãofosforiladas, enquanto a RppH removeria um pirofosfato da extremidade 5' de moléculas trifosforiladas, deixando um monofosfato em 5'. Assim, amostras de RNA total foram tratadas com a RppH (25 U) em uma reação com volume final de 50 uL e 500 ng de RNA. Foi realizada uma incubação por 30 min a 37°C. Após o término dessa reação, foi adicionada a T4 PNK (10 U) e ATP 1 mM, seguindo-se incubação por 30 min a 37°C. O tampão utilizado em ambas as reações foi o *T4 PNK buffer*. Ao final, as enzimas foram inativadas pela adição de EDTA 500 mM pH8,0 e incubação por 20 min a 65°C, seguindo-se purificação do RNA total com *MinElute PCR Purification kit (Qiagen)*. O protocolo de purificação foi seguido conforme sugerido pelo fabricante, sendo que foi utilizado acetato de sódio 3M, pH 5,0 para corrigir o pH após adição do tampão inicial do *kit*. Ao final, a amostra de RNA foi eluída em 10 μL de tampão Tris-HCl 10 mM, pH 8,5.

Em seguida, foi iniciado o protocolo de preparação de bibliotecas de sRNAs, seguindo conforme sugerido pelo fabricante. O RNA total tratado com RppH e T4 PNK foi utilizado para ligação de adaptadores à extremidade 3', que pode ser realizada de duas formas: incubar a reação a 25°C por 1 hora ou incubar a reação a 16°C por 18 horas, sendo que esta última condição pode aumentar a eficiência na ligação dos adaptadores a RNA metilados. Dessa forma, a amostra foi submetida as duas condições paralelamente e as demais etapas do protocolo foram realizadas como recomendado pelo fabricante, exceto pela utilização de 15 ciclos na etapa de amplificação dos fragmentos por PCR. Após o término do protocolo, as bibliotecas foram novamente purificadas utilizando o *MinElute PCR Purification kit (Qiagen)* e a seleção por tamanho dos fragmentos foi realizada com as *Agencourt AMPure XP beads* (*Beckman Coulter Life Sciences*). Ao final, a amostra purificada e selecionada foi eluída em 15 µL de água. Seguiu-se a análise de distribuição dos fragmentos da biblioteca por eletroforese capilar no equipamento *2100 BioAnalyzer* utilizando-se o *High Sensitivity DNA kit (Agilent Technologies*) e quantificação por qPCR absoluto com descrito no item 3.5.5.

3.5.5. Quantificação de bibliotecas de cDNA por qPCR absoluto

A quantificação das bibliotecas de cDNA foi realizada por meio de qPCR (PCR quantitativo) absoluto utilizando o *Kapa Library Quantification kit (Kapa Biosystems)* para plataformas de sequenciamento *Illumina*. O kit dispõe de amostras de DNA com concentrações conhecidas para gerar uma curva padrão que varia entre 20 pM a 0,0002 pM.

Antes do ensaio de qPCR a concentração das bibliotecas de cDNA foi estimada baseando-se na quantificação por espectrofotômetro *NanoDrop*, tamanho médio dos fragmentos da biblioteca obtido por eletroforese capilar e uma tabela de conversão ng/µL em nM (protocolo *Nextera DNA Sample Preparation Guide* da *Illumina*), e devidamente diluídas para concentrações nos limites da curva padrão. Os ensaios de qPCR foram realizados no equipamento *7500 Real Time PCR System (Applied Biosystems)* a 95°C por 5 minutos, seguindo-se 35 ciclos de 95°C por 30 segundos e 60°C por 45 segundos. A análise foi realizada em com *software* do próprio equipamento que gera curvas de amplificação para cada amostra e fornece o cálculo do CT (*Cycle Threshold*). Os valores médios de CT foram utilizados para calcular a concentração baseado no tamanho médio dos fragmentos de cada biblioteca.

3.6. Sequenciamento de DNA no equipamento MiSeq (Illumina)

O sequenciamento das bibliotecas de cDNA foi realizado no equipamento *MiSeq* (*Illumina*) com o *MiSeq Reagent Kit v2* de 500 ciclos com a estratégia de sequenciamento *Paired-End* que resulta em sequências (*reads*) de 250 nucleotídeos para cada extremidade de um fragmento sequenciado. No máximo 8 bibliotecas de cDNA com índices distintos e compatíveis foram sequenciadas em uma mesma corrida. Antes de iniciar o sequenciamento, a concentração das bibliotecas a serem sequenciadas foi ajustada para 4 nM e o mesmo volume de cada uma delas foi misturado em um único tubo, seguindo-se adição de volume igual de NaOH 0,2N e incubações a temperatura ambiente por 5 minutos e a 95°C por 1 minuto. Após essa etapa de desnaturação, foi realizada a diluição com tampão fornecido pelo kit para concentração final 6 a 10 pM, seguindo-se aplicação de todo volume (600 μL) do *pool* de bibliotecas no cartucho de sequenciamento.

O *workflow* escolhido para a análise após a corrida foi o *Generate FastQ*. Tal formato de arquivo de saída permite obter as sequências geradas na corrida (*read*1 e *read*2), acoplado com a informação da qualidade para cada nucleotídeo sequenciado.

O sequenciamento das bibliotecas de sRNAs foi realizado com menor número de ciclos (50 ciclos) e com a estratégia single-end.

Os sequenciamentos foram realizados, sempre que possível, no equipamento *MiSeq* (*Illumina*) instalado no Centro Avançado de Tecnologias em Genômica (CATG) do Instituto

de Química da Universidade de São Paulo. Porém, algumas poucas corridas foram realizadas no *MiSeq* instalado no Centro de Facilidades de Apoio à Pesquisa (CEFAP) do Instituto de Ciências Biomédicas da Universidade de São Paulo.

3.7. Análises bioinformáticas e estatísticas

Após o término do sequenciamento, são gerados arquivos em formato FastQ contendo as sequências (*read*1 e *read*2) com informação do escore de qualidade (*Q score*) para cada nucleotídeo, atribuído pelo *software* do próprio *MiSeq*. Os arquivos FastQ gerados para cada biblioteca foram submetidos a análise com *software FASTQC* (Andrews *et al.*, 2014) que gera plotagens baseadas no *Q score* por base sequenciada, distribuição de *Q score* médio de cada sequência obtida, conteúdo de adaptadores remanescentes, entre outras informações.

O mapeamento das sequências obtidas por RNA-Seq nas respectivas sequências genômicas de referência foi realizado utilizando o *software CLC Genomics Workbench*, versão 6.5⁴ com o módulo *RNA-Seq analysis*. Foram utilizados arquivos em formato *gbk*, o qual contém as informações das sequências de todos os genes anotados dos genomas de referência. Duas opções de mapeamento foram realizadas, gerando listas de genes com valores de expressão normalizados por FPKM (*Fragments per kilobase transcript per million reads*) (Mortazavi *et al.*, 2008) ou com valores brutos de contagem de quantas sequências mapearam em cada gene.

Análises de correlação de Pearson entre as réplicas biológicas foram realizadas com os valores de expressão por FPKM utilizando a função *cor.test* do pacote *stats* do *software* R (R Core Team, 2013). Para as análises estatísticas de expressão diferencial entre genes de transcritomas diferentes, foi usado o pacote *DESeq2* (Love *et al.*, 2014) do *software* R (R Core Team, 2013), usando como dados de entrada os valores brutos de contagem de sequências para cada gene. Como parâmetros para definir um gene como diferencialmente expresso entre duas condições distintas foi estabelecido ter um valor de padj < 0,05 e um valor de $log_2FoldChange > |1|$. A função *plotMA* do mesmo pacote foi usada para gerar os gráficos de dispersão dos genes baseado em seus respectivos valores de expressão média e razão de expressão. A função *heatmap.2* do pacote *gplots* foi utilizada para gerar h*eatmaps*, enquanto que a função *radarchart* do pacote *fmsb* foi usada para gerar os *Radar Charts*, ambas no *software* R (R Core Team, 2013).

⁴ http://www.clcbio.com

O software Rockhopper2 (Tjaden, 2015) foi utilizado para análises de regiões não traduzidas (*Untranslated regions* ou *UTR*) e definição de operons.

Análises comparativas de genes dos genomas de *X. fastidiosa* foram realizadas utilizando-se a plataforma IMG/ER (*Integrated Microbial Genomes-Expert Review*) (Markowitz *et al.*, 2012). Para pesquisa de identidade/similaridade de sequências de DNA ou de proteínas também foram utilizados os programas *BLASTn*, *BLASTp e BLASTx* (Altschul et al., 1990) (em bases públicas de dados tais como *GenBank* do NCBI (*National Center for Biotechnology Information*⁵), *Xylella fastidiosa Genome Project*⁶ e *Xylella fastidiosa - Pierce's Disease Strain Genome Project*⁷.

Análises funcionais dos transcritomas foram realizadas com o *software MinPath (Ye e Doak, 2009)* que analisa quais as vias metabólicas estão representadas nos transcritomas, sendo considerado os genes com valores de FPKM acima de 10. Os dados também foram submetidos a classificação funcional com base nas categorias do *Gene Ontology* (Ashburner et al., 2000; Gene Ontology Consortium, 2004) e vias *KEGG* (Kanehisa e Goto, 2000; Kanehisa et al., 2016) para avaliação de vias enriquecidas (P > 0,05) em determinada condição quando comparada a outra. Para isso foram utilizados os genes diferencialmente expressos e o *web-software BayGO* (Vencio *et al.*, 2006).

As sequências obtidas com o sequenciamento de bibliotecas de sRNAs foram submetidas a montagem com *software MIRA* (Chevreux *et al.*, 1999) para que fossem gerados *contigs* representativos das sequências completas dos candidatos a sRNAs, além de fornecer informações do número de *reads* usados para gerar cada possível sequência completa de sRNA. As sequências dos sRNAs de ambas as bibliotecas foram comparadas entre si por meio da ferramenta BLAST (Altschul *et al.*, 1990), de modo a identificar quais candidatos foram encontrados nas duas bibliotecas. O mesmo procedimento foi aplicado na comparação das sequências das sequências dos sRNAs.

⁵ http://www.ncbi.nlm.nih.gov

⁶ http://aeg.lbi.ic.unicamp.br/xf/

⁷ http://aeg.lbi.ic.unicamp.br/world/xfpd/

4. Resultados e Discussão

Como já mencionado, este trabalho teve como objetivo sequenciar, analisar e comparar os transcritomas de diferentes cepas de *Xylella fastidiosa*, no início e fim da fase exponencial de crescimento em meio rico PWG e em meio mínimo PIM6. Para a descrição do repertório completo de genes que são expressos ou diferencialmente expressos nestas condições foi utilizada a técnica de RNA-Seq.

O meio PWG é um meio complexo e completo utilizado corriqueiramente no cultivo de *X. fastidiosa*. O meio PIM6 foi desenvolvido como uma alternativa para reproduzir a composição da seiva do xilema, que sabidamente é pobre em nutrientes. De modo geral, essa bactéria tem crescimento lento no cultivo *in vitro*, com tempo de duplicação >9 horas, dependendo do meio de cultivo. Nas plantas, *X. fastidiosa* coloniza exclusivamente o lúmen dos vasos do xilema e é considerada um fitopatógeno generalista em razão da ampla gama de hospedeiros em que já foi encontrada (Chatterjee *et al.*, 2008; Janse e Obradovic, 2010; Almeida e Nunney, 2015).

As principais etapas envolvidas na obtenção dos transcritomas de diferentes cepas de *X. fastidiosa* estão esquematizadas na Figura 1. Os resultados obtidos serão apresentados e discutidos em duas partes:

A) Descrição e comparação dos transcritomas das cepas 9a5c e Temeculal no início e fim da fase exponencial de crescimento nos meios PWG e PIM6. Estas duas cepas são consideradas cepas de referência e apresentam diferenças genômicas e fenotípicas no cultivo *in vitro*. Além de análises de expressão gênica diferencial e descrição de rotas e vias metabólicas enriquecidas em cada condição experimental, o sequenciamento do transcritoma destas duas cepas foi utilizado na definição das regiões não-traduzidas dos genes (5'UTR e 3'UTR) e identificação de operons preditos nestes genomas. Também foram identificados transcritos que potencialmente são sRNAs na cepa 9a5c.

B) Descrição e comparação dos transcritomas de oito cepas isoladas de citros (cepas 9a5c, J1a12, U24d e Fb7), de cafeeiro (cepa 3124), de hibisco (cepa Hib4), ameixeira (cepa Pr8x) e de videira (cepa Temecula1), no início e fim da fase exponencial de crescimento no meio PWG. Ao contrário das cepas Temecula1 e 9a5c, as outras seis cepas são ainda pouco estudadas, mas também apresentam diferenças genômicas e fenotípicas *in vitro*. A análise comparativa de transcritomas de cepas de *X*. *fastidiosa* que foram isoladas de diferentes espécies de plantas busca associar eventuais



diferenças na expressão gênica aos seus respectivos fenótipos e especificidade de hospedeiro.

Figura 1. Etapas para obtenção dos transcritomas de X. fastidiosa por RNA-Seq.

Parte A) Descrição e comparação dos transcritomas das cepas 9a5c e Temecula1

4.1. Curva de crescimento das cepas 9a5c e Temecula1 em meio PWG e PIM6

O crescimento das cepas 9a5c e Temecula1 foi avaliado em meio PWG e PIM6 por 21 dias, com o objetivo de melhor definir os tempos de coleta, representativos da fase inicial e final do crescimento exponencial. Como esperado, em ambas as cepas foi observado um maior crescimento no meio rico PWG em comparação ao meio mínimo PIM6, sendo que a fase estacionária foi atingida muito mais cedo nesse meio (Figura 2).

A cepa 9a5c aparentemente sustentou um melhor crescimento no meio PWG do que a cepa Temecula1, uma vez que alcançou uma densidade celular mais alta $(DO_{600nm}=1,4)$ ao atingir a fase estacionária (~10 dias). A cepa Temecula1 cresceu rapidamente no início do cultivo $(DO_{600nm} 0,05 \text{ para } 0,3 \text{ em } 3 \text{ dias})$ porém atingiu mais cedo a fase estacionária (7 dias), com a metade do valor de densidade celular da cepa 9a5c $(DO_{600nm}=0,7)$ (Figura 2).

As curvas de crescimento (Figura 2) mostram um modesto aumento da população de células no início do cultivo no meio PIM6 para ambas as cepas, embora com uma baixa densidade celular ($DO_{600nm}=0,1$) ao alcançar a fase estacionária (3 dias), muito antes do observado no meio PWG. Portanto, ainda que o PIM6 supostamente mimetize a composição do fluido xilemático, este meio não é adequado para cultivos de longo prazo.

Comparações entre o crescimento em diferentes meios de cultivo com composições distintas foram analisadas em outros estudos. Curvas de crescimento foram obtidas para o meio definido 3G10-R e o meio complexo PW. No meio 3G10-R foi observado um atraso para atingir a fase estacionária, com uma fase exponencial pouco evidente, enquanto que no meio PW a fase estacionária foi alcançada com poucos dias (Ciraulo *et al.*, 2010). Ao contrário, no presente estudo, o crescimento no meio definido PIM6 atingiu mais cedo a fase estacionária do que no meio complexo PWG, o qual levou mais tempo até atingi-la com mais de 10 dias, ambos com fases exponenciais definidas. Tal resultado corrobora o observado em um estudo com o crescimento *in vitro* de *Xylella* em seiva do xilema de videiras, no qual foi observada uma antecipação

da chegada à fase estacionária em comparação com um meio complexo artificial (Zaini *et al.*, 2009).

Para obter quantidade de células para extração de RNA total suficiente para o RNA-Seq, foram realizados novos ensaios em meio PIM6 iniciando-se o cultivo com densidades celulares mais altas (dados não mostrados). Desta forma, foi definido que os experimentos para obtenção das amostras de RNA total de células cultivadas neste meio iniciariam com a DO_{600nm} 0,25.

Estas análises de crescimento guiaram a definição do tempo e as condições de cultivo das cepas 9a5c e Temecula1 para a extração de RNA total para a análise comparativa do transcritoma nestes dois meios. Baseando-se nas curvas de crescimento, foi decido pela coleta no 1° e 3° dias (início da fase exponencial) nos meios PIM6 e PWG, respectivamente, para ambas as cepas. Para as amostras do final da fase exponencial, foram definidos o 3° dia para o meio PIM6 em ambas as cepas e o 7° e 10° dias para o meio PWG em Temecula1 e 9a5c, respectivamente (Figura 2).

Como mostrado na Figura 3, pode-se observar uma grande variação nos fenótipos do cultivo *in vitro* em PIM6 ou PWG de ambas as cepas, quanto a formação de agregados em solução (células planctônicas) e de agregados aderidos à parede do frasco (células em biofilme). Os diferentes fenótipos apresentados pelas duas cepas em situações de cultivo semelhantes é indicativo de prováveis diferenças em seus respectivos perfis de expressão gênica.



Figura 2. Curva de crescimento das cepas 9a5c e Temecula1. As células das cepas 9a5c (painel superior) e Temecula1 (painel inferior) foram crescidas em tubos cônicos de 50mL com DO_{600nm} inicial de 0,05 nos meios PWG e PIM6, a 28°C e 170 rpm. Medições da DO_{600nm} foram realizadas nos 1°, 3°, 7°, 10°, 15° e 21° dias de crescimento. Barras verticais representam o desvio padrão da média de três réplicas técnicas.



Figura 3. Fenótipo do cultivo das cepas 9a5c e Temecula1 dos dias escolhidos para a extração de RNA total. (A) cepa 9a5c com 1 dia em PIM6; (B) cepa 9a5c com 3 dias em PIM6; (C) cepa 9a5c com 3 dias em PWG; (D) cepa 9a5c com 10 dias em PWG; (E) cepa Temecula1 com 1 dia em PIM6; (F) cepa Temecula1 com 3 dias em PIM6; (G) cepa Temecula1 com 3 dias em PWG; (H) cepa Temecula1 com 7 dias em PWG. DO_{600nm} inicial de 0,25 para no cultivo em meio PIM6 e 0,05 no cultivo em PWG. As imagens mostram uma de três réplicas biológicas para cada condição. Frascos incubados a 28°C e 170 rpm.

4.2. Obtenção dos transcritomas por RNA-Seq

O primeiro requerimento crítico para o sucesso da técnica de RNA-Seq é a obtenção das amostras de RNA total com elevado grau de pureza e integridade. A pureza das preparações de RNA é avaliada pela medida da absorbância a 230, 260 e 280 nm, sendo desejável a obtenção de razões A_{260nm}/A_{280nm} e $A_{260nm}/A_{230nm} > 1,8$. A integridade é analisada por eletroforese capilar no equipamento *2100 Bioanalyzer* (*Agilent Technologies*) o qual calcula o *RNA Integrity Number* (RIN) que deve ser >8. A Tabela 3 lista estes parâmetros de qualidade para todas as amostras de RNA extraídas das cepas 9a5c e Temecula1, que de modo geral foram bastante adequados, especialmente para as preparações originadas de PWG. No caso das preparações de meio PIM6 foi necessário o uso do *LifeGuard* (*MoBio Laboratories*) para obtenção de RIN > 7,0. Além disso, a completa remoção do DNA das preparações de RNA total foi confirmada por PCR com um par de *primers* específico para detecção de DNA de *X. fastidiosa* (dados não mostrados).

A Tabela 3 também inclui as quantificações das amostras de RNA total tanto pela medida da A_{260nm} no NanoDrop quanto por método fluorimétrico (reagente *RiboGreen*) que é mais preciso. Note que a medição da A_{260nm} medida no NanoDrop superestimou a concentração do RNA em relação as determinações com o reagente *RiboGreen*.

Após as análises de qualidade e pureza das preparações de RNA, foi realizada a depleção de moléculas de rRNAs para obtenção de uma fração enriquecida em mRNAs, seguindo-se análise por eletroforese capilar. Esta etapa também é crítica, uma vez que os rRNAs constituem mais de 90% do RNA total bacteriano. Para todas as amostras indicadas na Tabela 3, a eficiência na remoção de rRNAs foi alta quando foi utilizado 3-5µg de RNA total (dados não mostrados).

Finalmente, as amostras de RNA purificadas e depletadas de rRNAs foram utilizadas para a preparação das bibliotecas de cDNA. Para todas elas, foi obtido um excelente rendimento, quanto a concentração como na qualidade e pureza (Tabela 3). A concentração das bibliotecas foi determinada qPCR absoluto em que são utilizados iniciadores complementares aos adaptadores utilizados na confecção das bibliotecas. Além da quantificação precisa (Tabela 3), este ensaio também avalia a eficiência do

protocolo de preparação das bibliotecas, uma vez que somente serão amplificados os fragmentos que tiveram os adaptadores corretamente ligados em suas extremidades. Outro parâmetro relevante é o tamanho médio dos fragmentos de cDNA da biblioteca, o qual é definido em eletroforese capilar.

As 24 bibliotecas de cDNA foram sequenciadas em cinco corridas no equipamento *MiSeq* (*Illumina*) utilizando a estratégia de *Paired-End* com um *kit* para 500 ciclos. Dessa forma, esperava-se sequências de até 250 bases correspondentes a cada extremidade do fragmento sequenciado. O rendimento destas corridas foi excelente, com quantidades de sequências muito próximas ou até maiores do que o previsto pelo fabricante, que é de 25 milhões de sequências (Tabela 4). Além disso, uma alta porcentagem das sequências apresentou alta qualidade de sequenciamento, aumentando a confiabilidade nos dados, o que será discutido na próxima seção. A densidade de *clusters* de fragmentos formados na *flow cell* na etapa inicial do sequenciamento excedeu os valores tidos como seguros (Tabela 4), os quais segundo o fabricante devem estar entre 1000 e 1200. Porém, felizmente este fato não interferiu no rendimento do sequenciamento.

C	Meio de	e Tempo	NanoDrop	A260/	A260/	DIN	RiboGreen	NanoDrop	A260/	A260/	Tamanho	qPCR	Dánlian
Ce	cultivo	(dias)	(ng/µL)	280	230	NI N	(ng/µL)	(ng/µL)	280	230	médio (pb)	(n M)	Kepiica
			329,8	2,07	2,48	8,1	348,6	79,5	1,80	1,97	390	113,6	1
		1	199,2	2,09	2,42	7,1	188,9	68,3	1,83	2,29	366	191,6	2
	PIM6		239,0	2,11	2,41	7,4	118,8	76,8	1,81	2,25	457	224,8	3
	1 11/10		245,9	2,08	2,40	7,9	165,1	82,0	1,80	2,26	375	364,7	1
		3	258,8	2,12	2,46	8,6	201,1	74,0	1,80	2,26	376	296,5	2
, L			246,1	2,10	2,39	8,0	140,3	84,5	1,81	2,20	469	169,5	3
0a			294,8	2,12	1,32	8,4	296,8	58,1	1,80	2,32	361	218,1	1
		3	265,3	2,09	2,19	7,3	275,3	61,6	1,83	2,26	407	146,6	2
	PWC		261,9	2,11	2,45	8,5	147,4	93,1	1,82	2,25	461	251,0	3
	1.00		328,2	2,10	2,47	8,4	292,7	35,3	1,78	1,96	355	58,9	1
		10	258,6	2,13	2,42	8,2	252,7	63,3	1,82	2,20	365	83,1	2
			273,3	2,11	2,42	8,6	168,0	80,7	1,84	2,32	471	198,0	3
			239,7	2,11	2,47	8,1	265,4	41,1	1,78	2,28	356	146,0	1
		1	232,1	2,09	2,24	7,8	163,9	52,1	1,85	2,25	385	171,1	2
	DIM6		251,0	2,10	2,44	8,1	170,7	88,8	1,83	2,19	450	io (pb)(nM)Kep) 390 113,61 366 191,62 157 224,83 375 364,71 376 296,52 469 169,53 361 218,11 407 146,62 461 251,03 355 58,91 365 83,12 471 198,03 356 146,01 385 171,12 450 213,63 353 248,41 345 288,82 429 199,03 410 250,81 384 271,02 443 230,93 388 116,41 365 190,32 443 206,03	3
	F HVIO		241,1	2,10	2,44	8,0	197,9	58,2	1,79	2,09	353	248,4	1
·		3	240,5	2,10	2,35	8,4	184,3	59,4	1,79	2,13	345	288,8	2
ula			227,9	2,10	2,39	8,7	109,6	69,2	1,83	2,07	429	199,0	3
med			264,5	2,12	2,42	8,3	177,6	85,7	1,84	2,25	410	250,8	1
Ē		3	267,6	2,11	2,17	8,2	209,8	70,2	1,79	2,08	384	376 296,5 469 169,5 361 218,1 407 146,6 461 251,0 355 58,9 365 83,1 471 198,0 356 146,0 385 171,1 450 213,6 353 248,4 345 288,8 429 199,0 410 250,8 384 271,0 443 230,9 388 116,4 365 190,3 443 206,0	2
	DWC		266,2	2,13	2,46	8,6	167,5	91,4	1,84	2,19	443	230,9	3
	1 WG		447,6	2,09	2,45	8,3	471,5	64,4	1,79	1,72	388	116,4	1
		7	253,6	2,08	2,37	8,0	235,9	53,9	1,80	2,26	365	190,3	2
			261,1	2,12	2,45	8,9	147,6	76,5	1,83	2,17	443	206,0	3
								1					

Tabela 3. Dados das amostras de RNA das duas cepas nas condições de cultivo indicadas (painel esquerdo) e de suas respectivas bibliotecas de cDNA (painel direito).

Corrido	Total de	Sequências	Número de	Densidade de	%
Corrida	sequências*	filtradas**	bibliotecas	clusters	>Q30***
1	25.687.600	16.908.052	4	1439	77,7
2	29.269.792	23.373.882	5	1607	75,3
3	27.988.934	23.358.634	4	1547	77,7
4	26.881.580	22.864.072	5	1451	77,7
5	24.908.752	22.505.602	8	1303	81,9

Tabela 4. Resumo do rendimento das cinco corridas contendo as bibliotecas das cepas 9a5c e Temecula1.

Paired-End reads.* **Filtragem que remove sequências com baixa qualidade. *>Q30 indica sequências de alta qualidade.

4.3. Análise da qualidade das sequências dos transcritomas

Antes de realizar o mapeamento das sequências nos genomas de referência e outras análises mais aprofundadas, foram realizadas análises para verificar a qualidade das sequências obtidas a partir de todas as bibliotecas de cDNA.

Como exemplo, será discutido o resultado das análises das sequências da primeira réplica da amostra de RNA da cepa 9a5c crescida por 1 dia em meio PIM6 (Figura 4). A análise de qualidade com *software FASTQC* (Andrews *et al.*, 2014) foi feita separadamente com sequências de cada uma das extremidades no sequenciamento *Paired-End* (Figura 4A e 4B), com as sequências de ambas as extremidades já pareadas (Figura 4C) e com as sequências após o *quality trimming* (>Q30) (Figura 4D) realizado com o *software CLC Genomics Worbench*. Este processo consiste em encurtar sequências deletando parte de suas bases com escores de qualidade inferiores ao *threshold* desejado ou mesmo remover sequências inteiras com valores médios de qualidade abaixo deste *threshold*.

As análises de qualidade de base indicam que à medida que o tamanho da sequência aumenta, o seu escore médio de qualidade diminui, ou seja, sequências mais longas tendem a apresentar um maior número de bases com escores de qualidade inferiores (Figura 4). Este efeito é menor para as sequências da primeira extremidade (*read1*) (Figura 4A) do que para a segunda extremidade (*read2*) (Figura 4B). Esta variação é intrínseca da tecnologia utilizada (*MiSeq/Illumina*), na qual o rendimento dos reagentes sofre uma queda durante a segunda metade da corrida de 500 ciclos, afetando muito mais a qualidade das sequências da segunda extremidade. Mesmo assim, a análise das sequências pareadas (Figura 4C) mostrou que a menor qualidade do *read2* não

prejudica a qualidade do *read1*. A análise das sequências submetidas ao *quality trimming* mostrou, como esperado, somente sequências acima do escore Q30 (Figura 4D). Porém, trata-se de um critério muito elevado de qualidade uma vez praticamente todas as sequências foram encurtadas, o que poderia prejudicar o mapeamento correto no genoma. Assim, foi definido que para as análises de expressão diferencial subsequentes seriam utilizadas as sequências sem realizar o processo de *trimming*.

A estratégia de realizar ou não trimming de sequências de RNA-Seq é controversa. Williams e colaboradores (Williams et al., 2016) avaliaram o impacto do trimming na análise de expressão diferencial. Para tal, eles geraram conjuntos de dados de RNA-Seq de quatro amostras de neurônios sensoriais de larvas de Drosophila melanogaster, e usaram três algoritmos para o trimming baseado na qualidade de base com diferentes parâmetros de O escore. Com os parâmetros de trimming mais agressivos, mais de 10% dos genes tiveram mudanças significativas nos seus níveis de expressão. Observaram este mesmo padrão com outros dois conjuntos de dados de RNA-Seq e com *pipelines* alternativos de análise de expressão genica. Uma fonte considerável das diferenças observadas pode ser atribuída ao alinhamento de reads curtos resultantes do trimming mais agressivo. Eles ainda realizaram uma comparação entre dados de RNA-Seq e microarranjos, e sugerem que realizar nenhum trimming ou trimming moderado resulta em estimativas de expressão gênica bem mais acuradas. Em resumo, estes autores concluem que o trimming agressivo pode influenciar fortemente a acurácia da estimativa de níveis de expressão gênica, o que subsequentemente impacta na identificação de genes diferencialmente expressos. De todo modo, eles sugerem que deve ser avaliado em cada caso se o trimming será ou não benéfico para o tipo de análise a ser realizada (Williams et al., 2016).

De modo geral, todos os transcritomas sequenciados neste trabalho apresentaram padrões de qualidade semelhantes ao exemplo mostrado na Figura 4, evidenciando a robustez da técnica de RNA-Seq e confiabilidade dos dados obtidos.



Figura 4. Análise da qualidade por base das sequências da primeira réplica da cepa 9a5c cultivada por 1 dia em PIM6. (A) análise das sequências da primeira extremidade (*read1*); (B) análise das sequências da segunda extremidade (*read2*); (C) análise com as sequências pareadas (*read1+read2*); (D) análise após o *quality trimming* (Q>30). As análises foram realizadas usando o *software FASTQC*. O eixo das abcissas contém os tamanhos das sequências (1 até 250), enquanto que o eixo das ordenadas contém os valores de Q escore (de 0 a 40).

4.4. Mapeamento das sequências dos transcritomas nos genomas de referência

Após análise de qualidade das sequências, foi realizado o mapeamento nos respectivos genomas de referência. Para tal os arquivos em formato FastQ contendo as sequências e seus respectivos escores de qualidade por base para o sequenciamento de ambas extremidades, foram importados para o *software CLC Genomics Workbench*. O mapeamento foi realizado pelo pareamento do *read1* e do *read2* de cada fragmento ao genoma de referência (no caso genomas das cepas 9a5c e Temecula1, incluindo seus respectivos plasmídeos), o que possibilita a obtenção do tamanho da sequência que separa as extremidades pareadas de cada fragmento mapeado. A Tabela 5 mostra que em todos os transcritomas foi obtida uma alta porcentagem de mapeamento, variando entre 70-90,3%. Vale lembrar que o genoma *X. fastidiosa* tem tamanho de ~2,8 Mpb e que a cepa 9a5c tem 2 plasmídeos (pXF51 de ~51 kbp e pXF1.3 de ~1,3 kpb) enquanto

que cepa Temecula1 possui apenas o pXFPD1.3 de ~1.3 kpb (Simpson *et al.*, 2000; Van Sluys *et al.*, 2003).

O número de sequências por biblioteca variou de 5 a 18 milhões (Tabela 5). Esta variação reflete tanto o número de bibliotecas que foram combinadas em uma corrida de sequenciamento bem como o seu rendimento, mas todas apresentaram cobertura plenamente satisfatória, sendo que a quantidade de fragmentos sequenciados está dentro do recomendado para análises de RNA-Seq para genomas bacterianos. Por exemplo, em um estudo que explorou dados de transcritomas para duas réplicas biológicas de *Escherichia coli* obtidas no início e no final do crescimento, concluiu que 2 a 3 milhões de sequências por amostra são suficientes para se obter e identificar genes diferencialmente expressos (*fold change*>2) com alta significância estatística, no caso de dados de duas réplicas biológicas altamente correlatas. Tais as análises foram realizadas com diferentes profundidades de número de sequências, sugerindo, inclusive, que um aumento excessivo da profundidade de sequenciamento pode ser prejudicial para um mapeamento acurado dos transcritos com relevância biológica (Haas *et al.*, 2012).

Сера	Meio de cultivo	Tempo (dias)	Total de sequências	Tamanho médio (bases)	Distância entre paired reads (bases)	Sequências mapeadas	Réplica
			9.687.644	173	70 a 248	7.897.614 (81,5%)	1
		1	7.649.084	157	63 a 245	5.276.706 (70,0%)	2
	PIM6		5.899.510	160	73 a 249	5.170.546 (87,6%)	3
	r IIvio		8.386.060	158	73 a 251	7.219.356 (86,1%)	1
		3	9.421.628	157	68 a 250	8.116.162 (86,1%)	2
Sc			5.705.434	160	72 a 249	5.003.758 (87,7%)	3
9a			6.308.458	163	78 a 241	4.484.794 (71,1%)	1
		3	13.830.460	161	74 a 241	9.995.040 (72,3%)	2
	PWG		5.231.176	163	74 a 249	4.523.362 (86,5%)	3
	1.00		8.608.030	168	70 a 241	7.170.120 (83,3%)	1
		10	18.589.024	161	71 a 246	14.220.660 (76,5%)	2
			6.029.866	161	71 a 250	5.178.372 (85,9%)	3
			5.593.228	158	73 a 244	4.615.180 (82,5%)	1
		1	10.209.216	152	69 a 241	8.472.134 (83,0%)	2
la1	PIM6		5.302.846	167	75 a 300	4.754.520 (89,7%)	3
ecu	1 11010		9.916.280	156	69 a 240	8.500.780 (85,7%)	1
em		3	7.976.518	153	68 a 249	6.955.478 (87,2%)	2
-			5.186.310	161	74 a 248	4.560.958 (87,9%)	3
	PWG	3	12.094.328	166	71 a 241	9.511.234 (78,6%)	1

Tabela 5. Resumo do rendimento de cada biblioteca sequenciada.

	9.624.394	162	71 a 250	8.373.524 (87,0%)	2
	5.742.484	171	75 a 300	5.152.250 (89,7%)	3
	7.397.116	163	62 a 244	6.145.576 (83,1%)	1
7	7.922.810	152	67 a 248	6.698.052 (84,5%)	2
	5.336.466	164	75 a 297	4.815.956 (90,3%)	3

4.5. Análise de correlação entre réplicas biológicas e entre diferentes transcritomas

Com o mapeamento das sequências de cada transcritoma nos genomas de referência das cepas 9a5c e Temecula1 foram obtidos os valores de expressão baseados na normalização por FPKM (*Fragments per kilobase per million*) (Mortazavi *et al.*, 2008), a qual considera o número de sequências mapeadas em cada gene, o tamanho do transcrito e o tamanho total do transcritoma.

Estes valores de expressão por FPKM foram utilizados para análises de correlação de Pearson entre as três réplicas biológicas de cada condição, comparadas duas a duas. Tal análise fornece a informação se elas são reprodutíveis e poderiam ser de fato tratadas como réplicas, e, portanto, ser utilizadas nas análises de expressão diferencial e seus respectivos testes estatísticos. O teste de correlação de Pearson é um teste paramétrico robusto que mede o grau de relação linear entre duas variáveis quantitativas (Pearson, 1931a; Pearson, 1931b; Norman, 2010). Todos os valores de correlação de Pearson para todas as comparações entre réplicas foram maiores que 0,9 (Tabela 6), indicando que todas as réplicas podem ser tratadas como tal. As plotagens entre valores de FPKM entre duas réplicas de uma mesma condição demonstram a alta correlação entre as réplicas biológicas (Figura 5). As plotagens combinando os outros dois pares de réplicas de mesma condição confirmam esta alta correlação (dados não apresentados).

