

UNIVERSIDADE FEDERAL DE PELOTAS

Programa de Pós-Graduação em Biotecnologia



Dissertação

**Abordagens bioinformáticas no estudo  
pangenômico de *Xanthomonas campestris***

**Amanda Munari Guimarães**

Pelotas, 2020

Amanda Munari Guimarães

Abordagens bioinformáticas no estudo pangenômico de *Xanthomonas campestris*

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciências (área do Conhecimento: Bioinformática).

Orientador: Dr. Luciano da Silva Pinto  
Coorientador: Dr. Frederico Schmitt Kremer

Pelotas, 2020

Universidade Federal de Pelotas / Sistema de Bibliotecas  
Catalogação na Publicação

G963a Guimarães, Amanda Munari

Abordagens bioinformáticas no estudo pangenômico de *Xanthomonas campestris* / Amanda Munari Guimarães ; Luciano da Silva Pinto, orientador ; Frederico Schmitt Kremer, coorientador. — Pelotas, 2020.

89 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Biotecnologia, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2020.

1. Fitopatígeno. 2. Fator de virulência. 3. Genoma núcleo. 4. Genômica microbiana. I. Pinto, Luciano da Silva, orient. II. Kremer, Frederico Schmitt, coorient. III. Título.

CDD : 574.2322

### **Banca Examinadora**

Prof. Dr Luciano da Silva Pinto (UFPel, CDTec)

Dr. Ismail Teodoro de Souza Junior

Dr. Frederico Schmitt Kremer

Prof<sup>a</sup>. Dr<sup>a</sup>. Patricia Diaz (UFPel, CDTec)

Para minha mãe, com gratidão e amor.  
Dedico.

## **Agradecimentos**

Primeiramente, agradeço à Universidade Federal de Pelotas, a qual possibilitou o meu aperfeiçoamento intelectual e científico, através do contato com os professores, alunos, técnicos e funcionários.

Agradeço todos os integrantes do grupo BioProLab que me acompanham há 6 anos, e sempre me motivaram, deram apoio, foram amigos e, principalmente, me ensinaram muito sobre a biotecnologia e sobre trabalhar em grupo. Em especial, o Luciano, meu orientador, que depositou confiança em mim e acima de tudo, me incentivou a ser uma pesquisadora independente. Meu coorientador Frederico, que se tornou um grande amigo e sem o qual esse trabalho não seria possível. Tem minha eterna gratidão pela amizade, ensinamentos e troca de conhecimentos científicos.

Aos meus pais, por me darem a oportunidade de viver, por terem me dado amor e carinho, por terem permitido que eu crescesse com conforto, boas oportunidades e rodeada de pessoas maravilhosas. Sou eternamente grata à minha mãe pelas noites mal dormidas, pelas madrugadas nos plantões, pelos abraços nos momentos mais críticos, pelas comemorações nos momentos de alegria e por sempre apoiarem minhas escolhas, fazendo o possível, e às vezes o impossível, para eu conquistar meus sonhos. Por ter me ensinado a ser responsável, ter caráter, ser uma pessoa que sempre busca o melhor em cada oportunidade.

Aos meus irmãos, Rafael, Pâmela, Maria Eduarda e Ana Paula. Vocês me ensinaram o que é compartilhar, o que é amar alguém tão intensamente, a ter paciência, a ser uma pessoa melhor para dar o exemplo para as manas menores, a saber aceitar a vida e compreender que o tempo que vivemos juntos não diz nada sobre o quanto nos amamos e o quanto somos especiais uns para os outros. Ao Rafael, especialmente, que fez parte da minha educação, sendo meu exemplo, cuidando de mim de forma tão carinhosa e única, e acima de tudo por ter sido meu melhor amigo durante 16 anos da minha vida.

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

## Resumo

GUIMARÃES, Amanda. **Abordagens bioinformáticas no estudo pangenômico de *Xanthomonas campestris***. 2020. 83f. Dissertação - Programa de Pós-Graduação em Biotecnologia. Universidade Federal de Pelotas, Pelotas.

*Xanthomonas campestris* é uma espécie de bactéria fitopatogênica gram-negativa que possui diferentes patovares, os quais apresentam especificidade ao hospedeiro, doença e modo de infecção. Os hospedeiros dessa espécie são de importância econômica, como a uva, couve, pimentão e outras plantas do gênero *Brassica*. Na presente dissertação, são apresentadas diferentes estratégias de bioinformática na análise genômica dos patovares de *X. campestris*. Os genomas foram obtidos no banco de dados NCBI, e reanotados estruturalmente utilizando a ferramenta GENIX. A análise de pangenoma foi realizada através do GET\_HOMOLOGUES, a qual possibilitou identificar grupos ortólogos, além de estabelecer o *core*, *soft*, *shell* e *cloud* genoma. A anotação metabólica dos patovares foi realizada utilizando as ferramentas antiSMASH e BlastKoala, bem como os genes diferencialmente distribuídos relacionados a processos de patogenia e fatores de virulência, como efetores TAL, goma xantana e fator DSF foram analisados através do alinhamento de sequências. Sendo assim, através das abordagens bioinformáticas utilizadas neste trabalho, foi possível identificar características genômicas, as quais são relevantes para o entendimento da patogenia de cada patovar, bem como identificar genes com potencial biotecnológico.

**Palavras-chave:** fitopatógeno, fator de virulência, genoma núcleo, genômica microbiana

## Abstract

GUIMARÃES, Amanda. **Bioinformatics approaches in the pangenomic study of *Xanthomonas campestris***. 2020. 83f. Dissertation - Programa de Pós-Graduação em Biotecnologia. Universidade Federal de Pelotas, Pelotas.

*Xanthomonas campestris* is a gram-negative phytopathogenic bacteria that has different pathovars, which present specificity to the host, disease and mode of infection. The hosts of this species are of economic importance, such as grapes, kale, peppers, and other plants of the genus Brassica. In this dissertation, different bioinformatics strategies are presented in the genomic analysis of *X. campestris* pathovars. The genomes were obtained from the NCBI database, and structurally re-annotated using the GENIX tool. The pangenome analysis was performed using the GET\_HOMOLOGUES, which made it possible to identify orthologous groups. Also was possible to establish the core, soft, shell, and cloud genome. The metabolic annotation of the pathovars was performed using the antiSMASH, and BlastKoala tools, as well as the differentially distributed genes related to pathogenesis processes and virulence factors, such as TAL effectors, xanthan gum and DSF factor, were analyzed through sequence alignment. Through the bioinformatics approaches used in this work, it was possible to identify genomic characteristics, which are relevant for understanding the pathogenesis of each pathovar, as well as identifying genes with biotechnological potential.

**Keywords:** phytopathogen, virulence factor, core genome, microbial genomics

## Lista de Figuras

<b>Figura 1.</b> Sintomas causados pelo gênero <i>Xanthomonas</i> em diferentes tecidos de plantas .....	4
<b>Figura 2.</b> Esquema representativo dos sistemas de secreção de proteínas em <i>Xanthomonas</i> spp.....	5
<b>Figura 3.</b> Esquema da rota e biossíntese de xantana em <i>Xanthomonas campestris</i> .....	10
<b>Figura 4.</b> Mapa do <i>cluster rpf</i> em <i>X. campestris</i> , e organização transcricional envolvida na sinalização, biossíntese e percepção do fator de difusão de sinal.....	13
<b>Figura 5.</b> Diagrama de Venn ilustrando as categorias de um pangenoma.....	17
<b>Figura 6.</b> Gráfico representativo do número de clusters gênicos do pangenoma de <i>Xanthomonas campestris</i> .....	31
<b>Figura 7.</b> Diagrama de Venn dos grupos ortólogos do pangenoma de <i>Xanthomonas campestris</i> .....	31
<b>Figura 8.</b> Gráfico descrevendo as características do pangenoma de <i>Xanthomonas campestris</i> .....	32
<b>Figura 9.</b> <i>Heatmap</i> da análise de identidade de nucleotídeos entre os genomas dos patovares de <i>Xanthomonas campestris</i> .....	33
<b>Figura 10.</b> Gráfico das rotas metabólicas anotadas pelo <i>software</i> BlastKOALA para cada patovar de <i>Xanthomonas campestris</i> .....	35
<b>Figura 11.</b> <i>Heatmap</i> dos metabólitos secundários preditos pelo AntiSMASH para os genomas dos patovares de <i>Xanthomonas campestris</i> .....	37
<b>Figura 12.</b> <i>Heatmap</i> dos genes de biossíntese de xantana presentes nos genomas dos patovares de <i>Xanthomonas campestris</i> .....	41
<b>Figura 13.</b> <i>Heatmap</i> dos genes de virulência presentes nos genomas dos patovares de <i>Xanthomonas campestris</i> .....	44
<b>Figura 14.</b> Diagrama de Venn das proteínas efetoras preditas para o <i>core</i> genoma de <i>Xanthomonas campestris</i> .....	45
<b>Figura 15.</b> <i>Heatmap</i> dos genes de DSF presentes nos genomas dos patovares de <i>Xanthomonas campestris</i> .....	47

## Lista de Tabelas

<b>Tabela 1.</b> Lista dos genomas representativos de <i>Xanthomonas campestris</i> .....	28
<b>Tabela 2.</b> Número de genes anotados utilizando o <i>software</i> Genix em comparação com a anotação disponível no NCBI para cada genoma dos patovares.....	29
<b>Tabla 3.</b> Cobertura da anotação metabólica pelo <i>software</i> BlastKOALA para os genomas de <i>Xanthomonas campestris</i> .....	34
<b>Tabela 4.</b> Relação dos genes alvos de TALES preditos pelo <i>software</i> TargeTALE...38	
<b>Tabela 5.</b> Relação dos genes da rota de biossíntese de xantana em <i>Xanthomonas campestris</i> .....	38
<b>Tabela 6.</b> Genes de virulência presentes nos genomas de <i>X. campestris</i> .....	41
<b>Tabela 7.</b> Relação dos genes de DSF em <i>Xanthomonas campestris</i> .....	45

## Lista de Abreviaturas

DSF - *Diffusible Signal Fator* (Fator de Difusão de Sinal)

DNA – *Deoxyribonucleic acid* (Ácido desoxirribonucleico)

EPS - *Extracellular Polysaccharides* (Polissacarídeos Extracelulares)

FAA - *Fasta of Amino Acid*

FFN - *Fasta Feature Nucleotides*

FNA - *Fasta of Nucleic Acid*

GBK - Genbank

GFF - *gene feature format*

GO - *Gene Ontology* (Ontologia Gênica)

KEGG – *Kyoto Encyclopedia of Gene and Genomes* (Enciclopédia de genes e genomas de Kyoto)

Hrp - *Hypersensitive Pathogenicity Response* (Resposta Hipersensível de Patogenicidade)

LPS - Lipopolissacarídeos

MAMPs/PAMPs - *Pathogen-Associated Molecular Patterns* (Padrões Moleculares Associados aos Patógenos)

Mb - Megabase

NCBI - *National Center of Biotechnology Information* (Centro Nacional de Informação de Biotecnologia)

NGS - *Next Generation Sequencing* (Sequenciamento de Nova Geração)

ORF - *Open Read Frame* (Quadro Aberto de Leitura)

PTI - *PAMPs Triggered Immunity* (Imunidade Desencadeada por PAMPs)

pv - Patovar

T1SS - *Type 1 Secretion System* (Sistema de Secreção do Tipo 1)

T2SS - *Type 2 Secretion System* (Sistema de Secreção do Tipo 2)

T3SS - *Type 3 Secretion System* (Sistema de Secreção do Tipo 3)

T4SS - *Type 4 Secretion System* (Sistema de Secreção do Tipo 4)

T5SS - *Type 5 Secretion System* (Sistema de Secreção do Tipo 5)

T6SS - *Type 6 Secretion System* (Sistema de Secreção do Tipo 6)