Сера	Meio	Tempo (dias)	Réplicas	Valor de correlação de Pearson
			1 e 2	0,998
		1	1 e 3	0,998
	DIM6		2 e 3	0,994
	I INIO		1 e 2	0,999
		3	1 e 3	0,999
2			2 e 3	0,999
9a£			1 e 2	0,993
		3	1 e 3	0,970
	DWC		2 e 3	0,958
	FWG		1 e 2	0,999
		10	1 e 3	0,995
			2 e 3	0,994
			1 e 2	0,987
		1	1 e 3	0,996
	DIM6		2 e 3	0,986
	r IIvio		1 e 2	0,996
1		3	1 e 3	0,998
cula			2 e 3	0,998
mee			1 e 2	0,992
Te		3	1 e 3	0,894
	DWC		2 e 3	0,932
	T WU		1 e 2	0,999
		7	1 e 3	0,998
			2 e 3	0,997

Tabela 6. Valores de correlação de Pearson entre as réplicas de uma mesma condição. Análise foi realizada com os valores de FPKM de cada transcritoma.



Figura 5. **Gráficos de correlação entre pares de réplicas da mesma condição baseado nos valores de FPKM transformados em logaritmo na base 10**. (A) 9a5c no início da fase exponencial em PIM6; (B) 9a5c no final da fase exponencial em PIM6; (C) 9a5c no início da fase exponencial em PWG; (D) 9a5c no final da fase exponencial em PWG; (E) Temecula1 no início da fase exponencial em PIM6; (F) Temecula1 no final da fase exponencial em PIM6; (G) Temecula1 no início da fase exponencial em PWG; (H) Temecula1 no final da fase exponencial em PWG.

Os valores de FPKM também foram usados para uma análise de *heatmap* com todas as réplicas de ambas as cepas (Figura 6). Os valores foram transformados em logaritmo na base 10 (log10), de modo a minimizar os efeitos de genes com valores muito altos de expressão gênica, o que levaria a uma dificuldade de visualização dos genes com baixa expressão. Trata-se de outra forma de avaliar a similaridade entre as réplicas e amostras, uma vez que elas são agrupadas em um dendrograma. O resultado (Figura 6) mostra a separação dos transcritomas em quatro grupos distintos, os quais correspondem exatamente às diferentes cepas e meios de cultura. Entretanto, são observadas duas discrepâncias: a terceira réplica biológica do transcritoma de 9a5c no início do crescimento em PWG agrupou com seus transcritomas de fase tardia e a terceira réplica biológica do transcritoma da Temecula1 no final do crescimento em PWG agrupou com seus transcritomas de fase inicial nos transcritoma (Figura 6).

Com estes mesmos valores de FPKM em log10 de todos os transcritomas, foi realizada uma análise para verificar quais condições seriam mais similares entre si. Para isso, foram obtidas médias dos valores de FPKM das réplicas de uma mesma condição e plotadas em um *heatmap* com dendograma (Figura 7). Os transcritomas agruparam de acordo a mesma cepa e mesmo meio, com exceção apenas do transcritoma da cepa Temecula1 na fase tardia em PWG, que agrupou com transcritomas em PIM6 (Figura 7).

Em resumo, apesar das discrepâncias apontadas acima, consideramos que os dados obtidos são suficientemente robustos para análises subsequentes.



Figura 6. *Heatmap* com valores de FPKM em log10 de todas as réplicas das condições analisadas em ambas as cepas. As colunas representam os transcritomas indicados. As linhas representam cada gene analisado (1584 genes). O dendrograma no topo da figura mostra o agrupamento dos transcritomas. O gráfico na parte superior direita da figura representa uma gradação de cores baseando-se nos valores de FPKM, sendo que na escala quanto mais próximo do vermelho menor a expressão e, quanto mais próximo do branco maior a expressão. Para a geração da imagem foi usada a função *heatmap.2* do pacote *gplots* utilizando a plataforma *R*.



Figura 7. *Heatmap* com as médias dos valores de FPKM em log10 de todas as réplicas para cada condição analisada em ambas as cepas. As colunas representam os transcritomas indicados. As linhas representam cada gene analisado (1584 genes). O dendrograma no topo da figura mostra o agrupamento dos transcritomas. O gráfico na parte superior direita da figura representa uma gradação de cores baseando-se nos valores médios de FPKM, sendo que na escala quanto mais próximo do vermelho menor a expressão e, quanto mais próximo do branco maior a expressão. Para a geração da imagem foi usada a função *heatmap.2* do pacote *gplots* utilizando a plataforma *R*.

4.6. Identificação de regiões não-traduzidas (UTRs) e operons

Dentre as muitas vantagens da análise de expressão gênica por RNA-Seq em comparação com os microarranjos, está a possibilidade de definir as coordenadas de início de transcrição dos genes e também de descrever a distribuição de genes em operons, e, portanto, sendo uma forma de validar predições da anotação de genomas. Considerando que sequências dos transcritomas apresentaram boa qualidade e alta correlação entre réplicas, elas foram utilizadas para a definição de regiões 5' e 3' não-traduzidas (5'UTR e 3'UTR) e para identificação de operons preditos nestes genomas.

Para isso foi realizado um mapeamento utilizando os arquivos FastQ de todos os transcritomas de uma mesma cepa utilizando o *software Rockhopper2* (Tjaden, 2015). Para esta análise, todas as sequências dos transcritomas de uma mesma cepa foram reunidas e, portanto, o número de sequências utilizadas na análise aumentou extraordinariamente. Esta junção de transcritomas partiu do pressuposto que as coordenadas do início e final da transcrição não devem variar entre transcritomas de modo importante em bactérias.

A análise com as sequências dos transcritomas da cepa 9a5c revelou 1142 5'UTRs (41,3% dos genes analisados) e 1281 3'UTRs (46,3% dos genes analisados). Enquanto isso, a análise com as sequências da cepa Temecula1 definiu 800 5'UTRs (38,3%) e 972 3'UTRs (46,5%). A análise também foi realizada mapeando as sequências no plasmídeo de 9a5c, o pXF51, sendo possível definir 33 5'UTRs (51,6% dos genes analisados) e 30 3'UTRs (46,9% dos genes analisados). As UTRs encontradas definem, portanto, as coordenadas de início e final da transcrição destes genes com base em dados de RNA-Seq. As tabelas suplementares S1 e S2 descrevem as coordenadas destas regiões UTRs encontradas para as cepas 9a5c (genoma e plasmídeo pXF51) e Temecula1, respectivamente. Este tipo de resultado poderá ser melhor explorado quando for integrada com a anotação de cada um destes dois genomas.

Com esta análise não foi possível definir os limites de tamanho de todos os transcritos dos genes de ambas as cepas. Provavelmente, genes com baixa expressão e, portanto, menor abundância de transcritos, tiveram suas definições de UTRs inviabilizadas ou dificultadas. Este problema talvez seja resolvido com novos

sequenciamentos das mesmas bibliotecas, aumentando ainda mais a cobertura do genoma, ou mesmo sequenciar novas réplicas das mesmas condições.

Além da análise de UTRs, o mesmo *software* foi usado para descrever quantos e quais operons poderiam ser definidos com base na presença de transcritos policistrônicos. Novamente foram usadas todas as sequências dos transcritomas de cada cepa. O número de operons encontrados foi de 545 e 386 para as cepas 9a5c e Temecula1, respectivamente (Tabela 7). Dentre os operons encontrados, a maioria deles foi predita com apenas dois genes, embora também tenham sido encontrados operons com mais de 10 genes (Tabela 7).

A predição a partir de dados de RNA-Seq foi comparada com o banco de dados $DOOR^2$ (*Database of prOkaryotic OpeRons, version 2.0*) (Dam *et al.*, 2007; Mao *et al.*, 2009). Este banco contém mais de 1,3 milhões de operons preditos em mais de 2000 genomas procarióticos. Trata-se de uma predição computacional baseada em características genômicas que incluem distância intergênica, conservação da vizinhança, distância filogenética, informação de pequenos motivos de DNA, similaridade entre termos GO (*Gene Ontology*) de pares de genes e razão de comprimento entre pares de genes. Como pode ser observado na Tabela 7, as quantidades de operons de ambas as predições. Outros genomas de *X. fastidiosa* presentes no DOOR² mostraram quantidades próximas do número de operons observados em Temecula1. Todos os genomas são de cepas norte-americanas, com 409, 429 e 437 operons para as cepas M12, M23 e GB514, respectivamente (Tabela 7). O maior número de operons em 9a5c pode ser justificado por apresentar um genoma maior do que as outras cepas citadas.

A predição de operons também foi realizada mapeando as sequências obtidas por RNA-Seq na sequência do plasmídeo de 9a5c, o pXF51. Foram encontrados 14 operons, sendo que novamente a maioria apresentou apenas dois genes, o que foi similar ao predito pelo banco de dados DOOR (Tabela 7).

No único estudo de RNA-Seq com cepas de *Xylella fastidiosa*, uma análise de estruturas de operons também foi realizada. Foram sequenciados transcritomas da cepa Temecula1 cultivada em meio PD2, sem ou com suplementação de cálcio (Parker *et al.*, 2016). Também foi utilizado o *software Rockhooper2*, o qual identificou 389 operons, quantidade muito próxima dos 386 operons encontrados aqui para a cepa Tem1, com conteúdo de 2 a 15 genes em ambos os estudos. Entretanto, nem todos os operons foram encontrados exatamente iguais em ambas as análises, sendo vistos 320 operons em

comum. Portanto, 66 operons foram encontrados somente na análise do presente estudo, enquanto que 69 somente na análise de Parker e colaboradores. Sabe-se que a regulação da transcrição de operons pode variar com a condição de cultivo e com determinado estímulo ou sinal (no caso, diferentes meios e suplementação nutricional). A maioria dos operons "exclusivos" das duas análises corresponde às mesmas regiões, porém com diferenças no conteúdo de genes presentes no policistron, ou seja, foram vistos genes a mais ou a menos em determinado transcrito em uma condição em comparação com a outra. Trata-se de uma comprovação em larga escala da regulação diferencial da transcrição de mRNAs policistrônicos, sugerindo a presença de regiões promotoras, TSSs e regiões de término da transcrição alternativos.

Apesar dos dados de RNAseq gerados no presente trabalho terem sido obtidos a partir do sequenciamento de bibliotecas de cDNA sem informação de orientação de fita, todos os operons preditos nas duas cepas contém genes próximos e com a mesma orientação, conforme checado por verificação manual. As tabelas suplementares S3 e S4 apresentam a descrição de todos os operons encontrados com a predição por RNA-Seq das cepas 9a5c e Temecula1, respectivamente.

e o buildo de	uuu05 2001	Número de genes por operon									
			2	3	4	5	6	7	8	9	≥10
Сера	Método	Total									
0.5	RNA-Seq	545	318	111	57	23	11	8	4	5	8
9a5c	DOOR	546	323	117	60	22	6	4	5	3	6
9a5c	RNA-Seq	14	7	1	2	2	0	1	1	0	0
pXF51	DOOR	13	7	1	3	0	0	1	1	0	0
T 1.1	RNA-Seq	386	209	89	45	20	9	4	2	2	6
Temeculat	DOOR	400	233	84	42	22	7	4	2	1	5
M12	DOOR	409	243	89	33	21	9	5	3	1	5
M23	DOOR	420	238	88	46	19	13	7	3	1	5
M23 plasmídeo	DOOR	9	5	2	0	0	0	1	0	0	1
GB514	DOOR	426	263	80	43	17	13	5	1	1	3
GB514 plasmídeo	DOOR	11	7	3	0	1	0	0	0	0	0

Tabela 7. Predição de estruturas de operons de cepas de *X. fastidiosa*, utilizando o *software Rockhopper2* e o bando de dados $DOOR^2$. As colunas indicam o número de genes no operon.

4.7. Análise dos transcritomas baseada nos valores de FPKM

4.7.1. Panorama da expressão gênica das cepas 9a5c e Temecula1

Inicialmente foram utilizados os valores de FPKM para as comparações de expressão gênica dentro do mesmo transcritoma, uma vez que não é adequado usá-los para comparar abundância de transcritos entre transcritomas distintos.

As listas de genes e seus respectivos valores de expressão em FPKM, para cada transcritoma (dados não apresentados), foram ordenadas e os genes então divididos em quatro categorias, sendo que a categoria mais representada foi de genes com FPKM entre 100 e 0. Isso mostra que apesar de terem sido observados valores muito altos de FPKM, a maioria dos genes (41,7-79,8%) apresentou-se com baixa expressão, em todos os transcritomas analisados (Tabela 8).

Foi possível observar genes com valores de expressão igual a zero em todos os transcritomas (6,3 – 9,6%). Dentre estes, estão a grande maioria dos genes de tRNAs pois estes transcritos são perdidos durante o procedimento de preparação das bibliotecas. Os tRNAs detectados possuem valores de FPKM baixos. Em resumo, estas observações confirmam a completa ausência de contaminação com DNA nas preparações de RNA total que foram utilizadas no RNA-Seq.

Transcritos de rRNAs foram detectados em poucos transcritomas, mas também com baixíssima expressão (ex: rRNA 23S com FPKM de 0,10 em duas réplicas de Temecula1_PWG_7d). Um dos parâmetros utilizados no mapeamento foi que o fragmento sequenciado somente seria contado caso mapeasse exclusivamente em um único local do genoma. Caso o fragmento fosse mapeado em dois ou mais locais no genoma, ele não seria contado e seria considerado não-mapeado. Isto evitaria qualquer mapeamento espúrio em locais de genes parálogos ou regiões repetidas, ambos eventos possíveis nos genomas de *X. fastidiosa*. Uma vez que os genomas dessa bactéria possuem dois *clusters* de genes de rRNAs, pouquíssimos fragmentos foram contados nessa estratégia de mapeamento. Isso não quer dizer que a depleção dos rRNAs funcionou completamente, uma vez que poderíamos encontrar fragmentos para rRNAs no conjunto dos não-mapeados. Porém devido às altas porcentagens de genes com FPKM acima de zero, em todas as condições testadas (90,4 - 93,7%) (Tabela 8), pode-se evidenciar que a depleção foi satisfatória, levando ao enriquecimento da fração de

mRNAs. Além disso, esta alta porcentagem de genes expressos mostra que no genoma de *X. fastidiosa* a grande maioria de seus genes está ativa, mesmo que em níveis basais.

Dentre os genes com expressão igual a zero estão principalmente aqueles com função hipotética ou provenientes de regiões de prófagos. Este dado sugere que estes genes sejam parte de regiões remanescentes de fagos incompletos e inativos, ou que a condição de cultivo não tenha estimulado sua expressão. Apesar da maioria dos genes relacionados a fagos estarem silenciados, alguns deles apresentaram valores positivos de expressão por FPKM.

Dentre os demais genes que não estão aparentemente expressos em todos os transcritomas de ambas as cepas nas condições investigadas estão genes codificadores de uma *DNA-(apurinic or apyrimidinic site) lyase*, uma permease de um sistema de transporte do tipo ABC e do fator de tradução EF-Tu. Na cepa 9a5c, ainda foi encontrado silenciado um gene da DNA primase, para uma hidrolase dependente de metal e para uma proteína componente do sistema de secreção do tipo IV (VirD4). Na cepa Temecula1, três dos quatro genes codificadores da toxina *Zonula occludens (Zot)* não estão expressos. Contudo, a ausência de expressão destes genes citados pode refletir artefatos gerados pela estratégia de análise, que exclui mapeamentos que não são únicos, sendo que estes genes podem estar duplicados no genoma de *X. fastidiosa*. Verificações manuais serão necessárias para investigar estes casos pontuais.

As cepas 9a5c e Temecula1 também possuem plasmídeos em seus genomas (Simpson *et al.*, 2000; Marques *et al.*, 2001; Van Sluys *et al.*, 2003) e, dessa forma, as sequências obtidas nos transcritomas também foram mapeadas em suas respectivas sequências. Para os pequenos plasmídeos de ambas as cepas, o pXF1.3, não foi detectada expressão de seus genes em nenhuma condição.

Em relação ao plasmídeo pXF51 da cepa 9a5c (Marques *et al.*, 2001), na maioria dos transcritomas todos os 69 genes preditos estão expressos. As únicas exceções foram duas réplicas de transcritomas de células crescidas em PIM6 no início da fase exponencial, nos quais um gene que codifica uma proteína hipotética teve valor zero de FPKM (pXF51_00066). Todos os genes expressos mostraram altos valores de expressão em ambos os meios de cultivo, concentrando-se nas categorias "acima de 10.000" e "entre 10.000 e 1.000" (Tabela 9). Esta alta expressão pode ser reflexo do plasmídeo estar, eventualmente, presente em mais de uma cópia.

Сера	Meio de cultivo	Tempo (dias)	Total de genes*	>10.000	Entre 10.000 e 1.000	Entre 1.000 e 100	Entre 100 e 0	Zero ⁺	Genes expressos (%)	Réplica
			2581	8	69	842	1460	202	92,2	1
		1	2581	7	44	688	1634	208	91,9	2
	PIM6		2581	6	97	1152	1153	173	93,3	3
	1 11/10		2581	4	45	1161	1185	186	92,8	1
		3	2581	4	44	1212	1139	182	92,9	2
9 95c			2581	6	47	1201	1164	163	93,7	Réplica 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1
Jase			2581	9	83	807	1465	217	91,6	1
		3	2581	6	109	868	1416	182	92,9	2
	PWC		2581	7	84	1088	1223	179	93,1	3
	IWG		2581	4	43	800	1547	187	92,8	1
		10	2581	6	61	775	1569	170	93,4	2
			2581	6	80	1039	1274	182	92,9	3
			2420	9	112	778	1289	232	90,4	1
		1	2420	7	81	775	1352	205	91,5	2
	PIM6		2420	7	119	1085	1009	200	91,7	3
	1 11/10		2420	2	64	999	1160	195	91,9	1
		3	2420	2	88	1083	1053	194	expressos (%) R 92,2 91,9 93,3 92,8 92,9 93,7 91,6 92,9 93,1 92,8 92,9 93,1 92,8 93,4 92,9 93,1 92,8 93,4 92,9 90,4 91,5 91,7 91,9 92,0 91,3 91,3 91,5 91,7 91,0 90,7 91,4 91,4	2
Temecula1			2420	4	74	1020	1112	210	91,3	3
Temecular			2420	8	107	805	1289	211	91,3	1
		3	2420	7	111	1014	1082	206	91,5	2
	PWG		2420	6	125	999	1088	202	91,7	3
	100		2420	2	70	666	1464	218	91,0	1
		7	2420	2	12	248	1932	226	90,7	2
			2420	6	97	929	1180	208	91,4	3

Tabela 8. Número de genes contabilizados por faixa de valores de FPKM e porcentagem de genes expressos em cada transcritoma das cepas 9a5c e Temecula1

*total de genes analisado, subtraindo-se os genes de rRNAs e tRNAs; ⁺genes com valor zero de FPKM, subtraindo-se os genes de rRNAs e tRNAs.

Meio de cultivo	Tempo (dias)	Total de genes	>10.000	Entre 10.000 e 1.000	Entre 1.000 e 100	Entre 100 e 0	Zero	Genes expressos (%)	Réplica
		69	23	44	1	0	1	98,5	1
	1	69	29	38	2	0	0	100	2
PIM6		69	29	38	1	0	1	98,5	3
1 11010	3	69	23	44	2	0	0	100	1
		69	22	45	1	1	0	100	2
		69	20	47	2	0	0	100	3
		69	25	42	2	0	0	100	1
	3	69	25	41	2	1	0	100	2
PWG		69	24	43	1	1	0	100	3
		69	24	41	4	0	0	100	1
	10	69	24	41	4	0	0	100	2
		69	28	39	1	1	0	100	3

Tabela 9. Número de genes do plasmídeo pXF51 contabilizados por faixa de valores de FPKM e porcentagem de genes expressos em cada transcritoma da cepa 9a5c.

4.7.2. Descrição dos genes mais expressos

A partir das listas de genes e seus respectivos valores de expressão gênica normalizados por FPKM foram destacados os dez genes mais expressos em todas as condições e réplicas das cepas 9a5c (Tabela 10) e Temecula1 (Tabela 11).

Em todos os transcritomas o gene que apresentou o maior valor de FPKM codifica o ncRNA que é componente da *Bacterial RNase P class A. RNase P* é uma ribonucleoproteína envolvida principalmente no processamento da extremidade 5' de tRNAs, mas que também atua em outras moléculas de RNA. Ela é encontrada em todos os domínios da vida, sendo considerada ribonucleoproteína *housekeeping* o que justifica os altos valores de expressão do ncRNA observados em todos os transcritomas e condições analisadas. Além do ncRNA de 350-400 nucleotídeos, a RNaseP é composta uma subunidade proteica de aproximadamente 14 kDa (Evans *et al.*, 2006; Altman, 2011; Mondragon, 2013), sendo que o ncRNA atua como a parte catalítica desta ribozima (Guerrier-Takada *et al.*, 1983; Reiter *et al.*, 2010; Masquida e Westhof, 2011; Mondragon, 2013). Altos níveis de expressão deste ncRNA também foram verificados em um estudo com o fitopatógeno *Xanthomonas campestris*, atingindo valores próximos de um milhão de sequências mapeadas para este transcrito (Liu *et al.*, 2013), de modo semelhante ao que verificamos nos transcritomas de *X. fastidiosa*.
Outro ncRNA também foi está entre os transcritos mais abundantes em todos os transcritomas é o RNA 6S. Ele foi observado nas primeiras dez primeiras posições dos rankings em 19 dos 24 transcritomas analisados, enquanto que nos demais é encontrado entre os 15 genes mais expressos. O RNA 6S desempenha um papel regulador na transcrição pela interação com a RNA polimerase ligada ao σ^{70} em resposta à mudanca da fase exponencial de crescimento para a fase estacionária (Wassarman e Storz, 2000). Além disso, o 6S mostrou-se essencial para a sobrevivência celular durante a fase estacionária (Trotochaud e Wassarman, 2004). Mesmo durante a fase estacionária tardia, na qual a maioria das moléculas de RNA polimerase- σ^{70} está complexada com o RNA 6S, nem todos os promotores por ela reconhecidos estão inibidos. Consequentemente à ligação do RNA 6S a RNA polimerase- σ^{70} , a expressão de genes regulados pela RNA polimerase- σ^{s} , importante durante a fase estacionária e condições de stress, é aumentada (Wassarman, 2007). Interessantemente, o RNA 6S está entre os mais expressos nos transcritomas de todas as réplicas de 9a5c e Temecula1 extraídas próximas da fase estacionária, tanto em meio PIM6 (3 dias para ambas) como em meio PWG (10 dias para 9a5c e 7 dias para Temecula1). Por outro lado, para ambas as cepas, o 6S parece ligeiramente menos abundante nos transcritomas de amostras extraídas de PWG no início da fase exponencial. Observamos que para algumas réplicas do crescimento em PIM6 em sua fase exponencial inicial, o 6S é encontrado em altos níveis. Isto pode ser devido ao fato de que o meio mínimo constituiria uma situação de estresse nutricional, o que poderia estimular a expressão deste ncRNA para inibir a expressão de genes regulados pelo σ^{70} e, consequentemente, liberar a expressão de genes de resposta a estresses ou de fase estacionária. X. fastidiosa possui três fatores sigma alternativos, σ^{32} , $\sigma^{E} e \sigma^{54}$, sendo que os dois primeiros relacionados a resposta a estresses (Koide et al., 2006; da Silva Neto et al., 2007; da Silva Neto et al., 2008).

Diversos trabalhos de análise de transcritomas por meio de RNA-Seq já reportaram a presença destes dois ncRNAs (RNAseP e 6S), em algumas condições sendo altamente expressos, em diversas espécies bacterianas tais como *Helicobacter pylori* (Sharma et al., 2010), *Vibrio splendidus* (Toffano-Nioche et al., 2012), *Yersinia pestis* (Yan et al., 2013), *Neisseria gonorrhoeae* (McClure et al., 2014) e *Pseudomonas aeruginosa* (Dotsch *et al.*, 2012; Wurtzel *et al.*, 2012). Além disso, essas moléculas também foram encontradas em transcritomas de fitopatógenos como *Pseudomonas syringae* (Filiatrault et al., 2010), *Xanthomonas oryzae* patovar *oryzae* (Liang *et al.*, 2011) e *Pectobacterium atrosepticum* (Kwenda *et al.*, 2016).

Outro destaque entre os dez genes mais expressos nos 24 transcritomas analisados são os genes de diversas bacteriocinas. Dentre elas estão as três colicinas V descritas no genoma de *X. fastidiosa* (XF0262, XF0263 e XF0264) (Simpson *et al.*, 2000; Pashalidis *et al.*, 2005) e novas microcinas, previamente anotadas como proteínas hipotéticas, para quais há evidencia de função e expressão em *X. fastidiosa* (Rodrigo R. Duarte e Aline M. da Silva, dados não publicados). Em conjunto, as observações indicam que, mesmo em um crescimento *in vitro*, estas cepas ainda continuam a produzir toxinas que supostamente tem papel de controlar a população bacteriana no xilema (Pashalidis *et al.*, 2005; Koide *et al.*, 2006; Zaini *et al.*, 2008; Fogaca *et al.*, 2010).

Observamos entre os genes com maiores valores de RPKM, o que codifica a proteína Ax21, primeiramente descrita em Xanthomonas oryzae pv. oryzae. Nesta bactéria, foi sugerido que Ax21 está relacionada à percepção de quórum e com a regulação da expressão de genes ligados a motilidade, virulência e formação de biofilme. A proteína estaria diretamente interagindo com o sistema Rax, o qual seria responsável por sulfatar Ax21 e formar um poro através das duas membranas por onde a proteína seria lançada ao meio extracelular. Outras proteínas deste mesmo sistema teriam a função de sensor e de regulador de resposta de acordo com as alterações de concentração de Ax21. Além disso, esta proteína seria responsável pela ativação da resposta imune inata de seu hospedeiro vegetal, uma vez que interagiria com o receptor Xa21 em arroz (Han et al., 2011). Entretanto, não há interação comprovada de Ax21 com o sistema Rax o que indica que ela não seria sulfatada (Bahar et al., 2014). Além disso, esses autores sugerem que Ax21 não seria difusível e secretada para o meio extracelular, mas sim transportada para o periplasma pelo sistema Sec e localizada na membrana externa. Ela somente seria secretada através de vesículas de membrana externa (Bahar et al., 2014). Ax21 foi renomeada como Omp1X, sendo confirmada como uma de membrana externa, mais especificamente um tipo único de porina com domínio estrutural de barril- β , possuindo diversos papéis em X. oryzae, incluindo motilidade e formação de biofilme (Park et al., 2014). A verificação de que o gene ortólogo de Omp1X /Ax21 tem alta expressão em X. fastidiosa o torna um candidato interessante para futuros estudos.

A lista de transcritos mais abundantes inclui ainda as chaperonas GroL e GroES e também Hfq, que é uma chaperona de RNA. Estes transcritos codificam proteínas reconhecidamente abundantes nas células de um modo geral, mas sua presença pode

indicar que estas condições de cultivo *in vitro* são estressantes para *X. fastidiosa*. Além disso, transcritos de bacterioferritina e peroxiredoxina também estão entre os mais abundantes, reforçando esta hipótese. Genes de proteínas de choque frio e os fatores σ^{70} , e σ^{32} (RpoH) também apresentaram alta expressão. Outros genes encontrados abundantemente expressos codificam proteínas para pilinas como FimA, proteínas de membrana como *OmpW*, peptidases e proteases como Clp, entre outros (Tabelas 10 e 11). Interessantemente, os genes com função hipotética XF9a_01460, XF9a_01737, XF9a_01177/XFTem_00614 também aparecem altamente expressos em vários dos transcritomas e, portanto, são bons candidatos para futuros estudos posteriores para caracterização de suas respectivas funções.

Quanto aos genes do plasmídeo pXF51 de 9a5c, observamos que em todos os transcritomas analisados e em todas as condições e réplicas, o transcrito mais abundante codifica para uma *phage-related protein* (pXF51_00036) (Tabela S5). Buscas em bancos de dados, mostraram que essa proteína teria similaridade com a toxina *RelE*, a qual faz parte de um sistema de toxina/antitoxina de manutenção de plasmídeos na célula (Lee *et al.*, 2014). Em *X. fastidiosa*, o sistema DinJ/RelE codificado no cromossomo da cepa Temecula1 controla a proliferação e a população bacteriana na colonização da planta, e o nocaute deste sistema resulta em um fenótipo de hipervirulência (Burbank e Stenger, 2017).

Meio de	Tempo	1	2	3	4	5	6	7	8	9	10
cultivo	(dias)										
		RNase P	Colicina V (XF0263)	Ax21 (XF1803)	Colicina V (XF0264)	Omp W	prepilin- type domain	Hipotética (1460)	GroL	Omp	Conserved protein
PIM6	1	RNase P	Colicina V (XF0263)	Hipotética (1460)	6S RNA	Omp W	Colicina V (XF0264)	Ax21 (XF1803)	GroL	Omp	Cold shock proteins
		RNase P	Omp W	Colicina V (XF0263)	Hipotética (1460)	Ax21 (XF1803)	GroL	Colicina V (XF0264)	prepilin-type domain	6S RNA	FimA
		RNase P	6S RNA	Colicina V (XF0263)	Hipotética (1737)	Hipotética (1460)	prepilin- type domain	Cold shock proteins	Ax21 (XF1803)	OmpW	Bacteriocina (XF1217)
	3	RNase P	Colicina V (XF0263)	6S RNA	Hipotética (1737)	Prepilin-type domain	Hipotética (1460)	Cold shock proteins	Bacteriocina (XF1217)	Bacterio- ferritina	Ax21 (XF1803)
		RNase P	6S RNA	Hipotética (1737)	Colicina V (XF0263)	Prepilin-type domain	Hipotética (1460)	OmpW	Ax21 (XF1803)	Cold shock proteins	Bacteriocina (XF1217)
		RNase P	Colicina V (XF0264)	Colicina V (XF0263)	Ax21 (XF1803)	Hipotética (1737)	Hipotética (1460)	OmpW	Bacteriocina (XF1217)	Bacterio- ferritin	Bacteriocina (XF1218)
	3	RNase P	Colicina V (XF0264)	Ax21 (XF1803)	Colicina V (XF0263)	Cold shock Proteins	Hipotética (1460)	Peptidase (XF0531)	Bacteriocina (XF1217)	6S RNA	Bacteriocina (XF1218)
PWG		RNase P	Omp W	Colicina V (XF0263)	Ax21 (XF1803)	Hipotética (1737)	Hipotética (1177)	Hipotética (1460)	Colicina V (XF0264)	Omp precursor (XF1024)	Bacteriocina (XF1217)
		RNase P	Colicina V (XF0263)	6S RNA	Hipotética (1737)	Colina V (XF0264)	Bacterio- ferritina	Molecular chaperone	Ax21 (XF1803)	Omp W	Hipotética (1460)
	10	RNase P	6S RNA	Colicina V (XF0263)	Cold shock proteins	Bacteriocina (XF1217)	Colicina V (XF0264)	Hipotética (1737)	Bacterio- ferritina	Hipotética (1460)	Bacteriocina (XF1307_2)
		RNase P	6S RNA	Colicina V (XF0263)	Omp W	Ax21 (XF1803)	Hipotética (1737)	Colicina V (XF0264)	Bacterio- ferritina	Bacteriocina (XF1306)	Hipotética (1177)

Tabela 10. Lista dos dez genes mais expressos nos transcritomas da cepa 9a5c, baseando-se em valores de FPKM.

Meio de cultivo	Tempo (dias)	1	2	3	4	5	6	7	8	9	10
		RNase P	Ax21 (PD1063)	Omp W	Prepilin- type domain	Cold shock Proteins	6S RNA	Colicina V (PD0216)	Bacteriocina	GroL	Proteína ribossomal
	1	RNase P	6S RNA	Omp W	Ax21 (PD1063)	GroES	Molecular chaperone	GroL	prepilin-type domain	Cold shock proteins	Colicina V (PD0216)
PIM6		RNase P	Ax21 (PD1063)	6S RNA	Omp W	Prepilin-type domain	Colicina V (PD0216)	FimA	Cold shock proteins	Omp	Hipotética (614)
		RNase P	6S RNA	OmpW	Omp precursor (PD0312)	Omp	Colicina V (PD0215)	Clp protease, subunit ClpP	Molecular chaperone	RNA chaperone Hfq	Peroxire- doxina
	3	RNase P	6S RNA	Colicina V (PD0215)	OmpW	Membrane protein	RNA chaperone Hfq	Clp protease, subunit ClpP	Omp	Molecular chaperone	Omp precursor (PD0312)
		RNase P	6S RNA	Competen- ce protein	Colicina V (PD0215)	Omp W	Proteína de membrana	RNA chaperone Hfq	Hipotética (614)	Cold shock proteins	Omp precursor (PD0312)
		RNase P	Omp W	Colicina V (PD0216)	Ax21 (PD1063)	Cold shock Proteins	Bacteriocina (PD0497)	Bacteriocina	Colicina V (PD0217)	Bacteriocina	Bacteriocina (PD0556)
PWG -	3	RNase P	Omp W	Ax21 (PD1063)	Omp precursor (PD0312)	Colicina V (PD0216)	Hipotética (614)	Bacteriocina (PD0497)	Cold shock proteins	Proteína de membrana	6S RNA
		RNase P	Omp W	Hipotética (614)	Ax21 (PD1063)	FimA	Bacteriocina (PD0497)	Colicina V (PD0216)	Omp precursor (PD0312)	Cold shock proteins	Competence protein
		RNase P	Colicina V (PD0216)	Molecular chaperone	6S RNA	Colicina V (PD0217)	Fator sigma RpoH	Proteína de membrana	Clp protease, subunit ClpP	Bacterio- ferritina	Proteína ribossomal
	7	RNase P	6S RNA	Molecular chaperone	Fator sigma RpoH	Bacterio- ferritina	Clp protease, subunit ClpP	Peroxire- doxina	GroES	DNA-binding protein H-NS	Fator Sigma 70
		RNase P	Colicina V (PD0216)	Omp W	Colicina (PD0217)	Ax21 (PD1063)	Hipotética (614)	6S RNA	Bacteriocina	Bacteriocina (PD0497)	Cold shock proteins

Tabela 11. Lista dos dez genes mais expressos nos transcritomas da cepa Temecula1, baseando-se em valores de FPKM.

4.7.3. Análise de rotas metabólicas expressas

Os *rankings* de genes gerados a partir de dados de expressão gênica, normalizados por FPKM, também foram usados para uma análise global de quais rotas metabólicas estariam ativas nas condições utilizadas no presente trabalho. Para esta análise foi utilizado o *software MinPath* (*Minimal set of Pathways*) (Ye e Doak, 2009), o qual utiliza um método de parcimônia para reconstrução de rotas biológicas a partir de predições de famílias proteicas. Desta forma, é obtida uma estimativa mais conservativa e confiável das rotas biológicas de um grupo de dados específico.

Foram analisados todos os transcritomas das cepas 9a5c e Temecula1, sendo considerados efetivamente expressos os genes que apresentaram valores de expressão de FPKM maiores que 10. Os resultados foram obtidos com a notação de que "1" significa presença da rota naquele transcritoma e "0" significa ausência da mesma (Tabela S6).

Analisando os dados com mais detalhes, pode-se observar a presença de rotas essenciais para crescimento e proliferação. Dentre elas, rotas do metabolismo energético, como glicólise/gliconeogênese, ciclo do ácido tricarboxílico (TCA), via das pentoses-fosfato e glioxilato, fosforilação oxidativa, metabolismo de piruvato e ubiquinona, entre outras. O metabolismo de outros carboidratos também foi observado como amido, sacarose, frutose e manose, embora não tenha sido encontrada a rota de metabolismo de galactose.

Todas as rotas de manutenção e transmissão da informação genética como replicação do DNA, transcrição, tradução, síntese de aminoacil-tRNAs estão presentes em todos os transcritomas. Além disso, os sistemas de reparo do DNA por excisão de bases ou nucleotídeos e por *mismatch* também estão presentes, assim como a rota de recombinação homóloga. O metabolismo de purinas e pirimidinas também foi encontrado em ambas as cepas. Porém a, rota de metabolismo de nucleotídeos de açúcar para síntese de precursores de gliconjugados, está ausente.

As rotas de exportação de proteínas e dos sistemas de secreção do tipo II e IV estão ativas em todos os transcritomas. Outras rotas de transporte também foram observadas como transportadores ABC, sistema de dois componentes e sistema de fosfotransferase de açúcar (PTS).

Rotas relacionadas à síntese de membranas e parede celular como o metabolismo de glicerofosfolipídeos, biossíntese de peptidioglicano e lipopolissacarídeo (LPS),

também estão ativas. Por outro lado, as rotas de biossíntese de ácidos graxos, saturados e insaturados, e esteroides também estão ativas nas duas cepas. Já as rotas de metabolismo de ácidos graxos e glicerolipídeos estão presentes em 9a5c mas ausentes em Temecula1.

Diversas vitaminas e coenzimas como tiamina, riboflavina, vitamina B6, biotina, porfirinas, ácido lipóico, nicotinato e nicotinamida, pantotenato e co-enzima A também tiveram suas rotas representadas nos transcritomas. Entretanto o metabolismo de ascorbato (vitamina C) não foi encontrado em Temecula1, somente em 9a5c, e a rota de metabolismo de inositol não foi encontrado em nenhuma delas.

Interessantemente, rotas para a biossíntese de antibióticos e poliquetídeos foram encontradas ativas como as de vancomicina, ansamicina, tetraciclina e estreptomicina. Não foram encontradas as rotas para biossíntese de penicilina cefalosporinas e de resistência a beta-lactâmicos.

Finalmente, as rotas de metabolismo de nitrogênio, enxofre, glutationa e grupos amino através do ciclo da uréia foram encontradas ativas. Entretanto, nem todas as rotas de biossíntese e metabolismo de aminoácidos estão presentes nos transcritomas. As rotas para metabolismo de alanina, aspartato, glicina, serina, treonina, arginina, prolina, histidina, fenilalanina e metionina estão ativas. Porém, as rotas de metabolismo de glutamato, cisteína, tirosina, glutamina e glutamato, seleno e cianoaminoácidos, estão inativas. Para os aminoácidos lisina, valina, leucina e isoleucina somente estão ativas as rotas de biossíntese, ao contrário de suas rotas de degradação. O metabolismo de triptofano foi encontrado ativo somente nas duas primeiras réplicas dos transcritomas do início da fase exponencial em PIM6 e PWG e na segunda réplica do final do crescimento em PWG, todos na cepa 9a5c. Nos outros transcritomas, esta rota está inativa. Coincidentemente, a rota de metabolismo de metano foi encontrada inativa nestes mesmos transcritomas acima mencionados, porém ativa nos demais. Ainda não é possível especular qualquer relação entre estes essas duas observações. A provável inatividade de rotas de síntese de alguns aminoácidos talvez seja uma justificativa para esta bactéria ter crescimento lento e dependente da suplementação destes aminoácidos no meio de cultivo.

Em resumo, os transcritomas das cepas 9a5c e Temecula1 mostraram poucas divergências em relação às suas rotas metabólicas ativas e mostraram que as rotas essenciais para seu crescimento e sobrevivência estão presentes.

4.8. Análises de expressão gênica diferencial

4.8.1. Comparações entre os transcritomas

Após a descrição dos genes mais expressos em um mesmo transcritoma baseada em valores de FPKM, foram realizadas comparações entre valores de expressão de um mesmo gene entre transcritomas distintos. Para tal, foi usado o pacote *DESeq2* (Love *et al.*, 2014), o qual exige como dados de entrada valores de contagem brutos de quantas sequências foram mapeadas para determinado gene. Não é recomendado utilizar valores de expressão já normalizados por FPKM, uma vez que o *pipeline* do pacote realiza uma normalização interna. A obtenção das listas de genes com contagem bruta também foi realizada pelo mapeamento das sequências no software *CLC Genomics Workbench*.

As análises comparativas foram realizadas em três níveis: 1) comparações entre transcritomas da mesma cepa e mesmo meio, em fases de crescimento diferentes, como por exemplo, entre os transcritomas da cepa 9a5c crescida em PIM6 no início e no final da fase exponencial; 2) comparações entre transcritomas da mesma cepa, em meios diferentes, mas em fases de crescimento equivalentes, como por exemplo, entre o transcritoma de 9a5c no início da fase exponencial em PIM6 com o transcritoma de 9a5c no início da fase exponencial em PWG; 3) comparações entre transcritomas de cepas diferentes, porém em mesmo meio e fase de crescimento, como por exemplo, entre o transcritoma de 9a5c e o transcritoma de Temecula1 ambas no início da fase exponencial em PIM6. Os parâmetros utilizados para considerar um gene como significativamente diferencialmente expresso foram um valor de *p*, ajustado pela correção FDR (*False Discovery Rate*) (Benjamini e Hochberg, 1995), menor do que 0,05 e um valor de razão de expressão (*fold change*) maior que 2. A Tabela 12 apresenta um resumo de quantos genes foram encontrados dentro desses parâmetros em cada comparação.

Como uma forma gráfica de mostrar uma visão geral da dispersão de todos os genes em cada transcritoma, o pacote *DESeq2* (Love *et al.*, 2014) também realiza plotagens, tendo os valores de expressão média entre todas as condições analisadas no eixo das abscissas e os seus respectivos valores de razão de expressão no eixo das ordenadas (Figura 8). Estes gráficos apresentam em pontos vermelhos os genes considerados diferencialmente expressos de acordo com os parâmetros acima citados e, em cinza, os genes cuja expressão não variou.

Níveis de	Total de	Condição 1	Condição 2	Genes DE*	Genes up-
comparação	genes	(c1)	(c2)	(p≤0,05 ; <i>FC</i> > 2,0)	regulated (c1/c2)
	2639	9a5c_PIM6_1d	9a5c_PIM6_3d	83 (3,1%)	79/4
1	2639	9a5c_PWG_3d	9a5c_PWG_10d	0	0
•	2476	Tem_PIM6_1d	Tem_PIM6_3d	61 (2,5%)	50/11
	2476	Tem_PWG_3d	Tem_PWG_7d	1 (0,04%)	0/1
	2639	9a5c_PIM6_1d	9a5c_PWG_3d	4 (0,15%)	0/4
2	2639	9a5c_PIM6_3d	9a5c_PWG_10d	20 (0,8%)	3/17
-	2476	Tem_PIM6_1d	Tem_PWG_3d	1 (0,04%)	1/0
	2476	Tem_PIM6_3d	Tem_PWG_7d	6 (0,24%)	2/4
	1882	9a5c_PIM6_1d	Tem_PIM6_1d	102 (5,4%)	54/48
3	1882	9a5c_PIM6_3d	Tem_PIM6_3d	238 (12,6%)	115/123
U U	1882	9a5c_PWG_3d	Tem_PWG_3d	58 (3,1%)	42/16
	1882	9a5c_PWG_10d	Tem_PWG_7d	95 (5,0%)	46/49

Tabela 12. Resumo do número de genes diferencialmente expressos (p<0,05; *fold change*>2) em todas as comparações.

*DE: diferencialmente expressos. Tem= Temecula1

No primeiro nível de comparação foi possível observar um maior número de genes diferencialmente expressos nas comparações entre transcritomas de células em meio PIM6 do que entre os transcritomas de células em PWG, para ambas as cepas. Dentre estes genes, a maioria está mais expressa no início da fase exponencial. Dos genes considerados diferencialmente expressos nas comparações das duas cepas, 31 são comuns entre elas, correspondendo a 37,3% e 50,8% nas cepas 9a5c e Temecula1, respectivamente. Todos estes genes estão mais expressos na mesma fase de crescimento nas duas cepas, sendo 28 genes no início do crescimento e 3 no final. Foram encontrados genes relacionados a proteínas ribossomais, chaperonas, biogênese dos pili longo e curto e genes com função hipotética sendo positivamente regulados no início da fase exponencial estão um gene de uma colicina V, de uma fumarase e um gene com função hipotética.

Estes resultados podem sugerir que as células cultivadas em meio mínimo enfrentam, em menor intervalo de tempo, mudanças no ambiente, por exemplo, um esgotamento dos nutrientes, e que resulta na alteração do transcritoma. As observações de uma maior atividade de síntese proteica associada à atividade de chaperonas corrobora uma resposta de adaptação a um ambiente estressante. Esta análise pode ser extrapolada ao ambiente natural da bactéria, a qual vive constantemente em um ambiente de escassez nutricional, seja no xilema da planta ou no cibário do inseto, tendo que rapidamente regular a expressão de seus genes para sobreviver.

Nas comparações entre transcritomas em meio PWG somente um gene, de uma colicina V, foi encontrado diferencialmente expresso na comparação de transcritomas da cepa Temecula1, com maior expressão no final da fase exponencial. Este gene é o mesmo encontrado como positivamente regulado no final do crescimento em PIM6 em ambas as cepas. O fato de somente um gene ser encontrado diferencialmente expresso nas comparações em PWG pode ser justificado devido às células em meio rico encontrarem um ambiente favorável desde o início do crescimento até o final da fase exponencial, não necessitando de alterações significativas na expressão de seus genes.

Nas comparações entre meios de cultivo, no segundo nível de análises, foi observado um maior número de genes diferencialmente expressos entre os transcritomas do final da fase exponencial, para ambas as cepas. Dentre estes genes considerados diferencialmente expressos, a maioria deles foi positivamente regulada nos transcritomas em meio PWG. Entretanto, nenhum gene foi comum entre as duas cepas. Na cepa 9a5c, vários genes relacionados a proteínas ribossomais, um gene responsável pela formação do poro para a externalização do pilus longo, um gene para uma proteína de transporte pela membrana, um gene para colicina V e três genes para bacteriocinas, além de genes com função hipotética, foram positivamente regulados no final da fase exponencial em PIM6 estão um gene para uma proteína translocase e dois genes com função hipotética. Para a cepa Temecula1, dois genes foram positivamente regulados no final do crescimento em PIM6, codificador de uma proteína ribossomal e uma proteína de um sistema de estabilização de plasmídeo. Por sua vez, no meio PWG quatro genes para proteínas hipotéticas foram identificados.

Dentre os poucos genes encontrados diferencialmente expressos nas comparações entre os transcritomas de fase inicial nos dois meios, três bacteriocinas e uma bacterioferritina foram positivamente reguladas em PWG na cepa 9a5c. O único gene diferencialmente expresso na cepa Temecula1 foi positivamente regulado no início do crescimento em PIM6 e codifica para uma proteína hipotética.

Em dois estudos utilizando a tecnologia de microarranjos para análise de expressão gênica, foram realizadas comparações dos transcritomas da cepa 9a5c cultivada em dois meios diferentes (Travensolo *et al.*, 2009; Ciraulo *et al.*, 2010). Assim como no presente estudo, um dos meios apresentava composição especificamente

desenvolvida para o melhor crescimento de Xylella em comparação com outros meios complexos (Lemos et al., 2003; Leite et al., 2004). Vários genes foram encontrados diferencialmente expressos entre as duas condições nos dois estudos, porém somente quatro genes em comum com as comparações aqui apresentadas. Travensolo e colaboradores compararam os transcritomas de 9a5c nos meios XDM₂ (meio definido) e BCYE (meio complexo), após cultivo de 4 dias a 30°C e agitação de 140 rpm. Somente o gene pilQ (XF0373), o qual codifica para a proteína formadora de poro na membrana para externalização do pilus longo, foi encontrado em comum com a comparação entre os transcritomas de 9a5c de final da fase exponencial em ambos os meios. Nos dois trabalhos, este gene apresentou-se mais expressos nos meios complexos, BCYE e PWG, embora com diferentes razões de expressão (1,83 e 3,73, respectivamente) (Travensolo et al., 2009). No segundo estudo, novamente a cepa 9a5c teve seus transcritomas analisados, agora nos meios 3G10-R (meio definido) e PW (meio complexo), após cultivo a 28°C e 100 rpm por 13 e 3 dias, respectivamente. Comparando os resultados deste trabalho com os genes diferencialmente expressos obtidos na comparação entre os transcritomas de início da fase exponencial do presente estudo, um gene codificando para uma bacterioferritina (XF0395) foi encontrado positivamente regulado nos meios complexos PW e PWG, com razões de expressão 2,33 e 6,15, respectivamente. Ao comparar com os genes diferencialmente expressos no final da fase exponencial, dois genes foram encontrados em comum. Um deles, codificando para uma proteína ribossomal (XF1164) mostrou concordância com o presente estudo, sendo mais expresso nos meios complexos com diferentes razões de expressão (1,88 para o PW e 5,66 para o PWG). Entretanto, outro gene, codificando para um colicina V, foi visto mais expresso no meio definido 3G10-R (razão de expressão igual a 17,15) e mais expresso no meio complexo PWG no presente estudo (razão de expressão igual a 4,92) (Ciraulo et al., 2010). Desta forma, sugere-se que embora as composições dos meios sejam semelhantes, quaisquer diferenças de nutrientes podem influenciar nas comparações entre eles, além de variações nas condições de cultivo, densidade celular, passagem da cepa utilizada, entre outros fatores, dificultando as comparações entre estudos.

A observação de uma maior quantidade de genes positivamente regulados em PWG poderia ser explicada pelo fato de que as células estariam metabolicamente mais ativas neste meio devido à abundância de nutrientes. Entretanto, o objetivo do presente trabalho era identificar possíveis genes com expressão aumentada no meio mínimo PIM6, o qual simula a condição nutricional do fluido xilemático do hospedeiro vegetal. Porém, somente foram encontrados três genes positivamente regulados neste meio em cada cepa, quando comparados com o meio PWG, o que nos leva a concluir que não há diferenças significativas na expressão dos genes nos dois meios de cultivo quando comparamos pontualmente fases de crescimento semelhantes. Por outro lado, há diferenças na expressão gênica ao longo do crescimento em cada meio, conforme descrito nos parágrafos acima.

No terceiro e último nível de comparação foram analisados os valores de expressão somente de genes compartilhados entre as cepas 9a5c e Temecula1, não levando em consideração genes exclusivos de cada uma delas. Nestas comparações, podemos observar um maior número de genes diferencialmente expressos nas comparações entre transcritomas de células cultivadas em PIM6, embora também tenha sido encontrado um número razoável de genes diferencialmente expressos nas comparações entre transcritomas de PWG. Dentre estes genes, nas comparações de início do crescimento em ambos os meios, foi observado um maior número de genes positivamente regulados na cepa 9a5c. Ao contrário, a cepa Temecula1 apresentou um maior número de genes positivamente regulados nas duas comparações de final de crescimento nos dois meios (Tabela 12).

Quando são comparados os genes diferencialmente expressos entre os transcritomas das duas cepas cultivadas em PIM6, observamos que 84 genes são comuns às comparações de início e final do crescimento. Tal quantidade de genes corresponde a 82,4% de todos os genes considerados diferencialmente expressos no início do crescimento e 35,3% no final. Destes genes em comum, 46 foram positivamente regulados na cepa 9a5c e 38 na cepa Temecula1. Todos os genes foram positivamente regulados para a mesma cepa nas duas fases de crescimento, ou seja, tais genes mantém sua expressão desde o início até o final do crescimento. Dentre os genes mais expressos em 9a5c, vários codificam proteínas com função hipotética, um gene para uma adesina da extremidade do pilus longo, genes relacionados a bacteriófagos, genes para proteínas de sistema de toxina/antitoxina e estabilização de plasmídeos, DNA polimerase I, componentes do sistema de transporte ABC, entre outros. Os genes regulados positivamente na cepa Temecula1 codificam para proteínas relacionadas principalmente a bacteriófagos e com função hipotética, embora também tenham sido encontrados genes relacionados à síntese do pilus curto, proteínas de membrana, a

desidrogenases, uma proteína ribossomal, proteínas de transporte pela membrana, entre outros.

Nas comparações entre os transcritomas das fases de crescimento em PWG foram observados 41 genes regulados durante todo o crescimento, correspondendo a 70,7% e 43,2% de todos os genes considerados diferencialmente expressos no início e no final do crescimento, respectivamente. Trinta genes foram encontrados mais expressos na cepa 9a5c e 11 na cepa Temecula1. Todos eles foram positivamente regulados para a mesma cepa nas duas fases de crescimento, mantendo sua expressão desde o início até o final. Dentre os genes positivamente regulados para a cepa 9a5c, há predominância de genes de proteínas relacionadas a bacteriófagos ou com função hipotética, embora deva ser destacada a maior expressão do gene da DNA polimerase I, de uma adesina da extremidade do pilus longo, uma proteína quinase relacionada à sinalização, entre outras. Enquanto isso, os genes positivamente regulados para a cepa Temecula1 foram quase todos de proteínas de fagos e com função hipotética, com exceção de uma aminotransferase e uma desidrogenase.

Este alto número de genes diferencialmente expressos que foi observado justifica outras comparações entre transcritomas de cepas diferentes, uma vez que mesmo os genes ortólogos desses genomas parecem apresentar regulação distinta na mesma condição de cultivo. Além disso, em todas as comparações entre cepas foi possível observar altos valores de razão de expressão (*fold change*) entre genes dos transcritomas comparados. Tal resultado demonstra que além de apresentar um maior número de genes diferencialmente expressos, as comparações entre cepas também apresentam maiores diferenças na expressão, sugerindo que cada cepa possui um panorama único de expressão gênica em resposta a determinada condição de cultivo.



Figura 8. Gráficos de dispersão dos genes analisados em cada comparação. No eixo das abscissas estão os valores de expressão média; no eixo das ordenadas estão os valores de razão de expressão em log2. Os pontos em vermelho representam genes com valores de *p* ajustado menores do que 0,05; os pontos em cinza representam genes com valores de *p* ajustado maiores que 0,05. Cada gráfico representa uma comparação: (A) 9a5c-PIM6-1d e 9a5c-PIM6-3d; (B) 9a5c-PWG-3d e 9a5c-PWG-10d; (C) 9a5c-PIM6-1d e 9a5c-PWG-3d; (D) 9a5c-PIM6-3d e 9a5c-PWG-10d; (E) Tem-PIM6-1d e Tem-PIM6-3d; (F) Tem-PWG-3d e Tem-PWG-7d; (G) Tem-PIM6-1d e Tem-PIM6-3d e Tem-PIM6-3d e Tem-PWG-7d; (I) 9a5c-PIM6-1d e Tem-PIM6-1d; (J) 9a5c-PIM6-3d e Tem-PIM6-3d; (K) 9a5c-PWG-3d e Tem-PWG-3d; (L) 9a5c-PIM6-1d e Tem-PWG-7d. Tem= Temecula1

4.8.2. Categorização funcional dos genes diferencialmente expressos

4.8.2.1. Análise por ontologia gênica

Foram realizadas análises globais funcionais dos genes considerados diferencialmente expressos (p<0,05e fold change>2), utilizando o software online BayGO (Vencio et al., 2006). Primeiramente, tais genes foram categorizados baseandose nas suas ontologias gênicas (Gene Ontology ou GO) (Ashburner et al., 2000; Gene Ontology Consortium, 2004).

Iniciamos as análises com os transcritomas do nível 1 de comparação. Nas fases iniciais do crescimento exponencial de ambas as cepas em meio PIM6 quando comparado com sua fase final, foi possível observar um enriquecimento de categorias *GO* relacionadas à transcrição e tradução (Tabela 13). Dentre as categorias em comum estão biossíntese de proteínas, biogênese e montagem do ribossomo e transcrição pela atividade de RNA polimerase. A cepa 9a5c apresentou como categorias exclusivas algumas relacionadas ao enovelamento proteico, atividade de fatores de elongação da tradução, ligação a tRNA, atividade de nucleotidiltransferase, transporte proteico intracelular e secreção de proteínas. Por sua vez, a cepa Temecula1 apresentou duas categorias exclusivas enriquecidas relacionadas à atividade de metalopeptidades e metaloendopeptidases. Entretanto, na fase final do crescimento exponencial não foi encontrada nenhuma categoria enriquecida nas comparações das duas cepas.

Na comparação entre as diferentes fases de crescimento no meio PWG na cepa 9a5c, já havíamos mencionado que não foi encontrado nenhum gene diferencialmente expresso. Na cepa Temecula1, foi encontrado apenas um gene.

ente transcritomas de diferentes fases de cresentento has cepas yase e renceutar.								
Categorias GO u	Categorias GO up-regulated no início da fase exponencial em meio PIM6 na cepa 9a5c							
Código GO	Descrição	Р	G^*	G90% IC				
GO:0006457	protein folding	0,01	0,61	[0,61;0,61]				
GO:0051082	unfolded protein binding	0,01	0,74	[0,74;0,74]				
GO:0003723	RNA binding	0	0,82	[0,82;0,82]				
GO:0006412	protein biosynthesis	0	0,89	[0,88;0,90]				
GO:0003746	translation elongation factor activity	0,03	0,81	[0,81;0,81]				
GO:0006414	translational elongation	0	0,86	[0,86;0,86]				
GO:0003735	structural constituent of ribosome	0	0,95	[0,95;0,96]				
GO:0005622	Intracelular	0	0,87	[0,86;0,88]				
GO:0005840	Ribosome	0	0,95	[0,95;0,96]				
GO:0030529	ribonucleoprotein complex	0	0,95	[0,95;0,95]				
GO:0000049	tRNA binding	0	0,88	[0,88;0,88]				
GO:0015935	small ribosomal subunit	0	0,99	[0,99;0,99]				
GO:0019843	rRNA binding	0	0,96	[0,96;0,96]				

Tabela 13. Caracterização funcional baseada em categorias *GO* dos genes diferencialmente expressos entre transcritomas de diferentes fases de crescimento nas cepas 9a5c e Temecula1.

GO:0003899	DNA-directed RNA polymerase activity	0	0,91	[0,91;0,91]			
GO:0006350	Transcription	0	0,83	[0,83;0,83]			
GO:0006351	transcription, DNA-dependent	0	0,97	[0,97;0,97]			
GO:0016779	nucleotidyltransferase activity	0,02	0,68	[0,68;0,68]			
GO:0006986	response to unfolded protein	0,02	0,82	[0,82;0,82]			
GO:0045449	regulation of transcription	0,04	0,74	[0,70;0,80]			
GO:0006886	intracellular protein transport	0	0,90	[0,90;0,90]			
GO:0009306	protein secretion	0,04	0,64	[0,64;0,64]			
GO:0015450	protein translocase activity	0	0,90	[0,90;0,90]			
GO:0015934	large ribosomal subunit	0	0,93	[0,93;0,93]			
GO:0042254	ribosome biogenesis and assembly	0	1,00	[1,00;1,00]			
Categorias GO up-regulated no início da fase exponencial em meio PIM6 na cepa Temecula1							
Código GO	Descrição	P	G^*	G90% IC			
GO:0003723	RNA binding	0	0,78	[0,75;0,81]			
GO:0006412	protein biosynthesis	0	0,89	[0,87;0,90]			
GO:0003735	structural constituent of ribosome	0	0,94	[0,93;0,94]			
GO:0005622	Intracelular	0	0,85	[0,83;0,87]			
GO:0005840	Ribosome	0	0,94	[0,92;0,94]			
GO:0030529	ribonucleoprotein complex	0	0,92	[0,92;0,93]			
GO:0015935	small ribosomal subunit	0,01	0,82	[0,82;0,82]			
GO:0019843	rRNA binding	0	0,90	[0,90;0,90]			
GO:0003899	DNA-directed RNA polimerase activity	0	0,82	[0,82;0,82]			
GO:0006350	Transcription	0,04	0,73	[0,67;0,83]			
GO:0006351	transcription, DNA-dependent	0	0,94	[0,94;0,94]			
GO:0004222	metalloendopeptidase activity	0,04	0,70	[0,70;0,70]			
GO:0042254	ribosome biogenesis and assembly	0,01	1,00	[1,00;1,00]			
GO:0008237	metallopeptidase activity	0,03	0,80	[0,80;0,80]			
GO:0006364	rRNA processing	0,03	0,72	[0,72;0,72]			

*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").

A categorização funcional dos genes diferencialmente expressos também foi realizada nas comparações de nível 2, entre os meios de cultivo e em uma mesma fase do crescimento bacteriano. Dentre todas as comparações, somente foram observadas categorias enriquecidas no final do crescimento em PWG na cepa 9a5c quando comparado com o final do crescimento em PIM6 (Tabela 14). Tais categorias estão relacionadas exclusivamente a atividade de síntese e transporte de proteínas, incluindo categorias de formação de ribossomos.

Tabela	14.	Caracterizaçã	ío funciona	l baseada	em	categorias	GO	dos	genes	diferencialmente	expressos
entre tra	anscr	ritomas de dif	erentes mei	os de cult	ivo 1	na cepa 9a5	c.				

Categorias GO up-regulated no meio PWG no final da fase exponencial na cepa 9a5c							
Código GO	Descrição	Р	G^*	G90% IC			
GO:0003723	RNA binding	0	0,90	[0,90;0,90]			
GO:0006412	protein biosynthesis	0	0,97	[0,96;0,97]			
GO:0003735	structural constituent of ribosome	0	0,98	[0,98;0,99]			
GO:0005622	Intracelular	0	0,97	[0,96;0,97]			
GO:0005840	Ribosome	0	0,98	[0,97;0,99]			
GO:0030529	ribonucleoprotein complex	0	0,98	[0,98;0,99]			
GO:0019843	rRNA binding	0,01	0,96	[0,96;0,96]			
GO:0008565	protein transporter activity	0	0,85	[0,83;0,90]			

*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").

Como descrito anteriormente, no último nível de comparação envolvendo as duas cepas nas mesmas condições foi encontrado um maior número de genes diferencialmente expressos, os quais também foram categorizados funcionalmente (Tabela 15). Na comparação entre transcritomas do início do crescimento em PIM6, a cepa 9a5c apresentou um enriquecimento de categorias relacionadas à ligação, integração e recombinação de DNA, além de atividades de helicase e ligase e uma categoria de fosforilação de aminoácidos. Por sua vez, a cepa Temecula1 na mesma fase de crescimento em PIM6 mostrou enriquecimento de categorias relacionadas à síntese e transporte de proteínas. Ainda no meio PIM6, mas na comparação de transcritomas no final da fase exponencial, foram encontradas enriquecidas categorias relacionadas à atividade de metiltransferases e metilação de DNA na cepa 9a5c, além novamente das categorias de fosforilação de aminoácidos e integração de DNA, também enriquecidas no início do seu crescimento. Enquanto isso, a cepa Temecula1 apresentou o maior número de categorias enriquecidas, com 28 no total. Elas se dividem em quatro importantes processos celulares: transcrição (atividade de RNA polimerase, regulação da transcrição, atividade de fator sigma), metabolismo de ATP (transporte de prótons/hidrogênio, atividade de ATP sintase e ATPase), defesa (atividade de lisozima, citólise, catabolismo de peptidioglicano e parede celular, resposta de defesa contra bactérias) e síntese proteica (biogênese de ribossomo, enovelamente e transporte). O maior número de categorias enriquecidas na cepa Temecula1 significa uma atividade metabólica mais intensa no meio PIM6, possivelmente para lidar com o estresse nutricional encontrado neste meio. De fato, o meio PIM6 foi desenvolvido baseado na composição nutricional do fluido xilemático de videiras na América do Norte (Michelle Igo, comunicação pessoal), o que pode justificar uma melhor adaptação desta cepa neste meio. De qualquer forma, o uso do PIM6 para o crescimento de outras cepas que não tenham videiras como hospedeiros não é inviabilizada, devido ao fato de ainda constituir um meio pobre em nutrientes e que se assemelha com composições de fluidos xilemáticos de outras plantas.

Tabela 15. Caracterização funcional baseada em categorias *GO* dos genes diferencialmente expressos entre transcritomas em meio PIM6.

Categorias GO I				
Código GO	Descrição	Р	G^*	G90% IC
GO:0005524	ATP binding	0,01	0,49	[0,41;0,60]
GO:0003677	DNA binding	0,01	0,72	[0,60;0,81]
GO:0006310	DNA recombination	0,02	0,85	[0,78;0,92]
GO:0003676	nucleic acid binding	0,04	0,62	[0,51;0,77]

GO:000/386	holicano activity	0	0.70	[0 74:0 86]	
GO:0004380	netcuse uctivity	0	0,79	[0,74,0,80]	
CO:0016874	ligase activity	0.04	0,00	[0, 30, 0, 30]	
GO:0015074	DNA integration	0,04	0,48	[0,40,0,32]	
Cotogonias CO	UNA integration		0,98	[0,94,0,99]	
Categorias GO	Deservição	D	C^*	C000/ IC	
	Descrição	<u>r</u>	<u>G*</u>	[0 20:0 60]	
GO:0000412	protein diosynthesis	0,05	0,45	[0, 39; 0, 60]	
GO:0003077	DNA binaing	0	0,62	[0,49;0,77]	
GO:0005733	structural constituent of ribosome	0,01	0,00	[0,00;0,73]	
GO:0005022	Diharawa	0,01	0,00	[0, 54; 0, 70]	
GO:0005840	Ribosome	0,01	0,00	[0,60;0,69]	
GO:0030329	Protoin turnen out activity	0	0,07	[0, 60; 0, 75]	
GU:0008363		0,01	0,72	[0,00;0,82]	
Categorias GO	<i>up-regulated</i> na <u>cepa 9a5c</u> no <u>final</u> da fase exponencial		G th		
Código GO	Descrição	P	<i>G</i> *	G90% IC	
GO:0000287	magnesium ion binding	0	0,52	[0,52;0,52]	
GO:0006306	DNA-methylation	0,02	0,95	[0,76;0,99]	
GO:0008170	N-methyltransferase activity	0,04	0,95	[0,78;0,99]	
GO:0009007	Site-specific DNA-methyltransferase (adenine-specific)	0	1,00	[1,00;1,00]	
	activity				
GO:0006468	Protein amino acid phosphorylation	0,04	0,53	[0,53;0,53]	
GO:0015074	DNA integration	0,03	0,91	[0,77;0,97]	
Categorias GO	<i>up-regulated</i> na <u>cepa Temecula1</u> no <u>final</u> da fase exponen	cial			
Código GO	Descrição	P	G^*	G90% IC	
GO:0003677	DNA binding	0,04	0,41	[0,29;0,52]	
GO:0003735	structural constituent of ribosome	0	0,58	[0,53;0,63]	
GO:0005622	Intracelular	0	0,47	[0,39;0,55]	
GO:0005840	Ribosome	0	0,58	[0,53;0,64]	
GO:0030529	ribonucleoprotein complex	0	0.60	[0.56:0.67]	
GO:0015935	small ribosomal subunit	0	0.73	[0,73:0,73]	
GO:0019843	rRNA binding	0	0.68	[0.64:0.73]	
GO:0003899	DNA-directed RNA polymerase activity	0	0.84	[0.84:0.84]	
GO:0006350	Transcription	0.04	0.59	[0.51:0.72]	
GO:0006355	regulation of transcription DNA-dependent	0	0.50	[0, 40:0, 59]	
GO:0003796	lysozyme activity	Õ	1.00	[0, 92.1, 00]	
GO:0009253	nentidoolycan catabolism	0 04	0.90	[0,76:0.96]	
GO:0016998	cell wall catabolism	0.02	0,90	[0,76;0,96]	
GO:0019835	Cytalysis	0.02	0,90	[0, 76; 0, 93]	
GO:0012033	defense response to hacteria	0,02	1.00	[0, 70, 0, 0, 90]	
GO:0042742	response to unfolded protein	0,02	0.60	[0, 50, 1, 00]	
GO:0000980	ATP synthesis coupled proton transport	0	0,09	[0,09,0,09]	
GO:0015380	All synthesis coupled proton transport	0 02	0,73	[0, 73, 0, 73]	
CO:0010403	hydrogen transporting ATP synthesis activity rotational	0,02	0,09	[0,09,0,09]	
00.0040955	nyarogen-transporting ATF synthase activity, rotational	0,02	0,75	[0,75,0,75]	
CO:0046061	mechanism hydrogen transporting ATPage activity rotational	0.02	0.72	[0 72:0 72]	
00.0040901	nyarogen-transporting ATT use activity, rotational	0,02	0,75	[0,75,0,75]	
CO.0006252	mechanism thereasing initiation	0	0.00	[0,0,0,0,00]	
GO:0006352		0	0,90	[0,90;0,90]	
GO:0016987	sigma fator activity	0,02	0,90	[0,90;0,90]	
GO:0015031	protein transport	0	0,01	[0,01;0,61]	
GO:0015934	large ribosomal subunit	0	0,77	[0, 7, 0, 7, 7]	
GO:0008462	enaopeptidase Cip activity	0	1,00	[0,95;1,00]	
GO:0015992	proton transport	0,02	0,73	[0, 73; 0, 73]	
GO:0016820	nyarolase activity, acting on acid anhydrides, catalyzing	0	0,90	[0,90;0,90]	
00.0015050	transmembrane movement of substances	0.01	0 77		
GO:0015078	hydrogen ion transporter activity	0,01	0,77	[0,77;0,77]	
*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").					

Os transcritomas das duas cepas crescidas em meio PWG também apresentaram diferenças. No início da fase exponencial somente a cepa 9a5c apresentou categorias enriquecidas, mostrando maior atividade de ligação a DNA e ácidos nucleicos, ligação de ATP, atividade de helicase e fosforilação de aminoácidos. Ao final do crescimento exponencial no meio PWG, pudemos encontrar quase as mesmas categorias enriquecidas no início em 9a5c, apenas substituindo fosforilação de aminoácidos por atividade de ligase. Enquanto isso, a cepa Temecula1 novamente mostrou alta atividade metabólica no final do crescimento, desta vez em PWG. Dentre as categorias super-representadas estão o metabolismo de carbohidratos, transcrição e, novamente, defesa (atividade de lisozima, citólise, catabolismo de peptidioglicano e parede celular, resposta de defesa contra bactérias), as mesmas encontradas no final do crescimento em PIM6 (Tabela 16).

Tabela 16. Caracterização funcional baseada em categorias *GO* dos genes diferencialmente expressos entre transcritomas em meio PWG.

entre transertomas em meio r w G.							
Categorias GO	Categorias GO up-regulated na cepa 9a5c no início da fase exponencial						
Código GO	Descrição	P	G^*	G90% IC			
GO:0005524	ATP binding	0	0,74	[0,67;0,80]			
GO:0003677	DNA binding	0,03	0,62	[0,46;0,80]			
GO:0003676	Nucleic acid binding	0	0,77	[0,69;0,84]			
GO:0004386	Helicase activity	0	0,87	[0,83;0,92]			
GO:0006468	Protein amino acid phosphorylation	0	0,92	[0,92;0,92]			
Categorias GO	up-regulated na <u>cepa 9a5c</u> no <u>final</u> da fase exponencial						
Código GO	Descrição	P	G^*	G90% IC			
GO:0005524	ATP binding	0	0,69	[0,60;0,77]			
GO:0003677	DNA binding	0,01	0,70	[0,61;0,84]			
GO:0003676	Nucleic acid binding	0	0,74	[0,66;0,85]			
GO:0004386	Helicase activity	0	0,86	[0,82;0,91]			
GO:0016874	Ligase activity	0	0,63	[0,55;0,67]			
Categorias GO	<i>up-regulated</i> na <u>cepa Temecula1</u> no <u>final</u> da fase exponen	cial					
Código GO	Descrição	P	G^*	G90% IC			
GO:0003677	DNA binding	0,03	0,64	[0,50;0,77]			
GO:0003824	Catalytic activity	0	0,51	[0,44;0,66]			
GO:0005975	Carbohydrate metabolismo	0,03	0,70	[0,57;0,84]			
GO:0016798	Hydrolase activity, acting on glycosyl bonds	0	0,90	[0,83;0,95]			
GO:0008152	Metabolismo	0,03	0,52	[0,44;0,64]			
GO:0003899	DNA-directed RNA polymerase activity	0	0,93	[0,93;0,93]			
GO:0016779	Nucleotidyltransferase activity	0	0,78	[0,72;0,85]			
GO:0003796	Lysozyme activity	0	1,00	[0,98;1,00]			
GO:0009253	Peptidoglycan catabolism	0	0,98	[0,94;0,99]			
GO:0016998	Cell wall catabolism	0	0,98	[0,94;0,99]			
GO:0019835	Cytolysis	0	0,97	[0,94;0,98]			
GO:0042742	Defense response to bactéria	0	1,00	[0,98;1,00]			

*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").