TBL – (*Feature*)Table

Xcc - *Xanthomonas campestris* pv. *campestris*

## Sumário

1 INTRODUÇÃO GERAL.....	1
2 REVISÃO BIBLIOGRÁFICA .....	3
2.1 Gênero <i>Xanthomonas</i> .....	3
2.2 <i>Xanthomonas campestris</i> .....	6
2.3 Fatores de virulência em <i>Xanthomonas campestris</i> .....	8
2.3.1 Xantana.....	9
2.3.2 Proteínas efetoras e efetores TAL.....	11
2.3.3 Fator de Difusão de Sinal.....	12
2.4 Abordagens bioinformáticas para análise de procariotos.....	14
2.4.1 Análise funcional e bancos de ontologia.....	15
2.4.2 Pangenômica.....	16
2.4.3 Genômica comparativa .....	18
3 HIPÓTESE E OBJETIVOS.....	21
3.1 HIPÓTESE.....	21
3.2 OBJETIVO GERAL.....	21
3.3 OBJETIVOS ESPECÍFICOS.....	21
4 MÉTODOS.....	22
4.1 Obtenção dos genomas de <i>Xanthomonas campestris</i> .....	22
4.2 Re-anotação estrutural de cada gene.....	22
4.3 Análise dos grupos ortólogos.....	23
4.4 Análise pangenômica estrutural.....	24
4.5 Cálculo de identidade média de nucleotídeos.....	24
4.6 Anotação de rotas metabólicas.....	24
4.7 Análise de metabólitos secundários.....	25
4.8 Predição dos efetores TAL.....	25
4.9 Análise dos genes de biossíntese de xantana.....	25
4.10 Análise dos genes de virulência.....	26
4.11 Predição de proteínas efetoras.....	26
4.12 Análise dos genes DSF.....	27
5 RESULTADOS.....	28

5.1 Obtenção dos genomas de <i>Xanthomonas campestris</i> .....	28
5.2 Re-anotação estrutural de cada gene.....	29
5.3 Análise dos grupos ortólogos.....	30
5.4 Análise pangenômica estrutural.....	32
5.5 Cálculo de identidade média de nucleotídeos.....	32
5.6 Anotação de rotas metabólicas.....	34
5.7 Análise de metabólitos secundários.....	35
5.8 Predição dos efetores TAL.....	37
5.9 Análise dos genes de biossíntese de xantana.....	38
5.10 Análise dos genes de virulência.....	41
5.11 Predição de proteínas efetoras.....	44
5.12 Análise dos genes do DSF.....	45
6 DISCUSSÃO.....	48
7 DISCUSSÃO GERAL E PERSPECTIVAS.....	56
8 CONCLUSÃO GERAL.....	57
8 REFERÊNCIAS.....	58
10 ANEXOS.....	70
Anexo A. Pipeline do <i>software</i> GENIX.....	70
Anexo B. Visão geral do algoritmo do <i>software</i> GET_HOMOLOGUES.....	71
Anexo C. Visão geral do algoritmo do <i>software</i> OrthoMCL.....	72
Anexo D. Proteínas efetoras presentes em todos os patovares de <i>Xanthomonas campestris</i> .....	73

## 1 INTRODUÇÃO GERAL

*Xanthomonas* é um gênero de bactérias fitopatogênicas Gram-negativas em forma de bastonete e flageladas, que abarca uma ampla gama de espécies, as quais podem ser subdivididas em diversos patovares (variante patogênica) conforme sua especificidade pelo hospedeiro, doença e modo de infecção. Esse gênero possui uma grande gama de fatores de virulência, dentre os quais se destacam as adesinas, enzimas extracelulares de degradação de parede celular e outros carboidratos como os lipopolissacarídeos e exopolissacarídeos. Além disso, apresenta uma grande variedade de hospedeiros, como os citros, uvas e algodão, os quais têm como agente etiológico, respectivamente, *X. citri* subsp. *citri*, *X. campestris* pv. *viticola* e *X. axonopodis* pv. *mabacearum*.

A espécie *X. campestris* é a mais complexa do gênero, uma vez que possui diversos patovares, sendo estes caracterizados por sua alta especificidade ao hospedeiro, sintomas e doença causada, indicando sua alta diversidade patogênica. Considerando essa variedade, essa espécie é responsável por um grande impacto na produtividade e na economia de plantas como uva, tomate, pimentão, rabanete, banana, dentre outras. Esse fitopatógeno utiliza de certas abordagens para infectar seus hospedeiros, como os sistemas de secreção de proteínas, proteínas efetoras, resposta hipersensível e de patogenicidade, sistema de difusão de sinal e a goma xantana.

As ferramentas da genômica moderna possibilitam a caracterização dos organismos em níveis genéticos, fisiológicos, de eventos evolutivos e rotas metabólicas. Nesse contexto, o sequenciamento de nova geração permitiu um aumento expressivo no volume de dados genômicos em bancos públicos, como o GenBank. Atualmente, existem 34 genomas completos de *X. campestris* depositados no banco de dados. Considerando esse volume de informações, estudos comparativos possibilitam compreender a plasticidade e as características

destes genomas. Em vista disso, os estudos de pangenoma e de genômica comparativa auxiliam na compreensão e na elucidação de características genéticas de diferentes isolados de uma mesma espécie ou gênero. Essas duas abordagens permitem elucidar aspectos evolutivos, adaptativos e interações entre patógeno-hospedeiro.

A presente dissertação é composta por abordagens bioinformáticas distintas, referentes ao estudo das relações entre genoma e patogenicidade dos patovares de *Xanthomonas campestris*. Será descrito o pangenoma de *X. campestris*, dando ênfase para a relação genômica dos patovares. Bem como, será descrita a distribuição de genes de virulência e rotas metabólicas, a fim de entender rotas e genes específicos a cada patovar.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Gênero *Xanthomonas*

O gênero *Xanthomonas* é um grupo diversificado e economicamente importante, que conta com mais de 30 espécies de bactérias gram-negativas fitopatogênicas. As espécies de *Xanthomonas* (spp.) são, tipicamente, em forma de bastonete com um único flagelo polar, são aeróbios obrigatórios. A formação de colônias desse gênero geralmente é amarela devido à presença de pigmento xantomonadina ligado à membrana. Os membros do gênero causam doenças em pelo menos 124 espécies monocotiledôneas e 268 espécies de dicotiledôneas, com destaque para *Xanthomonas citri*, *Xanthomonas oryzae* e *Xanthomonas campestris* pv. *campestris*, as quais têm como hospedeiro, respectivamente, citros (laranja, limão e lima), arroz e crucíferas (brócolis e couve-flor) (Royer et al., 2013; Rodriguez-R et al., 2012). Além disso, muitas espécies do gênero se dividem em diferentes patovares, conforme a especificidade de hospedeiro, sintomas, doença e modo de infecção. Em geral, esses fitopatógenos causam uma variedade de sintomas, incluindo necrose, cancrios, manchas, podridão, hipertrofia e hiperplasia, além de afetar uma variedade de tecidos da planta, incluindo folhas, hastes, caules, flores, brotos, sementes e frutas (Figura 1) (Jacques et al., 2016).

O processo de infecção ocorre, primeiramente, nos tecidos foliares, onde através de feridas e abertura dos estômatos e hidatódios a bactéria penetra o parênquima mesofílico. A primeira interação molecular hospedeiro-patógeno ocorre enquanto a bactéria ainda está nas superfícies dos tecidos, no qual, através de diversos sensores, há o reconhecimento do ambiente (Parkinson et al., 2007; Ryan et al., 2009).

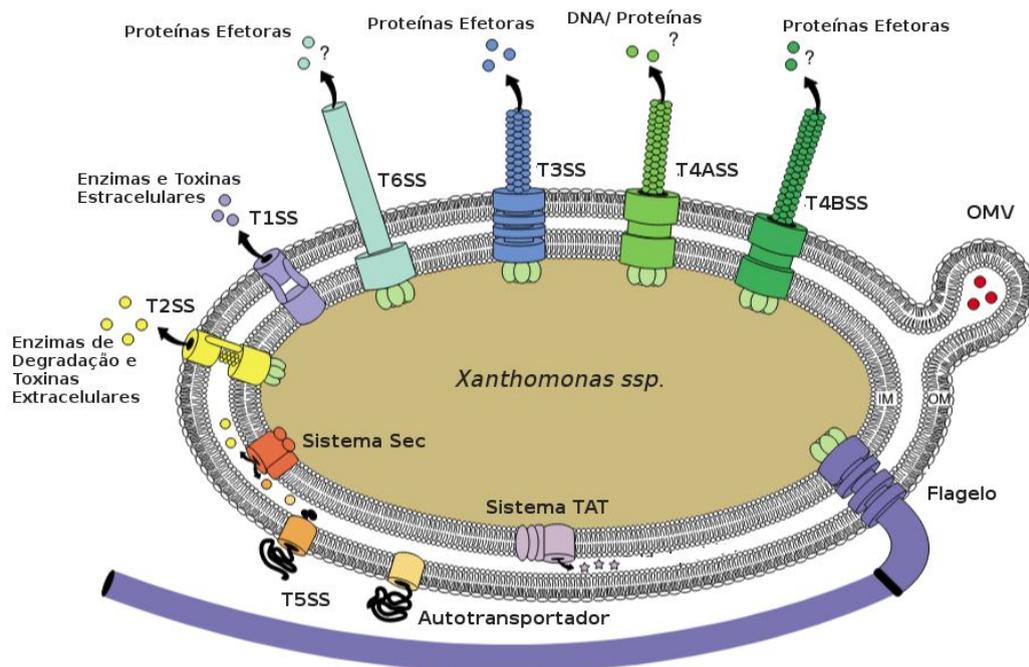


**Figura 1.** Sintomas causados pelo gênero *Xanthomonas* em diferentes tecidos de plantas. (A) Cancro cítrico causado por *X. citri* em folhas de laranja. (B) Cancro cítrico causado pela *X. citri* subsp. *citri* no fruto da laranja. (C) Podridão negra causada pela *X. campestris* pv. *campestris* em couve. Adaptado de Vicente & Holub, (2013).

Sabe-se que os microrganismos enfrentam diversas barreiras até infectar de fato as plantas, dentre elas está a parede celular da planta, a qual apresenta receptores extracelulares que reconhecem padrões moleculares associados aos patógenos (MAMPs/PAMPs—*Pathogen-Associated Molecular Patterns*) (Gürlebeck; et al., 2006). Esse reconhecimento acarreta numa resposta imune basal desencadeada por PAMPs (*PTI – PAMPs Triggered Immunity*) e que, geralmente, limitam a ação do patógeno. No entanto, as espécies do gênero *Xanthomonas* desenvolveram mecanismos de evadir essa defesa do hospedeiro utilizando proteínas secretadas através de seis sistemas altamente especializados e conservados, chamado de Sistema de Secreção do Tipo I, II, III, IV, V e VI, os quais diferem significativamente em sua composição e função, e no reconhecimento de substratos (Figura 2) (Ryan et al., 2011). Estas proteínas são usualmente denominadas “efetoras”.

O T1SS é um complexo de proteína heterotrimérica que consiste em um cassete de ligação de ATP na membrana interna, um canal de proteína na membrana externa e proteína de fusão na membrana que liga os componentes de membrana interna e externa. Em *X. oryzae* pv. *oryzae* mostrou-se necessário para

a obtenção de uma resposta de resistência em grãos de arroz que carregam o gene *Xa21* de resistência à doença. Já o T2SS, é principal sistema de secreção de toxinas e enzimas extracelulares tais como proteases, lipases e enzimas degradantes da parede celular que podem contribuir para a interação do hospedeiro-patógeno. Os principais fatores de patogenicidade da maioria das bactérias do gênero *Xanthomonas* é o T3SS, sendo ele codificado pelo grupo dos genes *hrp/hrc* e permite a secreção de, aproximadamente, 25 proteínas. Dentre essas, estão as proteínas efetoras que são direcionadas para o interior das células do hospedeiro (Cianciotto & White, 2017). O T4SS está relacionado ao sistema de conjugação bacteriana, secretando no meio extracelular DNA e proteínas. O T5SS é responsável pela secreção de adesinas. Por fim o T6SS, o qual ainda não é bem caracterizado.



**Figura 2.** Esquema representativo dos sistemas de secreção de proteínas em *Xanthomonas* spp. Os sistemas T2S e T5S dependem do sistema Sec ou TAT para o transporte de proteínas através da membrana interna. Os sistemas T3S, T4S e T6S estão associados às estruturas de pilus extracelulares para translocar proteínas para a célula hospedeira. Somente em alguns casos, a secreção de proteínas depende da formação de vesículas da membrana externa (OMV). IM, membrana interna; OM, membrana externa; TAT, translocação gêmea-arginina. Adaptado de Büttner & Bonas, 2010.

A caracterização dos fatores de virulência e patogenicidade no gênero *Xanthomonas* foi possível devido a estudos genômicos, os quais revelaram que os plasmídeos variam de tamanho e carregam genes de virulência e genes de resistência; *cassetes* de genes e integrons são responsáveis pela grande diversidade de genomas no gênero (Gonçalves & Rosato, 2002). Análises de pangenoma do gênero *Xanthomonas* revelaram que é do tipo aberto, uma vez que existem muitos elementos móveis dentro de cada população, e a adição de genomas à análise fornece a descoberta de novos genes (Couger et al., 2015). A compreensão desses elementos gerais do gênero permite relacionar fatores genéticos que refletem na filogenia e na classificação das espécies, bem como permite entender os processos individuais de interação hospedeiro-patógeno.

## **2.2 *Xanthomonas campestris***

A espécie *X. campestris* é a mais complexa dentro do gênero, além de ser uma das cinco principais fitobactérias com maior relevância, uma vez que possui uma grande diversidade de patovares, os quais possuem especificidade de hospedeiro, sintomas e doenças (Schatschneider et al., 2017). Sabe-se que as doenças causadas por essa espécie possuem diversos tipos de transmissão, como por exemplo solo contaminado, resíduos de plantas infectadas e pelo transporte de sementes contaminadas. Outra característica importante da espécie está relacionada a sua sobrevivência no solo, a qual pode ser de, aproximadamente, 40 dias no inverno e 20 dias no verão. Em algumas condições, a bactéria pode sobreviver em sementes por mais de um ano, sendo assim, as sementes servem como reservatório que facilita o transporte do patógeno para diversas áreas. Além disso, a bactéria pode dispersar por curtas distâncias através do vento, de insetos, aerossol, irrigação, chuva e por equipamentos (Rukayadi et al., 2000; López et al., 1999).

Dentre os 14 patovares de *X. campestris* analisados neste trabalho, somente quatro deles apresentam estudos e materiais descritivos das suas características

individuais, como sintomas, doença, processos de infecção e características da patogenia. Dentre esses, está *X. campestris* pv. *campestris* (*Xcc*) o qual é o patovar mais bem descrito dentro da espécie. Esse patovar é o agente etiológico da doença conhecida como podridão negra, a qual afeta plantas da família *Brassicaceae*, incluindo a planta modelo *Arabidopsis thaliana*. A podridão negra foi elucidada pela primeira vez por Garman (1894), sendo uma doença responsável por causar sérios danos à lavoura, e tem como sintomas lesões amarelas em forma de V, com o vértice voltado para o centro do tecido. Os sintomas podem evoluir, provocando necroses, quedas foliares, murcha e apodrecimento, e os vasos do xilema ficam escurecidos. A infecção de *Xcc* em plantas hospedeiras ocorre, inicialmente, através dos hidatódios ou por feridas, que resulta na invasão dos espaços intracelulares do parênquima mesófilo. Em geral, *Xcc* é um agente etiológico em ambientes de clima quentes e úmidos, consequentemente, afeta regiões tropicais, subtropicais e regiões continentais úmidas (Vicente & Holub, 2013).

*X. campestris* pv. *musacearum*, o qual é responsável por causar a murcha bacteriana da banana (falsa banana) *Ensete ventricosum*. O processo de infecção ocorre, primeiramente, nos tecidos florais, entretanto o processo ainda não é bem compreendido (Tripathi et al., 2010). No entanto, antes da manifestação de sinais externos e morte da planta, pode ocorrer o amadurecimento prematuro e o apodrecimento dos frutos (Studolme et al., 2010). O primeiro relato da doença foi em 1968 na Etiópia, atualmente a doença está confirmada em 21 distritos da Uganda, do Norte de Kiwu, República Democrática do Congo, Ruanda, Tanzânia, Burundi e no Quênia. Já no Brasil, a doença nunca foi relatada e é considerada inexistente. Sabe-se que essa doença gera impactos negativos nas comunidades rurais desses países, onde milhões de pessoas dependem da banana como alimento básico e fonte de renda (Ocimati et al., 2018).

Já o patovar *X. campestris* pv. *viticola* o agente etiológico do cancro bacteriano da videira (*Vitis vinifera* L.) produtoras de uvas finas de mesa. Os

sintomas são lesões primárias nas folhas com pontos necróticos de 1 a 2 mm de diâmetro, algumas vezes causando a morte de extensas áreas do limbo foliar, obstrução do tecido vascular nos galhos, nós e talos. Dessa forma há o impedimento do fluxo de seiva, conseqüentemente há a interferência negativa no crescimento vegetativo e reprodutivo da videira. Os sintomas variam conforme a tolerância à doença e as condições ambientais. A disseminação pode ocorrer através de restos de cultura infectados espalhados pelo pomar ou aderidos em roupas, veículos, mas principalmente em materiais utilizados na colheita dos frutos de forma inadequada e não desinfestadas (Silva et al., 2012; Trindade et al., 2007). Por fim, *X. campestris* pv. *raphani* responsável por causar mancha foliar do rabanete (*Raphanus sativus*), o primeiro relato da doença foi em 1930 nos Estados Unidos. Os sintomas são lesões necróticas nas folhas, pecíolos e caules (Vicente et al., 2006).

O processo de patogenia de *X. campestris* é caracterizado, primeiramente, pela interação patógeno-hospedeiro nas superfícies aéreas da planta, onde são secretados uma série de proteínas efetoras, incluindo estruturas filamentosas chamadas Resposta Hipersensível de Patogenicidade (Hrp), as quais são anexas ao sistema de secreção, e através do Fator de Sinal Difusível (DSF), o qual é dependente do agrupamento de genes Rpf, permitindo a comunicação célula-célula e criação de biofilmes nas superfícies da planta. Para a efetividade da infecção os fatores de virulência desempenham um papel importante no processo. Sendo eles, enzimas líticas que atacam a parede celular da planta, excreção de proteases, amilases, celulasas, lipases que ajudam a diminuir os mecanismos de defesa da planta e síntese de polissacarídeo extracelular (EPS) (Lonjon et al., 2017; Scheibner et al., 2017).

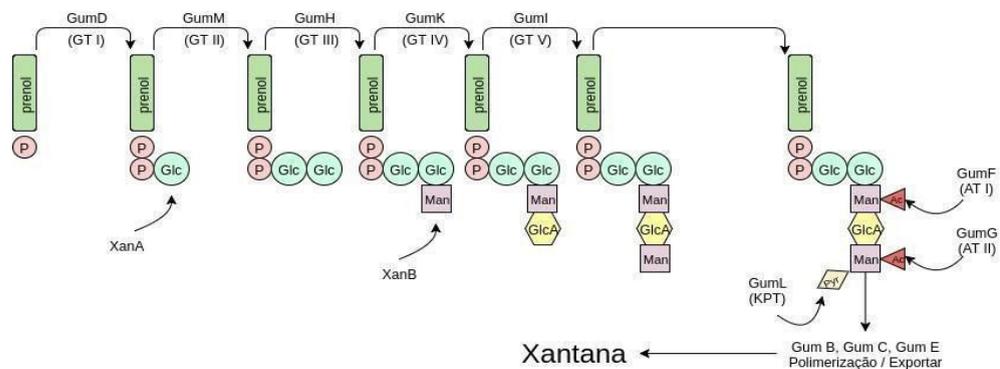
### **2.3 Fatores de Virulência em *Xanthomonas campestris***

O patógeno necessita de estratégias que permitam ao máximo o fornecimento de nutrientes pelo hospedeiro para sua conseqüente colonização e

multiplicação. Para que a bactéria se estabeleça no hospedeiro e cause a doença, um processo complexo de atividades coordenadas de genes bacterianos, ligados à patogenicidade e virulência, é exigido. Um dos principais objetivos da patologia vegetal é a identificação de fatores de virulência, visto que identificá-los e entendê-los contribuem para a elucidação da interação hospedeiro-patógeno (Zhang et al., 2019; Büttner & Bonas, 2010). Nesse contexto, sabe-se que a infecção bem sucedida dos fitopatógenos dependem da combinação de diferentes sistemas de secreção, ou de sistemas de transporte de proteínas no meio extracelular da bactéria. A genética molecular extensiva de *X. campestris* identificou os principais determinantes da patogenicidade e fatores de virulência, dentre os quais se destacam as adesinas, enzimas extracelulares de degradação de parede celular e outros carboidratos, lipopolissacarídeos e exopolissacarídeos (EPS), sistemas de translocação de proteínas, e proteínas efetoras (Lima et al., 2008; Maximiano et al., 2017). Estes fatores, normalmente, contribuem nas modificações, e, por vezes, no comprometimento da homeostase do tecido invadido.

### 2.3.1 Xantana

O EPS está relacionado com a sobrevivência em ambientes hostis, proteção contra a desidratação, concentração de minerais e nutrientes, redução do contato com moléculas hidrofóbicas ou carregadas e fixação na superfície do vegetal hospedeiro. Em *X. campestris* o EPS é denominado de xantana, sendo esse um heteropolissacarídeo sintetizado por meio de processo fermentativo aeróbico, que utiliza a glicose ou sacarose como principal substrato. A biosíntese da goma xantana envolve algumas rotas metabólicas e genes como *xanA*, *xanB* e o *operon gum* (Figura 3) (Katzen et al., 1998; Vojnov et al., 2001).



**Figura 3.** Esquema da rota da biossíntese de xantana em *Xanthomonas campestris*. Os componentes das ligações lipídicas são representados por: Glc, glucose; Man, manose; GlcA, ácido glucurônico; Ac, grupo acetilo; Pyr, grupo pirúvico. A designação de cada proteína é seguida pela sua função proposta da seguinte forma: GT, glicosiltransferase; AT, acetiltransferase; KPT, transferase de piruvato cetal. As setas indicam a possível ordem de polimerização e os genes envolvidos em cada etapa. Adaptado de Katzen *et al.*, (1998).

Evidências experimentais sugerem que a xantana suprime as respostas de defesa basais das plantas, como a deposição de calose na parede celular da planta. Isso ocorre, presumivelmente, por quelação de íons de cálcio divalentes que estão presentes no apoplasto da planta, sendo esses íons necessários para a ativação das respostas de defesa das plantas (Yun *et al.*, 2006; Aslam *et al.*, 2008).

Além disso, outros estudos têm demonstrado a participação da xantana na formação de biofilmes (Dow *et al.*, 2003; Rigano *et al.*, 2007; Torres *et al.*, 2007). Biofilmes são estruturas caracterizadas pelo elevado grau de organização de uma população bacteriana, na qual bactérias se ligam umas às outras, em superfícies bióticas ou abióticas. Essa estrutura organizacional se dá por meio de uma matriz polimérica extracelular que consiste principalmente em EPS, proteínas e lipídios (Sutherland, 2001). A formação do biofilme fornece proteção contra os antibióticos e às respostas de defesa do hospedeiro, bem como permite a sobrevivência epifítica bacteriana antes da colonização do espaço intercelular da planta, além de permitir a fixação nos vasos do xilema (Stoodley *et al.*, 2002).