4.8.2.2. Análise por vias KEGG

Os genes considerados diferencialmente expressos também foram categorizados em relação as suas vias metabólicas que estariam mais representadas e enriquecidas (vias KEGG, *Kyoto Encyclopedia of Genes and Genomes*) (Kanehisa e Goto, 2000; Kanehisa *et al.*, 2016).

No primeiro nível de comparação, entre fases de crescimento no mesmo meio, somente foram encontradas vias enriquecidas no início de crescimento em PIM6. Dentre elas estão vias relacionadas ao processamento da informação genética, tradução, ribossomo e RNA polimerase em ambas as cepas. Uma via relacionada ao exporte e enovelamento de proteínas foi encontrada enriquecida na cepa 9a5c, enquanto uma via relacionada a fatores de tradução na cepa Temecula1 (Tabela 17).

Tabela 17. Caracterização funcional baseada em vias KEGG dos genes diferencialmente expressos entre transcritomas de diferentes fases de crescimento nas cepas 9a5c e Temecula1.

Vias KEGG up-regulated no início da fase exponencial em meio PIM6 na cepa 9a5c						
Código KEGG	Descrição	P	<i>G</i> *	G90% IC		
1200	genetic information processing	0	0,85	[0,84;0,86]		
1220	Translation	0	0,82	[0,81;0,83]		
3100	protein folding and associated processing	0,02	0,71	[0,71;0,71]		
3010	Ribosome	0	0,93	[0,93;0,94]		
3020	RNA polymerase	0,01	0,85	[0,85;0,85]		
Vias KEGG up-	<i>regulated</i> no <u>início</u> da fase exponencial em <u>meio PIM6</u> na	cepa <u>T</u>	'emecul	a <u>1</u>		
Código KEGG	Descrição	P	G^*	G90% IC		
1200	genetic information processing	0	0,90	[0,87;0,91]		
1220	Translation	0	0,90	[0,88;0,91]		
3014	other translation factors	0,02	0,74	[0,74;0,74]		
3010	Ribosome	0	0,93	[0,92;0,94]		
3020	RNA polymerase	0,02	0,78	[0,78;0,78]		

*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").

As vias KEGG também foram classificadas nas comparações entre meios de cultivo, sendo encontradas vias enriquecidas somente no final do crescimento em PWG na cepa 9a5c. Três vias foram encontradas, relacionadas a processamento da informação genética, tradução e ribossomo (Tabela 18).

Tabela 18. Caracterização funcional base	ada em vias KEGG	dos genes diferenci	ialmente expressos entre
transcritomas de diferentes meios de culti	o na cepa 9a5c.		

Vias KEGG <i>up-regulated</i> no <u>meio PWG</u> no <u>final</u> da fase exponencial na cepa <u>9a5c</u>							
Código KEGG	Descrição	Р	G^*	G90% IC			
1200	genetic information processing	0,01	1,00	[1,00;1,00]			
1220	Translation	0,01	0,93	[0,90;0,93]			
3010	ribosome	0	0,97	[0,97;0,97]			

*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").

Nas comparações entres cepas, algumas vias foram encontradas enriquecidas. O início do crescimento em meio PIM6 estimulou na cepa 9a5c o enriquecimento da via de metabolismo de carboidratos, enquanto que na cepa Temecula1 foi observado um enriquecimento das vias de metabolismo de lipídeos, tradução e ribossomo. Já no final do crescimento em PIM6, a cepa 9a5c mostrou uma maior expressão de genes de vias relacionadas ao metabolismo de aminoácidos (fenilalanina, tirosina, triptofano e histidina), transportadores ABC e a via das pentoses-fosfato. Por sua vez, a cepa Temecula1 manteve um enriquecimento de vias de tradução e ribossomo, além de vias relacionadas com transcrição, síntese de ATP, enovelamento proteico, metabolismo de tirosina e transportadores acoplados a íons (Tabela 19).

Tabela 19. Caracterização funciona	l baseada em vias KEGG	dos genes diferencialmente e	xpressos entre
transcritomas em meio PIM6.			

Vias KEGG <i>up-regulated</i> na <u>cepa 9a5c</u> no <u>início</u> da fase exponencial								
Código KEGG	Descrição	P	G^*	G90% IC				
1110	Carbohydrate metabolism	0	0,70	[0,58;0,77]				
Vias KEGG up-1	<i>regulated</i> na <u>cepa Temecula1</u> no <u>início</u> da fase exponencia	al						
Código KEGG	Descrição	P	G^*	G90% IC				
1130	lipid metabolism	0,02	0,74	[0,67;0,82]				
1220	Translation	0,01	0,60	[0,55;0,70]				
3010	Ribosome	0	0,79	[0,77;0,85]				
Vias KEGG up-r	<i>regulated</i> na <u>cepa 9a5c</u> no <u>final</u> da fase exponencial							
Código KEGG	Descrição	P	<i>G</i> *	G90% IC				
1160	metabolism of other amino acids	0,01	0,43	[0,43;0,43]				
400	phenylalanine, tyrosine and tryptophan biosynthesis	0,04	0,55	[0,46;0,66]				
30	pentose phosphate pathway	0	0,74	[0,74;0,74]				
2010	ABC transporters, prokaryotic	0,03	0,56	[0,56;0,56]				
340	histidine metabolismo	0	0,80	[0,80;0,80]				
Vias KEGG up-1	<i>regulated</i> na <u>cepa Temecula1</u> no <u>final</u> da fase exponencial	l						
Código KEGG	Descrição	Р	G^*	G90% IC				
1200	genetic information processing	0	0,53	[0,46;0,59]				
1220	Translation	0,02	0,46	[0,41;0,51]				
1210	Transcription	0,01	0,61	[0,56;0,69]				
350	tyrosine metabolism	0,01	0,93	[0,86;0,97]				
3100	protein folding and associated processing	0	0,70	[0,67;0,75]				
3010	Ribosome	0	0,73	[0,70;0,77]				
2052	other ion-coupled transportes	0,03	0,57	[0,57;0,57]				
193	ATP synthesis	0,02	0,81	[0,81;0,81]				
3020	RNA polymerase	0	0,89	[0,89;0,89]				

*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").

Nas comparações do crescimento em PWG foram observadas somente duas vias enriquecidas, a de processamento da informação genética no início do crescimento na cepa 9a5c e a de metabolismo de lipídeos no final do crescimento na cepa Temecula1 (Tabela 20).

Vias KEGG up-regulated na cepa 9a5c no início da fase exponencial									
Código KEGG	Descrição	P	G^*	G90% IC					
1200	genetic information processing	0	1,00	[1,00;1,00]					
Vias KEGG up-r	<i>regulated</i> na <u>cepa Temecula1</u> no <u>final</u> da fase exponencia	ıl							
Código KEGG	Descrição	P	G^*	G90% IC					
1130	lipid metabolismo	0	0,78	[0,74;0,85]					

Tabela 20. Caracterização funcional baseada em vias KEGG dos genes diferencialmente expressos entre transcritomas em meio PWG.

*G é a medida gamma de associação estatística e G90% seu intervalo de credibilidade 90% ("error-bar").

4.8.2.3. Análise de genes relacionados à virulência e patogenicidade

Foi realizada uma análise dos níveis de expressão de genes relacionados à virulência em *X. fastidiosa* (Simpson *et al.*, 2000; Van Sluys *et al.*, 2003; Chatterjee *et al.*, 2008) eventualmente incluídos entre os genes diferencialmente expressos identificados nas comparações anteriormente descritas (Tabela 21).

No primeiro nível de comparação, somente foi possível observar genes diferencialmente expressos no início do crescimento em meio PIM6. Dentre eles, destacam-se genes relacionados com a biogênese do pilus longo, responsável pela motilidade twitching da bactéria (Meng et al., 2005), como o gene pilV em ambas as cepas. Já, os genes pilO e pilQ foram mais expressos somente na cepa 9a5c. Tal resultado sugere que um contato inicial com o xilema promoveria a expressão de genes relacionados à motilidade. O gene fimA relacionado com a formação de pilus curto (tipo chaperone-usher) (Li et al., 2007), responsável pela adesão ao substrato (parede dos vasos do xilema ou exoesqueleto do tubo digestório do inseto vetor), também foi diferencialmente expresso nas duas cepas analisadas. Outro gene observado em comum entre as comparações de 9a5c e Temecula1 codifica para uma protease dependente de zinco com função de chaperona. Dentre os genes diferencialmente expressos em somente uma das cepas estão os que codificam para a proteína regulatória LexA, um fator sigma 70, duas proteases e uma proteína de membrana externa na cepa 9a5c; na cepa Temecula1, foi encontrada uma oligopeptidase dependente de zinco e uma flavodoxina multimérica (WrbA).

Nas comparações entre as fases de crescimento em meio PWG, não foi encontrado nenhum gene diferencialmente expresso relacionado com a virulência da bactéria. Tal observação pode refletir o menor número de genes diferencialmente expressos no meio PWG em comparação com o meio PIM6. Obviamente, não é possível afirmar que tais genes de virulência não estão ativos no meio rico em nutrientes, somente é inferir que eles não apresentam mudança de expressão entre as duas fases de crescimento testadas, ao contrário do que acontece no meio PIM6.

No segundo nível de comparação, entre os transcritomas dos dois meios de cultivo, somente foi possível observar dois genes diferencialmente expressos relacionados à virulência, ambos nas comparações de final do crescimento. Na cepa 9a5c, o gene *pilQ*, responsável pela externalização do pilus longo, foi encontrado mais expresso no meio PWG. Na cepa Temecula1, transcritos de uma proteína regulatória relacionada com estabilização de ribossomos/ modulação do σ^{54} estariam mais abundantes no meio PIM6. Esta proteína tem sido relacionada a resposta a estresse ambientais e a virulência (Oosthuizen *et al.*, 2002; Ancona *et al.*, 2014). Nas comparações de início da fase exponencial, não foram observadas diferenças de expressão destes genes entre os meios de cultivo nas duas cepas analisadas.

No terceiro nível de comparação, entre os transcritomas das duas cepas nas mesmas condições, somente foram analisados os genes presentes em ambas. No início do crescimento em meio PIM6, foram encontrados cinco genes diferencialmente expressos. Dentre eles, um gene de uma adesina da extremidade do pilus longo (*PilY1*), uma histidina quinase e uma proteína com função regulatória (AraC-type) foram mais expressos na cepa 9a5c. Enquanto isso, a cepa Temecula1 mostrou maior expressão de um gene de uma proteína com domínio de clivagem do tipo pré-pilina (relacionada ao pilus longo) e de uma lipase. Todos estes genes mantiveram suas expressões diferenciais no final do crescimento em PIM6, inclusive no que diz respeito em qual cepa estariam mais expressos. Outros genes foram encontrados diferencialmente expressos nesta comparação de final do crescimento em PIM6. Na cepa 9a5c, foram observados os genes de uma glicosidade (enzima degradadora de parede vegetal), um regulador transcricional (família LuxR/UhpA), uma flavodoxina multimérica (WrbA), outro regulador transcricional e um gene para hemolisinas. Na cepa Temecula1, os genes de RpfG e RpfF foram mais expressas no final do crescimento em PIM6. A primeira seria um regulador de resposta sensível às concentrações de DSF e atuando como uma fosfodiesterase, degradando o segundo mensageiro intracelular c-di-GMP. Já *RpfF* seria responsável pela própria síntese de DSF (Chatterjee *et al.*, 2008). Juntamente com os genes do *cluster rpf*, foi observado um gene de diguanilato ciclase regulado positivamente em Temecula 1 quando comparada a cepa 9a5c. Em X. fastidiosa, o acúmulo de DSF (molécula sinalizadora produzida por genes deste cluster) modula negativamente os níveis de c-di-GMP, resultando em maior formação de biofilme

(Chatterjee *et al.*, 2010). Outros genes encontrados positivamente regulados em Temecula1 foram dois genes de fatores sigma (σ 70 e RpoH), de regulação de nitrogênio, subunidade ClpP da protease Clp, uma protease dependente de zinco com função de chaperona e uma proteína associada com grânulos de poli-hidroxialcanoatos.

As diferenças de expressão gênica nos transcritomas de células cultivadas em PWG foram poucas. Tanto no início da fase exponencial como no seu final, foi possível observar na cepa 9a5c uma maior expressão de um gene codificando uma adesina da extremidade do pilus longo (*PilY1*) e outro para uma histidina quinase relacionada a transdução de sinal. A cepa 9a5c também teve um gene para hemolisina mais expresso no início do crescimento. A cepa Temecula1 apenas observamos uma maior expressão de um gene de uma lipases no final do crescimento em PWG.

Estas observações sugerem diferenças relevantes na virulência de cada cepa e, possivelmente na resposta da bactéria ao ambiente do xilema, e poderão nortear futuros estudos sobre sistemas ou genes que foram destacados nas análises comparativa dos transcritomas que foram detalhadas aqui.

Anotação	Arrata año DAC	Descrição	Fold change		
XF/PD	Anotação IMG	Descrição	(<i>log2</i>)		
9a5c_PIM6	_1d x 9a5c_PIM	6_3d			
Biogênese d	o pilus longo				
XF0029	XF9a_00026	type IV pilus modification protein PilV	2,13		
XF0371	XF9a_00335	Tfp pilus assembly protein PilO	1,79		
XF0373	XF9a_00337	type IV pilus secretin (or competence protein) PilQ	1,73		
Biogênese d	o pilus curto				
XF0083	XF9a_00074	P pilus assembly protein, pilin FimA	2,04		
Funções reg	gulatórias				
XF0122	XF9a_00107	SOS regulatory protein LexA (EC:3.4.21.88)	1,64		
XF1350	XF9a_01234	RNA polymerase sigma factor, sigma-70 family	1,51		
Proteases					
XF0453	XF9a_00409	Membrane protease subunits, stomatin/prohibitin homologs (EC:3.4)	1,58		
XF1484	XF9a_01357	ATP-dependent protease HslVU, peptidase subunit (EC:3.4.25.2)	2,14		
XF2625	XF9a_02481	Zn-dependent protease with chaperone function	2,84		

 Tabela 21. Lista com genes diferencialmente expressos em todas as comparações e que estão relacionados com sistemas de virulência de X. fastidiosa.

		(EC:3.4.24)					
Outras fun	ções						
	3	Outer membrane protein and related					
XF0343	XF9a_00315	peptidoglycan-associated (lipo)proteins	1,91				
Temecula1	PIM6 1d x Tem	ecula1 PIM6 3d					
Biogênese d	lo pilus longo						
PD0020	XFTem_00023	type IV pilus modification protein PilV	2,30				
Biogênese d	lo pilus curto						
PD0062	- XFTem_00068	P pilus assembly protein, pilin FimA	2,28				
Funções re	gulatórias						
PD0422	XFTem_00475	Multimeric flavodoxin WrbA	1,99				
Proteases							
PD0856	XFTem_00970	Zn-dependent oligopeptidases (EC:3.4.15.5)	2,08				
		Zn-dependent protease with chaperone function	0.55				
PD1995	XFTem_02316	(EC:3.4.24)	2,57				
9a5c_PIM6	6_3d x 9a5c_PW0	5_10d					
Biogênese d	lo pilus longo						
		type IV pilus secretin (or competence protein)	1.0.6				
XF0373	XF9a_00337	PilQ	-1,86				
Temecula1	Temecula1_PIM6_3d x Temecula1_PWG_7d						
Funções re	gulatórias						
DD0(2)	VET 00722	ribosomal subunit interface protein	2.24				
PD0636	XF1em_00723	(sigma fator 54)	2,24				
9a5c_PIM6	5_1d x Temecula1	L_PIM6_1d					
Biogênese d	lo pilus longo						
XF1224	XF9a_01129	Tfp pilus assembly protein, tip-associated adhesin	1.07				
PD0502	XFTem_00565	PilY1	1,90				
XF2539	XF9a_02403	prepilin-type N-terminal cleavage/methylation	0.25				
PD1924	XFTem_02236	domain (PilA)	-2,35				
Funções re	gulatórias						
XF1254	XF9a_01154	AraC-type DNA-binding domain-containing	1.00				
PD0520	XFTem_00591	proteins	1,99				
XF2535	XF9a_02399	Sional transduction high dia a hin and	2.44				
PD1920	XFTem_02232	Signal transauction histiatine kinase	2,44				
Lipases							
XF2151	XF9a_02036	hypothetical protain (lingsa)	2 62				
PD1211	XFTem_01402	nypoinencai protein (upase)	-2,02				
9a5c_PIM6	5_3d x Temecula	L_PIM6_3d					
Enzimas de	egradadoras de pa	rede vegetal (CWDEs)					
XF0845	XF9a_00767	Beta-glucosidase-related glycosidases	1,63				

PD1829	XFTem_02128	(EC:3.2.1.21)	
Biogênese o	lo pilus longo		
XF1224	XF9a_01129	Tfp pilus assembly protein, tip-associated adhesin	2 50
PD0502	XFTem_00565	PilY1	2,50
XF2539	XF9a_02403	prepilin-type N-terminal cleavage/methylation	-2 94
PD1924	XFTem_02236	domain (PilA)	2,74
Cluster rpf			
XF1113	XF9a_01026	Response regulator containing a CheY-like	1 57
PD0405	XFTem_00457	receiver domain and an HD-GYP domain $(RpfG)$	-1,57
XF1115	XF9a_01028	Enoyl-CoA hydratase/carnithine racemase	1 / 9
PD0407	XFTem_00459	(RpfF)	-1,40
Fosfodieste	rase e diguanilato	ciclase	
XF2624	XF9a_02479	PAS domain S-box/diguanylate cyclase (GGDEF)	2 10
PD1994	XFTem_02315	domain	-2,19
Funções re	gulatórias		
VE0072	VE0. 00808	Response regulator containing a CheY-like	
AF0972	XF9a_00898	receiver domain and an HTH DNA-binding	1,53
FD0208	AFTell_00303	domain	
XF1133	XF9a_01043	Multimonia flavo dovin Wah A	1 4 4
PD0422	XFTem_00475	Mullimeric Jiavoaoxin WrbA	1,44
XF1254	XF9a_01154	AraC-type DNA-binding domain-containing	2.62
PD0520	XFTem_00591	proteins	2,02
XF1275	XF9a_01172	poly(hydroxyalkanoate) granule-associated	1.64
PD0534	XFTem_00606	protein	-1,04
XF1350	XF9a_01234	DNA nohumanaga sigma faatan sigma 70 famih	1 46
PD0593	XFTem_00672	KIVA polymeruse sigma jacior, sigma-70 jamily	-1,40
XF1463	XF9a_01338	Transariational manufators	1 69
PD0682	XFTem_00771	Transcriptional regulators	1,08
XF1843	XF9a_01728	Nituogan nagulatan protain DII	1 75
PD1025	XFTem_01193	Nurogen regulatory protein F II	-1,75
XF2535	XF9a_02399	Sional transduction high dia a high and	2.06
PD1920	XFTem_02232	Signal transauction histiatine kindse	2,00
XF2691	XF9a_02538	-learne etc Dr II	1.00
PD2048	XFTem_02377	alternative sigma jactor kpoh	-1,00
Hemolisina	S		
XF1280	XF9a_01174	Hemolysins and related proteins containing CBS	2.4.4
PD0536	XFTem_00610	domains	2,44
Lipases			
XF2151	XF9a_02036	how other is all and it. (1)	0.01
PD1211	XFTem_01402	nypotnetical protein (lipase)	-2,81

_

Proteases			
XF1187	XF9a_01095	ATP-dependent Clp protease, proteolytic subunit	1.54
PD0472	XFTem_00531	<i>ClpP</i> (<i>EC</i> :3.4.21.92)	-1,34
XF2625	XF9a_02481	Zn-dependent protease with chaperone function	1.60
PD1995	XFTem_02316	(EC:3.4.24)	-1,00
9a5c_PWG	_3d x Temecula1	_PWG_3d	
Biogênese d	lo pilus longo		
XF1224	XF9a_01129	Tfp pilus assembly protein, tip-associated adhesin	2.51
PD0502	XFTem_00565	PilY1	2,51
Funções reg	gulatórias		
XF2535	XF9a_02399	Circuit (and destine histidies history	2.55
PD1920	XFTem_02232	Signal transauction histiathe kinase	2,33
Hemolisina	S		
XF1280	XF9a_01174	Hemolysins and related proteins containing CBS	1.90
PD0536	XFTem_00610	domains	1,02
9a5c_PWG	_10d x Temecula	1_PWG_7d	
Biogênese d	lo pilus longo		
XF1224	XF9a_01129	Tfp pilus assembly protein, tip-associated adhesin	2.00
PD0502	XFTem_00565	PilY1	2,00
Funções reg	gulatórias		
XF2535	XF9a_02399	Cional tanana du stisu bisti din s bin sas	2.08
PD1920	XFTem_02232	Signal transauction histiathe kinase	2,98
Lipases			
XF2151	XF9a_02036	hum other is all must size (lin as a)	2.74
PD1211	XFTem_01402	nypoineucai proiein (upase)	-2,74

4.9. Identificação de pequenos RNAs (sRNAs) expressos pela cepa 9a5c

É crescente o interesse na identificação de pequenos RNAs regulatórios em bactérias (sRNA) e no estudo de sua biogênese e dos mecanismos de atuação (Morris e Mattick, 2014; Nitzan *et al.*, 2017). Na tentativa de descrever o conteúdo de sRNAs em *X., fastidiosa*, e concomitantemente estimar seus níveis de expressão, foi realizado o sequenciamento de bibliotecas enriquecidas em sRNAs preparadas da cepa 9a5c cultivada no meio rico PWG (início do crescimento exponencial). Segundo o protocolo de preparação de bibliotecas de sRNA, a etapa de ligação dos adaptadores é passível de modificação. Deve-se escolher entre incubar a amostra por um menor período de tempo à temperatura ambiente (1 hora a 25°C, biblioteca A) ou, aumentar o tempo de

incubação em uma temperatura inferior (18 horas a 16°C, biblioteca B). A diferença estaria no fato de que diferentes tipos de sRNAs poderiam ter diferentes eficiências na ligação dos adaptadores. Uma vez que se trata da primeira tentativa em analisar sRNAs de *X. fastidiosa*, foi decidido realizar o sequenciamento de bibliotecas preparadas com os dois protocolos de incubação.

Após o sequenciamento, foi realizada a etapa de filtragem dos dímeros de adaptadores que possam ter sido formados e do corte (*trimming*) de sequências de adaptadores das sequências dos cDNA obtidos. Em seguida, as sequências obtidas foram usadas para montagem de transcritos, gerando sRNAs candidatos (Tabelas S7 e S8). Foram encontrados 1154 e 1167 sRNAs candidatos para as bibliotecas A e B, respectivamente. Conforme sugerido no protocolo, a biblioteca B, incubada por um período de tempo maior, teve um maior número de sRNAs obtidos, apesar da diferença ser de somente 13 sRNAs. Entretanto, podemos observar a presença de transcritos montados com tamanhos fora dos limites descritos na literatura para sRNAs bacterianos, que teriam comprimentos entre 50 e 500 nucleotídeos. Após a retirada dessas sequências que estão fora dos limites, foram definidos 1126 e 1143 sRNAs para as bibliotecas A e B, respectivamente, com tamanhos médios de cerca de 80 pares de bases (Tabela 22).

Os valores de cobertura dos transcritos, isto é, a razão entre "número total de bases das sequências (*reads*)" pelo "número de bases do sRNA montado", variou de pouco mais de 3x até cerca de 6500x, contando ambas as bibliotecas. O número de *reads* usados para a montagem de cada sRNA candidato variou entre 10 e 11700, valores que podem indicar uma medida superficial de expressão daquele candidato.

Biblioteca	Total de sRNAs	Total (50 a 500)	Tamanho (média) (pb)	Cobertura (média)	N.° de <i>reads</i> (média)
A (1h: 25°C)	1154	1126	50 - 486	3,4 - 4441,9	10 - 11202
			(81)	(28,4)	(83,9)
B (16h• 18°C)	1167	1143	50 - 476	3,3 - 6581,5	10 - 11773
D (101, 10 C)	1107	1145	(82,5)	(33,9)	(113,8)

Tabela 22. Resultados da montagem dos sRNAs candidatos. Os valores de tamanho, cobertura e número de *reads* correspondem ao total de sRNAs com tamanho entre 50 e 500.

Em seguida, foram realizados alinhamentos das sequências das duas bibliotecas entre si e contra o genoma da cepa 9a5c, utilizando a ferramenta BLAST (Altschul *et al.*, 1990). No primeiro caso, foi possível observar 321 alinhamentos com alta porcentagem de identidade (>96%) entre as sequências dos sRNAs candidatos das duas bibliotecas, correlacionando os sRNAs em comum (Tabela S9). Apesar da grande quantidade de sRNAs candidatos observados nas duas bibliotecas, apenas cerca de 27% foram encontrados em ambas, de acordo com este alto valor de identidade considerado. Pode-se sugerir que a diferença no tempo de incubação das preparações refletiu nos diferentes perfis de sRNAs candidatos observados.

Ao realizar o alinhamento dos sRNAs candidatos com o genoma de 9a5c, foram obtidas coordenadas genômicas exatas que definem a região de onde eles foram transcritos (Tabelas S10 e S11). Em ambas as bibliotecas, foram encontrados sRNAs que mapearam em mais de um local no genoma, podendo estar próximos ou mesmo distantes. Trata-se de uma informação interessante, uma vez que o genoma de *Xylella* apresenta diversas regiões repetidas, o que podem ter aumentado o número de cópias destes sRNAs. Segundo o protocolo de preparação de bibliotecas utilizado, é preservada a informação da direção da transcrição, ou seja, de qual fita do DNA o transcrito originou-se. Dessa forma, será possível definir sRNAs candidatos que estejam antisenso a regiões contendo ORFs já conhecidas. Tal correlação entre ORFs e sRNAs deverá ser feita futuramente, juntamente com a busca por possíveis moléculas de mRNAs que possam ser alvo de regulação por estes sRNAs.

Parte B) Descrição e comparação dos transcritomas de oito cepas de Xylella fastidiosa

Nesta segunda parte do trabalho serão apresentados os estudos dos transcritomas de oito cepas de *X. fastidiosa* em duas fases do crescimento em meio PWG. Como detalhado na seção anterior (parte A) foi observado maior quantidade de genes diferencialmente expressos na comparação de transcritomas de duas cepas (9a5c e Temecula1) cultivadas nas mesmas condições, do que na comparação entre meios ou fases de crescimento de uma mesma cepa. Estas observações sugerem uma regulação gênica diferente entre genes ortólogos destas duas cepas, o que motivou a ampliação do trabalho para descrever e comparar transcritomas de seis outras cepas de *X. fastidiosa*.

Dessa forma, foram realizadas análises dos transcritomas de outras seis cepas isoladas de hospedeiros vegetais ou locais distintos (Tabela 2): J1a12, U24d e Fb7, isolados de citros; 3124, isolado de cafeeiro; Hib4, isolado de hibisco e Pr8x isolado de ameixeira. Ao contrário das cepas Temecula1 e 9a5c, estas seis cepas são ainda pouco estudadas, mas também apresentam diferenças genômicas e fenotípicas *in vitro* (Pierry, 2012; Santana, 2012). Somadas às análises dos seus genomas, as análises dos transcritomas serão importantes para explicar os fenótipos observados. Além disso, a comparações entre os transcritomas tem o potencial de apontar para determinantes relevantes para a especificidade de hospedeiro, bem como de revelar novos genes associados colonização e virulência.

Assim como, no caso das cepas 9a5c e Temecula1, foram sequenciados os transcritomas de células no início e no final do crescimento exponencial no meio rico PWG. Este meio foi escolhido porque resulta em melhor rendimento do crescimento populacional e maior facilidade na obtenção de amostras íntegras de RNA total. Foram realizadas as curvas de crescimento para orientar a escolha do tempo de coleta das células para extração da amostra de RNA total. Os procedimentos de extração de RNA total, depleção do rRNA, preparo das bibliotecas de cDNA, sequenciamento do transcritoma e análises de dados foram realizados como descrito na parte A deste trabalho, seguindo o fluxograma mostrado na Figura 1.

4.10. Curva de crescimento populacional de seis cepas de X. fastidiosa

O crescimento populacional das seis cepas no meio PWG foi analisado por 21 dias (Figura 9), assim como realizado para as cepas 9a5c e Temecula1 (Figura 2). As cepas isoladas de citros, J1a12 e U24d, apresentaram certo atraso no crescimento nos primeiros dias de cultivo quando comparado com as outras cepas analisadas. Consequentemente, ambas atingiram o final da fase exponencial tardiamente, porém com valores semelhantes de densidade celular. A cepa Pr8x, isolada de ameixeira, apresentou valores altos de densidade celular ao final do ensaio, assim como observado para a cepa 9a5c. A cepa 3124 apresentou um perfil mais característico de crescimento bacteriano, com suas fases bem definidas e atingindo a fase estacionária com tempo semelhante ao observado para a cepa J1a12.

A cepa Fb7, isolada de citros na Argentina, atingiu a fase estacionária mais cedo, (~8 dias) além de apresentar um crescimento populacional menor ($DO_{600nm} < 1,0$). Tal fato também foi observado para a cepa Temecula1, a qual parece ter atingido a fase estacionária com 7 dias. Já o perfil de crescimento populacional da cepa Hib4, isolada de hibisco, parece apresentar um decréscimo moderado com 15 dias, porém retoma após 21 dias e ultrapassa ligeiramente $DO_{600nm} > 1,0$.

As curvas de crescimento foram utilizadas para guiar a escolha do dia de coleta para cada uma das seis cepas. As amostras do início do crescimento exponencial foram obtidas no 3° dia para as cepas Fb7, 3124, Hib4 e Pr8x, no 5° dia para a cepa J1a12 e 7° dia para a cepa U24d. As amostras do final do crescimento exponencial foram obtidas no 7° dia para as cepas Fb7, 3124 e Hib4, no 10° dia para J1a12 e Pr8x e no 15° dia para U24d (Figura 9). Todos os cultivos foram iniciados com $DO_{600nm} = 0,05$.

Assim como observado para as cepas 9a5c e Temecula1, os fenótipos *in vitro* de cultivo das seis cepas analisadas nesta etapa também mostraram grande variação (Figura 10). As cepas J1a12, Fb7 e Hib4 mostraram fenótipos de menor agregação de células planctônicas quando comparados com as outras cepas. Apesar disso, em fases tardias de crescimento, apresentaram formação de biofilme aderido ao frasco. A cepa U24d apresentou a maior deposição de células aderidas em biofilme.



Figura 9. Curva de crescimento de seis cepas de *X. fastidiosa*. (A) cepa J1a12, (B) cepa U24d, (C) cepa Fb7, (D) cepa 3124, (E) cepa Hib4 e (F) cepa Pr8x foram cultivadas em tubos cônicos de 50mL com DO_{600nm} inicial de 0,05 no meio PWG, a 28° C e 170 rpm. Medições da DO_{600nm} foram realizadas nos 1°, 3°, 7°, 10°, 15° e 21° dias de cultivo. Barras verticais representam o desvio padrão da média de três réplicas técnicas.



Figura 10. Fenótipo de crescimento *in vitro* no meio PWG dos dias escolhidos para a extração de RNA total. (A) cepa J1a12 com 5 dias; (B) cepa J1a12 com 10 dias; (C) cepa U24d com 7 dias; (D) cepa U24d com 15 dias; (E) cepa Fb7 com 3 dias; (F) cepa Fb7 com 7 dias; (G) cepa 3124 com 3 dias; (H) cepa 3124 com 7 dias; (I) cepa Hib4 com 3 dias; (J) cepa Hib4 com 7 dias; (K) cepa Pr8x com 3 dias; (L) cepa Pr8x com 10 dias. O cultivo foi iniciado com DO_{600nm} de 0,05. As imagens mostram uma de duas réplicas biológicas para cada condição. Frascos incubados a 28°C e 170 rpm.

4.11. Obtenção dos transcritomas por RNA-Seq

As extrações das amostras de RNA total das seis cepas, quantificações e análises de qualidade e integridade foram realizadas exatamente como descrito na parte A (item 4.3). A Tabela 23 lista os parâmetros de qualidade para todas as amostras de RNA extraídas das seis cepas, os quais, de modo geral, estão plenamente adequados para realização do RNA-Seq. Também para estas amostras, a completa remoção do DNA destas preparações de RNA total foi confirmada por PCR com um par de *primers* específico para detecção de DNA de *X. fastidiosa* (dados não mostrados).

As preparações de RNA total foram submetidas a depleção do rRNAs para enriquecimento em moléculas de mRNAs, seguindo-se análise por eletroforese capilar. Para quase todas as amostras indicadas na Tabela 23, a eficiência na remoção de rRNAs foi muito bem-sucedida (dados não mostrados). Entretanto, em algumas réplicas das cepas Hib4 e U24d a depleção do rRNA foi incompleta, o que só foi verificado após o sequenciamento do transcritoma. Assim, novas réplicas foram obtidas, sendo que apenas os parâmetros destas réplicas constam da Tabela 23.

As amostras purificadas e depletadas de rRNAs foram então utilizadas para a preparação das bibliotecas de cDNA. Assim como observado para as bibliotecas das cepas 9a5c e Temecula1, para todas elas, foi obtido um excelente rendimento, quanto a concentração como na qualidade, pureza e tamanho médio do inserto (Tabela 23). As 30 bibliotecas de cDNA foram sequenciadas em nove corridas no equipamento *MiSeq (Illumina)* utilizando a estratégia de *Paired-End* com um *kit* para 500 ciclos. Este total de bibliotecas inclui 6 bibliotecas das cepas Hib4 e U24d que foram refeitas devido a falha na etapa depleção do rRNA como mencionado acima. Assim como as corridas anteriores (Tabela 3), o rendimento da maioria dessas corridas foi excelente, com quantidades de sequências muito próximas ou até maiores do que o previsto pelo fabricante, que é de 25 milhões de sequências (Tabela 24). Dados das corridas 1 e 3 (Tabela 3) foram novamente citados na Tabela 24, uma vez que incluíram bibliotecas da cepa Fb7.

Cabe destacar que a corrida 11 gerou cerca de 20 milhões de sequências devido a uma densidade de *clusters* mais baixa que as demais. Por outro lado, densidades muito altas também podem prejudicar, como aconteceu na corrida 8, quando a porcentagem de sequências de alta qualidade apresentou uma queda em relação as outras corridas. Essas variações na densidade de *clusters* provavelmente refletem pequenos erros na quantificação das bibliotecas, sendo que a concentração que foi utilizada nestes sequenciamentos variou de 6 a 8 pM.