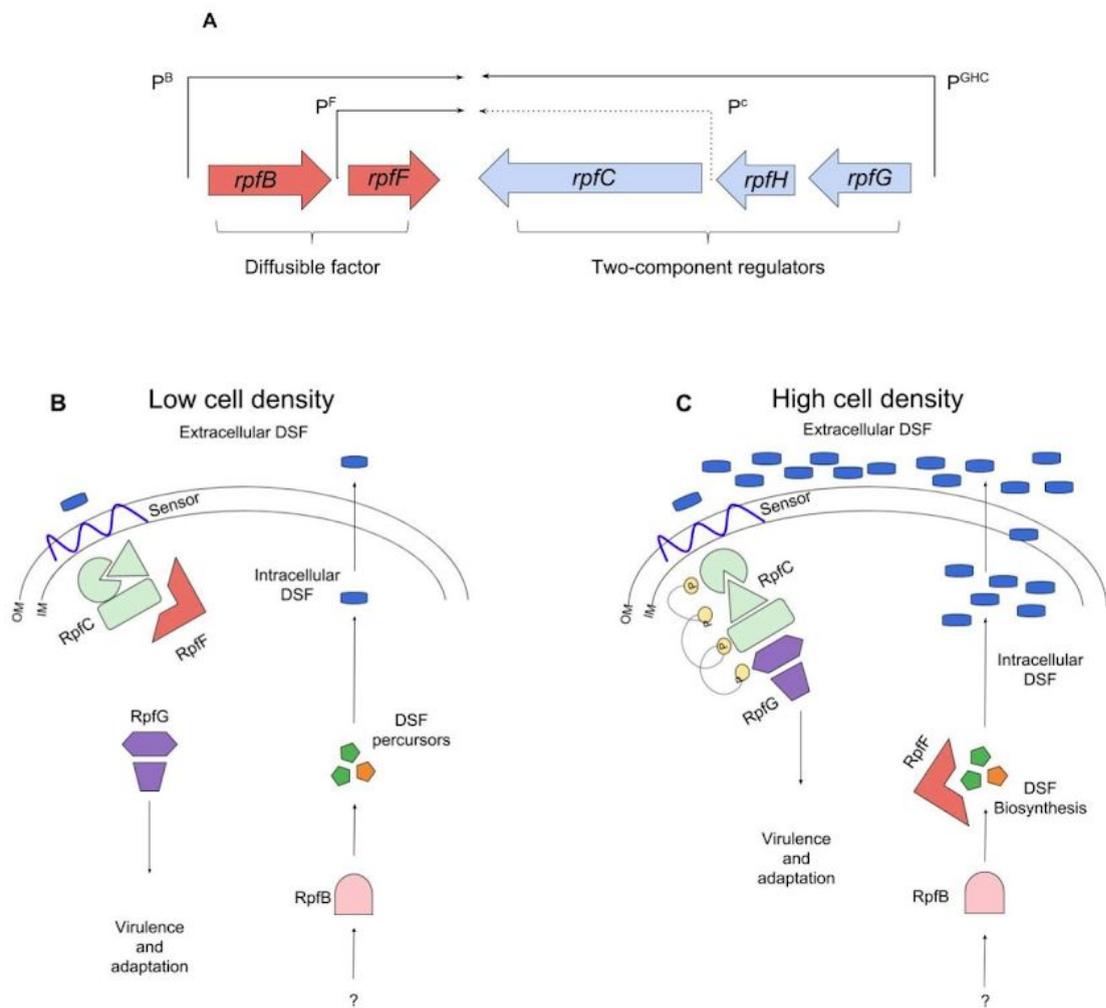
### 2.3.2 Proteínas efetoras e efetores TAL

Estudos envolvendo bactérias Gram-negativas, com sistema de T3SS, revelaram que estas bactérias secretam proteínas no interior das células hospedeiras, as quais desencadeiam reações de hipersensibilidade. O processo de infecção e desenvolvimento de *X. campestris* se inicia no espaço extracelular da planta hospedeira, por isso existe o mecanismo de translocação de proteínas efetoras para o citoplasma do hospedeiro. Essas proteínas se acumulam no núcleo e são denominadas efetoras pois modulando a expressão de genes alvo, bem como possuem sítios de reconhecimento de promotores de genes, no hospedeiro, que uma vez expressos contribuem para desencadear a doença. A estrutura dessas proteínas permite que sítios, comuns a todas efetoras, interajam com proteínas do hospedeiro e com gene alvos. As proteínas efetoras são multidomínio, apresentam de localização nuclear (NLS) na porção C-terminal, um N-terminal necessário à maquinaria de T3SS e um domínio de ativador transcricional ácido semelhante aos fatores de transcrição (Leng et al. 2019; Yan et al., 2019).

Os efetores TAL (*Transcricional activator like*) são proteínas importantes, produzidas pela maioria das espécies do gênero *Xanthomonas*, incluindo a espécie *X. campestris*. Essas proteínas são fatores de virulência, sendo elas secretadas pela via do T3SS (Bochtler, 2012). A principal função dos efetores TAL está relacionada com a modulação transicional de genes através do reconhecimento específico de regiões do genoma do hospedeiro. A atividade específica dos efetores TAL ocorre devido a presença de domínios centrais denominados *repeat di-residues* (RVDs), os quais contêm repetições em *tandem* de 34 aminoácidos. Em *X. campestris* os efetores TAL atuam diretamente no núcleo celular do hospedeiro, modulando genes envolvidos em funções como transporte de açúcar e processos epigenéticos (Scholze & Boch, 2010).

### 2.3.3 Fator de Difusão de Sinal

*X. campestris* usa o mecanismo de sinalização célula a célula, conhecido como sensor de quorum (QS), para coordenar comportamentos e atividades em resposta ao seu ambiente, como densidade populacional, bioluminescência, transferência de plasmídeo e motilidade da superfície. Essa regulação compõe mecanismo de produção, transdução de sinal, ativação de expressão de genes alvos e detecção de pequenas moléculas de difusão de sinal, também chamadas de autoindutores (Deng et al., 2011; Ryan et al., 2011). Desde a década de 1980, uma gama de QS tem sido ilustrada em microrganismos, indicando vários tipos de sinais QS e diferentes mecanismos de transdução de sinal. Três tipos principais de autoindutores foram identificados em bactérias Gram-negativas: (1) lactonas N-acil-homoserina (AHLs) compostas por um anel de homo-serina-lactose contendo cadeias acila C4-C18 e se comportam como uma molécula de sinal intra espécie; (2) outra é a 4,5-di-hidroxi-2,3-pentadiona (DPD), uma molécula de sinal interespecies; (3) o terceiro é chamado de fatores de difusão de sinal (DSF), descoberto pela primeira vez em *X. campestris* (Barber et al., 1997; Wang et al., 2004). Além de seus papéis principais na sinalização intra-espécies, também foi reconhecido a importância do sinal DSF nos processos de virulência, formação de biofilme, motilidade, tolerância ao estresse, adaptação e síntese de EPS. Análises genéticas e bioquímicas com *X. campestris* pv. *campestris* (Xcc) estabeleceu que a síntese e percepção do sinal DSF requerem produtos do agrupamento de genes *rpf* (regulação de fatores de patogenicidade) (Figura 4) (He et al., 2006; Zhou et al., 2015).



**Figura 4.** Mapa do *cluster rpf* em *X. campestris*, e organização transcricional envolvida na sinalização, biossíntese e percepção de DSF. (A) O *rpfG*, *rpfH* e *rpfC* são transcritos como um operon de um promotor  $P^{GHC}$ , e o *rpfC* possui seu próprio promotor fraco,  $P^C$ ; *rpfB* e *rpfF* são transcritos como um *operon* a partir de um promotor *upstream* do *rpfB*,  $P^B$ , no entanto, o *rpfF* possui seu próprio promotor,  $P^F$ . As proteínas RpfF e RpfB são responsáveis pela síntese do DSF, enquanto RpfG e RpfC compõem um sistema regulador de dois componentes. (B) quando o tamanho da população bacteriana e a concentração extracelular de DSF são baixas, o RpfC mantém uma conformação compacta, formando um complexo com RpfF e nenhuma transferência de fósforo é iniciada, o que impede síntese efetiva do sinal DSF. (C) quando a população bacteriana e a concentração extracelular de DSF são altas, o sinal extracelular de acúmulo de DSF interage com o RpfC, induzindo as alterações de conformação no sensor, que iniciam a autofosforilação e o fósforo para RpfG e liberam RpfF. Este evento aumenta a biossíntese e altera a expressão do regulador DSF, induzindo a expressão do EPS e das enzimas extracelulares. A seta indica fluxo ou geração de sinal. O ponto de interrogação indica precursores desconhecidos.

## 2.4 Abordagens bioinformáticas para análise genômica de procariotos

Sequenciamento é o processo que determina a ordem de uma sequência nucleotídica do DNA ou RNA de um organismo. Essa técnica teve início em 1975 com Sanger e Coulson (Sanger & Coulson, 1975). No entanto, foi com o advento dos sequenciamentos de nova geração que houve um aumento expressivo no número de dados genômicos. Com o grande fluxo de dados gerados tanto pelo sequenciamento, como pelas análises genômicas, surgiu a necessidade de tratar esses dados e armazená-los de forma virtual, ou seja, armazenar dados biológicos utilizando ferramentas de informática. Nesse contexto, a bioinformática, atualmente, é centro da interpretação dos dados genômicos, uma vez que possibilitou a criação de banco de dados e ferramentas que aperfeiçoam o processo de análises. Juntamente com a bioinformática, os bioinformatas exercem papel crucial na evolução da pesquisa, pois a análise dos dados gerados é provavelmente o próximo gargalo na pesquisa biológica (Lucaciu et al., 2019).

As abordagens da bioinformática e da biologia molecular permitem a prospecção *in silico* de informações, a partir de dados genômicos, possibilitando a exploração dos recursos genéticos com a finalidade de obter produtos biotecnológicos como antibióticos, agentes terapêuticos, probióticos, enzimas, polímeros, além de propiciar melhores prognósticos e prevenções de doenças emergentes (McInerney et al., 2008).

Nesse sentido a caracterização de organismos procariotos é de extrema relevância possuindo, atualmente, diferentes focos, por exemplo, entendimento da genética e fisiologia, eventos evolutivos, bioquímica, desenvolvimento de novos compostos biotecnológicos e patogênese de doenças (Handelsman, 2004; Woyke et al., 2010). A genômica de microrganismo conta com um imenso banco, na base de dados *Nucleotide* do NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) constam, aproximadamente, 52.291.489 entradas de genomas para bactérias. Já no banco GOLD ([www.genomesonline.org](http://www.genomesonline.org)), atualmente, estão registrados cerca de 136.778 projetos

genomas para Procariotos. Essa grande quantidade de dados possibilitou o surgimento de novas linhas de pesquisas, tais como a genômica funcional e proteômica, as quais permitem a análise comparativa dos genomas bacterianos de forma a identificar mecanismos adaptativos, aspectos evolutivos e produtos biotecnológicos.

### 2.4.1 Análise funcional e bancos de ontologia

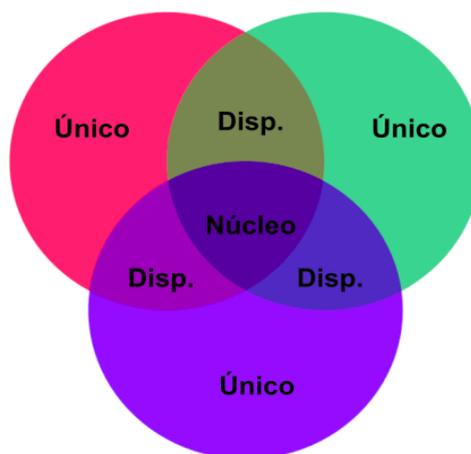
A anotação de genomas refere-se à identificação de seus elementos estruturais e funcionais. Normalmente, esse processo é realizado após o sequenciamento e subsequente montagem de um genoma, sendo a anotação uma etapa obrigatória para que o depósito dos dados seja realizado em bancos de dados. Apesar da identificação funcional fornecer informações sobre o produto gênico, ela não apresenta informações sobre os processos biológicos da proteína (König et al., 2018). Nesse contexto, a anotação funcional fornece dados mais complexos, permitindo a identificação da função proteica, bem como rotas metabólicas a ela relacionada.

A ontologia genica permite estabelecer relações padronizadas entre componentes celulares, funções moleculares, funções bioquímicas e processos biológicos. Essa relação é feita através de um vocabulário com termos que definem essas descrições (Ryan et al., 2009). Nesse caso, o *Cluster of Ortholog Groups* (COG), por exemplo, é um banco de dados que possui nomenclatura padronizada para processos biológicos em microrganismos. A existência desse tipo de banco de dados permite maior padronização na anotação de genomas, bem como fornece informações dos produtos gênicos que podem ser usados para comparação entre organismos. Além do COG, bancos de ontologia incluem o banco *Gene Ontology* (GO), o qual fornecem uma classificação útil para funções proteicas e contém mais de 40.000 termos, e o *KEGG Ontology* (KO) que agrega dados ontológicos a rotas metabólicas.

Existem diversas ferramentas que permitem a análise funcional a partir de bancos de dados de ontologia. Dentre elas destacam-se o BLAST2GO (Conesa & Götz, 2008), o qual mapeia as funções através da associação dos hits mais relevantes do BLAST (Altschul et al., 1990), o QuickGO que relaciona as funções a proteínas do UniProt, KAAS e BlastKOALA (Kanehisa et al., 2016) que permitem anotar os genomas a partir de ontologias metabólicas.

### 2.4.2 Pangenômica

Uma das abordagens da genômica é a análise do pangenoma, a qual teve como fundador do termo o pesquisador Rino Rappulio que em parceria com por Tettelin et al. (2005) descreveu essa análise, pela primeira vez, através do estudo do pangenoma de *Streptococcus agalactiae*. Por pangenômica compreende-se a comparação entre diferentes isolados de uma mesma espécie ou de um mesmo gênero com a finalidade de entender aspectos evolutivos, adaptativos e interações entre patógeno-hospedeiro (Glasner et al., 2008). O termo pangenoma está relacionado ao número total de genes não redundantes presentes num conjunto de dados. Esse conjunto de dados é dividido em três categorias: I) Genoma núcleo ou *core genome*; II) Genoma acessório; e III) Genes espécie-específicos (Figura 5) (Muzzi & Donati, 2011).



**Figura 5.** Diagrama de Venn ilustrando as categorias de um pangenoma. A região de intersecção (Núcleo) representa os genes que são compartilhadas por todos os isolados. As região que representam o genoma acessório ou dispensável, compreende aos genes compartilhado entre dois ou mais isolados, mas não por todos. Nas regiões indicadas com “Único” demonstram os genes espécie-específico, ou seja, genes exclusivos a um isolado.

Por genoma núcleo entende-se o conjunto de genes que estão presentes em todos os genomas dos isolados em questão. De uma forma geral, esses genes comuns aos genomas estão relacionados à manutenção dos aspectos básicos da biologia do organismo e são principalmente relacionadas à replicação, tradução e manutenção da homeostase celular. Dessa forma, quanto mais distante os organismos estão filogeneticamente menores é a quantidade de genes em comum, logo menor é o genoma núcleo. Por outro lado, genomas com relação filogenética acabam por compartilhar mais genes e, conseqüentemente, apresentam um genoma do núcleo maior (Udaondo et al., 2016).

O genoma acessório, por sua vez, representa um conjunto de genes compartilhados por alguns isolados, mas não por todos. Esses genes, geralmente, estão ligados a funções específicas relacionadas à sobrevivência, virulência e resistência a antibióticos, por conseqüência refletem as características do estilo de vida do organismo. O genoma acessório, possivelmente, é oriundo da transferência horizontal de genes e da evolução parafilética.

Os genes espécie-específicos são aqueles presente em somente um isolado, os quais desempenham funções adaptativas, e estão relacionados a virulência e patogenicidade em organismos patogênicos, em organismos não patogênicos esses genes podem estar relacionados a rotas metabólicas (Machado et al., 2019).

Além da divisão *core* genoma, genoma acessório e genes espécie-específicos, em alguns trabalhos o pangenoma pode ser dividido em *soft*, *shell* e *cloud* genoma. O *soft* genoma representa os genes presentes em 95% dos genomas. Já o *shell* genoma é a definição para os genes compartilhados entre 10% e 88% dos genomas. Por fim, o *cloud* genoma que agrupa genes presentes em menos de 10% dos genomas (Bezuidt et al., 2016)

Tendo em vista o conhecimento de genoma núcleo, genoma acessório e genes espécie-específicos, duas denominações podem ser dadas a um pangenoma tendo em vista a distribuição cumulativa dos genes. Essa distribuição cumulativa é realizada através da permutação dos genomas analisados gerando uma inferência do grau de “aumento” no número de genes a medida que genomas são adicionados à análise. O resultado dessa permutação informa a característica do pangenoma, sendo ele aberto ou fechado. O pangenoma aberto, significa que a probabilidade de um novo genoma sequenciado fornece novos genes, ainda não descritos para o grupo em análise, é alta, sendo este caso comum em organismos que possuem um pangenoma grande e que novos genes podem ser identificados. O pangenoma fechado, se aplica para organismos que vivem isolados ou restritos a um nicho, essa situação dificulta a transferência horizontal de genes, dessa forma o pool de genes não está em expansão (Marschall et al., 2016).

### 2.4.3 Genômica comparativa

Os microrganismos são os seres vivos mais bem adaptados, numerosos e antigos do planeta e capazes de colonizar todos os nichos biológicos. A existência deles é essencial para manter o equilíbrio ecológico, além de representarem uma fonte importante de recurso genético para a biotecnologia e o desenvolvimento da bioeconomia do século 21 (Allen & Banfield, 2005). As estratégias que utilizam desses recursos permitem o desenvolvimento de novos fármacos, aplicações diversas nas áreas da saúde, agricultura, indústria e meio ambiente.

O genoma consiste em toda informação genética armazenada em um organismo na forma de DNA, que pode ser transmitido de uma geração à outra, englobando tanto genes, unidades indispensáveis que contêm um código para produção dos aminoácidos e sequências reguladoras de expressão, quanto sequências não-codificadoras, ou seja, genoma é a totalidade do DNA presente em um organismo. Estudos que buscam por informações contidas nos genomas dos organismos são de suma importância, uma vez que possibilitam caracterizar genes relacionados a doenças, rotas metabólicas e processos celulares (Li & Burmeister, 2005).

A genômica é responsável pela caracterização de genomas de organismos, fornecendo um catálogo completo dos genes e das proteínas presentes no mesmo. Considera-se que o surgimento dessa área tenha sido no final da década de 1980 com o projeto de sequenciamento do genoma humano. Sendo, portanto, um ramo recente na ciência com uma crescente expansão (Jansen & Nap, 2001). No entanto, independentemente do real momento do seu surgimento, compreende-se que desde a descoberta da dupla hélice de DNA até o desenvolvimento de técnicas de sequenciamento essa área pode ser implementada definitivamente como um campo da ciência moderna (Lockhart & Winzeler, 2000).

Após o surgimento da genômica, esta foi subdividida em diversas subáreas, como genômica funcional, estrutural, metagenômica e epigenômica, dentre outras. A genômica funcional tem como foco compreender funções e interações de genes e proteínas. Já a genômica estrutural disponibiliza ferramentas que permitem a predição e determinação de estruturas tridimensionais de proteínas, assim como a organização e estrutura dos genes. A metagenômica por sua vez é responsável pela análise da diversidade de microrganismos presentes em determinada amostra e, análise funcional de genes de organismos em ambientes específicos. Por fim, a epigenômica busca entender as modificações epigenéticas no material genético (Morozova & Marra, 2008).

### 3 HIPÓTESE E OBJETIVOS

#### 3.1 HIPÓTESE

As diferentes estratégias de bioinformática no estudo genômico dos patovares da bactéria *X. campestris* pode contribuir para a prospecção e caracterização de genes diferencialmente distribuídos, os quais podem estar associados a patogenicidade e possuir potencial biotecnológico.

#### 3.2 OBJETIVO GERAL

Analisar as características do pangenoma dos patovares *arecae*, *azadirachtae*, *campestris*, *centellae*, *durante.*, *incanae*, *leeana*, *musacearum*, *raphani*, *thespesiae*, *viticola*, *vitiscarnosae*, *vitistrifoliae* e *vitiswoodrowii* de *X. campestris*.

Identificar genes e rotas metabólicas diferencialmente distribuídos em cada patovar, e compartilhados por todas as cepas, visando a identificação de características moleculares relacionadas à patogenia.

#### 3.3 OBJETIVO ESPECÍFICO

De forma a alcançar os objetivos gerais, os objetivos específicos são:

- Obter a padronização dos dados a partir da reanotação estrutural dos genomas.
- Obter os genomas núcleo e acessório através da análise de pangenoma para prospectar e caracterizar genes diferencialmente distribuídos entre os patovares.

## 4. MÉTODOS

### 4.1 Obtenção dos genomas de *Xanthomonas campestris*

Os genomas foram obtidos através do banco de dados GenBank presente na plataforma do NCBI. Para fazer a pesquisa dos genomas foi utilizada a *tag* de filtragem: *Xanthomonas campestris* [organism] AND ((*whole genome shotgun* [TITLE] AND *sequencing project* [TITLE]) OR *complete genome* [TITLE]). Fez-se o *download* dos genomas, em seguida foi realizada a métrica N50 através de um script em Python para selecionar um genoma representativo de cada patovar, considerando a diferença de genomas disponíveis para cada um deles.

### 4.2 Re-anotação estrutural de cada gene

Para a re-anotação estrutural dos genes de cada um dos 14 genomas selecionados, utilizou-se a ferramenta GENIX (Kremer et al., 2016) através da linha de comando do sistema operacional LINUX. Esse software utiliza como arquivo de entrada *.fasta*, além disso sua abordagem é através da combinação de ferramentas bioinformáticas como Prodigal (Hyatt et al., 2010), BLASTp (Altschul et al., 1990), tRNAscan-SE (Lowe & Eddy, 1997), RNAmmer (Lagesen et al., 2007), Aragorn (Laslett, 2004), HMMER (Eddy, 2011), BLASTn e INFERNAL (Nawrocki et al., 2009), RFam (Griffiths-Jones et al., 2003), Antifam (Eberhardt et al., 2012) e o conjunto de dados não redundante gerado pelo CD-HIT (Li & Godzik, 2006) Material Complementar 1. Os arquivos de saída dessa ferramenta são no formato GenBank (*.gbk*), GFF e formato Feature Table (*.tbl*), FASTA com arquivos contendo as sequências das proteínas (*.faa*), features não codificadas (*.ffn*) e sequências de DNA codificadas (*.fna*).