Cono	Tempo	NanoDrop	A260/	A260/	DIN	RiboGreen	NanoDrop	A260/	A260/	Tamanho	qPCR	Dánlico
Cepa	(dias)	(ng/µL)	280	230	NIN	$(ng/\mu L)$	(ng/µL)	280	230	médio (pb)	(nM)	Replica
	5	229,8	2,15	0,89	8,2	142,3	84,7	1,85	2,19	412	138,4	1
I1917	5	267,1	2,14	2,26	8,5	157,7	89,6	1,87	2,21	417	110,1	2
51412	10	251,9	2,15	2,18	8,7	161,2	87,2	1,86	2,11	388	141,8	1
	10	238,7	2,15	1,59	8,2	145,9	92,9	1,84	2,15	404	159,8	2
	7	255,7	2,14	2,26	8,8	149,4	89,3	1,86	2,17	401	136,5	1
1124d	/	280,7	2,13	1,81	9,4	166,2	85,6	1,87	2,29	422	144,4	2
0 2 4u	15	177,3	2,10	1,77	9,2	100,5	45,1	1,84	2,20	368	106,4	5
	15	185,3	2,10	1,44	9,1	117,9	41,4	1,80	1,96	352	90,7	6
	3	255,4	2,13	2,13	9,3	141,1	74,4	1,85	2,22	369	202,9	1
Fh7	3	246,9	2,14	1,07	9,4	131,3	78,0	1,87	2,34	370	148,4	2
F 07	7	368,2	2,16	1,87	8,0	335,3	60,7	1,84	1,91	381	94,0	1
	/	201,1	2,10	2,13	8,4	224,6	50,6	1,82	2,29	342	191,5	2
	3	266,5	2,15	1,42	8,5	155,9	93,2	1,84	2,21	383	128,9	1
3124	5	262,9	2,15	2,04	8,8	152,3	98,3	1,84	2,16	402	132,7	2
3124	7	255,1	2,14	2,20	9,1	151,5	85,7	1,85	2,20	401	134,2	1
	/	260,6	2,15	2,22	9,5	154,8	87,9	1,85	2,20	438	117,1	2
	2	268,7	2,15	1,93	8,1	155,8	88,7	1,86	2,14	407	152,8	2
Hib/	3	281,8	2,12	2,31	8,6	175,3	93,9	1,82	1,93	415	222,9	3
11104	7	193,1	2,07	2,03	8,9	126,7	76,2	1,86	2,14	382	177,4	3
	7	176,6	2,08	1,42	8,6	111,2	73,1	1,86	2,27	352	211,1	4
	2	252,5	2,14	2,22	9,2	152,9	85,1	1,87	2,09	416	138,8	1
Pr8v	5	252,5	2,15	2,18	9,7	146,2	73,1	1,90	2,05	389	122,6	2
1104	10	262,1	2,16	1,15	9,0	152,8	85,1	1,90	2,15	397	149,2	1
	10	267,5	2,15	2,23	9,1	157,2	73,3	1,90	2,08	383	133,6	2

Tabela 23. Dados das amostras de RNA de cada uma das seis cultivadas em PWG (painel esquerdo) e de suas respectivas bibliotecas de cDNA (painel direito).

Corrida	Total de sequências*	Sequências filtradas**	Número de bibliotecas	Densidade de clusters	% >Q30***
1+	25.687.600	16.908.052	4	1439	77,7
3+	27.988.934	23.358.634	4	1547	77,7
6	22.602.532	20.644.205	6	1196	80,1
7	29.875.986	23.164.145	5	1654	74,6
8	30.020.946	24.186.651	5	1630	62,3
9	29.033.867	23.293.348	5	1591	75,1
10	27.364.241	23.518.326	5	1471	77,3
11	20.793.330	19.006.836	5	1099	79,7
12	22.873.995	19.779.191	2	1245	74,1

Tabela 24. Resumo do rendimento de nove corridas contendo as bibliotecas das cepas J1a12, U24d, Fb7, 3124, Hib4 e Pr8x.

Paired ends reads.* **Filtragem que remove sequências com baixa qualidade. *>Q30 indica sequências de alta qualidade. + Corridas citadas na Tabela 3 mas que contém bibliotecas da cepa Fb7.

4.12. Análises da qualidade das sequências dos transcritomas

As sequências obtidas nos sequenciamentos de todos os transcritomas das seis cepas foram analisadas quanto à qualidade de suas bases, seguindo a mesma estratégia descrita no item 4.3. De modo geral, quase todos os transcritomas sequenciados nesta parte do apresentaram padrões de qualidade semelhantes ao exemplo mostrado anteriormente (Figura 4), como exemplificado pela análise das sequências da biblioteca da primeira réplica da cepa 3124 crescida por 3 dias em meio PWG (Figura 11A-C).

No entanto, foi verificado que as bibliotecas sequenciadas na corrida 8 (3124_3d1, Pr8x_3d1 e Pr8x_10d1) apresentaram uma maior quantidade de sequências com escores de qualidade acima de Q15 (Figura 11D), o que reflete o fato dessa corrida ter apresentado menor porcentagem de sequências com alta qualidade.


Figura 11. Análise da qualidade por base das sequências do transcritomas $3124_3d1 e Pr8x_3d1$. Análise das sequências da primeira extremidade (*read1*) (A), da segunda extremidade (*read2*) (B) e as sequências pareadas (*read1+read2*) (C) do transcritoma da primeira réplica da cepa 3124 cultivada por 3 dias em PWG. (D) análise das sequências pareadas (*read1+read2*) de Pr8x. As análises foram realizadas usando o *software FASTQC*. O eixo das abcissas contém os tamanhos das sequências (1 até 250), enquanto que o eixo das ordenadas contém os valores de *Q* escore (de 0 a 40).

4.13. Mapeamento das sequências dos transcritomas nos genomas de referência

O *software CLC Genomics Workbench* foi utilizado para o mapeamento das sequências nos genomas de referência exatamente como descrito no item 4.4. Todas as cepas utilizadas neste estudo já tiveram seus genomas sequenciados e montados em trabalhos anteriores (Pierry, 2012; Santana, 2012). Portanto, cada transcritoma foi adequadamente mapeado na sua respectiva sequência genômica e anotação gênica para obtenção de valores de expressão baseados na normalização por FPKM.

O número de sequências por biblioteca variou refletindo o número de bibliotecas sequenciadas por corrida (Tabela 25), e assim como observado para os transcritomas de 9a5c e Temecula1 (item 4.4.), estão dentro do recomendado para análise de expressão

diferencial (Haas *et al.*, 2012). As porcentagens de sequências mapeadas (66,4 a 88,9%) foram consideradas satisfatórias para as análises subsequentes.

Сера	Tempo (dias)	Total de sequências	Tamanho médio (bases)	Distância entre paired reads (bases)	Sequências mapeadas	Réplica
11912	5	8.094.534	156	63 a 240	6.752.880 (83,4%)	1
	3	9.982.832	157	65 a 249	8.064.520 (80,8%)	2
J1412	10	9.683.066	150	62 a 240	7.195.622 (74,3%)	1
	10	7.957.756	150	63 a 249	5.792.488 (72,8%)	2
	7	10.279.226	151	60 a 240	8.997.494 (87,5%)	1
11244	/	9.332.698	150	59 a 240	7.635.964 (81,8%)	2
0 24 u	15	19.078.998	152	64 a 250	15.067.468 (79,0%)	5
	15	19.890.124	146	63 a 249	17.282.894 (86,9%)	6
	2	5.783.996	159	71 a 240	4.809.678 (83,2%)	1
Fb7	3	9.142.576	157	70 a 240	6.984.550 (76,4%)	2
FU/	7	7.743.594	163	66 a 241	6.458.294 (83,4%)	1
		10.071.298	147	67 a 241	8.943.750 (88,8%)	2
	2	9.733.762	158	67 a 249	8.445.092 (86.8%)	1
2124	3	9.546.034	153	63 a 240	8.236.080 (86,3%)	2
5124	7	8.542.486	160	63 a 235	6.554.300 (76,7%)	1
		11.366.984	149	63 a 240	10.105.710 (88,9%)	2
	2	8.676.050	155	62 a 249	6.783.638 (78,2%)	2
Hib4	3	7.658.814	160	68 a 248	6.161.342 (80,5%)	3
1104	7	8.571.260	155	64 a 248	7.132.760 (83,2%)	3
	/	6.979.646	155	67 a 249	5.719.766 (82,0%)	4
	3	10.111.206	159	62 a 238	7.770.606 (76,9%)	1
DrQv	3	7.377.760	154	65 a 249	5.832.910 (79,1%)	2
FIOX	10	8.900.358	163	63 a 232	5.913.106 (66,4%)	1
	10	8.509.818	149	64 a 249	6.681.236 (78,5%)	2

 Tabela 25. Resumo do rendimento de cada biblioteca sequenciada.

4.14. Análise de correlação das réplicas biológicas e entre diferentes transcritomas

As análises de transcritomas relatadas na parte A altos valores de correlação entre as três réplicas biológicas para cada amostra quando comparadas aos pares. Este resultado justificou a decisão de sequenciar somente duas réplicas de cada condição no caso dos transcritomas das seis cepas descritas nesta parte da tese.

Sabe-se que um maior número de réplicas eleva o poder estatístico nas análises de expressão diferencial, principalmente na detecção de genes com baixa expressão. Para contornar eventuais imprecisões de análises de expressão baseadas em duas réplicas, Schurch e colaboradores (Schurch *et al.*, 2016) recomendam que seja estabelecido um limite ao *fold change* diferente da hipótese nula do *default*, que seria 0, o que é possível quando é utilizada a ferramenta *DESeq2*. Estes autores realizaram um estudo com 48 réplicas biológicas de duas linhagens de *Saccharomyces cerevisiae* e utilizaram 11 ferramentas de análise de expressão diferencial. Dentre elas, *DESeq2* foi considerada uma das 5 melhores ferramentas, uma vez que é capaz de controlar o índice de falsos positivos, mantendo-o perto ou abaixo de 5% (conforme utilizado neste trabalho, padj<0,05) independentemente do número de réplicas ou do valor limite de *fold change* para a hipótese nula sendo testada. Neste modo, a ferramenta testa se o valor medido para o *fold change* de um determinado gene é consistente em estar abaixo do valor limite escolhido, do que ser consistente com zero, fornecendo um mecanismo natural de incorporar um limite ao *fold change* de uma forma estatisticamente significante.

Assim como para os transcritomas de 9a5c e Temecula1, também foram realizadas análises de correlação de Pearson entre as duas réplicas de uma mesma condição, utilizando valores de expressão por FPKM. Os valores de correlação das réplicas foram sempre acima de 0,9 (Tabela 26), indicando que podem ser tratadas como tal. As plotagens de valores de FPKM entre duas réplicas e uma mesma condição demonstram a alta correlação entre as réplicas biológicas (Figura 12).

Сера	Meio	Tempo (dias)	Valor de correlação de Pearson
I1a12	PWG	5	0,984
91a12	PWG	10	0,995
11244	PWG	7	0,991
0240	PWG	15	0,995
Eb7	PWG	3	0,999
107	PWG	7	0,999
3124	PWG	3	0,992
5124	PWG	7	0,990
Hib/	PWG	3	0,926
11104	PWG	7	0,955
Pr8v	PWG	3	0,997
1101	PWG	10	0,994

Tabela 26. Valores de correlação de Pearson entre as réplicas de uma mesma condição. Análise foi realizada com os valores de FPKM de cada transcritoma.



Figura 12. Gráficos de correlação entre réplicas da mesma condição baseado nos valores de FPKM transformados em logaritmo na base 10. Os gráficos correspondem à correlação entre réplicas de transcritomas em PWG das cepas: J1a12 início (A) e final (B) da fase exponencial; U24d início (C) e final (D) da fase exponencial; Fb7 início (E) e final (F) da fase exponencial; 3124 início (G) e final (H) da fase exponencial; Hib4 início (I) e final (J) da fase exponencial; Pr8x início (K) e final (L) da fase exponencial.

Com os valores de FPKM, transformados em log10, de todos os transcritomas analisados foram gerados dois *heatmaps*, um separado por réplicas (Figura 13) e outro com valores médios de cada condição (Figura 14). Estas análises incluíram os transcritomas de 9a5c e Temecula1 obtidos do crescimento em meio PWG para comparação com os transcritomas das outras seis cepas.

O *heatmap* baseado nos valores de cada réplica separadamente mostrou que as cepas Fb7, J1a12, Pr8x e Hib4 tiveram todos os seus transcritomas agrupados lado a lado, inclusive separados de acordo com a mesma fase de crescimento (Figura 13).

Entretanto, as cepas 9a5c, U24d, 3124 e Temecula1 não seguiram o mesmo padrão. Em alguns casos, as réplicas de mesma fase do crescimento se mantiveram agrupadas, como para as cepas 9a5c e 3124. Já na cepa U24d, uma das réplicas de início do crescimento agrupou-se distante do grupo de outras réplicas da mesma cepa. Os transcritomas de Temecula1, a única cepa norte-americana e da subespécie fastidiosa, apresentaram localizações isoladas dos grupos maiores, comportando-se com "grupo externo". Os dois transcritomas de início de crescimento de Temecula1 localizaram-se externamente ao grupo de todos os transcritomas de Hib4 e os de fase inicial das cepas 9a5c e 3124. A primeira réplica do transcritoma de Temecula1 na fase final comportou-se com "grupo externo" ao grupo contendo todos os transcritomas de Pr8x, três réplicas de U24d e os transcritomas de fase final de 9a5c. Já a segunda réplica do transcritoma de Temecula1 na fase final posicionou-se externamente ao grupo contendo todos os transcritomas de Fb7, os transcritomas de 3124 na fase final e uma réplica do transcritoma de início do crescimento de U24d. Os transcritomas da cepa J1a12 foram os que mostraram maior diferença em relação aos outros, uma vez que se localizaram isoladamente em um único grupo (Figura 13). De modo geral, foi possível observar uma coerência nos agrupamentos das réplicas baseados nos valores de FPKM, uma vez que quase todas elas foram encontradas próximas, corroborando os altos valores de correlação.

O *heatmap* baseado nos valores médios (log10) de FPKM de cada condição mostrou um agrupamento das condições similar ao observado no *heatmap* anterior. As cepas Fb7, J1a12, Hib4 e Pr8x apresentaram seus transcritomas de início e final do crescimento agrupados lado a lado. As cepas 9a5c e 3124 tiveram os transcritomas de início do crescimento posicionados em um mesmo grupo. Por outro lado, o transcritoma da cepa 9a5c no final do crescimento agrupou com o da cepa U24d, também da fase final do crescimento. Já a condição de fase inicial de U24d agrupou-se com a de fase final de 3124, tendo a condição de Temecula1 em fase final localizada próxima. Ainda, Temecula1 teve sua condição de início do crescimento localizada externamente ao grupo que inclui Hib4 (Figura 14).



Figura 13. *Heatmap* com valores de FPKM em log10 de todas as réplicas das condições analisadas em todas as cepas. As colunas representam os transcritomas, devidamente nomeadas. As linhas representam cada gene analisado (1341 genes). O dendrograma no topo da figura mostra o agrupamento dos transcritomas em questão. O gráfico na parte superior direita da figura representa uma gradação de cores baseando-se nos valores de FPKM, sendo que na escala quanto mais próximo do vermelho menor a expressão e, quanto mais próximo do branco maior a expressão. Para a geração da imagem foi usada a função *heatmap.2* do pacote *gplots* utilizando a plataforma *R*.



Figura 14. *Heatmap* com as médias dos valores de FPKM em log10 de todas as réplicas para cada condição analisada em todas as cepas. As colunas representam as condições, devidamente nomeadas. As linhas representam cada gene analisado (1341 genes). O dendrograma no topo da figura mostra o agrupamento das condições em questão. O gráfico na parte superior direita da figura representa uma gradação de cores baseando-se nos valores médios de FPKM, sendo que na escala quanto mais próximo do vermelho menor a expressão e, quanto mais próximo do branco maior a expressão. Para a geração da imagem foi usada a função *heatmap.2* do pacote *gplots* utilizando a plataforma R.

4.15. Análise dos transcritomas baseada nos valores de FPKM

4.15.1. Panorama da expressão gênica das seis cepas

As listas de genes e seus respectivos valores de expressão em FPKM (dados não apresentados) para cada transcritoma foram devidamente ordenadas e os genes divididos em quatro categorias (Tabela 27) de acordo com intervalo de valores de FPKM. Assim como observado nos transcritomas de 9a5c e Temecula1, a categoria mais representada foi a de genes com FPKM entre 100 e 0, evidenciando que a maioria dos genes apresenta baixa expressão (45,6 - 76,6%).

Genes com valores de expressão igual a zero também foram observados nestes transcritomas (5,9 – 13,3%). Dentre esses genes, estão a grande maioria dos genes de tRNAs, pois, como mencionado anteriormente, tais transcritos são perdidos durante o procedimento de preparação das bibliotecas. Estas observações confirmam a completa ausência de contaminação com DNA nas preparações de RNA total que foram utilizadas no RNA-Seq.

Também como observado anteriormente (parte A, item 4.7.1), sequências de rRNAs foram detectados com baixíssima expressão em poucos transcritomas. Conforme já mencionado, o parâmetro utilizado no mapeamento permite a contagem do fragmento apenas caso ele mapeie exclusivamente em um único local do genoma. Uma vez que também observamos altos valores de porcentagens de mapeamento das sequências em um único local no genoma em todos os transcritomas (86,2 - 93,9%), pode-se considerar a depleção de rRNA satisfatória (Tabela 27).

A maioria dos genes encontrados com expressão igual a zero são aqueles com função hipotética ou provenientes de regiões de prófagos, assim como verificado para os transcritomas analisados na parte A (item 4.7.1.). Vale lembrar que o conteúdo de fagos nos diferentes genomas varia em número e composição de genes (Santana, 2012). Novamente, podemos sugerir que tais genes possivelmente fazem parte de regiões remanescentes de fagos incompletos e inativos, embora poderiam eventualmente ser expressos em outras condições de cultivo. Entretanto, alguns genes relacionados a fagos apresentaram valores positivos de expressão por FPKM em todos os transcritomas. Destacamos a observação de que genes codificadores da toxina Zonula occludens (Zot) também não estão expressos nas cepas 3124 e Hib4, assim como visto em Temecula1. Nas outras cepas, todos os genes para Zot estariam expressos.

Nas cepas de citros U24d e J1a12, os genes para uma DNA primase e para a proteína VirD4 também foram encontrados silenciados, assim como em 9a5c. Um gene codificador de uma flavodoxina multimérica (*WrbA*) foi encontrado com expressão zero em U24d. Já na cepa J1a12, dois genes para subunidades da protease Clp e dois genes para proteínas de montagem do pilus longo (*PilA*) não estão expressos. Entretanto, outras duas cópias de *pilA* mostraram-se fracamente expressas. Na outra cepa de citros, Fb7, genes de proteínas de exportação de grupo heme e uma acetiltransferases foram encontrados silenciados.

Assim como visto para a cepa U24d, as cepas 3124 e Pr8x mostraram expressão zero para o gene da flavodoxina multimérica (*WrbA*). Na cepa 3124, genes de uma possível hemaglutinina e de uma metilase de DNA não estão expressos. Já em Pr8x, um gene de uma permease para arabinose e outro de GTPase relacionada a fatores de elongação da tradução estariam silenciados.

Na cepa Hib4 foi observado que genes codificadores de acetiltransferases, proteínas de biossíntese de grupo heme, glicosil transferases, uma esterase predita, uma metilase envolvida na biossíntese de ubiquinona e metaloproteases dependentes de zinco (elastase) também não estão expressos. Porém, outras duas cópias de elastases estariam expressas. Por fim, genes da proteína de retração do pilus longo (*PilT*) e uma hemaglutinina também apresentaram valores de expressão iguais a zero.

As cepas que tiveram seus transcritomas sequenciados, exceto a cepa 3124, possuem plasmídeos (Pierry, 2012; Santana, 2012). As sequências obtidas nos transcritomas também foram mapeadas contra as sequências destes plasmídeos e o nível de expressão desses genes foi classificado baseado nos valores de FPKM. A grande maioria dos genes mostrou altos valores de expressão em todas as condições e cepas, concentrando-se nas categorias "acima de 10.000" e "entre 10.000 e 1.000" (Tabela 28). A alta expressão destes genes pode estar relacionada ao número de cópias destes plasmídeos.

Nos plasmídeos pXF27 (J1a12), pXF39 (Pr8x) e nos transcritomas da fase inicial em pXF64 (Hib4) a categoria mais representada foi a de genes com FPKM maiores que 10.000. Enquanto isso, os outros plasmídeos e os transcritomas de fase final em pXF64, a categoria com maior número de genes foi a de genes com FPKM entre 10.000 e 1.000 (Tabela 28).

Quase todos os genes dos plasmídeos foram encontrados com valores de FPKM diferentes de zero, com algumas exceções (Tabela 28). No pXF39 da cepa Fb7 foi anotado o ncRNA *traJ-II*, porém tais transcritos não foram detectados em nenhum dos transcritomas analisados. Trata-se de um motivo *cis*-regulatório encontrado *in silico* em proteobactérias, porém ainda sem função definida (Weinberg *et al.*, 2010). Os outros poucos genes com expressão igual a zero são classificados com funções hipotéticas. Alguns deles se repetem nos transcritomas da mesma cepa, como por exemplo, o primeiro gene do pXF51 de U24d, os genes 50 e 52 no pXF64, e o genes 1, 21 e 57 do pXF39 de Pr8x.

Сера	Tempo (dias)	Total de genes*	>10.000	Entre 10.000 e 1.000	Entre 1.000 e 100	Entre 100 e 0	Zero ⁺	Genes expressos (%)	Réplica
J1a12	5	2784	9	89	929	1422	335	88,0	1
	C C	2784	11	104	901	1434	334	88,0	2
51412	10	2784	2	64	767	1621	330	88,1	1
	10	2784	4	30	709	1697	344	87,6	2
	7	2623	5	32	691	1711	184	93,0	1
U24d	,	2623	4	16	388	1996	219	91,7	2
0240	15	2623	9	63	987	1410	154	94,1	5
	15	2623	6	45	759	1651	162	93,8	6
	3	2659	4	87	611	1706	251	90,6	1
Fb7	C C	2659	4	74	552	1819	209	92,1	2
	7	2659	3	35	340	2037	243	90,9	1
		2659	4	38	474	1979	162	93,9	2
	3	2674	6	95	735	1578	260	90,3	1
3124	5	2674	8	98	772	1540	256	90,4	2
5124	7	2674	3	40	510	1836	285	89,3	1
		2674	3	35	530	1840	266	90,1	2
	3	2765	8	89	808	1504	356	87,1	2
Hib4	5	2765	10	117	949	1321	368	86,2	3
11104	7	2765	10	126	929	1356	344	87,1	3
	,	2765	6	130	1003	1262	364	86,4	4
	3	2603	12	72	811	1506	202	92,2	1
Pr8v	C C	2603	11	70	738	1563	221	91,5	2
IIVA	10	2603	8	70	784	1531	210	91,9	1
	10	2603	6	36	686	1672	203	92,2	2

Tabela 27. Número de genes contabilizados por faixa de valores de FPKM e porcentagem de genes expressos em cada transcritoma.

*total de genes analisado, subtraindo-se os genes de rRNAs e tRNAs. ⁺genes com valor zero de RPKM, subtraindo-se os genes de rRNAs e tRNAs.

Plasmideo	Tempo (dias)	Total de genes	>10.000	Entre 10.000 e 1.000	Entre 1.000 e 100	Entre 100 e 0	Zero	Genes expressos (%)	Réplica
	5	38	31	6	1	0	0	100	1
pXF27	5	38	26	11	1	0	0	100	2
(J1a12)	10	38	30	7	1	0	0	100	1
	10	38	28	9	0	0	1	97,4	2
	5	73	25	42	5	0	1	98,6	1
pXF51	5	73	27	38	7	0	1	98,6	2
(J1a12)	10	73	29	38	6	0	0	100	1
	10	73	26	40	7	0	0	100	2
	7	70	27	40	1	1	1	98,6	1
pXF51	/	70	32	35	1	0	2	97,1	2
(U24d)	15	70	29	38	2	0	1	98,6	5
	15	70	26	41	2	0	1	98,6	6
	2	51	21	26	3	0	1	98,0	1
pXF39	3	51	21	26	3	0	1	98,0	2
(Fb7)	7	51	23	25	2	0	1	98,0	1
		51	23	27	0	0	1	98,0	2
	2	67	33	29	3	0	2	97,0	2
pXF64	3	67	32	30	3	0	2	97,0	3
(Hib4)	7	67	28	34	3	0	2	97,0	3
	/	67	30	33	2	0	2	97,0	4
	3	57	32	20	3	0	2	96,5	1
pXF39	5	57	33	20	2	0	2	96,5	2
(Pr8x)	10	57	32	22	0	0	3	94,7	1
	10	57	35	19	1	0	2	96,5	2

Tabela 28. Número de genes dos plasmídeos indicados contabilizados por faixa de valores de FPKM, e porcentagem de genes expressos.

4.15.2. Descrição dos genes mais expressos

A partir das listas de genes e seus respectivos valores de expressão gênica normalizados por FPKM foram destacados os dez genes mais expressos em todas as condições e réplicas das seis cepas em questão (Tabelas 29 e 30).

Em quase todos os transcritomas o gene que apresentou maior expressão codifica para o ncRNA da endoribonuclease *RNase P* (Altman, 2011), assim como observado para os transcritomas das cepas 9a5c e Temecula1 (item 4.7.2). As exceções foram as duas réplicas dos transcritomas de fase final do crescimento em Hib4, nos quais o gene mais expresso codifica uma colicina V (XF0263). O ncRNA da *RNase P* foi encontrado, respectivamente, nas 6^a e 2^a posições dos ordenamentos das réplicas 1 e 2 de Hib4.

O ncRNA 6S também foi encontrado nas dez primeiras posições em 18 dos 24 transcritomas analisados. As exceções foram dois transcritomas de fase inicial (um em J1a12 e um em 3124) e todos os transcritomas de Hib4. Nestes, sua posição no ranqueamento variou de 11^a até 47^a. Por se tratar de um RNA relacionado à mudança da fase exponencial para a fase estacionária, em geral ele ocupa posições mais altas nos *rankings* de FPKM dos transcritomas do final do crescimento, assim como visto para as cepas 9a5c e Temecula1.

De modo similar ao observado com os transcritomas de 9a5c e Temecula1 (Parte 4.7.2), em todos os 24 transcritomas analisados estão vários genes codificadores de bacteriocinas. Entre eles estão os ortólogos de três colicinas V descritas no genoma de X. fastidiosa (XF0262, XF0263 e XF0264) (Simpson et al., 2000), sendo que a XF0263 figura na primeira posição dos rankings dos transcritomas do final do crescimento em Hib4. Também procuramos a identidade dos genes anotados como hipotéticos e estes codificam várias novas microcinas preditas em genomas de X. fastidiosa (Rodrigo R. Duarte e Aline M. da Silva, dados não publicados). Além daquelas indicadas nos 9a5c transcritomas de Temecula1 e foram detectados transcritos de bacteriocinas/microcinas exclusivas das cepas J1a12, 3124 e Fb7. Desta forma, modificamos a anotação destes genes para "bacteriocina". A observação de que o arsenal de toxinas expressos nas cepas 9a5c e Temecula1 também é abundantemente encontrado em outras seis cepas de X. fastidiosa, é um forte indício de que esta pode ser uma característica importante e comum neste fitopatógeno. Isto sugere um comportamento agressivo desta bactéria em seus nichos naturais, para disputar recursos (espaço e nutrientes) com outros microrganismos presentes nos seus hospedeiros e/ou vetores.

Outra característica comum entre todas as cepas é detecção do gene de Ax21/Omp1X nas primeiras posições dos *rankings* de FPKM dos transcritomas das seis cepas, principalmente com maior expressão no início do crescimento.

Outros genes também encontrados entre os dez mais expressos em todos os transcritomas codificam proteínas relacionadas a estresses, sejam as *cold shock proteins* ou a *heat shock protein Hsp20*. Genes de proteínas de membrana externa, como a *OmpW*, também estão entre os mais expressos em todas as cepas e condições. Bacterioferritinas foram encontradas altamente expressas nos transcritomas das cepas J1a12, U24d, Fb7 e 3124, assim como haviam sido vistas em 9a5c e Temecula1. Além disso, as cepas J1a12 e Fb7 apresentaram em comum a alta expressão de genes de uma

cisteína protease e de duas subunidades da protease Clp (ClpP e ClpX). O gene de uma peroxiredoxina está entre os dez mais expressos em J1a12. Já em Fb7, três genes de proteínas ribossomais, um de uma lisozima, um de uma peptidase e dois genes de lipases (especificamente no final do crescimento) estão abundantemente expressos. Mais uma vez, genes com função hipotética também aparecem altamente expressos em vários dos transcritomas, os quais poderão ser analisados em futuros estudos (Tabelas 29 e 30).

Os genes dos plasmídeos de todas as cepas também foram ranqueados baseados em seus valores de FPKM (Tabela S12). O plasmídeo pXF51 de U24d apresentou o gene 37, o qual codifica uma proteína relacionada a fagos, sempre na primeira colocação em todos os transcritomas desta cepa. Os outros genes observados entre os dez primeiros do *ranking* codificam hidrolases putativas, proteína *VirB5*, transglicosilases, uma proteína de regulação transcricional e a proteína *TrbJ* (relacionada à transferência conjugativa) (Tabela S12).

A cepa J1a12 apresenta dois plasmídeos, o pXF27 e o pXF51. O primeiro mostrou alta expressão dos genes 13 (proteína do sistema de secreção do tipo IV) e 23 (função hipotética) nos transcritomas de início do crescimento. Já nos transcritomas do seu final, o gene 8 (função hipotética) foi o primeiro colocado nos *rankings*, seguido também pelo gene 23. Outros genes altamente expressos codificam proteína inibidora de crescimento (provavelmente uma toxina para manutenção do plasmídeo na célula), proteínas do sistema de secreção do tipo IV e transposases. No pXF51, o gene 37 (proteína com função hipotética) foi o primeiro colocado em todos os transcritomas desta cepa. Os outros genes observados entre os dez mais expressos no pXF51 da cepa J1a12 codificam para a proteína *VirB5*, uma proteína relacionada a fagos e uma proteína de regulação transcricional (Tabela S12).

No plasmídeo encontrado cepa Fb7, o gene 37 (proteína relacionada a fagos) foi encontrado como mais expresso em todos os transcritomas desta cepa. O segundo colocado codifica uma metil-ester carboxilesterase, seguido de diversos genes com função hipotética, reguladores transcricionais, proteína da família *ParB*, a interferase *MazF* e a proteína *TrbJ* (Tabela S12). No plasmídeo pXF39 da cepa Pr8x o gene 19 (função hipotética) foi encontrado como mais expresso em todos as condições analisadas nesta cepa. Outros genes observados nas dez primeiras posições codificam para proteínas do sistema de secreção do tipo IV, como a *VirB2, VirB3, VirB5* e *VirD2*, além de reguladores transcricionais, proteínas relacionadas a fagos e diversas proteínas

hipotéticas (Tabela S12). Finalmente, o pXF64 da cepa Hib4 apresentou o gene 4 (função hipotética) como mais expresso nos transcritomas da fase inicial de crescimento. Já nos transcritomas de final do crescimento, o gene 35 (aciltransferase) e o 42 (função hipotética) foram encontrados na primeira posição nas réplicas analisadas. Genes codificando para proteínas hipotéticas, proteínas relacionadas a fagos, recombinases e aciltransferases também figuram entre os dez genes mais expressos (Tabela S12).

Сера	Tempo (dias)	1	2	3	4	5	6	7	8	9	10
J1a12	5	RNase P	Omp W	Bacteriocina (XF1217)	Cold shock proteins	Bacteriocina (XF1218)	6S RNA	Bacteriocina (11430)	Hipotética (XF1287)	Ax21 (XF1803)	Bacterio- ferritina
	5	RNase P	Bacteriocina (11430)	Bacteriocina (XF1305)	Cold shock proteins	Ax21 (XF1803)	Bacteriocina (XF1217)	Omp W	Colicina V (XF0263)	Bacteriocina (XF1218)	Hipotética (XF1255)
		RNase P	6S RNA	Cold shock proteins	Hipotética (XF1255)	Hipotética (XF1287)	Bacterio- ferritina	Omp W	Bacteriocina (XF1217)	Cysteine protease	Clp protease subunit ClpP
	10	RNase P	Colicina V (XF0262)	6S RNA	Hipotética (XF1255)	Heat shock protein Hsp20	Cold shock proteins	Clp regulatory subunit ClpX	Bacterio- ferritina	Peroxiredo- xina	Bacteriocina (XF1217)
U24d -	-	RNase P	Colicina V (XF0263)	Colicina V (XF0264)	6S RNA	Hipotética (1759)	Bacteriocina (XF1305)	Hipotética (1480)	Bacteriocina (XF1217)	Bacteriocina (XF1306)	Bacteriocina (XF1307_2)
	1	RNase P	6S RNA	Colicina V (XF0263)	Colicina V (XF0264)	Bacteriocina (XF1305)	Hipotética (1759)	Ax21 (XF1803)	Bacteriocina (XF1306)	Bacteriocina (XF1217)	Hipotética (1480)
	15	RNase P	6S RNA	Colicina V (XF0264)	Colicina V (XF0263)	Hipotética (1759)	Bacteriocina (XF1217)	Bacterio- ferritina	OmpW	Bacteriocina (XF1306)	Bacteriocina (XF1305)
		RNase P	6S RNA	Colicina V (XF0263)	Colicina V (XF0264)	Hipotética (1759)	Hipotética (1480)	Bacterio- ferritina	Bacteriocina (XF1217)	Heat shock protein Hsp20	Bacteriocina (XF1306)
	2	RNase P	Cold shock protein	Ax21 (XF1803)	Bacterio- ferritina	Omp	Proteína ribossomal S13P	Cysteine protease	6S RNA	Hipotética (111798)	Proteína ribossomal S3P
Fb7 _	5	RNase P	Cold shock protein	Ax21 (XF1803)	6S RNA	Omp	Cysteine protease	Hipotética (111798)	Bacterio- ferritina	Proteína ribossomal S3P	Proteína ribossomal S13P
		RNase P	6S RNA	Colicina V (XF0262)	Secretory lipase	Lysozyme	Hipotética (111798)	Peptidase	Cold shock protein	Secretory lipase	Cysteine protease
	7	RNase P	6S RNA	Cold shock protein	Colicina V (XF0262)	Hipotética (111798)	Lysozyme	Clp protease subunit ClpX	Clp protease subunit ClpP	Proteína ribossomal L29	Cysteine protease

Tabela 29. Lista dos dez genes mais expressos nos transcritomas das cepas J1a12, U24d e Fb7, baseando-se em valores de FPKM.