### 4.3 Análise dos grupos ortólogos

A partir dos genes re-annotados, para a análise dos grupos ortólogos usou-se as ferramentas GET\_HOMOLOGUES (Contreras-Moreira & Vinuesa, 2013), OrthoMCL e PanSeq (Laing et al., 2010). Primeiramente, utilizou-se o software GET\_HOMOLOGUES, o qual tem como função identificar clusters de proteínas e sequências de nucleotídeos homólogos em sequências similares, identificar grupos ortólogos de genes flanqueados por *open reading frames* (ORFs) ao longo do genoma, além de definir o pan e core genoma, o algoritmo está descrito no Material complementar 2. O arquivo de entrada para essa ferramenta é no formato .gbk. Após instalar todos os módulos necessários para o funcionamento da ferramenta, foi selecionado quais parâmetros seriam usados, dentre os diversos disponíveis pelo programa. Aqueles selecionados foram: -t=1 (selecionar grupos ortólogos que estão presentes em pelo menos um dos genomas, e configurações padrão).

Em seguida utilizou-se o software OrthoMCL, o qual utiliza o proteoma no formato .fasta como entrada dos dados. A visão geral do algoritmo está representada no Material complementar 3. Após instalar todos requisitos, seguiu-se os passos determinados: (1) Instalar o banco de dados; (2) Realizar o *download* e instalar o programa mcl, obtido através do site <http://www.micans.org/mcl/src/mcl-latest.tar.gz>; (3) Configurar o OrthoMCL; (4) Rodar o módulo orthomclInstallSchema com a finalidade de adicionar o Schema ao banco de dados; (5) Rodar o módulo orthomclAdjustFasta para gerar arquivos .fasta com os parâmetros requeridos; (6) Rodar o módulo orthomclFilterFasta com a finalidade de filtrar as proteínas baseado no tamanho e na porcentagem de stop codons, gerando dois arquivos: goodProteins.fasta e poorProteins.fasta; (7) Rodar all-v-all NCBI BLAST para o arquivo goodProteins.fasta; (8) Rodar o módulo orthomclBlastParser através do NCBI BLAST; (9) Rodar o módulo orthomclLoadBlast; (10) Rodar o módulo orthomclPairs, o qual tem como finalidade encontrar pares de proteínas; (11) Rodar o módulo orthomclDumpPairsFiles; (12)

Rodar o programa mcl para o arquivo mcl\_input.txt file criado no passo 11; (13)  
Rodar o módulo orthomclMclToGroups.

Por fim, a ferramenta PanSeq. Foi necessário ajustar as configurações de diretório, quantos núcleos seriam utilizados do computador, mínima região sintênica a ser considerada, sendo, neste caso, 500 pb, tamanho do fragmento nesse caso o mínimo é de 500 pb, limiar de identidade de 85%, quantos genomas seriam analisados (14) e o *runmode* para pan.

#### 4.4 Análise pangenômica estrutural

Para determinar se as análises pangenômicas de *X. campestris* pertencem a classificação de pangenoma aberto ou fechado, utilizou-se a ferramenta de bioinformática PanGP (Zhao et al., 2014). Esse software utiliza como arquivos de entrada no formato .fasta, além disso para utilizá-lo foi necessárias outras ferramentas bioinformáticas como PGAP (Zhao et al., 2012), OrthoMCL e PanOCT (Fouts et al., 2012), as quais foram devidamente instaladas.

#### 4.5 Cálculo de identidade média de nucleotídeos

O cálculo para ANI (*Average Nucleotide Identity*) foi realizado utilizando o pacote de scripts em Python getANI (<https://github.com/biopro/getANI>), através da linha de comando do ambiente LINUX. Foram selecionados os melhores *hits* usando um limite mínimo de 70% de identidade. Com base na identidade gerada pelo cálculo de ANI, um *heatmap* foi gerado utilizando o pacote em Python Seaborn (<https://seaborn.pydata.org/>) com a mensuração de distância de Minkowski.

#### 4.6 Anotação de rotas metabólicas

A anotação das rotas metabólicas foi realizada com o genoma de cada patovar desse estudo de forma individual, utilizando a ferramenta BlastKoala

(Kanehisa et al., 2016). Os arquivos dos genomas re-annotados pelo GENIX no formato .faa contendo as sequências de aminoácidos foram submetido, um a um, na plataforma online do programa. Para os parâmetros “insira o grupo de taxonomia do seu genoma” foi selecionada a opção “*Taxonomy ID*” preenchendo o campo com o código taxonômico da espécie *X. campestris* (339) segundo o banco de dados do NCBI, por fim o parâmetro “banco de dados de genes” foi selecionada a opção “*species\_prokaryotes*”. Com os campos selecionados cada genoma de cada patovar foi submetido.

#### **4.7 Análise de metabólitos secundários**

Para análise dos metabólitos secundários utilizou-se a ferramenta AntiSMASH (Blin et al., 2019). Os arquivos dos genomas re-annotados pelo GENIX no formato .fasta contendo as sequências de nucleotídeos foram submetido, um a um, na plataforma online do programa. Todos os parâmetros de “*extra features*” foram selecionados.

#### **4.8 Predição dos efetores TAL**

A predição dos efetores TAL foi realizada com a ferramenta de uso *web* TargTALE, a qual pode ser acessada pela *url*: <http://bioprolab.ufpel.edu.br/targeTALE/>. Os 14 genomas (Tabela 1) de *X. campestris* foram analisados individualmente, seguindo a orientação da ferramenta. O formato do arquivo de entrada utilizado foi .fasta, e o campo “Promoterome” foi preenchido com o promoteroma de *Brassica oleracea*, a qual é um dos principais hospedeiros de *X. campestris*.

#### **4.9 Análise dos genes de biossíntese de xantana**

Para determinar a distribuição de genes relacionados à biossíntese de xantana nos genomas dos 14 patovares de *X. campestris* (Tabela 1) foi realizado

um alinhamento de sequências. Para isso, fez-se uma revisão da literatura para identificar os genes relacionados à biossíntese da xantana. Após identificá-los, fez-se o *download* das sequências através do banco de dados UniProt. Por fim, um alinhamento múltiplo foi realizado por linha de comando no ambiente LINUX utilizando a ferramenta Blastp, a qual permitiu alinhar sequências de aminoácidos a fim de identificar a presença dos genes de biossíntese de xantana descritos na bibliografia nos genomas dos patovares.

#### **4.10 Análise dos genes de virulência**

Para análise dos fatores de virulência, a primeira etapa foi realizar uma revisão bibliográfica para elencar os genes relacionados à virulência da *X. campestris*. Após identificá-los, fez-se o *download* das sequências através do banco de dados UniProt. Por fim, um alinhamento múltiplo foi realizado por linha de comando no ambiente LINUX utilizando a ferramenta Blastp, a qual permitiu alinhar sequências de aminoácidos a fim de identificar a presença dos genes de virulência descritos na bibliografia nos genomas dos patovares.

#### **4.11 Predição de proteínas efetoras**

A predição de proteínas efetoras foi realizada com os softwares EffectiveDB (Eichinger et al., 2016) e GET\_HOMOLOGUES. Primeiramente, utilizou-se a ferramenta EffectiveDB através da linha de comando, a qual usa como arquivo de entrada no formato .fasta, nesse caso foram utilizados os genomas de todos patovares re-annotados pelo GENIX no formato .fasta. A partir dos resultados do EffectiveDB, utilizou-se o GET\_HOMOLOGUES para determinar a ocorrência dessas proteínas nos genomas dos patovares, de forma a determinar quais proteínas efetoras fazem parte do *core* genoma, quais são do genoma acessório e quais são específicas a cada patovar.

## 4.2 Análise dos genes do DSF

Para realizar a análise dos DSFs, a primeira etapa foi revisar a bibliografia para elencar os e genes relacionados a DSF de da *X. campestris*. Após identificá-los, fez-se o *download* das sequências através do banco de dados UniProt. Por fim, um alinhamento múltiplo foi realizado por linha de comando no ambiente LINUX utilizando a ferramenta Blastp, a qual permitiu alinhar sequências de aminoácidos a fim de identificar a presença dos genes de virulência descritos na bibliografia nos genomas dos patovares.

## 5 RESULTADOS

### 5.1 Obtenção dos genomas de *Xanthomonas campestris*

A pesquisa pela disponibilidade de genomas de *X. campestris* resultou em 82 genomas, sendo 18 genomas completos e 64 genomas rascunho. Foi realizado o *download* dos 82 genomas, dentre os quais 1 genoma disponível para o patovar *arecae*, 1 genoma do patovar *azadiractae*, 55 genomas do patovar *campestris*, 1 genoma do patovar *centellae*, 1 genoma do patovar *durantae*, 3 genomas do patovar *incanae*, 1 genoma do patovar *leeana*, 11 genomas do patovar *musacearum*, 3 genomas do patovar *raphani*, 1 genoma do patovar *thespesiae*, 1 genoma do patovar *viticola*, 1 genoma do patovar *vitiscarnosae*, 1 genoma do patovar *vitistrifoliae* e 1 genoma do patovar *vitiswoodrowii*. Na tabela 1 estão listados os 14 genomas selecionados pela métrica N50, os respectivos patovares, hospedeiros, códigos de acesso e referência.

**Tabela 1.** Lista dos genomas representativos de *Xanthomonas campestris*.

Patovar	Hospedeiro	Código de Acesso (Genbank)	Referência
<i>X. campestris</i> pv. <i>arecae</i>	<i>Areca catechu</i>	JSBP01000001.1	Não publicado
<i>X. campestris</i> pv. <i>azadiractae</i>	<i>Azadiracta indica</i>	LOKS01000001.1	Não publicado
<i>X. campestris</i> pv. <i>campestris</i>	Espécies de <i>Brassica</i>	AM920689.1	(Vorhölter et al. 2008)
<i>X. campestris</i> pv. <i>centellae</i>	<i>Centella asiatica</i>	LOJR01000001.1	Não publicado
<i>X. campestris</i> pv. <i>durantae</i>	<i>Duranta repens</i>	LOKP01000001.1	Não publicado
<i>X. campestris</i> pv. <i>incanae</i>	<i>Mathiola incana</i>	ATNN01000001.1	Não publicado
<i>X. campestris</i> pv. <i>leeana</i>	<i>Leea edgeworthii</i>	LOJY01000001.1	Não publicado
<i>X. campestris</i> pv. <i>musacearum</i>	Falsa banana	AKBE01000001.1	(Wasukira et al., 2012)
<i>X. campestris</i> pv. <i>raphani</i>	Rabanete ( <i>Raphanus sativus</i> )	CP002789.1	(Bogdaove et al., 2011)
<i>X. campestris</i> pv. <i>thespesiae</i>	<i>Thespesie populnea</i>	LOJU01000001.1	Não publicado
<i>X. campestris</i> pv. <i>viticola</i>	Uva ( <i>Vitis vinifera</i> L.)	NWTJ01000001.1	Não publicado

<i>X. campestris</i> pv. <i>vitiscarnosae</i>	<i>Vitis carnosa</i>	LOKI01000001.1	Não publicado
<i>X. campestris</i> pv. <i>Vitistrifoliae</i>	<i>Vitis trifolia</i>	LOKH01000001.1	Não publicado
<i>X. campestris</i> pv. <i>vitiswoodrowii</i>	<i>Vitis woodrowii</i>	LOKG01000001.1	Não publicado

## 5.2 Re-anotação estrutural de cada gene

A re-anotação estrutural dos genomas de *X. campestris* utilizando a ferramenta GENIX apresentou diferença no número de genes anotados, em comparação com o método de anotação utilizado pelos autores responsáveis pelo depósito do genoma no banco de dados do NCBI. Essa comparação pode ser visualizada na tabela 2.

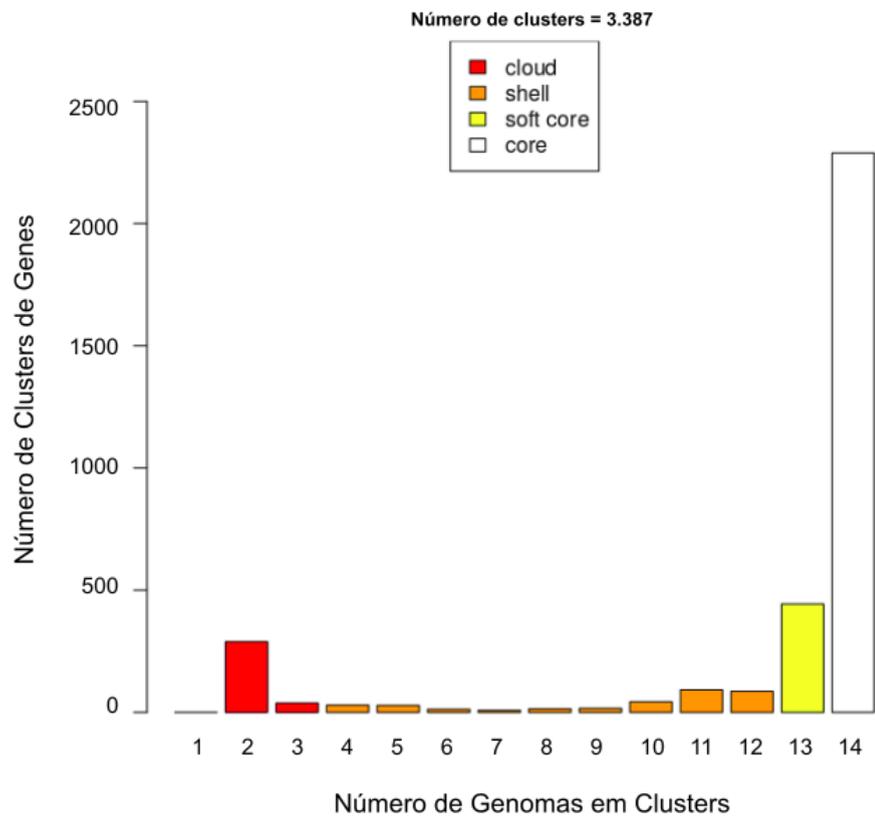
**Tabela 2.** Número de genes anotados pelo *software* Genix em comparação com a anotação disponível no NCBI para cada genoma dos patovares.

Patovar	Genes anotados por outros softwares	Genes anotados pelo Genix
<i>X. campestris</i> pv. <i>aracae</i>	4290	4129
<i>X. campestris</i> pv. <i>azadirachtae</i>	4498	4354
<i>X. campestris</i> pv. <i>campestris</i>	4453	4282
<i>X. campestris</i> pv. <i>centellae</i>	4473	4366
<i>X. campestris</i> pv. <i>durantae</i>	4672	4618
<i>X. campestris</i> pv. <i>incanae</i>	genoma não anotado	4098
<i>X. campestris</i> pv. <i>leeana</i>	4170	4062
<i>X. campestris</i> pv. <i>musacearum</i>	4060	4030
<i>X. campestris</i> pv. <i>raphani</i>	4520	4095
<i>X. campestris</i> pv. <i>therpesiae</i>	4154	4052
<i>X. campestris</i> pv. <i>viticola</i>	genoma não anotado	4212
<i>X. campestris</i> pv. <i>vitiscarnosae</i>	4257	4185
<i>X. campestris</i> pv. <i>vitistrifoliae</i>	4325	4227
<i>X. campestris</i> pv. <i>vitiswoodrowii</i>	4234	4147

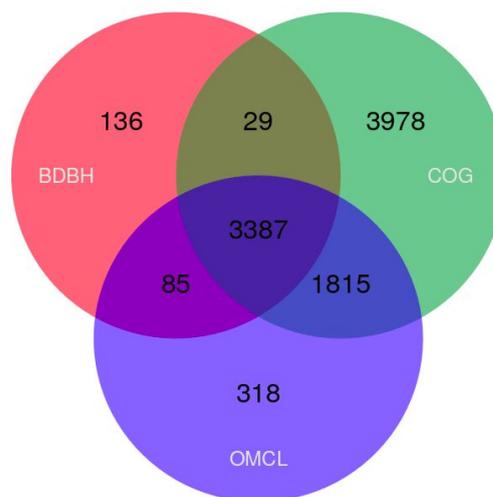
### 5.3 Análise dos grupos ortólogos

Para avaliar os grupos ortólogos, determinar o core genoma e o genoma acessório foram utilizadas três abordagens através dos softwares PanSeq, OrthoMCL e GET\_HOMOLOGUES. Os resultados do PanSeq revelaram regiões genômicas relacionadas ao pangenoma com tamanho de ~11 Megabases (Mb), enquanto as regiões genômicas conservadas entre as cepas, ou seja, o core genoma compreende o tamanho de 1.2 Mb.

A terceira análise, através do GET\_HOMOLOGUES, demonstrou um número de 3.387 grupos ortólogos consensos para os diferentes algoritmos utilizados pela ferramenta, sendo 5.605 identificados através do próprio OrthoMCL (reimplementação), 9.209 a partir do COG e 3.637 por meio do BDBH (Figura 7). Dos *clusters* de grupos ortólogos consenso, o core genoma apresenta 2.337 *cluster*, o *soft core* possui 493 *clusters*, ou seja, são *clusters* presentes em pelo menos 95% dos genomas, já o genoma *shell* possui 363 *clusters*, por fim o genoma *cloud* apresenta 194 *clusters* gênicos (Figura 6).



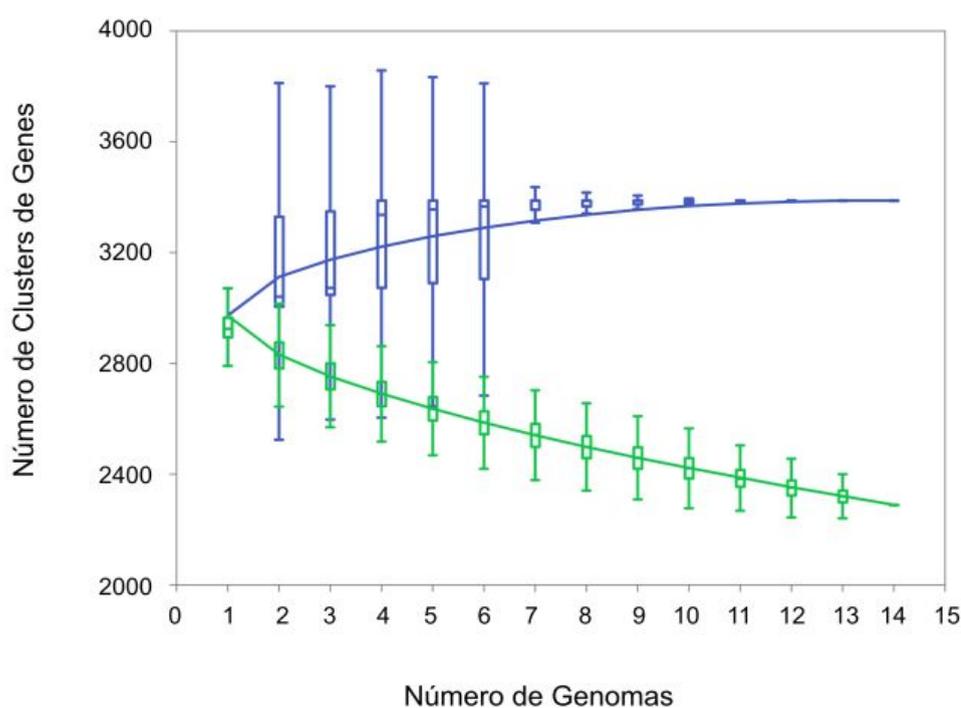
**Figura 6.** Gráfico representativo do número de clusters gênicos do pangenoma de *Xanthomonas campestris*.



**Figura 7.** Diagrama de venn dos grupos ortólogos do pangenoma de *Xanthomonas campestris*. Em rosa a predição dos grupos ortólogos pelo algoritmo BDBH, em verde os grupos ortólogos preditos pelo COG, em roxo os grupos ortólogos gerados pelo OrthoMCL.

## 5.4 Análise pangenômica estrutural

Os resultados gerados pela ferramenta PanGP são referentes aos 14 genomas analisados. Para o número de grupos ortólogos no pangenoma obteve-se ~2250 a 3900 grupos, já para o core genoma ~1750 a 2900 grupos. Sobre a característica do pangenoma, tem-se que esse estudo é do tipo pangenoma fechado (Figura 8).

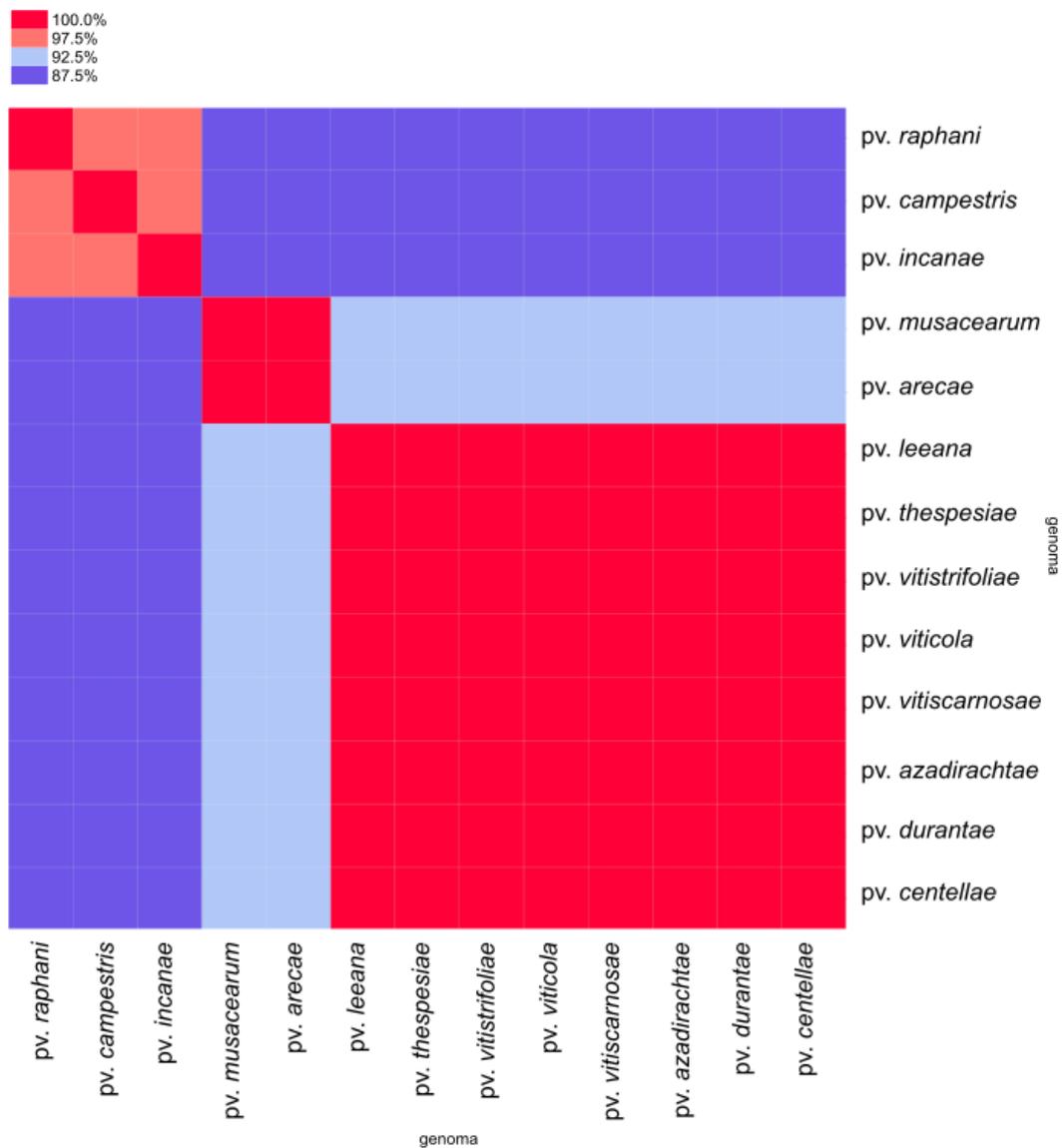


**Figura 8.** Gráfico descrevendo as características do pangenoma de *Xanthomonas campestris*. No eixo x está o número genomas utilizados, no eixo y o número de grupos ortólogos. Os dados em azul representam o pangenoma e os dados em verde representa o core genoma.