Cepa	Tempo (dias)	1	2	3	4	5	6	7	8	9	10
		RNase P	Colicina V	Hipotética	Ax21	Bacteriocina	Cold shock	Hipotética	OmpW	Bacteriocina	Bacterio-
	3	11110001	(XF0263)	(1481)	(XF1803)	(XF1217)	proteins	(1207)	omp ()	(XF1218)	ferritina
	5	PNase P	Colicina V	Cold shock	Ax21	Bacteriocina	Hipotética	Hipotética	Bacteriocina	65 RNA	Bacteriocina
2124		Kivase I	(XF0263)	proteins	(XF1803)	(XF1217)	(1481)	(1207)	(XF1218)	05 KIA	(1802)
5124		DNaga D	Colicina V	65 DNA	Hipotética	Bacteriocina	Bacteriocina	Hipotética	Cold shock	Bacterio-	Bacteriocina
	-	KINASE P	(XF0263)	05 KNA	(1207)	(XF1217)	(1802)	(236)	proteins	ferritina	(XF1305)
	1	DN D	Colicina V	Hipotética	CC DNA	Hipotética	Bacteriocina	Bacteriocina	OW	Colicina V	Bacteriocina
		KNase P	(XF0263)	(1207)	05 KNA	(236)	(1802)	(XF1217)	Ompw	(XF0262)	(XF1305)
	2	RNase P	Colicina V	Ax21	Bacteriocina	Bacteriocina		Cold shock	Bacteriocina	Bacteriocina	Bacteriocina
			(XF0263)	(XF1803)	(XF1217)	(XF1218)	Ompw	proteins	(XF1305)	(XF1306)	(XF1219)
	3		Colicina V	Ax21	Bacteriocina	Bacteriocina		Cold shock	Bacteriocina	Colicina V	Hipotética
		RNase P	(XF0263)	(XF1803)	(XF1217)	(XF1218)	Ompw	proteins	(XF1305)	(XF0264)	(989)
HID4	7	Colicina V	Colicina V	Cold shock	Bacteriocina	Bacteriocina	DNasa D	Acyl carrier	Bacteriocina	Bacteriocina	Bacteriocina
		(XF0263)	(XF0264)	proteins	(XF1217)	(XF1218)	Kinase P	protein	(XF1219)	(XF1306)	(XF1305)
		Colicina V	D Nasa D	Colicina V	Bacteriocina	Bacteriocina	Cold shock	Bacteriocina	Ax21	Bacteriocina	Bacteriocina
		(XF0263)	KINdSE F	(XF0264)	(XF1217)	(XF1218)	proteins	(XF1307_2)	(XF1803)	(XF1219)	(XF1306)
		DNasa D	Colicina V	Hipotética	Bacteriocina	OmnW	Ax21	Colicina V	Bacteriocina	Bacteriocina	65 DNA
	2	KINASE F	(XF0263)	(1189)	(XF1217)	Ompw	(XF1803)	(XF0264)	(XF1305)	(XF1218)	05 KINA
	3	DNasa D	Colicina V	Hipotética	Colicina V	Bacteriocina	65 DNA	OmpW	Ax21	Bacteriocina	Cold shock
Dugy		KINASE F	(XF0263)	(1189)	(XF0264)	(XF1217)	05 KNA	Ompw	(XF1803)	(XF1305)	proteins
Frox		DNasa D	Colicina V	65 DNA	Hipotética	Bacteriocina	Bacteriocina	OmpW	Cold shock	Bacteriocina	Colicina V
	10	KINASE F	(XF0263)	05 KNA	(1189)	(XF1305)	(XF1217)	Ompw	proteins	(XF1307_2)	(XF0264)
	10	D Nasa D	Colicina V	Colicina V	65 DNA	Hipotética	Bacteriocina	Bacteriocina	Bacteriocina	Bacteriocina	Cold shock
		RNase P	(XF0263)	(XF0264)	US KINA	(1189)	(XF1305)	(XF1217)	(XF1307_2)	(XF1306)	proteins

Tabela 30. Lista dos dez genes mais expressos nos transcritomas das cepas 3124, Hib4 e Pr8x, baseando-se em valores de FPKM.

4.15.3. Análise de rotas metabólicas expressas

Assim como realizado para os transcritomas das cepas 9a5c e Temecula1 (item 4.7.3), os valores de expressão gênica normalizados para FPKM de todos os transcritomas das seis cepas descritos nesta parte da tese, também foram submetidos a análises de rotas metabólicas ativas, utilizando o *software MinPath*. Os genes utilizados nesta análise apresentaram valores de expressão de FPKM maiores que 10, sendo que a notação "1" significa presença da rota naquele transcritoma e "0" significa a sua ausência (Tabela S13).

A maioria das observações encontradas na análise das cepas da parte A também foi vista na análise das outras seis cepas da parte B, mostrando que as principais rotas metabólicas de sobrevivência e manutenção da vida estão presentes em todas elas (Tabela S13). Isto significa que nenhuma cepa analisada teria alguma deficiência importante no metabolismo celular que possa prejudicar o seu adequado crescimento.

As rotas relacionadas à obtenção de energia, metabolismo de carboidratos, manutenção e transmissão da informação genética, metabolismo de purinas e pirimidinas, exportação de proteínas e outras rotas de transporte de moléculas, formação de membranas e parede celular, metabolismo de vitaminas, coenzimas, nitrogênio e enxofre, biossíntese e metabolismo de alguns aminoácidos, estão presentes e ativas nas cepas analisadas.

Da mesma forma, rotas metabólicas que foram consideradas inativas nas cepas 9a5c e Temecula1 também foram detectadas nos transcritomas das outras seis cepas. Dentre elas, o metabolismo de inositol, galactose, açúcares presentes em nucleotídeos, além das rotas de metabolismo e degradação dos aminoácidos glutamato, cisteína, tirosina, valina, leucina, isoleucina e lisina.

Entretanto, algumas particularidades puderam ser observadas. A rota para o metabolismo de triptofano foi vista inativa em quase todos os transcritomas, com exceção de uma réplica de U24d_15d e uma réplica de 3124_7d. Nestes mesmos transcritomas, a rota para o metabolismo de metano foi observada inativa, enquanto que foi vista ativa nos outros transcritomas. Esta mesma relação também pode ser observada em alguns transcritomas da cepa 9a5c, descrita na parte A. No transcritoma de uma das réplicas de U24d_7d a rota de metabolismo de enxofre mostrou-se inativa.

As rotas metabólicas observadas nos transcritomas da cepa Fb7 foram as que apresentaram um maior número de diferenças em relação aos outros transcritomas. A

rota de degradação de *N*-glicano mostrou-se inativa em Fb7 e ativa nas outras cepas. Entretanto, a rota de degradação de estruturas de glicanos mostrou-se ativa em Fb7 e inativa nos outros transcritomas. Apesar de serem capazes de degradar estruturas de glicanos, os genes para a rota de sua biossíntese aparentemente estão ausentes nesta cepa. Além disso, as rotas para biossíntese de zeatina e de alcalóides também estariam ausentes nesta cepa. A rota relacionada ao sistema de secreção do tipo IV não foi encontrada em Fb7 e Pr8x. Estas mesmas duas cepas também diferiram das outras ao apresentarem a rota de resistência a beta-lactâmicos ativa.

Em resumo, a análise de rotas metabólicas das oito cepas de *X. fastidiosa* mostrou um perfil metabólico geral para este fitopatógeno, no qual as rotas essenciais para a sua sobrevivência nos seus nichos biológicos estão "ativas". Foi observada alta concordância entre as cepas e algumas poucas diferenças observadas, que não gerariam vantagens ou desvantagens significativas à determinada cepa. Vale lembrar que este tipo de análise leva em consideração somente se a rota está ativa e não diz respeito à intensidade da expressão gênica, a qual será analisada e comparada a seguir.

4.16. Análises de expressão gênica diferencial

A intensidade da expressão gênica foi comparada entre os transcritomas analisados, utilizando o mesmo pacote *DESeq2* (Love *et al.*, 2014), sendo que os dados de entrada foram os valores de contagem brutos de quantas sequências mapearam em cada gene.

Uma vez que para as seis cepas foram sequenciados transcritomas em um único meio de cultivo (PWG), foram realizados apenas dois níveis de comparação. A nomenclatura dos níveis definida na parte A (item 4.8.1) foi mantida: 1) comparações entre transcritomas da mesma cepa, mas em fases de crescimento diferentes, como por exemplo, entre os transcritomas da cepa J1a12 no início e no final da fase exponencial; 3) comparações entre transcritomas de cepas diferentes, porém na mesma fase de crescimento, como por exemplo, entre o transcritoma de J1a12 e o transcritoma de Fb7, ambos do início da fase exponencial. No nível 3, foram incluídas as comparações entre os transcritomas da cepa 9a5c e Temecula1 com os das outras seis cepas.

Assim como nas análises discutidas na parte A, os parâmetros utilizados para considerar um gene como significativamente diferencialmente expresso foram um valor de *p*, ajustado pela correção FDR (*False Discovery Rate*) (Benjamini e Hochberg, 1995), menor do que 0,05 e um valor de razão de expressão (*fold change*) maior que 2. A Tabela 31 apresenta o número de genes diferencialmente expressos dentro desses parâmetros, em cada comparação. Os gráficos de dispersão baseados nos valores de expressão média dos genes entre as condições analisadas foram gerados, mas devido ao alto número de comparações realizadas, será mostrado um exemplo deste gráfico para cada nível (Figura 15).

Tabela 31. Resumo do número de genes diferencialmente expressos (p<0,05; *fold change*>2) em todas as comparações.

	Total do	Condição 1	Condição 2	Genes DE*	Genes up-	Genes DE	
Níveis	genes	(c1)	(c2)	(p≤0,05;	regulated	em	
	genes	((1)	((2)	<i>FC</i> >2,0)	(c1/c2)	comum	
	2841	J1a12_PWG_5d	J1a12_PWG_10d	2 (0,07%)	2/0	-	
1	2679	U24d_PWG_7d	U24d_PWG_15d	1 (0,04%)	1/0	-	
	2714	Fb7_PWG_3d	Fb7_PWG_7d	14 (0,5%)	10/4	-	
	2729	3124_PWG_3d	3124_PWG_7d	11 (0,4%)	10/1	-	
	2822	Hib4_PWG_3d	Hib4_PWG_7d	3 (0,11%)	3/0	-	
	2660	Pr8x_PWG_3d	Pr8x_PWG_10d	0	0	-	
	2316	9a5c_PWG_3d	J1a12_PWG_5d	63 (2,7%)	44/19	55	
	2316	9a5c_PWG_10d	J1a12_PWG_10d	85 (3,7%)	55/30		
	2536	9a5c_PWG_3d	U24d_PWG_7d	4 (0,2%)	1/3	0	
	2536	9a5c_PWG_10d	U24d_PWG_15d	5 (0,2%)	0/5	0	
	2192	9a5c_PWG_3d	Fb7_PWG_3d	108 (4,9%)	63/45	60	
	2192	9a5c_PWG_10d	Fb7_PWG_7d	100 (4,6%)	46/54	. 07	
	2284	9a5c_PWG_3d	3124_PWG_3d	42 (1,8%)	32/10	37	
	2284	9a5c_PWG_10d	3124_PWG_7d	53 (2,3%)	35/18		
	2206	9a5c_PWG_3d	Hib4_PWG_3d	120 (5,4%)	83/37	105	
	2206	9a5c_PWG_10d	Hib4_PWG_7d	145 (6,6%)	93/52	- 105	
3	2268	9a5c_PWG_3d	Pr8x_PWG_3d	50 (2,2%)	28/22	39	
5	2268	9a5c_PWG_10d	Pr8x_PWG_10d	63 (2,8%)	29/34		
	2265	U24d_PWG_7d	3124_PWG_3d	76 (3,4%)	48/28	44	
	2265	U24d_PWG_15d	3124_PWG_7d	69 (3,0%) 37/32			
	2198	U24d_PWG_7d	Fb7_PWG_3d	183 (8,3%)	79/104	117	
	2198	U24d_PWG_15d	Fb7_PWG_7d	167 (7,6%)	66/101	- 117	
	2321	U24d_PWG_7d	J1a12_PWG_5d	94 (4,0%)	55/39	83	
	2321	U24d_PWG_15d	J1a12_PWG_10d	147 (6,3%)	75/72	- 85	
	2206	U24d_PWG_7d	Hib4_PWG_3d	126 (5,7%)	92/34	118	
	2206	U24d_PWG_15d	Hib4_PWG_7d	263 (11,9%)	137/126	- 110	
	2262	U24d_PWG_7d	Pr8x_PWG_3d	42 (1,9%)	25/17	38	
	2262	U24d_PWG_15d	Pr8x_PWG_10d	98 (4,3%)	40/58		

1874	U24d_PWG_7d	Tem_PWG_3d	129 (6,9%)	79/50	54
1874	U24d_PWG_15d	Tem_PWG_7d	87 (4,6%)	35/52	54
2274	J1a12_PWG_5d	3124_PWG_3d	119 (5,2%)	75/44	96
2274	J1a12_PWG_10d	3124_PWG_7d	104 (4,6%)	65/39	80
2186	J1a12_PWG_5d	Fb7_PWG_3d	139 (6,4%)	73/66	60
2186	J1a12_PWG_10d	Fb7_PWG_7d	84 (3,8%)	27/57	09
2199	J1a12_PWG_5d	Hib4_PWG_3d	179 (8,1%)	104/75	152
2199	J1a12_PWG_10d	Hib4_PWG_7d	209 (9,5%)	118/91	155
2249	J1a12_PWG_5d	Pr8x_PWG_3d	95 (4,2%)	48/47	80
2249	J1a12_PWG_10d	Pr8x_PWG_10d	106 (4,7%)	55/51	80
1856	J1a12_PWG_5d	Tem_PWG_3d	120 (6,5%)	70/50	52
1856	J1a12_PWG_10d	Tem_PWG_7d	60 (3,2%)	25/35	55
2163	Fb7_PWG_3d	3124_PWG_3d	163 (7,5%)	91/72	05
2163	Fb7_PWG_7d	3124_PWG_7d	114 (5,3%)	71/43	95
2117	Fb7_PWG_3d	Hib4_PWG_3d	225 (10,6%)	153/72	162
2117	Fb7_PWG_7d	Hib4_PWG_7d	216 (10,2%)	154/62	102
2178	Fb7_PWG_3d	Pr8x_PWG_3d	160 (7,3%)	94/66	108
2178	Fb7_PWG_7d	Pr8x_PWG_10d	142 (6,5%)	79/63	100
2192	3124_PWG_3d	Hib4_PWG_3d	148 (6,8%)	86/62	129
2192	3124_PWG_7d	Hib4_PWG_7d	185 (8,4%)	107/78	12)
2244	3124_PWG_3d	Pr8x_PWG_3d	63 (2,8%)	22/41	42
2244	3124_PWG_7d	Pr8x_PWG_10d	48 (2,1%)	18/30	72
1879	3124_PWG_3d	Tem_PWG_3d	118 (6,3%)	56/62	50
1879	3124_PWG_7d	Tem_PWG_7d	58 (3,1%)	16/42	50
2229	Hib4_PWG_3d	Pr8x_PWG_3d	134 (6,0%)	44/90	131
2229	Hib4_PWG_7d	Pr8x_PWG_10d	221 (9,9%)	68/153	151
1887	Hib4_PWG_3d	Tem_PWG_3d	145 (7,7%)	56/89	102
1887	Hib4_PWG_7d	Tem_PWG_7d	132 (7,0%)	35/97	102
1872	Pr8x_PWG_3d	Tem_PWG_3d	95 (5,1%)	62/33	47
1872	Pr8x_PWG_10d	Tem_PWG_7d	60 (3,2%)	28/32	יד /
1799	Tem_PWG_3d	Fb7_PWG_3d	206 (11,5%)	91/115	53
1799	Tem_PWG_7d	Fb7_PWG_7d	58 (3,2%)	18/40	55

*DE: diferencialmente expressos. Tem= Temecula1



Figura 15. Gráficos de dispersão dos genes analisados em uma comparação de cada nível. No eixo das abscissas estão os valores de expressão média; no eixo das ordenadas estão os valores de razão de expressão em log2. Os pontos em vermelho representam genes com valores de p ajustado menores do que 0,05; os pontos em cinza representam genes com valores de p ajustado maiores que 0,05. Cada gráfico representa uma comparação: (A) Fb7-PWG-3d e Fb7-PWG-7d; (B) 9a5c-PWG-3d e J1a12-PWG-5d.

No primeiro nível de comparação, entre transcritomas da mesma cepa e em tempos de crescimento diferentes, foi possível observar poucos genes considerados diferencialmente expressos, com os parâmetros utilizados. A maior quantidade de genes observada foi na comparação entre os transcritomas da cepa Fb7, com 14 genes, sendo 10 positivamente regulados no início do crescimento exponencial. O mesmo padrão se manteve nas comparações das outras cepas, com a maioria dos genes modulados positivamente na fase inicial. Tais resultados são corroborados com as comparações entre os transcritomas de 9a5c e Temecula1 em meio PWG, evidenciando que neste meio rico em nutrientes, o ambiente de cultivo está favorável à manutenção do crescimento por um período prolongado e que no intervalo de tempo das amostras analisadas poucas mudanças na expressão gênica foram necessárias.

Dentre os genes positivamente regulados no início do crescimento em Fb7, podemos encontrar genes codificando proteínas ribossomais, proteínas relacionadas a transporte, uma proteína para a biossíntese do pilus longo (relacionado à motilidade da bactéria) e o gene Ax21. Interessantemente, dois ncRNAs foram encontrados mais expressos no final do crescimento, a RNaseP e o 6S. Dos 11 genes diferencialmente expressos na cepa 3124, 10 mostraram-se mais expressos no início do crescimento: genes da DNA polimerase III, uma RNA polimerase, proteínas ribossomais e de um transportador. Já na cepa J1a12, genes de uma GTPase putativa e de uma proteína com

função hipotética foram mais expressos no início do crescimento. A cepa U24d apresentou apenas um gene codificando para uma proteína hipotética, sendo mais expresso no início do crescimento. Um gene de uma proteína de membrana *OmpW*, um gene relacionado à biossíntese do pilus longo e um com função hipotética, foram encontrados positivamente regulados na cepa Hib4. A cepa Pr8x não mostrou nenhum gene diferencialmente expresso dentro dos parâmetros utilizados.

No terceiro nível de comparação, entre transcritomas de cepas diferentes, foi observado um alto número de genes diferencialmente expressos em todas as 54 comparações. Assim como nas comparações entre as cepas da parte A, foram considerados somente genes ortólogos entre o par de cepas analisado. Examinando um mesmo par de cepas comparadas, observa-se um maior número de genes diferencialmente expressos nas comparações entre transcritomas de início do crescimento, mais especificamente em 15 pares (total de 27). No que diz respeito às comparações com a cepa 9a5c, quase em todos os casos foi observada uma maior quantidade de genes diferencialmente expressos no final do crescimento (incluindo a comparação com a cepa Temecula1), com exceção da comparação com a cepa Fb7. No caso de Fb7, em todas as comparações, maior número de genes diferencialmente expressos foi sempre observado nas comparações entre transcritomas no início do crescimento. Nas outras cepas, não foram observados padrões, variando entre as comparações com cada cepa. Além disso, um grande número de genes foi diferencialmente expressos desde o início até o final da fase exponencial, em todas as comparações entre pares de cepas (Tabela 31). Assim como já mencionado, o meio rico PWG parece ser um ambiente relativamente estável e que exige poucas mudanças na expressão gênica durante o crescimento populacional bacteriano. Devido ao alto número de comparações e de genes diferencialmente expressos, somente os genes relacionados à virulência serão discutidos mais detalhadamente na seção seguinte.

Em resumo, foi observado um alto número de genes diferencialmente expressos nas comparações entre transcritomas de cepas distintas, além de altos valores de razão de expressão (*fold change*), assim como observado na parte A. Isto sugere que as maiores diferenças entre as cepas não estão necessariamente no seu genoma, mas na regulação da expressão do conteúdo gênico em resposta a determinado ambiente ou estímulo.

4.16.1. Análise de genes relacionados à virulência e patogenicidade

A Tabela S14 contém todos os genes diferencialmente expressos nas comparações entre os transcritomas de todas as cepas analisadas, que já foram relacionados aos sistemas de virulência já descritos em *X. fastidiosa* (Simpson *et al.*, 2000; Van Sluys *et al.*, 2003; Chatterjee *et al.*, 2008).

Dentre as comparações do nível 1, somente as cepas Fb7 e Hib4 apresentaram genes relacionados à virulência dentre os que foram definidos como diferencialmente expressos. Em ambos os casos, um gene relacionado à biogênese do pilus longo (XF2539 e XF2542, para Fb7 e Hib4, respectivamente) foi mais expresso no início do crescimento (Tabela S14).

Nas análises de nível 3, diversos genes de virulência foram encontrados entre todos os pares de cepas, com exceção da comparação entre os transcritomas de início do crescimento das cepas 9a5c e U24d (Tabela S14).

Para melhor visualização dos resultados, gráficos do tipo *Radar Chart* ou *Spider Chart* foram confeccionados de duas formas diferentes. A primeira tem como vértices as oito cepas que tiveram seus transcritomas analisados (Figura 16). Cada um dos *charts* representa um sistema de virulência, sendo que as linhas verdes e vermelhas dizem respeito às comparações de transcritomas no início e final do crescimento, respectivamente. Como todas as cepas foram comparadas entre si, um nível aumentado no gráfico significa que determinada cepa apresentou maior expressão daquele sistema do que outra com a qual foi comparada. Por exemplo, se encontramos um nível 4 no *chart,* aquela cepa teve um maior número de genes positivamente regulados a seu favor do que outras 4 cepas com as quais foi comparada (mínimo 0 e máximo 7). Este tipo de análise fornece um panorama de expressão destes sistemas, e que não leva em consideração os genes individualmente e nem seus valores de razão de expressão. Tais informações estão detalhadas na Tabela S14.



Figura 16. Gráficos do tipo *radar chart* com os resultados das comparações entre as cepas em relação aos sistemas de virulência analisados. As linhas verdes representam as comparações entre transcritomas no início do crescimento exponencial, enquanto que as linhas vermelhas representam as comparações entre transcritomas no final do crescimento exponencial. Cada gráfico representa um sistema de virulência, como segue: (A) Enzimas degradadoras de parede (CWDE), (B) Biogênese do pilus longo, (C) Biogênese do pilus curto, (D) Goma fastidiana, (E) Lipopolissacarídeos, (F) Funções regulatórias, (G) Adesinas autotransportadoras, (H) Hemolisinas, (I) Lipases, (J) Proteases. A cada nível aumentado no gráfico significa que aquela cepa apresentou maior expressão daquele sistema do que em comparação com outra cepa.

Neste primeiro grupo de *Radar charts*, foi observado uma maior expressão de enzimas degradadoras de parede (CWDE) nas cepas U24d e Pr8x, nas comparações do início do crescimento; enquanto isso, J1a12 e Fb7 apresentam maior expressão nas comparações da fase final (Figura 16A). As cepas restantes não apresentaram diferenças significativas quando comparadas com as outras.

No que diz respeito aos genes relacionados com a biogênese do pilus longo, as cepas 3124 e Temecula1 apresentaram maior expressão desse sistema do que outras quatro cepas, nas comparações no início da fase exponencial (Figura 16B). Foram seguidas pelas cepas 9a5c e Pr8x com 3 níveis neste *chart* de genes para o pilus longo. Entretanto, nas comparações de final do crescimento, a cepa Pr8x apresentou maiores níveis, apresentando maior expressão destes genes do que outras quatro cepas. A cepa U24d se juntou com a 9a5c na segunda colocação de expressão destes genes, atingindo o nível 3. Tais cepas com maior expressão de genes do pilus longo seriam, teoricamente, mais móveis do que as outras, dentro das condições de crescimento analisadas.

Enquanto isso, os genes para a biogênese do pilus curto mostraram maior expressão nas cepas Hib4 e Temecula1, nas comparações entre transcritomas do início do crescimento, com 4 níveis cada no *chart* (Figura 16C). Porém, no final da fase exponencial, a cepa U24d apresentou maior expressão destes genes do que outras quatro cepas, sendo que a cepa Hib4 caiu para 3 níveis. Nota-se que a cepa Fb7 não apresentou nenhum nível para este sistema de pilus curto, o qual está relacionado à adesão célulacélula e célula-substrato, o que corrobora com seu fenótipo sem formação de grumos planctônicos e a ausência de células aderidas no frasco de cultivo.

A seguir foram analisados os genes relacionados com a formação de goma fastidiana, um produto celular essencial para a formação do biofilme bacteriano. A cepa que apresentou maior atividade destes genes foi a Fb7, com 5 e 6 níveis nas comparações de início e final da fase exponencial, respectivamente (Figura 16D). Estes dados estão de acordo com a alta viscosidade de suas colônias em placas de cultivo, contrastando com o aspecto seco e áspero de outras cepas de *X. fastidiosa*. Além disso, nota-se claramente a turbidez no meio líquido ao cultivar Fb7. A segunda colocação ficou com a cepa J1a12, a qual também apresenta certa viscosidade em suas colônias, embora não seja como a cepa Fb7 (Figura 16D).

Outro fator de virulência analisado foi a produção de lipopolissacarídeos (LPS). LPSs são componentes da membrana externa, atuando como proteção da célula contra substâncias tóxicas e como modulador da resposta de defesa do hospedeiro (Newman *et al.*, 2001). Genes para a produção de LPS foram mais expressos nas cepas 9a5c e Fb7, com 4 níveis cada, nas comparações entre transcritomas da fase inicial do crescimento exponencial (Figura 16E). Interessantemente, as outras duas cepas que também mostraram uma expressão aumentada destes genes foram U24d e J1a12, também cepas isoladas de citros. Já nas comparações de transcritomas de final do crescimento exponencial, Pr8x e Temecula1 apresentaram níveis no *chart*, embora U24d se mantenha como a cepa com nível mais alto (Figura 16E).

Uma extensa e diversa categoria de genes importantes para a virulência é a daqueles relacionados a funções regulatórias, como reguladores transcricionais e fatores sigma. Nesta categoria, a cepa J1a12 se destacou nas comparações com as outras cepas, atingindo os níveis 7 e 6 nas comparações de início e final da fase exponencial, respectivamente. As outras cepas também mostraram expressões aumentadas em algumas comparações, atingindo até o nível 4 (Fb7 nas comparações de início do crescimento exponencial) (Figura 16F).

Além das adesinas localizadas nas extremidades dos pili longo e curto, *X. fastidiosa* ainda apresenta outros genes codificando para proteínas adesinas afimbriais, como hemaglutininas e adesinas da família dos autotransportadores triméricos (*Xad*) (Simpson *et al.*, 2000; Caserta *et al.*, 2010). São três genes codificando para Xads presentes no genoma de *Xylella* e eles mostraram-se mais expressos na cepa Fb7, atingindo três níveis tanto nas comparações de fase inicial como nas de fase final (Figura 16G). A única outra cepa que mostrou expressão aumentada de genes para Xads foi a cepa U24d nas comparações de início do crescimento.

Assim como pudemos observar altos valores de expressão de bacteriocinas e colicinas em todos os transcritomas analisados, as hemolisinas também constituem um importante grupo de toxinas atuando como fatores de virulência amplamente distribuídos entre bactérias patogênicas Gram-negativas (Simpson *et al.*, 2000; Linhartova *et al.*, 2010). Genes de hemolisinas mostraram-se mais expressos na cepa Fb7, alcançando o nível 4 tanto nas comparações de fase inicial como final (Figura 16H). A cepa 3124 também apresentou expressão aumentada nas comparações entre transcritomas de fase inicial e final.

Outros dois grupos de enzimas que são considerados importantes fatores de virulências em *X. fastidiosa* são as lipases e proteases. Uma das lipases de *X. fastidiosa* já foi descrita como importante fator de virulência para este fitopatógeno (Nascimento

et al., 2016). Neste estudo foi gerado um mutante para a lipase *LesA* (PD1703 e XF0357), o qual apresentou virulência reduzida ao ser inoculado em videiras quando comparado com a cepa selvagem. Além disso, foi descrito que esta proteína seria secretada associada a vesículas de membrana externa e consequentemente poderia causar seu efeito tóxico em partes da planta distantes de sua célula de origem. Dessa forma, lipases estariam associadas diretamente com a patogenicidade de *X. fastidiosa* (Nascimento *et al.*, 2016). Nas análises dos transcritomas foram detectados genes de lipases abundantemente expressos na cepa Fb7, atingindo os níveis máximos nos *charts* das duas comparações (Figura 16I). Além de vários genes para lipases estarem positivamente regulados nesta cepa, ainda foram observados altos valores de razão de expressão (Tabela S14). Outra cepa com alta expressão de lipases foi a J1a12, atingindo os níveis 5 e 6 nas comparações de fase inicial e final, respectivamente. Em relação às cepas restantes, Temecula1 também se destacou atingindo nível 3 nas duas comparações (Figura 16I).

As proteases também constituem um importante grupo de enzimas relacionadas à virulência bacteriana. Elas podem agir diretamente na integridade do tecido do hospedeiro, atuando na inativação de proteínas de defesa, além de lidar com o *turnover* de proteínas bacterianas que tenham sido geradas em resposta às condições adversas encontradas no hospedeiro (Lantz, 1997; Ingmer e Brondsted, 2009). Em relação aos genes codificando proteases em *X. fastidiosa* (Chen *et al.*, 2008; Leite *et al.*, 2013; Zhang *et al.*, 2015; Gouran *et al.*, 2016), foi observada sua alta expressão na cepa J1a12 nas comparações de início da fase exponencial, seguido por Fb7 e U24d (Figura 16J). Entretanto, nas comparações de fase final do crescimento, a cepa Fb7 se destaca atingindo o nível 6, seguido de J1a12 e 9a5c (Figura 16J).

O segundo grupo de *Radar Charts* apresentou como vértices dez categorias de genes de virulência (Figura 17). Neste caso, cada um dos *charts* representa uma cepa que teve seus transcritomas analisados. Novamente, as linhas verdes dizem respeito às comparações de transcritomas no início do crescimento e as vermelhas no seu final. Dessa forma, será possível analisar qualquer mudança significativa de expressão de algum sistema entre as fases do crescimento. O esquema de níveis no *chart* é o mesmo utilizado no primeiro grupo.

A cepa 9a5c apresentou um aumento na expressão de genes relacionados a proteases na fase final do crescimento, enquanto que pode ser observada uma diminuição nas categorias de LPS e genes regulatórios (Figura 17A). Já na cepa U24d,

níveis para biogênese dos pili longo e curto e para a produção de LPS foram aumentados no final do crescimento, com diminuição dos níveis das categorias de CWDEs, Xads, hemolisinas, lipases e proteases (Figura 17B). A cepa J1a12 apresentou mudanças de níveis com o aumento de CWDEs, goma fastidiana, LPS e lipases, além de diminuições de níveis nas categorias de biogênese do pilus longo e proteases (Figura 17C). A cepa Fb7 mostrou um aumento de genes relacionados a CWDEs, goma fastidiana e proteases na fase final do crescimento. Entretanto, foram observadas quedas em níveis das categorias relacionadas a LPS e funções regulatórias (Figura 17D). Na cepa 3124, mudanças foram vistas no aumento dos níveis de expressão de hemolisinas e lipases, embora tenham sido observadas diminuições nas categorias de biogênese do pilus longo, goma fastidiana, funções regulatórias e proteases (Figuras 17E). A cepa Hib4 foi a que apresentou menos alterações entre as duas fases do crescimento exponencial, mostrando somente a queda de um nível na categoria de biogênese do pilus curto (Figura 17F). Já a cepa Pr8x mostrou um aumento da expressão de CWDEs, biogênese do pilus longo e LPS, com queda nos níveis de hemolisinas (Figura 17G). Temecula1 mostrou um aumento de genes relacionados a funções regulatórias e proteases na fase final do crescimento, embora tenha sido vista uma diminuição nos níveis de genes para biogênese dos pili longo e curto (Figura 17H).



Figura 17. Gráficos do tipo *radar chart* com os resultados das comparações entre as cepas em relação aos sistemas de virulência analisados. As linhas verdes representam as comparações entre transcritomas no início do crescimento, enquanto que as linhas vermelhas representam as comparações entre transcritomas no final do crescimento. Cada gráfico representa uma cepa, como segue: (A) 9a5c, (B) U24d, (C) J1a12, (D) Fb7, (E) 3124, (F) Hib4, (G) Pr8x, (H) Temecula1. A cada nível aumentado no gráfico significa que aquela cepa apresentou maior expressão daquele sistema do que em comparação com outra cepa.

5. Considerações finais

O sequenciamento de transcritomas de diferentes cepas de *Xylella fastidiosa*, no início e fim da fase exponencial de crescimento no meio rico PWG e no meio mínimo PIM6 possibilitou uma descrição abrangente do repertório de genes que é expresso nas diferentes cepas e condições de cultivo. A tecnologia de RNA-Seq mostrou-se altamente reprodutível e robusta, fornecendo enorme quantidade de dados de ótima qualidade, e sem precedentes para *X. fastidiosa*. Além da definição de perfis transcricionais e análises de expressão diferencial entre transcritomas, foi possível utilizar os dados gerados para definir as coordenadas genômicas exatas para os transcritos, descrevendo suas regiões 5' e 3' não-traduzidas. Além disso, as estruturas de operons expressos nas condições investigadas também foram definidas para as cepas 9a5c e Temecula1. Pela primeira vez foi descrito o perfil de sRNAs expressos por *X. fastidiosa* (cepa 9a5c).

Os transcritomas mostraram que a maior parte dos genes é transcrita em todas as cepas, mesmo que seus transcritos sejam pouco abundantes. Em todos os transcritomas foram observados ncRNAs muito abundantes, como os da RNase P e do 6S. Outra observação interessante foi a elevada expressão de genes de bacteriocinas (colicinas, microcinas, hemolisinas) sugerindo que mesmo em um cultivo *in vitro*, sem a presença de espécies competidoras, a *X. fastidiosa* expressa seu arsenal de toxinas. As principais vias metabólicas de manutenção da sobrevivência da célula foram encontradas ativas em todas as cepas e transcritomas.

Os transcritomas das cepas 9a5c e Temecula1 na condição de crescimento em PIM6, um meio que mimetiza a composição da seiva do xilema, revelou um conjunto de genes que tem relação com a carência de nutrientes que a bactéria enfrenta no xilema. O crescimento da população microbiana em PIM6 foi gravemente afetado pela escassez nutricional, antecipando a fase estacionária e com menor concentração celular. Ainda assim, nesta condição foi observada a expressão de genes que parecem relevantes para manutenção da sobrevivência no inóspito ambiente do hospedeiro vegetal e/ou do inseto vetor.

As análises de expressão diferencial entre os transcritomas das duas fases de crescimento em um mesmo meio evidenciaram que o estresse gerado pela limitação nutricional do meio PIM6 exigiu mudanças mais drásticas na expressão gênica, devido ao maior número de genes diferencialmente expressos observados. Além disso, o contato inicial com este meio estimulou uma maior ativação da transcrição. As comparações entre os transcritomas em meio PWG não mostraram muitas mudanças na expressão de seus genes,

sugerindo que o meio rico em nutrientes apresentou condições adequadas para o crescimento desde o início do cultivo, as quais não sofreram muitas alterações até o final da fase exponencial. Quando comparamos transcritomas de diferentes meios, foram observadas maiores diferenças na fase final exponencial, mostrando que o contato inicial com ambos os meios ativa basicamente os mesmos genes e na mesma intensidade, com divergências sendo observadas após o cultivo prolongado.

As comparações entre os transcritomas de diferentes cepas mostraram que genes ortólogos respondem de maneiras distintas a uma mesma condição, evidenciando que aparentemente são regulados de formas diferentes.

A comparação da expressão de genes de fatores de virulência das cepas de *X*. *fastidiosa* revelou diferenças significativas interessantes. Não foi observada uma homogeneidade na resposta para estes sistemas, sendo que cada sistema de virulência analisado se mostrou mais ativo em determinadas cepas. Em alguns casos foi possível correlacionar estas particularidades aos fenótipos *in vitro* e virulência *in planta*.

As análises dos transcritomas evidenciaram vários genes candidatos que poderão ser futuramente investigados em maior detalhe quanto ao seu papel na biologia e na virulência de *X. fastidiosa*. Além disso, a transcritômica comparativa expôs diferenças relevantes na regulação gênica de cepas de hospedeiros vegetais distintos que podem estar relacionadas a especificidade ao hospedeiro.

6. Bibliografia

- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F. e et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651-6.
- Aguero, C. B., Uratsu, S. L., Greve, C., Powell, A. L., Labavitch, J. M., Meredith, C. P. e Dandekar, A. M. 2005. Evaluation of tolerance to Pierce's disease and Botrytis in transgenic plants of Vitis vinifera L. expressing the pear PGIP gene. Mol Plant Pathol 6:43-51.
- 3. Alencar, V. C., Barbosa, D., Santos, D. S., Oliveira, A. C., de Oliveira, R. C. e Nunes, L. R. 2014. Genomic Sequencing of Two Coffee-Infecting Strains of *Xylella fastidiosa* Isolated from Brazil. Genome Announc 2.
- Alencar, V. C., Jabes, D. L., Menegidio, F. B., Sassaki, G. L., de Souza, L. R., Puzer, L., Meneghetti, M. C., Lima, M. A., Tersariol, I. L., de Oliveira, R. C. e Nunes, L. R. 2017. Functional and Evolutionary Characterization of a UDP-Xylose Synthase Gene from the Plant Pathogen *Xylella fastidiosa*, Involved in the Synthesis of Bacterial Lipopolysaccharide. Biochemistry 56:779-792.
- 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 Nunney, L. 2015. How do plant diseases caused by *Xylella fastidiosa* emerge? Plant disease 99:1457-1467.
- 8. **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 the Entomological Society of America 99:884-890.
- 9. Altman, N. S. e Hua, J. 2006. Extending the loop design for two-channel microarray experiments. Genet Res 88:153-63.
- 10. Altman, S. 2011. Ribonuclease P. Philos Trans R Soc Lond B Biol Sci 366:2936-41.
- 11. 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.
- 12. **Amanifar, N., Taghavi, M., Izadpanah, K. e Babaei, G.** 2014. Isolation and pathogenicity of *Xylella fastidiosa* from grapevine and almond in Iran. Phytopathologia Mediterranea 53:318-327.
- 13. Ancona, V., Li, W. e Zhao, Y. 2014. Alternative sigma factor RpoN and its modulation protein YhbH are indispensable for *Erwinia amylovora* virulence. Mol Plant Pathol 15:58-66.
- 14. **Anders, S. e Huber, W.** 2010. Differential expression analysis for sequence count data. Genome Biol 11:R106.
- 15. Andrews, S., Lindenbaum, P., Howard, B. e Ewels, P. 2014. FastQC High Throughput Sequence QC Report, 0.11.2 ed.
- Araujo, W. L., Marcon, J., Maccheroni, W., Jr., Van Elsas, J. D., Van Vuurde, J. W. e Azevedo, J. L. 2002. Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. Appl Environ Microbiol 68:4906-14.

- 17. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H. *et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25-9.
- 18. Auger, J. G., Shalla, T. A. e Kado, C. I. 1974. Pierce's disease of grapevines: evidence for a bacterial etiology. Science 184:1375-7.
- 19. **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.
- 20. **Backus, E. A., Shugart, H. J., Rogers, E. E., Morgan, J. K. e Shatters, R.** 2015. Direct Evidence of Egestion and Salivation of *Xylella fastidiosa* Suggests Sharpshooters Can Be "Flying Syringes". Phytopathology 105:608-20.
- 21. Bahar, O., Pruitt, R., Luu, D. D., Schwessinger, B., Daudi, A., Liu, F., Ruan, R., Fontaine-Bodin, L., Koebnik, R. e Ronald, P. 2014. The *Xanthomonas* Ax21 protein is processed by the general secretory system and is secreted in association with outer membrane vesicles. PeerJ 2:e242.
- Balleza, E., Lopez-Bojorquez, L. N., Martinez-Antonio, A., Resendis-Antonio, O., Lozada-Chavez, I., Balderas-Martinez, Y. I., Encarnacion, S. e Collado-Vides, J. 2009. Regulation by transcription factors in bacteria: beyond description. FEMS Microbiol Rev 33:133-51.
- Barbosa, D., Alencar, V. C., Santos, D. S., de Freitas Oliveira, A. C., de Souza, A. A., Coletta-Filho, H. D., de Oliveira, R. S. e Nunes, L. R. 2015. Comparative genomic analysis of coffee-infecting *Xylella fastidiosa* strains isolated from Brazil. Microbiology 161:1018-33.
- 24. **Battesti, A., Majdalani, N. e Gottesman, S.** 2011. The RpoS-mediated general stress response in *Escherichia coli*. Annu Rev Microbiol 65:189-213.
- 25. **Baumgartner, K. e Warren, J. G.** 2005. Persistence of *Xylella fastidiosa* in riparian hosts near northern California vineyards. Plant disease 89:1097-1102.
- 26. **Beier, D. e Gross, R.** 2006. Regulation of bacterial virulence by two-component systems. Curr Opin Microbiol 9:143-52.
- 27. **Benjamini, Y. e Hochberg, Y.** 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57:289-300.
- 28. Beretta, M. J. G., Harakava, R., Chagas, C. M., Derrick, K. S., Barthe, G. A., Ceccardi, T. L., Lee, R. F., Paradela, O., Sugimori, M. e Ribeiro, I. A. 1996. First report of *Xylella fastidiosa* in coffee. Plant disease 80:821.
- 29. Bergsma-Vlami, M., van de Bilt, J. L. J., Tjou-Tam-Sin, N. N. A., van de Vossenberg, B. T. L. H. e Westenberg, M. 2015. *Xylella fastidiosa* in *Coffea arabica* ornamental plants imported from Costa Rica and Honduras in the Netherlands. Journal of Plant Pathology 97:395.
- 30. Berisha, B., Chen, Y. D., Zhang, G. Y., Xu, B. Y. e Chen, T. A. 1998. Isolation of Peirce's disease bacteria from grapevines in Europe. European Journal of Plant Pathology 104:427-433.
- 31. **Bhattacharyya, A., Stilwagen, S., Ivanova, N., D'Souza, M., Bernal, A.** *et al.* 2002. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. Proc Natl Acad Sci U S A 99:12403-8.
- 32. Bird, A. 2007. Perceptions of epigenetics. Nature 447:396-8.
- 33. **Bonaldo, M. F., Lennon, G. e Soares, M. B.** 1996. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 6:791-806.

- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P. et al. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29:365-71.
- 35. **Breaker, R. R.** 2011. Prospects for riboswitch discovery and analysis. Mol Cell 43:867-79.
- 36. **Browning, D. F. e Busby, S. J.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol 2:57-65.
- 37. **Burbank, L. P. e Stenger, D. C.** 2017. The DinJ/RelE Toxin-Antitoxin System Suppresses Bacterial Proliferation and Virulence of *Xylella fastidiosa* in Grapevine. Phytopathology 107:388-394.
- 38. Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., Eddy, S. R., Gardner, P. P. e Bateman, A. 2013. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 41:D226-32.
- 39. Canales, R. D., Luo, Y., Willey, J. C., Austermiller, B., Barbacioru, C. C. *et al.* 2006. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotechnol 24:1115-22.
- 40. Cariddi, C., Saponari, M., Boscia, D., De Stradis, A., Loconsole, G., Nigro, F., Porcelli, F., Potere, O. e Martelli, G. P. 2014. Isolation of a *Xylella fastidiosa* strain infecting olive and oleander in Apulia, Italy. Journal of Plant Pathology 96:425-429.
- 41. **Casadesus, J. e Low, D.** 2006. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev 70:830-56.
- 42. **Casadesus, J. e Low, D. A.** 2013. Programmed heterogeneity: epigenetic mechanisms in bacteria. J Biol Chem 288:13929-35.
- 43. 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.
- 44. **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*. Curr Microbiol 27:137-42.
- 45. **Chatterjee, S., Almeida, R. P. e Lindow, S.** 2008. Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. Annu Rev Phytopathol 46:243-71.
- 46. **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.
- 47. Chen, J., Civerolo, E., Tubajika, K., Livingston, S. e Higbee, B. 2008. Hypervariations of a protease-encoding gene, PD0218 (pspB), in *Xylella fastidiosa* strains causing almond leaf scorch and Pierce's disease in California. Appl Environ Microbiol 74:3652-7.
- 48. **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.
- 49. Chen, J., Huang, H., Chang, C. J. e Stenger, D. C. 2013. Draft Genome Sequence of *Xylella fastidiosa* subsp. *multiplex* Strain Griffin-1 from *Quercus rubra* in Georgia. Genome Announc 1.
- 50. **Chen, J., Wu, F., Zheng, Z., Deng, X., Burbank, L. P. e Stenger, D. C.** 2016. Draft Genome Sequence of *Xylella fastidiosa* subsp. *fastidiosa* Strain Stag's Leap. Genome Announc 4.

- 51. 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.
- 52. **Chevreux, B., Wetter, T. e Suhai, S.** 1999. Genome sequence assembly using trace signals and additional sequence information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99:45-46.
- 53. Ciraulo, M. B., Santos, D. S., Rodrigues, A. C., de Oliveira, M. V., Rodrigues, T., de Oliveira, R. C. e Nunes, L. R. 2010. Transcriptome analysis of the phytobacterium *Xylella fastidiosa* growing under xylem-based chemical conditions. J Biomed Biotechnol 2010;781365.
- 54. **Clarke, J. D. e Zhu, T.** 2006. Microarray analysis of the transcriptome as a stepping stone towards understanding biological systems: practical considerations and perspectives. Plant J 45:630-50.
- 55. **Clifford, J. C., Rapicavoli, J. N. e Roper, M. C.** 2013. A rhamnose-rich O-antigen mediates adhesion, virulence, and host colonization for the xylem-limited phytopathogen *Xylella fastidiosa*. Mol Plant Microbe Interact 26:676-85.
- 56. **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.
- 57. Collier, D. N., Hager, P. W. e Phibbs, P. V., Jr. 1996. Catabolite repression control in the Pseudomonads. Res Microbiol 147:551-61.
- 58. **Condon, C.** 2007. Maturation and degradation of RNA in bacteria. Curr Opin Microbiol 10:271-8.
- 59. **Condon, C. e Putzer, H.** 2002. The phylogenetic distribution of bacterial ribonucleases. Nucleic Acids Res 30:5339-46.
- 60. **Conway, T. e Schoolnik, G. K.** 2003. Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. Mol Microbiol 47:879-89.
- 61. Cursino, L., Galvani, C. D., Athinuwat, D., Zaini, P. A., Li, Y., De La Fuente, L., Hoch, H. C., Burr, T. J. e Mowery, P. 2011. Identification of an operon, Pil-Chp, that controls twitching motility and virulence in *Xylella fastidiosa*. Mol Plant Microbe Interact 24:1198-206.
- 62. Cussiol, J. R., Alves, S. V., de Oliveira, M. A. e Netto, L. E. 2003. Organic hydroperoxide resistance gene encodes a thiol-dependent peroxidase. J Biol Chem 278:11570-8.
- 63. 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.
- 64. **da Silva Neto, J. F., Koide, T., Gomes, S. L. e Marques, M. V.** 2007. The single extracytoplasmic-function sigma factor of *Xylella fastidiosa* is involved in the heat shock response and presents an unusual regulatory mechanism. J Bacteriol 189:551-60.
- 65. **da Silva Neto, J. F., Koide, T., Gomes, S. L. e Marques, M. V.** 2010. Global gene expression under nitrogen starvation in *Xylella fastidiosa*: contribution of the sigma54 regulon. BMC Microbiol 10:231.
- 66. 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.
- 67. **Dalman, M. R., Deeter, A., Nimishakavi, G. e Duan, Z. H.** 2012. Fold change and p-value cutoffs significantly alter microarray interpretations. BMC Bioinformatics 13 Suppl 2:S11.
- 68. **Dam, P., Olman, V., Harris, K., Su, Z. e Xu, Y.** 2007. Operon prediction using both genome-specific and general genomic information. Nucleic Acids Res 35:288-98.
- Dandekar, A. M., Gouran, H., Ibanez, A. M., Uratsu, S. L., Aguero, C. B., McFarland, S., Borhani, Y., Feldstein, P. A., Bruening, G., Nascimento, R., Goulart, L. R., Pardington, P. E., Chaudhary, A., Norvell, M., Civerolo, E. e Gupta, G. 2012. An engineered innate immune defense protects grapevines from Pierce disease. Proc Natl Acad Sci U S A 109:3721-5.
- 70. **Danhorn, T. e Fuqua, C.** 2007. Biofilm formation by plant-associated bacteria. Annu Rev Microbiol 61:401-22.
- 71. **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.
- 72. **Davis, M. J., Purcell, A. H. e Thomson, S. V.** 1978. Pierce's disease of grapevines: isolation of the causal bacterium. Science 199:75-7.
- 73. **De La Fuente, L., Montanes, E., Meng, Y., Li, Y., Burr, T. J., Hoch, H. C. e Wu, M.** 2007. Assessing adhesion forces of type I and type IV pili of *Xylella fastidiosa* bacteria by use of a microfluidic flow chamber. Appl Environ Microbiol 73:2690-6.
- 74. **De Lay, N., Schu, D. J. e Gottesman, S.** 2013. Bacterial small RNA-based negative regulation: Hfq and its accomplices. J Biol Chem 288:7996-8003.
- 75. 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.
- 76. de Souza, A. A., Takita, M. A., Coletta-Filho, H. D., Caldana, C., Goldman, G. H., Yanai, G. M., Muto, N. H., de Oliveira, R. C., Nunes, L. R. e Machado, M. A. 2003. Analysis of gene expression in two growth states of *Xylella fastidiosa* and its relationship with pathogenicity. Mol Plant Microbe Interact 16:867-75.
- 77. de Souza, A. A., Takita, M. A., Coletta-Filho, H. D., Caldana, C., Yanai, G. M., Muto, N. H., de Oliveira, R. C., Nunes, L. R. e Machado, M. A. 2004. Gene expression profile of the plant pathogen *Xylella fastidiosa* during biofilm formation in vitro. FEMS Microbiol Lett 237:341-53.
- 78. **Dekhtyar, M., Morin, A. e Sakanyan, V.** 2008. Triad pattern algorithm for predicting strong promoter candidates in bacterial genomes. BMC Bioinformatics 9:233.
- 79. **Deutscher, M. P.** 2003. Degradation of stable RNA in bacteria. J Biol Chem 278:45041-4.
- 80. **Deutscher, M. P.** 2006. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. Nucleic Acids Res 34:659-66.
- 81. **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.
- 82. **Dharmadi, Y. e Gonzalez, R.** 2004. DNA microarrays: experimental issues, data analysis, and application to bacterial systems. Biotechnol Prog 20:1309-24.
- 83. Dias Neto, E., Correa, R. G., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A. *et al.* 2000. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proc Natl Acad Sci U S A 97:3491-6.
- 84. **Doddapaneni, H., Francis, M., Yao, J., Lin, H. e Civerolo, E. L.** 2007. Genomewide analysis of *Xylella fastidiosa*: implications for detection and strain relationships. African Journal of Biotechnology 6:55-66.
- 85. Dotsch, A., Eckweiler, D., Schniederjans, M., Zimmermann, A., Jensen, V., Scharfe, M., Geffers, R. e Haussler, S. 2012. The *Pseudomonas aeruginosa*

transcriptome in planktonic cultures and static biofilms using RNA sequencing. PLoS One 7:e31092.

- 86. Dourado, M. N., Santos, D. S., Nunes, L. R., Costa de Oliveira, R. L., de Oliveira, M. V. e Araujo, W. L. 2015. Differential gene expression in *Xylella fastidiosa* 9a5c during co-cultivation with the endophytic bacterium *Methylobacterium mesophilicum* SR1.6/6. J Basic Microbiol 55:1357-66.
- 87. **EFSA.** 2016. Update of a database of host plants of *Xylella fastidiosa:* 20 November 2015. EFSA Journal 14:4378.
- Elbeaino, T., Valentini, F., Abou-Kubaa, R., Moubarak, P., Yaseen, T. e Digiaro, M. 2014. Multilocus sequence typing of *Xylella fastidiosa* isolated from olive affected by "olive quick decline syndrome" in Italy. Phytopathologia Mediterranea 53:533-542.
- 89. **Esakova, O. e Krasilnikov, A. S.** 2010. Of proteins and RNA: the RNase P/MRP family. RNA 16:1725-47.
- 90. **Evans, D., Marquez, S. M. e Pace, N. R.** 2006. RNase P: interface of the RNA and protein worlds. Trends Biochem Sci 31:333-41.
- 91. Fedatto, L. M., Silva-Stenico, M. E., Etchegaray, A., Pacheco, F. T. H., Rodrigues, J. L. M. e Tsai, S. M. 2006. Detection and characterization of protease secreted by the plant pathogen *Xylella fastidiosa*. Microbiological Research 161:263-272.
- 92. Feklistov, A., Sharon, B. D., Darst, S. A. e Gross, C. A. 2014. Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol 68:357-76.
- 93. **Filiatrault, M. J.** 2011. Progress in prokaryotic transcriptomics. Curr Opin Microbiol 14:579-86.
- 94. Filiatrault, M. J., Stodghill, P. V., Bronstein, P. A., Moll, S., Lindeberg, M., Grills, G., Schweitzer, P., Wang, W., Schroth, G. P., Luo, S., Khrebtukova, I., Yang, Y., Thannhauser, T., Butcher, B. G., Cartinhour, S. e Schneider, D. J. 2010. Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. J Bacteriol 192:2359-72.
- 95. Filiatrault, M. J., Stodghill, P. V., Myers, C. R., Bronstein, P. A., Butcher, B. G., Lam, H., Grills, G., Schweitzer, P., Wang, W., Schneider, D. J. e Cartinhour, S. W. 2011. Genome-wide identification of transcriptional start sites in the plant pathogen *Pseudomonas syringae* pv. *tomato* str. DC3000. PLoS One 6:e29335.
- 96. Fogaca, A. C., Zaini, P. A., Wulff, N. A., da Silva, P. I., Fazio, M. A., Miranda, A., Daffre, S. e da Silva, A. M. 2010. Effects of the antimicrobial peptide gomesin on the global gene expression profile, virulence and biofilm formation of *Xylella fastidiosa*. FEMS Microbiol Lett 306:152-9.
- 97. **French, W. J. e Kitajima, E. W.** 1978. Occurrence of plum leaf scald in Brazil and Paraguay. Plant Disease Reporter 62:1035-1038.
- 98. **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.
- 99. Galperin, M. Y. 2004. Bacterial signal transduction network in a genomic perspective. Environ Microbiol 6:552-67.
- 100. Gambetta, G. A., Fei, J., Rost, T. L. e Matthews, M. A. 2007. Leaf scorch symptoms are not correlated with bacterial populations during Pierce's disease. J Exp Bot 58:4037-46.
- 101. **Gene Ontology Consortium.** 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32:258-261.

- 102. **Georg, J. e Hess, W. R.** 2011. cis-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev 75:286-300.
- 103. **Ghosh, S. e Deutscher, M. P.** 1999. Oligoribonuclease is an essential component of the mRNA decay pathway. Proc Natl Acad Sci U S A 96:4372-7.
- 104. Giampetruzzi, A., Chiumenti, M., Saponari, M., Donvito, G., Italiano, A., Loconsole, G., Boscia, D., Cariddi, C., Martelli, G. P. e Saldarelli, P. 2015a. Draft Genome Sequence of the *Xylella fastidiosa* CoDiRO Strain. Genome Announc 3.
- 105. Giampetruzzi, A., Loconsole, G., Boscia, D., Calzolari, A., Chiumenti, M., Martelli, G. P., Saldarelli, P., Almeida, R. P. e Saponari, M. 2015b. Draft Genome Sequence of CO33, a Coffee-Infecting Isolate of *Xylella fastidiosa*. Genome Announc 3.
- 106. Giampetruzzi, A., Morelli, M., Saponari, M., Loconsole, G., Chiumenti, M., Boscia, D., Savino, V. N., Martelli, G. P. e Saldarelli, P. 2016. Transcriptome profiling of two olive cultivars in response to infection by the CoDiRO strain of *Xylella fastidiosa* subsp. *pauca*. BMC Genomics 17:475.
- 107. **Gollnick, P. e Babitzke, P.** 2002. Transcription attenuation. Biochim Biophys Acta 1577:240-50.
- 108. **Gorke, B. e Stulke, J.** 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol 6:613-24.
- 109. **Gottesman, S.** 2004. The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annu Rev Microbiol 58:303-28.
- 110. Gouran, H., Gillespie, H., Nascimento, R., Chakraborty, S., Zaini, P. A., Jacobson, A., Phinney, B. S., Dolan, D., Durbin-Johnson, B. P., Antonova, E. S., Lindow, S. E., Mellema, M. S., Goulart, L. R. e Dandekar, A. M. 2016. The Secreted Protease PrtA Controls Cell Growth, Biofilm Formation and Pathogenicity in *Xylella fastidiosa*. Sci Rep 6:31098.
- 111. **Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A.** *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644-52.
- 112. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. e Eddy, S. R. 2003. Rfam: an RNA family database. Nucleic Acids Res 31:439-41.
- 113. **Gruber, A.** 2007. Expressed sequence tags, p. 141-167, Bioinformatics: Methods Express. Scion Publishing Limited.
- 114. **Guan, W., Shao, J., Davis, R. E., Zhao, T. e Huang, Q.** 2014a. Genome Sequence of a *Xylella fastidiosa* Strain Causing Sycamore Leaf Scorch Disease in Virginia. Genome Announc 2.
- 115. **Guan, W., Shao, J., Zhao, T. e Huang, Q.** 2014b. Genome Sequence of a *Xylella fastidiosa* Strain Causing Mulberry Leaf Scorch Disease in Maryland. Genome Announc 2.
- 116. **Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. e Altman, S.** 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35:849-57.
- 117. **Guilhabert, M. R. e Kirkpatrick, B. C.** 2005. Identification of *Xylella fastidiosa* antivirulence genes: hemagglutinin adhesins contribute a biofilm maturation to X. fastidios and colonization and attenuate virulence. Mol Plant Microbe Interact 18:856-68.
- 118. Güldumlaut~r, M. E., Çaglar, B. K., Castellano, M. A., Ünlü, L., Güran, S., Yilmaz, M. A. e Martelli, G. P. 2005. First report of almond leaf scorch in Turkey. Journal of Plant Pathology 87:246.

- 119. **Gutierrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C. e Merino, E.** 2009. Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. Microbiol Mol Biol Rev 73:36-61.
- 120. Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W. e Livny, J. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics 13.
- 121. Haelterman, R. M., Tolocka, P. A., Roca, M. E., Guzmán, F. A., Fernández, F. D. e Otero, M. L. 2015. First presumptive diagnosis of *Xylella fastidiosa* causing olive scorch in Argentina. Journal of Plant Pathology 97:393.
- 122. Hall-Stoodley, L., Costerton, J. W. e Stoodley, P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol 2:95-108.
- 123. Han, S. W., Lee, S. W. e Ronald, P. C. 2011. Secretion, modification, and regulation of Ax21. Curr Opin Microbiol 14:62-7.
- 124. Han, Y., Zhou, D., Pang, X., Song, Y., Zhang, L., Bao, J., Tong, Z., Wang, J., Guo, Z., Zhai, J., Du, Z., Wang, X., Zhang, X., Huang, P. e Yang, R. 2004. Microarray analysis of temperature-induced transcriptome of *Yersinia pestis*. Microbiol Immunol 48:791-805.
- 125. **Hardcastle, T. J. e Kelly, K. A.** 2010. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422.
- 126. **Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T. e Gerdes, K.** 2015. Recent functional insights into the role of (p)ppGpp in bacterial physiology. Nat Rev Microbiol 13:298-309.
- 127. Hengge, R., Grundling, A., Jenal, U., Ryan, R. e Yildiz, F. 2016. Bacterial Signal Transduction by Cyclic Di-GMP and Other Nucleotide Second Messengers. J Bacteriol 198:15-26.
- 128. **Henkin, T. M. e Yanofsky, C.** 2002. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. Bioessays 24:700-7.
- Hernandez-Martinez, R., de la Cerda, K. A., Costa, H. S., Cooksey, D. A. e Wong, F. P. 2007. Phylogenetic Relationships of *Xylella fastidiosa* Strains Isolated from Landscape Ornamentals in Southern California. Phytopathology 97:857-64.
- 130. **Hoch, J. A.** 2000. Two-component and phosphorelay signal transduction. Curr Opin Microbiol 3:165-70.
- 131. **Hopkins, D. L.** 1989. *Xylella fastidiosa*: xylem-limited bacterial pathogen of plants. Ann. Rev. Phytopathol. 27:271-290.
- 132. **Hopkins, D. L.** 2005. Biological control of Pierce's Disease in the vineyard with strains of *Xylella fastidiosa* benign to grapevine. Plant disease 89:1348-1352.
- 133. **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.
- 134. Huang, H. Y., Chang, H. Y., Chou, C. H., Tseng, C. P., Ho, S. Y., Yang, C. D., Ju, Y. W. e Huang, H. D. 2009. sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. Nucleic Acids Res 37:D150-4.
- 135. **Ingmer, H. e Brondsted, L.** 2009. Proteases in bacterial pathogenesis. Res Microbiol 160:704-10.
- 136. **Ionescu, M., Zaini, P. A., Baccari, C., Tran, S., da Silva, A. M. e Lindow, S. E.** 2014. *Xylella fastidiosa* outer membrane vesicles modulate plant colonization by blocking attachment to surfaces. Proc Natl Acad Sci U S A 111:E3910-8.
- 137. Irizarry, R. A., Warren, D., Spencer, F., Kim, I. F., Biswal, S. *et al.* 2005. Multiple-laboratory comparison of microarray platforms. Nat Methods 2:345-50.

- 138. **Jacob, F. e Monod, J.** 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318-56.
- 139. Jacques, M.-A., Denancé, N., Legendre, B., Morel, E., Briand, M., Mississipi, S., Durand, K., Olivier, V., Portier, P., Poliakoff, F. e Crouzillat, D. 2016. New coffee plant-infecting *Xylella fastidiosa* variants derived via homologous recombination. Appl Environ Microbiol 82:1556-1568.
- 140. Janga, S. C., Salgado, H., Collado-Vides, J. e Martinez-Antonio, A. 2007. Internal versus external effector and transcription factor gene pairs differ in their relative chromosomal position in Escherichia coli. J Mol Biol 368:263-72.
- 141. **Janse, J. D. e Obradovic, A.** 2010. *Xylella fastidiosa*: its biology, diagnosis, control and risks. Journal of Plant Pathology 92:S1.35-S1.48.
- 142. **Jousselin, A., Metzinger, L. e Felden, B.** 2009. On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends Microbiol 17:399-405.
- 143. **Kanehisa, M. e Goto, S.** 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30.
- 144. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. e Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457-62.
- 145. Kazmierczak, M. J., Wiedmann, M. e Boor, K. J. 2005. Alternative sigma factors and their roles in bacterial virulence. Microbiol Mol Biol Rev 69:527-43.
- 146. **Keene, J. D.** 2007. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8:533-43.
- 147. **Killiny, N. e Almeida, R. P.** 2009. Host structural carbohydrate induces vector transmission of a bacterial plant pathogen. Proc Natl Acad Sci U S A 106:22416-20.
- 148. **Killiny, N. e Almeida, R. P.** 2011. Gene regulation mediates host specificity of a bacterial pathogen. Environ Microbiol Rep 3:791-7.
- 149. Killiny, N., Prado, S. S. e Almeida, R. P. 2010. Chitin utilization by the insecttransmitted bacterium *Xylella fastidiosa*. Appl Environ Microbiol 76:6134-40.
- 150. Kin, T., Yamada, K., Terai, G., Okida, H., Yoshinari, Y., Ono, Y., Kojima, A., Kimura, Y., Komori, T. e Asai, K. 2007. fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. Nucleic Acids Res 35:D145-8.
- 151. 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.
- 152. 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.
- 153. Kolesov, G., Wunderlich, Z., Laikova, O. N., Gelfand, M. S. e Mirny, L. A. 2007. How gene order is influenced by the biophysics of transcription regulation. PNAS 104:13948-13953.
- 154. **Krugner, R., Sisterson, M. S., Chen, J., Stenger, D. C. e Johnson, M. W.** 2014. Evaluation of olive as a host of *Xylella fastidiosa* and associated sharpshooter vectors. Plant disease 98:1186-1193.
- 155. Kwenda, S., Gorshkov, V., Ramesh, A. M., Naidoo, S., Rubagotti, E., Birch, P. R. e Moleleki, L. N. 2016. Discovery and profiling of small RNAs responsive to stress conditions in the plant pathogen Pectobacterium atrosepticum. BMC Genomics 17:47.

- 156. **Labroussaa, F., Ionescu, M., Zeilinger, A., Lindow, S. e Almeida, R.** 2017. A chitinase is required for *Xylella fastidiosa* colonization of its insect and plant hosts. Microbiology.
- 157. Lacava, P. T., Andreote, F. D., Araújo, W. L. e Azevedo, J. L. 2006. Caracterização da comunidade bacteriana endofítica de citros por isolamento, PCR específico e DGGE. Pesq. agropec. bras. 41:637-642.
- 158. Lacava, P. T., Araujo, W. L., Marcon, J., Maccheroni, W., Jr. e Azevedo, J. L. 2004. Interaction between endophytic bacteria from citrus plants and the phytopathogenic bacteria *Xylella fastidiosa*, causal agent of citrus-variegated chlorosis. Lett Appl Microbiol 39:55-9.
- 159. Lantz, M. S. 1997. Are bacterial proteases important virulence factors? J Periodontal Res 32:126-32.
- 160. Lee, M. L., Kuo, F. C., Whitmore, G. A. e Sklar, J. 2000. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci U S A 97:9834-9.
- 161. 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.
- 162. Lee, M. W., Rogers, E. E. e Stenger, D. C. 2012. *Xylella fastidiosa* plasmid-encoded PemK toxin is an endoribonuclease. Phytopathology 102:32-40.
- 163. Lee, M. W., Tan, C. C., Rogers, E. E. e Stenger, D. C. 2014. Toxin-antitoxin systems mqsR/ygiT and dinJ/relE of *Xylella fastidiosa*. Physiological and Molecular Plant Pathology 87:59-68.
- 164. Leite, B., Andersen, P. C. e Ishida, M. L. 2004. Colony aggregation and biofilm formation in xylem chemistry-based media for *Xylella fastidiosa*. FEMS Microbiol Lett 230:283-90.
- 165. Leite, N. R., Faro, A. R., Dotta, M. A. O., Faim, L. M., Gianotti, A., Silva, F. H., Oliva, G. e Thiemann, O. H. 2013. The crystal structure of the cysteine protease Xylellain from *Xylella fastidiosa* reveals an intriguing activation mechanism. Febs Letters 587:339-344.
- Lemos, E. G., Alves, L. M. e Campanharo, J. C. 2003. Genomics-based design of defined growth media for the plant pathogen *Xylella fastidiosa*. FEMS Microbiol Lett 219:39-45.
- 167. Leu, L. S. e Su, C. C. 1993. Isolation, cultivation, and pathogenicity of *Xylella fastidiosa*, the causal bacterium of pear leaf scorch disease in Taiwan Plant disease 77:642-646.
- Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A. e Regev, A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 7:709-15.
- 169. Lewis, M. 2005. The lac repressor. C R Biol 328:521-48.
- 170. Li, L., Huang, D., Cheung, M. K., Nong, W., Huang, Q. e Kwan, H. S. 2013. BSRD: a repository for bacterial small regulatory RNA. Nucleic Acids Res 41:D233-8.
- 171. Li, W.-B., Pria, W. D., Teixeira, D. 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.
- 172. Li, W. B., Pria, W. D., Lacava, P. M., Qin, X. e Hartung, J. S. 2003. Presence of *Xylella fastidiosa* in Sweet Orange Fruit and Seeds and Its Transmission to Seedlings. Phytopathology 93:953-8.

- 173. 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.
- 174. 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.
- 175. Liang, H., Zhao, Y. T., Zhang, J. Q., Wang, X. J., Fang, R. X. e Jia, Y. T. 2011. Identification and functional characterization of small non-coding RNAs in *Xanthomonas oryzae* pathovar *oryzae*. BMC Genomics 12:87.
- 176. Lindow, S., Newman, K., Chatterjee, S., Baccari, C., Lavarone, A. T. e Ionescu, M. 2014. Production of *Xylella fastidiosa* diffusible signal factor in transgenic grape causes pathogen confusion and reduction in severity of Pierce's disease. Mol Plant Microbe Interact 27:244-54.
- 177. Linhartova, I., Bumba, L., Masin, J., Basler, M., Osicka, R., Kamanova, J., Prochazkova, K., Adkins, I., Hejnova-Holubova, J., Sadilkova, L., Morova, J. e Sebo, P. 2010. RTX proteins: a highly diverse family secreted by a common mechanism. FEMS Microbiol Rev 34:1076-112.
- 178. Liu, W., Yu, Y. H., Cao, S. Y., Niu, X. N., Jiang, W., Liu, G. F., Jiang, B. L., Tang, D. J., Lu, G. T., He, Y. Q. e Tang, J. L. 2013. Transcriptome profiling of *Xanthomonas campestris* pv. *campestris* grown in minimal medium MMX and rich medium NYG. Res Microbiol 164:466-79.
- 179. Lonetto, M. A., Brown, K. L., Rudd, K. E. e Buttner, M. J. 1994. Analysis of the *Streptomyces coelicolor* sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. Proc Natl Acad Sci U S A 91:7573-7.
- 180. Lopes, J. R. S., Daugherty, M. P. e Almeida, R. P. P. 2010. Strain origin drives virulence and persistence of *Xylella fastidiosa* in alfalfa. Plant pathology 59:963-971.
- 181. Lopes, S. A., Marcussi, S., Torres, S. C. Z., Souza, V., Fagan, C. e França, S. C. 2003. Weeds as alternative hosts of the citrus, coffee, and plum strains of *Xylella fastidiosa* in Brazil. Plant disease 87:544-549.
- 182. Love, M. I., Huber, W. e Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol 15:550.
- 183. **Mackie, G. A.** 2013. RNase E: at the interface of bacterial RNA processing and decay. Nat Rev Microbiol 11:45-57.
- 184. **Madan Babu, M. e Teichmann, S. A.** 2003. Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites. Trends Genet 19:75-9.
- 185. **Madden, L. V. e Wheelis, M.** 2003. The threat of plant pathogens as weapons against U.S. crops. Annu Rev Phytopathol 41:155-76.
- 186. **Magasanik, B.** 1961. Catabolite repression. Cold Spring Harb Symp Quant Biol 26:249-56.
- 187. **Malone, J. H. e Oliver, B.** 2011. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol 9:34.
- 188. Mandin, P., Toledo-Arana, A., d'Hérouel, A. F. e Repoila, F. 2013. RNA-mediated control of bacterial gene expression: role of regulatory non-coding RNAs. *In* R. A. Meyers (ed.), Encyclopedia of Molecular Cell Biology and Molecular Medicine: RNA Biology, Second edition ed. Wiley-VCH Verlag GmbH & Co. KGaA.
- 189. Mao, F., Dam, P., Chou, J., Olman, V. e Xu, Y. 2009. DOOR: a database for prokaryotic operons. Nucleic Acids Res 37:D459-63.

- 190. **Mardis, E. R.** 2008. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387-402.
- 191. **Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. e Gilad, Y.** 2008. RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509-17.
- 192. 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.
- 193. **Marques, M. V.** 2012. Biologia molecular e genética bacteriana. Sociedade Brasileira de Genética, Ribeirão Preto.
- 194. 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.
- 195. Masquida, B. e Westhof, E. 2011. RNase P: at last, the key finds its lock. RNA 17:1615-8.
- 196. McCann, H. C. e Guttman, D. S. 2007. Evolution of the type III secretion system and its effectors in plant-microbe interactions. New Phytologist 177:33-47.
- 197. **McClure, R., Tjaden, B. e Genco, C.** 2014. Identification of sRNAs expressed by the human pathogen *Neisseria gonorrhoeae* under disparate growth conditions. Front Microbiol 5:456.
- 198. **Mehta, A. e Rosato, Y. B.** 2001. Phylogenetic relationships of *Xylella fastidiosa* strains from different hosts, based on 16S rDNA and 16S-23S intergenic spacer sequences. Int J Syst Evol Microbiol 51:311-8.
- 199. Mendes, J. S., Santiago, A. S., Toledo, M. A., Horta, M. A., de Souza, A. A., Tasic, L. e de Souza, A. P. 2016. In vitro Determination of Extracellular Proteins from *Xylella fastidiosa*. Front Microbiol 7:2090.
- 200. 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.
- 201. Merfa, M. V., Niza, B., Takita, M. A. e De Souza, A. A. 2016. The MqsRA Toxin-Antitoxin System from *Xylella fastidiosa* Plays a Key Role in Bacterial Fitness, Pathogenicity, and Persister Cell Formation. Front Microbiol 7:904.
- 202. **Merino, E. e Yanofsky, C.** 2005. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. Trends Genet 21:260-4.
- 203. **Misra, T. K. e Apirion, D.** 1979. RNase E, an RNA processing enzyme from Escherichia coli. J Biol Chem 254:11154-9.
- 204. **Mizuno, T., Chou, M. Y. e Inouye, M.** 1984. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proc Natl Acad Sci U S A 81:1966-70.
- 205. Mondragon, A. 2013. Structural studies of RNase P. Annu Rev Biophys 42:537-57.
- 206. Monteiro, P. B., Teixeira, D. C., Palma, R. R., Garnier, M., Bove, J. M. e Renaudin, J. 2001. Stable transformation of the *Xylella fastidiosa* citrus variegated chlorosis strain with oriC plasmids. Appl Environ Microbiol 67:2263-9.
- 207. Moreau, Y., Aerts, S., De Moor, B., De Strooper, B. e Dabrowski, M. 2003. Comparison and meta-analysis of microarray data: from the bench to the computer desk. Trends Genet 19:570-7.
- 208. **Morita, T., Maki, K. e Aiba, H.** 2005. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev 19:2176-86.

- 209. Morozova, O., Hirst, M. e Marra, M. A. 2009. Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet 10:135-51.
- 210. Morris, K. V. e Mattick, J. S. 2014. The rise of regulatory RNA. Nat Rev Genet 15:423-37.
- 211. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. e Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621-8.
- 212. Muranaka, L. S., Giorgiano, T. E., Takita, M. A., Forim, M. R., Silva, L. F., Coletta-Filho, H. D., Machado, M. A. e de Souza, A. A. 2013. N-acetylcysteine in agriculture, a novel use for an old molecule: focus on controlling the plant-pathogen *Xylella fastidiosa*. PLoS One 8:e72937.
- 213. **Muranaka, L. S., Takita, M. A., Olivato, J. C., Kishi, L. T. e de Souza, A. A.** 2012. Global expression profile of biofilm resistance to antimicrobial compounds in the plant-pathogenic bacterium *Xylella fastidiosa* reveals evidence of persister cells. J Bacteriol 194:4561-9.
- 214. Mutz, K. O., Heilkenbrinker, A., Lonne, M., Walter, J. G. e Stahl, F. 2013. Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol 24:22-30.
- 215. Nagalakshmi, U., Waern, K. e Snyder, M. 2010. RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol Chapter 4:Unit 4 11 1-13.
- 216. Nagaraj, S. H., Gasser, R. B. e Ranganathan, S. 2006. A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 8:6-21.
- 217. Nascimento, R., Gouran, H., Chakraborty, S., Gillespie, H. W., Almeida-Souza, H. O., Tu, A., Rao, B. J., Feldstein, P. A., Bruening, G., Goulart, L. R. e Dandekar, A. M. 2016. The Type II Secreted Lipase/Esterase LesA is a Key Virulence Factor Required for *Xylella fastidiosa* Pathogenesis in Grapevines. Sci Rep 6:18598.
- 218. Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E. e Altuvia, S. 2009. A pH-responsive riboregulator. Genes Dev 23:2650-62.
- 219. 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.
- 220. Newman, M.-A., Dow, J. M. e Daniels, M. J. 2001. Bacterial lipopolysaccharides and plant-pathogen interactions. European Journal of Plant Pathology 107:95-102.
- 221. **Nitzan, M., Rehami, R. e Margalit, H.** 2017. Integration of bacterial small RNAs in regulatory networks. Annual Review of Biophysics 46.
- 222. Niza, B., Merfa, M. V., Alencar, V. C., Menegidio, F. B., Nunes, L. R., Machado, M. A., Takita, M. A. e de Souza, A. A. 2016. Draft Genome Sequence of 11399, a Transformable Citrus-Pathogenic Strain of *Xylella fastidiosa*. Genome Announc 4.
- 223. Nogaroto, V., Tagliavini, S., Gianotti, A., Mikawa, A., Barros, N. T., Puzer, L., Carmona, A. K., Costa, P. e Henrique-Silva, F. 2006. Recombinant expression and characterization of a *Xylella fastidiosa* cysteine protease differentially expressed in a nonpathogenic strain. FEMS Microbiol Lett 261:187-93.
- 224. Nookaew, I., Papini, M., Pornputtapong, N., Scalcinati, G., Fagerberg, L., Uhlen, M. e Nielsen, J. 2012. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. Nucleic Acids Res 40:10084-97.

- 225. Norman, G. 2010. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract 15:625-32.
- 226. Nudler, E. e Gottesman, M. E. 2002. Transcription termination and anti-termination in *E. coli*. Genes Cells 7:755-68.
- 227. Nunes, L. R., Rosato, Y. B., Muto, N. H., Yanai, G. M., da Silva, V. S., Leite, D. B., Goncalves, E. R., de Souza, A. A., Coletta-Filho, H. D., Machado, M. A., Lopes, S. A. e de Oliveira, R. C. 2003. Microarray analyses of *Xylella fastidiosa* provide evidence of coordinated transcription control of laterally transferred elements. Genome Res 13:570-8.
- 228. Nunney, L., Ortiz, B., Russell, S. A., Ruiz Sanchez, R. e Stouthamer, R. 2014a. The complex biogeography of the plant pathogen *Xylella fastidiosa*: genetic evidence of introductions and Subspecific introgression in Central America. PLoS One 9:e112463.
- 229. Nunney, L., Schuenzel, E. L., Scally, M., Bromley, R. E. e Stouthamer, R. 2014b. Large-scale intersubspecific recombination in the plant-pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. Appl Environ Microbiol 80:3025-33.
- 230. Nunney, L., Vickerman, D. B., Bromley, R. E., Russell, S. A., Hartman, J. R., Morano, L. D. e Stouthamer, R. 2013. Recent evolutionary radiation and host plant specialization in the *Xylella fastidiosa* subspecies native to the United States. Appl Environ Microbiol 79:2189-200.
- 231. Nunney, L., Yuan, X., Bromley, R., Hartung, J., Montero-Astua, M., Moreira, L., Ortiz, B. e Stouthamer, R. 2010. Population genomic analysis of a bacterial plant pathogen: novel insight into the origin of Pierce's disease of grapevine in the U.S. PLoS One 5:e15488.
- 232. Nunney, L., Yuan, X., Bromley, R. E. e Stouthamer, R. 2012. Detecting genetic introgression: high levels of intersubspecific recombination found in *Xylella fastidiosa* in Brazil. Appl Environ Microbiol 78:4702-14.
- 233. Oliver, J. E., Cobine, P. A. e De La Fuente, L. 2015. *Xylella fastidiosa* Isolates from Both subsp. *multiplex* and *fastidiosa* Cause Disease on Southern Highbush Blueberry (*Vaccinium* sp.) Under Greenhouse Conditions. Phytopathology 105:855-62.
- Oosthuizen, M. C., Steyn, B., Theron, J., Cosette, P., Lindsay, D., Von Holy, A. e Brozel, V. S. 2002. Proteomic analysis reveals differential protein expression by *Bacillus cereus* during biofilm formation. Appl Environ Microbiol 68:2770-80.
- Oshima, T., Aiba, H., Masuda, Y., Kanaya, S., Sugiura, M., Wanner, B. L., Mori, H. e Mizuno, T. 2002. Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12. Mol Microbiol 46:281-91.
- 236. Ozsolak, F. e Milos, P. M. 2011. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12:87-98.
- 237. Ozsolak, F., Platt, A. R., Jones, D. R., Reifenberger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M. e Milos, P. M. 2009. Direct RNA sequencing. Nature 461:814-8.
- 238. **Paget, M. S.** 2015. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules 5:1245-65.
- 239. **Park, H. J., Lee, S. W. e Han, S. W.** 2014. Proteomic and functional analyses of a novel porin-like protein in *Xanthomonas oryzae* pv. *oryzae*. J Microbiol 52:1030-5.
- 240. **Parker, J. K., Chen, H., McCarty, S. E., Liu, L. Y. e De La Fuente, L.** 2016. Calcium transcriptionally regulates the biofilm machinery of *Xylella fastidiosa* to promote continued biofilm development in batch cultures. Environ Microbiol 18:1620-34.

- 241. Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H. e Soldatov, A. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37:e123.
- 242. Parkinson, J. S. 1993. Signal transduction schemes of bacteria. Cell 73:857-71.
- 243. **Parsek, M. R. e Singh, P. K.** 2003. Bacterial biofilms: an emerging link to disease pathogenesis. Annu Rev Microbiol 57:677-701.
- 244. Pashalidis, S., Moreira, L. M., Zaini, P. A., Campanharo, J. C., Alves, L. M., Ciapina, L. P., Vencio, R. Z., Lemos, E. G., Da Silva, A. M. e Da Silva, A. C. 2005. Whole-genome expression profiling of *Xylella fastidiosa* in response to growth on glucose. OMICS 9:77-90.
- 245. **Pearson, E. S.** 1931a. The analysis of variance in cases of non-normal variation. Biometrika 23:114-133.
- 246. **Pearson, E. S.** 1931b. The test of significance for the correlation coefficient. Journal of the American Statistical Association 26:128-134.
- 247. **Pereira, P. S.** 2015. *Xylella fastidiosa* a new menace for Portuguese agriculture and forestry. Revista de Ciências Agrárias 38:149-154.
- 248. 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.
- 249. **Pierry, P. M.** 2012. Pirossequenciamento e análise comparativa de genomas do fitopatógeno *Xylella fastidiosa*. Departamento de Bioquímica, Universidade de São Paulo, São Paulo. 144.
- 250. **Pischimarov, J., Kuenne, C., Billion, A., Hemberger, J., Cemic, F., Chakraborty, T. e Hain, T.** 2012. sRNAdb: a small non-coding RNA database for gram-positive bacteria. BMC Genomics 13:384.
- 251. **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.
- Prado, S. S., Lopes, J. R. S., Demétrio, C. G. B., Borgatto, A. F. e Almeida, R. P. P. 2008. Host colonization differences between citrus and coffee isolates of *Xylella fastidiosa* in reciprocal inoculation. Scientia Agricola 65:251-258.
- 253. **Purcell, A. H. e Finlay, A.** 1979. Evidence for noncirculative transmission of Pierce's disease bacterium by sharpshooter leafhoppers. Phytopathology 69:393-395.
- 254. **Purcell, A. H., Saunders, S. R., Hendson, M., Grebus, M. E. e Henry, M. J.** 1999. Causal Role of *Xylella fastidiosa* in Oleander Leaf Scorch Disease. Phytopathology 89:53-8.
- 255. **R Core Team.** 2013. R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- 256. **Ramesh, A. e Winkler, W. C.** 2010. Magnesium-sensing riboswitches in bacteria. RNA Biol 7:77-83.
- 257. 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.
- 258. Rapicavoli, J. N., Kinsinger, N., Perring, T. M., Backus, E. A., Shugart, H. J., Walker, S. e Roper, M. C. 2015. O antigen modulates insect vector acquisition of the bacterial plant pathogen *Xylella fastidiosa*. Appl Environ Microbiol 81:8145-54.

- 259. 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.
- 260. **Regnier, P. e Arraiano, C. M.** 2000. Degradation of mRNA in bacteria: emergence of ubiquitous features. Bioessays 22:235-44.
- 261. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T. e Mondragon, A. 2010. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. Nature 468:784-9.
- 262. **Robinson, M. D., McCarthy, D. J. e Smyth, G. K.** 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-40.
- 263. **Rodrigues, C. M., de Souza, A. A., Takita, M. A., Kishi, L. T. e Machado, M. A.** 2013. RNA-Seq analysis of Citrus reticulata in the early stages of *Xylella fastidiosa* infection reveals auxin-related genes as a defense response. BMC Genomics 14:676.
- 264. **Rodríguez, C. M., Obando, J. J., Villalobos, W., Moreira, L. e Rivera, C.** 2001. First report of *Xylella fastidiosa* infecting coffee in Costa Rica. Plant disease 85:1027.
- 265. **Rogers, E. E.** 2012. Evaluation of Arabidopsis thaliana as a model host for *Xylella fastidiosa*. Mol Plant Microbe Interact 25:747-54.
- 266. **Rogers, E. E. e Backus, E. A.** 2014. Anterior foregut microbiota of the glassy-winged sharpshooter explored using deep 16S rRNA gene sequencing from individual insects. PLoS One 9:e106215.
- 267. **Rogers, E. E. e Stenger, D. C.** 2012. A conjugative 38 kB plasmid is present in multiple subspecies of *Xylella fastidiosa*. PLoS One 7:e52131.
- 268. **Rossetti, V. e De Negri, J. D.** 1990. Clorose variegada dos citros: revisão. Laranja 11:1-14.
- 269. Rossetti, V., Garnier, M., Bové, J. M., Beretta, M. J., Teixeira, A. R. R., Quaggio, J. A. e Negri, J. D. d. 1990. Occurrence of xylem-restricted bacteria in sweet orange trees affected by chlorotic variegation, a new citrus disease in Brazil. Comptes Rendus de l'Academie des Science 310:345-349.
- 270. **Ryan, R. P. e Dow, J. M.** 2011. Communication with a growing family: diffusible signal factor (DSF) signaling in bacteria. Trends Microbiol 19:145-52.
- 271. Saecker, R. M., Record, M. T., Jr. e Dehaseth, P. L. 2011. Mechanism of bacterial transcription initiation: RNA polymerase promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J Mol Biol 412:754-71.
- 272. **Santana, W. O.** 2012. Genômica comparativa de *Xylella fastidiosa*: diversidade do pangenoma e análise de genes de patogenicidade. Departamento de Bioquímica, Universidade de São Paulo, São Paulo. 152.
- 273. Santiago, A. D., Mendes, J. S., Dos Santos, C. A., de Toledo, M. A., Beloti, L. L., Crucello, A., Horta, M. A., Favaro, M. T., Munar, D. M., de Souza, A. A., Cotta, M. A. e de Souza, A. P. 2016. The Antitoxin Protein of a Toxin-Antitoxin System from *Xylella fastidiosa* Is Secreted via Outer Membrane Vesicles. Front Microbiol 7:2030.
- 274. Saponari, M., Boscia, D., Loconsole, G., Palmisano, F., Savino, V., Potere, O. e Martelli, G. P. 2014. New hosts of *Xylella fastidiosa* strain CoDiRO in Apulia. Journal of Plant Pathology 96:611.
- 275. **Saponari, M., Boscia, D., Nigro, F. e Martelli, G. P.** 2013. Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (southern Italy). Journal of Plant Pathology 95:668.

- 276. **Sayed, N., Jousselin, A. e Felden, B.** 2011. A cis-antisense RNA acts in trans in *Staphylococcus aureus* to control translation of a human cytolytic peptide. Nat Struct Mol Biol 19:105-12.
- 277. Schaad, N. W., Postnikova, E., Lacy, G., Fatmi, M. e Chang, C. J. 2004. *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp. [correction] *fastidiosa* [correction] subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. Syst Appl Microbiol 27:290-300.
- 278. Schena, M., Shalon, D., Davis, R. W. e Brown, P. O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-70.
- 279. Schmidtke, C., Abendroth, U., Brock, J., Serrania, J., Becker, A. e Bonas, U. 2013. Small RNA sX13: a multifaceted regulator of virulence in the plant pathogen *Xanthomonas*. PLoS Pathog 9:e1003626.
- 280. Schmidtke, C., Findeiss, S., Sharma, C. M., Kuhfuss, J., Hoffmann, S., Vogel, J., Stadler, P. F. e Bonas, U. 2011. Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. Nucleic Acids Res 40:2020-31.
- 281. Schmittgen, T. D., Zakrajsek, B. A., Mills, A. G., Gorn, V., Singer, M. J. e Reed, M. W. 2000. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. Anal Biochem 285:194-204.
- 282. Schreiber IV, H. L., 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.
- 283. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. e Ragg, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 7:3.
- 284. 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.
- 285. Schurch, N. J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hugues, T., Blaxter, M. e Barton, G. J. 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 22:839-851.
- 286. Sellner, L. N. e Turbett, G. R. 1998. Comparison of three RT-PCR methods. Biotechniques 25:230-4.
- 287. Seshasayee, A. S., Fraser, G. M. e Luscombe, N. M. 2010. Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. Nucleic Acids Res 38:5970-81.
- 288. Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R. e Cossart, P. 2013. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat Rev Microbiol 11:75-82.
- 289. Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R., Stadler, P. F. e Vogel, J. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464:250-5.
- 290. Sherwood, A. V. e Henkin, T. M. 2016. Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. Annu Rev Microbiol 70:361-74.

- 291. Shi, X. Y., Dumenyo, C. K., Hernandez-Martinez, R., Azad, H. e Cooksey, D. A. 2007. Characterization of regulatory pathways in *Xylella fastidiosa*: genes and phenotypes controlled by algU. Appl Environ Microbiol 73:6748-56.
- 292. Shi, X. Y., Dumenyo, C. K., Hernandez-Martinez, R., Azad, H. e Cooksey, D. A. 2009. Characterization of regulatory pathways in *Xylella fastidiosa*: genes and phenotypes controlled by gacA. Appl Environ Microbiol 75:2275-83.
- 293. Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P. e Hayashizaki, Y. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A 100:15776-81.
- 294. Shriner, A. D. e Andersen, P. C. 2014. Effect of oxygen on the growth and biofilm formation of *Xylella fastidiosa* in liquid media. Curr Microbiol 69:866-73.
- 295. Silva-Rocha, R. e de Lorenzo, V. 2010. Noise and robustness in prokaryotic regulatory networks. Annu Rev Microbiol 64:257-75.
- 296. Simpson, A. J., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M. *et al.* 2000. 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.
- 297. Slonim, D. K. e Yanai, I. 2009. Getting started in gene expression microarray analysis. PLoS Comput Biol 5:e1000543.
- 298. Smolka, M. B., Martins-de-Souza, D., Winck, F. V., Santoro, C. E., Castellari, R. R., Ferrari, F., Brum, I. J., Galembeck, E., Della Coletta Filho, H., Machado, M. A., Marangoni, S. e Novello, J. C. 2003. Proteome analysis of the plant pathogen *Xylella fastidiosa* reveals major cellular and extracellular proteins and a peculiar codon bias distribution. Proteomics 3:224-37.
- 299. Snyder, L., Peters, J. E., Henkin, T. M. e Champness, W. 2013. Molecular Genetics of Bacteria, 4th edition ed. ASM Press, Washington, DC.
- 300. Sorek, R. e Cossart, P. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet 11:9-16.
- 301. Sorek, R., Kunin, V. e Hugenholtz, P. 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol 6:181-6.
- 302. **Sousa, S., Marchand, I. e Dreyfus, M.** 2001. Autoregulation allows *Escherichia coli* RNase E to adjust continuously its synthesis to that of its substrates. Mol Microbiol 42:867-78.
- 303. Staron, A., Sofia, H. J., Dietrich, S., Ulrich, L. E., Liesegang, H. e Mascher, T. 2009. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. Mol Microbiol 74:557-81.
- 304. **Stenger, D. C., Lee, M. W., Rogers, E. E. e Chen, J.** 2010. Plasmids of *Xylella fastidiosa* mulberry-infecting strains share extensive identity and gene complement with pVEIS01 from the earthworm symbiont *Verminephrobacter eiseniae*. Physiological and Molecular Plant Pathology 74:238-245.
- 305. Stock, A. M., Robinson, V. L. e Goudreau, P. N. 2000. Two-component signal transduction. Annu Rev Biochem 69:183-215.
- 306. **Stoner, W. N.** 1952. A comparison between grape degeneration in Florida and Pierce's Disease in California The Florida Entomologist 35:62-68.
- 307. Storz, G., Vogel, J. e Wassarman, K. M. 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43:880-91.

- 308. **Strona, G., Carstens, C. J. e Beck, P. S.** 2017. Network analysis reveals why *Xylella fastidiosa* will persist in Europe. Sci Rep 7:71.
- 309. Su, C.-C., Chang, C. J., Chang, C.-M., Shih, H.-T., Tzeng, K.-C., Jan, F.-J., Kao, C.-W. e Deng, W.-L. 2013. Pierce's disease of grapevines in Taiwan: isolation, cultivation and pathogenicity of *Xylella fastidiosa*. Journal of Plant Pathology 161:389-396.
- 310. Su, C. C., Deng, W. L., Jan, F. J., Chang, C. J., Huang, H. e Chen, J. 2014. Draft Genome Sequence of *Xylella fastidiosa* Pear Leaf Scorch Strain in Taiwan. Genome Announc 2.
- 311. Su, C. C., Deng, W. L., Jan, F. J., Chang, C. J., Huang, H., Shih, H. T. e Chen, J. 2016a. *Xylella taiwanensis* sp. nov., causing pear leaf scorch disease. Int J Syst Evol Microbiol 66:4766-4771.
- 312. Su, X., Wellen, K. E. e Rabinowitz, J. D. 2016b. Metabolic control of methylation and acetylation. Curr Opin Chem Biol 30:52-60.
- 313. **Tang, F., Lao, K. e Surani, M. A.** 2011. Development and applications of single-cell transcriptome analysis. Nat Methods 8:S6-11.
- 314. **Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. e Conesa, A.** 2011. Differential expression in RNA-Seq: a matter of depth. Genome Res 21:2213-23.
- 315. **Ternan, N. G.** 2013. Small regulatory RNA molecules in bacteria. OA Microbiology 1:1-8.
- 316. **Thorne, E. T., Stevenson, J. F., Rost, T. L., Labavitch, J. M. e Matthews, M. A.** 2006. Pierce's disease symptoms: comparison with symptoms of water deficit and the impact of water deficits. Am. J. Enol. Vitic. 57.
- 317. **Tjaden, B.** 2015. De novo assembly of bacterial transcriptomes from RNA-Seq data. Genome Biol 16:1.
- 318. Toffano-Nioche, C., Nguyen, A. N., Kuchly, C., Ott, A., Gautheret, D., Bouloc, P. e Jacq, A. 2012. Transcriptomic profiling of the oyster pathogen *Vibrio splendidus* opens a window on the evolutionary dynamics of the small RNA repertoire in the Vibrio genus. RNA 18:2201-19.
- 319. Toledo, M. A., Schneider, D. R., Azzoni, A. R., Favaro, M. T., Pelloso, A. C., Santos, C. A., Saraiva, A. M. e Souza, A. P. 2011. Characterization of an oxidative stress response regulator, homologous to *Escherichia coli* OxyR, from the phytopathogen *Xylella fastidiosa*. Protein Expr Purif 75:204-10.
- 320. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J. e Pachter, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511-5.
- 321. **Travensolo, R. F., Carareto-Alves, L. M., Costa, M. V., Lopes, T. J., Carrilho, E.** e Lemos, E. G. 2009. *Xylella fastidiosa* gene expression analysis by DNA microarrays. Genet Mol Biol 32:340-53.
- 322. **Trotochaud, A. E. e Wassarman, K. M.** 2004. 6S RNA function enhances long-term cell survival. J Bacteriol 186:4978-85.
- 323. Van Horn, C., Chang, C. J. e Chen, J. 2017. De Novo Whole-Genome Sequence of *Xylella fastidiosa* subsp. multiplex Strain BB01 Isolated from a Blueberry in Georgia, USA. Genome Announc 5.
- 324. 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.

- 325. VanGuilder, H. D., Vrana, K. E. e Freeman, W. M. 2008. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 44:619-26.
- 326. Varani, A. M., Monteiro-Vitorello, C. B., Nakaya, H. I. e Van Sluys, M. A. 2013. The role of prophage in plant-pathogenic bacteria. Annu Rev Phytopathol 51:429-51.
- 327. Vencio, R. Z., Koide, T., Gomes, S. L. e Pereira, C. A. 2006. BayGO: Bayesian analysis of ontology term enrichment in microarray data. BMC Bioinformatics 7:86.
- 328. Vitreschak, A. G., Mironov, A. A., Lyubetsky, V. A. e Gelfand, M. S. 2008. Comparative genomic analysis of T-box regulatory systems in bacteria. RNA 14:717-35.
- 329. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. e Gelfand, M. S. 2004. Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet 20:44-50.
- 330. Voegel, T. M., Doddapaneni, H., Cheng, D. W., Lin, H., Stenger, D. C., Kirkpatrick, B. C. e Roper, M. C. 2013. Identification of a response regulator involved in surface attachment, cell-cell aggregation, exopolysaccharide production and virulence in the plant pathogen *Xylella fastidiosa*. Mol Plant Pathol 14:256-64.
- 331. **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.
- 332. Vogel, J. e Luisi, B. F. 2011. Hfq and its constellation of RNA. Nat Rev Microbiol 9:578-89.
- 333. **Wagner, E. G. e Romby, P.** 2015. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv Genet 90:133-208.
- 334. Wang, N., Li, J. L. e Lindow, S. E. 2012. RpfF-dependent regulon of *Xylella fastidiosa*. Phytopathology 102:1045-53.
- 335. Wang, Z., Gerstein, M. e Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57-63.
- 336. **Wassarman, K. M.** 2007. 6S RNA: a small RNA regulator of transcription. Curr Opin Microbiol 10:164-8.
- 337. Wassarman, K. M. e Storz, G. 2000. 6S RNA regulates *E. coli* RNA polymerase activity. Cell 101:613-23.
- 338. Waters, C. M. e Bassler, B. L. 2005. Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 21:319-46.
- 339. Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H. e Breaker, R. R. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol 11:R31.
- 340. 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, xylemlimited, fastidious plant bacteria related to *Xanthomonas* spp. International Journal of Systematic Bacteriology 37:136-143.
- 341. Willenbrock, H. e Ussery, D. W. 2004. Chromatin architecture and gene expression in *Escherichia coli*. Genome Biol 5:252.
- 342. Williams, C. R., Baccarella, A., Parrish, J. Z. e Kim, C. C. 2016. Trimming of sequence reads alters RNA-Seq gene expression estimates. BMC Bioinformatics 17:103.
- 343. Winkler, W. C. e Breaker, R. R. 2003. Genetic control by metabolite-binding riboswitches. Chembiochem 4:1024-32.
- 344. Winkler, W. C. e Breaker, R. R. 2005. Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 59:487-517.

- 345. Wolf, D. M., Fontaine-Bodin, L., Bischofs, I., Price, G., Keasling, J. e Arkin, A. P. 2008. Memory in microbes: quantifying history-dependent behavior in a bacterium. PLoS One 3:e1700.
- 346. **Wolf, J. B.** 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-Seq tutorial. Mol Ecol Resour 13:559-72.
- 347. Wurtzel, O., Yoder-Himes, D. R., Han, K., Dandekar, A. A., Edelheit, S., Greenberg, E. P., Sorek, R. e Lory, S. 2012. The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. PLoS Pathog 8:e1002945.
- 348. Yan, Y., Su, S., Meng, X., Ji, X., Qu, Y., Liu, Z., Wang, X., Cui, Y., Deng, Z., Zhou, D., Jiang, W., Yang, R. e Han, Y. 2013. Determination of sRNA expressions by RNA-seq in *Yersinia pestis* grown in vitro and during infection. PLoS One 8:e74495.
- 349. **Yanofsky, C.** 1981. Attenuation in the control of expression of bacterial operons. Nature 289:751-8.
- 350. **Yanofsky, C.** 2000. Transcription attenuation: once viewed as a novel regulatory strategy. J Bacteriol 182:1-8.
- 351. Ye, Y. e Doak, T. G. 2009. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol 5:e1000465.
- 352. Yuan, X., Morano, L., Bromley, R., Spring-Pearson, S., Stouthamer, R. e Nunney, L. 2010. Multilocus sequence typing of *Xylella fastidiosa* causing Pierce's disease and oleander leaf scorch in the United States. Phytopathology 100:601-11.
- 353. Zaini, P. A., Burdman, S., Igo, M. M., Parker, J. K. e De La Fuente, L. 2015. Fimbrial and Afimbrial Adhesins Involved in Bacterial Attachment to Surfaces, p. 492. *In* N. Wang, J. B. Jones, G. W. Sundin, F. F. White, S. A. Hogenhout, C. Roper, L. De La Fuente eJ. H. Ham (ed.), Virulence mechanisms of plant-pathogenic bacteria. American Phytopathological Society Press, St. Paul, Minnesota, EUA.
- 354. **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.
- 355. 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.
- 356. **Zhang, N. e Buck, M.** 2015. A perspective on the enhancer dependent bacterial RNA polymerase. Biomolecules 5:1012-9.
- 357. Zhang, S., Chakrabarty, P. K., Fleites, L. A., Rayside, P. A., Hopkins, D. L. e Gabriel, D. W. 2015. Three New Pierce's Disease Pathogenicity Effectors Identified Using *Xylella fastidiosa* Biocontrol Strain EB92-1. PLoS One 10:e0133796.
- 358. 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.
- 359. **Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. e Liu, X.** 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9:e78644.

Súmula curricular

Paulo Marques Pierry

Nascimento: 18/05/1987 – Santos / SP - Brasil

Formação Acadêmica

2005 - 2008	Graduação em Ciências Biológicas Licenciatura Bacharelado com ênfase em Biotecnologia Bacharelado com ênfase em Meio Ambiente Universidade Católica de Santos, UNISANTOS, Santos, São Paulo, Brasil
2009 -2012	Curso de Mestrado, Programa de Ciências Biológicas (Bioquímica). Universidade de São Paulo, USP, São Paulo, Brasil Orientadora: Profa. Dra. Aline Maria da Silva Bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo
2012 -2017	Curso de Doutorado, Programa de Ciências Biológicas (Bioquímica). Universidade de São Paulo, USP, São Paulo, Brasil Orientadora: Profa. Dra. Aline Maria da Silva Bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo

Formação complementar

2006-2008	Iniciação Científica
	Universidade Católica de Santos, UNISANTOS, Santos, São Paulo, Brasil
	Orientadoras: Profa. Dra. Adriana Florentino de Souza
	Profa. Dra. Kátia Maria Gomes Machado
	Profa. Dra. Rosângela Ballego Campanã
	Bolsista do Programa de Iniciação Científica (PROIN) - Universidade
	Católica de Santos

Prêmios

2008	Prêmio Bolsa Mérito pelo ótimo desempenho nas disciplinas do curso de Licenciatura e Bacharelado do curso de Ciências Biológicas da UNISANTOS.
2012	Menção honrosa pela participação no "Prêmio Pós-Graduação – Poster" na área de "Genômica" no 58° Congresso Brasileiro de Genética.
2014	Menção honrosa pela participação no "Prêmio Painel Pós-Graduação" na área de "Genômica e Bioinformática" no 60° Congresso Brasileiro de Genética.

Palestras e comunicações em congressos

Pierry, PM. Genômica e transcritômica no estudo do fitopatógeno *Xylella fastidiosa*. In: Seminários Gerais do Departamento de Bioquímica – IQ/USP, 2015, São Paulo - SP.

Pierry, PM; Martins-Jr, J; Zaini, PA; da Silva, AM. RNA-Seq Analysis Reveals Important Features of Gene Expression Signatures in *Xylella fastidiosa*. In: 115th General Meeting of American Society for Microbiology, 2015, New Orleans – Louisiana - EUA.

Pierry, PM; Martins-Jr, J; Zaini, PA; Igo, M; Lindow, SE; da Silva, AM. Variations in the transcriptomes of PD and CVC strains of *Xylella fastidiosa* cultivated in minimal and rich medium. In: 1° São Paulo Xanthomonadaceae (Xantho) Meeting, 2015, São Paulo - SP.

Zaini, PA; Souza, APS; Ionescu, M; **Pierry, PM**; Lindow, SE; da Silva, AM. Adhesion strategies of the phytopathogen *Xylella fastidiosa*. In: 44a. Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2015, Foz do Iguaçú - PR.

Pierry, PM; Martins-Jr, J; Zaini, PA; da Silva, AM. RNA-Seq profiling of *Xylella fastidiosa* provides novel insights into its virulence mechanisms. In: 60° Congresso Brasileiro de Genética, 2014, Guarujá - SP.

Martins-Jr, J; Santana, WO; Zaini, PA; **Pierry, PM**; Coletta-Filho, HD; de Souza, AA; Machado, MA; Kitajima, JP; da Silva, AM. Phylogenomics and identification of mobile genetic elements of South American *Xylella fastidiosa* strains. In: 58° Congresso Brasileiro de Genética, 2012, Foz do Iguaçú - PR.

Pierry, PM; Zaini, PA; Santana, WO; Coletta-Filho, HD; de Souza, AA; Machado, MA; Kitajima, JP; da Silva, AM. Pyrosequencing and genome comparison of two *Xylella fastidiosa* strains isolated from plum and hibiscus. In: 58° Congresso Brasileiro de Genética, 2012, Foz do Iguaçú - PR.

da Silva, AM; Santana, WO; **Pierry, PM**; da Silva, PIP; Beckedorff, FCF; Amaral, MS; Verjovski-Almeida, S; Coletta-Filho, HD; de Souza, AA; Machado, MA; Almeida, LGP; Vasconcelos, ATR. Comparative genomics of *Xylella fastidiosa* South American strains. In: Pierce's Disease Research Symposium, 2010, San Diego – California - EUA.

Pierry, PM; da Silva, PIP; Santana, WO; Beckedorff, FCF; Amaral, MS; Teixeira, DC; Almeida, LGP; Vasconcelos, ATR; Verjovski-Almeida, S.; da Silva, AM. Sequenciamento e comparação do genoma de uma cepa não-virulenta de *Xylella fastidiosa*. In: 56° Congresso Brasileiro de Genética, 2010, Guarujá - SP.

Pierry, PM; Campanha, RB; Machado, KMG; Souza, AF. Partial Purification and Characterization of Laccases from *Psilocybe castanella*. In: XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2009, Águas de Lindóia, SP.

Pierry, PM; Santi-Júnior, C; Souza, AF; Machado, KMG. Degradation of textile effluents by laccases. In: XXXVII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq e XI Congresso da PABMB, 2008, Águas de Lindóia, SP.

Pierry, PM; Souza, AF; Machado, KMG. Produção de atividade de lacase por *Psilocybe castanella* e avaliação de seu potencial de degradação de corantes têxteis reativos. In: 5° Congresso Brasileiro de Micologia, 2007, Recife.

Santi-Júnior, C; **Pierry, PM**; Souza, AF; Machado, KMG. Degradação de corantes têxteis reativos por lacases de fungos basidiomicetos. In: IV EPOA, Encontro sobre Aplicações Ambientais de Processos Oxidativos Avançados, 2007, Cubatão.

Participações em cursos

2016	Participação como monitor no "XI Curso de Inverno – Temas Avançados de Bioquímica e Biologia Molecular", realizado pelo IQ/USP.
2015	Participação como ouvinte no curso "The Business of Science", realizado durante o 115th General Meeting of American Society for Microbiology.
2015	Participação como ouvinte no 2° Minicurso Transcritômica com RNA-Seq, oferecido pelo Centro de Facilidades de Apoio à Pesquisa (ICB-USP).
2014	Participação como ouvinte no curso "Bioinformática: Análise de Dados de RNA-Seq via Bioconductor", realizado durante o 60° Congresso Brasileiro de Genética.
2013	Participação como monitor no "VIII Curso de Inverno – Temas Avançados de Bioquímica e Biologia Molecular", realizado pelo IQ/USP.
2013	Participação no treinamento "MiSeq System Operation and Nextera Sample Preparation", oferecido pela empresa Illumina.
2010	Participação como ouvinte no "Curso de Verão de Bioinformática", realizado pela Associação Brasileira de Bioinformática e Biologia Computacional.
2007	Participação como ouvinte no curso "Processos biotecnológicos aplicados à biorremediação", realizado durante o 5º Congresso Brasileiro de Micologia.

Revisões de periódicos

2015-Atual Atuação como revisor do periódico "BMC Bioinformatics".