## 5.5 Análise de identidade de nucleotídeos

Com a análise de ANI foi possível demonstrar que entre os 14 genomas analisados há três grupos clusterizados por identidade, ou seja, clusters com genomas que apresentam entre 97% e 100% de identidade. Sendo os patovares *campestris*, *raphani* e *incanae* pertencentes a um cluster. Os patovares

*musacearum* e *arecae* pertencentes a outro cluster, e os demais patovares *azadirachtae*, *centellae*, *durantae*, *leeana*, *thespesiae*, *viticola*, *vitiscarnosae*, *vitistrifoliae*, e *vitiswoodrowii* pertencentes a um terceiro cluster (Figura 9).



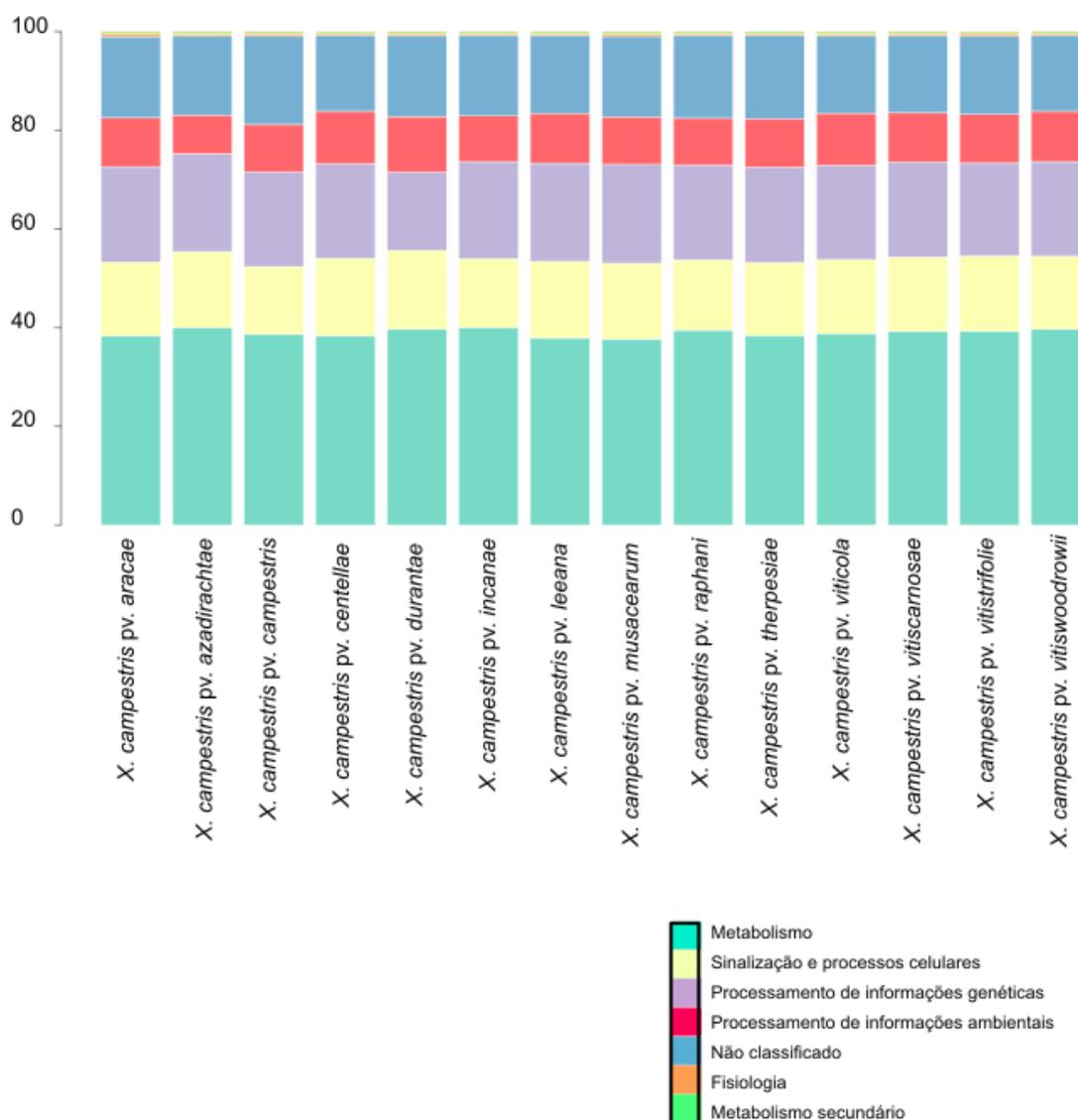
**Figura 9.** Heatmap da análise de identidade de nucleotídeos entre os genomas dos patovares de *Xanthomonas campestris*. Em vermelho os genomas com 100% de identidade, em vermelho claro os genomas com 95% de identidade, em azul escuro genomas com 87,5% de identidade, e em azul claro os genomas com 92.5% de identidade.

## 5.6 Anotação de rotas metabólicas

A análise de metabolismo utilizando a ferramenta BlastKoala possibilitou identificar rotas metabólicas nos 14 genomas analisados. A tabela 3 mostra a porcentagem do genoma de cada patovar anotado pela ferramenta, bem como o número de rotas metabólicas. As rotas anotadas foram: rotas de sinalização e processos celulares, processamento de informações do ambiente (resposta a alterações ambientais), processamento de informações genéticas (tradução, transcrição e replicação), metabolismo (rotas de catabolismo e anabolismo), fisiologia (manutenção da homeostase), metabolismo secundário e doenças humanas (Figura 10).

**Tabela 3.** Cobertura da anotação metabólica pelo *software* BlastKOALA para os genomas de *Xanthomonas campestris*.

Patovar	Cobertura do genoma	Número de anotações
<i>X. campestris</i> pv. <i>aracae</i>	52.3%	2.158
<i>X. campestris</i> pv. <i>azadirachtae</i>	51.5%	2.245
<i>X. campestris</i> pv. <i>campestris</i>	52.1%	2.232
<i>X. campestris</i> pv. <i>centellae</i>	51.7%	2.259
<i>X. campestris</i> pv. <i>durantae</i>	50.3%	2.323
<i>X. campestris</i> pv. <i>incanae</i>	53.2%	2.179
<i>X. campestris</i> pv. <i>leana</i>	54.0%	2.194
<i>X. campestris</i> pv. <i>musacearum</i>	53.3%	2.146
<i>X. campestris</i> pv. <i>raphani</i>	53.6%	2.193
<i>X. campestris</i> pv. <i>therpesiae</i>	54.0%	2.187
<i>X. campestris</i> pv. <i>viticola</i>	53.0%	2.273
<i>X. campestris</i> pv. <i>vitiscarnosae</i>	53.1%	2.219
<i>X. campestris</i> pv. <i>vitistrifoliae</i>	53.6%	2.222

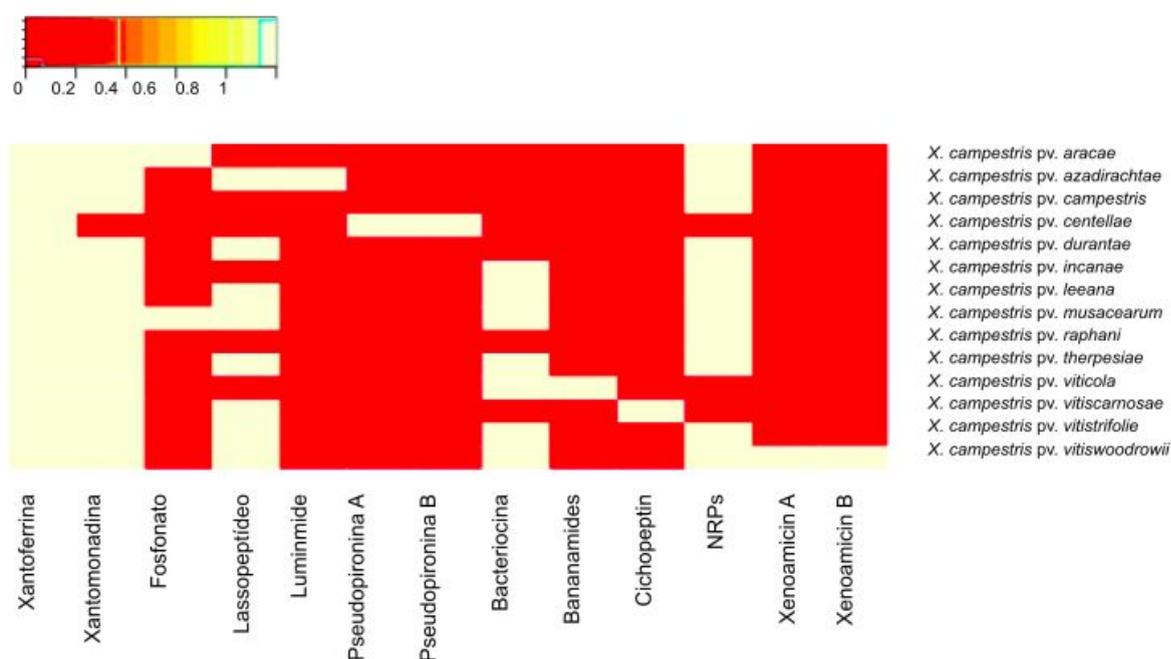


**Figura 10.** Gráfico das rotas metabólicas previstas pelo software BlastKoala para cada patovar de *Xanthomonas campestris*. As rotas metabólicas previstas pelo software BlastKoala foram relacionadas a processos celulares, processamento de informações do ambiente, processamento de informações genéticas, metabolismo, não classificado e fisiologia.

## 5.7 Análise de metabólitos secundários

A análise dos metabólitos secundários utilizando a ferramenta AntiSMASH possibilitou a identificação dos seguintes metabólitos: xantoferrina (ID: BG0001408), xantomonadina (ID: BG0000840), fosfanato (ID: não possui),

lassopeptídeo (ID: não possui), luminmide (ID: não possui), pseudopironina A (ID: BGC0001285), pseudopironina B (ID: BGC0001285), bacteriocina (ID: não possui), bananamides (ID: não possui), cichopectin (ID: não possui), nrps (ID: BGC0001128), xenoamicin A (ID: não possui), xenoamicin B (ID: não possui). A distribuição desses metabólitos nos genomas dos patovares está representada na figura 11, na qual é possível visualizar a presença de xantoferrina em todos os patovares, bem como a presença de xantomonadina em todos os patovares exceto no patovar *centellae*. O fosfonato está presente somente nos genomas dos patovares *leeana* e *aracae*. O metabólito lassopeptide está presente nos genomas de *azadirachtae*, *durantae*, *leeana*, *musacearum*, *thespesiae*, *vitiscarnosae*, *vitistrifoliae* e *vitiswoodrowii*. O metabólito luminmide se mostrou presente somente no genoma de *azadirachtae*. Os metabólitos pseudopironina A e pseudopironina B estão presentes somente no genoma do patovar *centellae*. A bacteriocina se mostrou presente nos genomas de *incanae*, *leeana*, *musacearum*, *thespesiae*, *viticola*, *vitistrifoliae* e *vitiswoodrowii*. Já o metabólito bananamides está presente somente do genoma do patovar *viticola*. O metabólito cichopectin está presente somente no genoma do patovar *vitiscarnosae*. O cluster de peptídeo sintetase não ribossômico (NRPS) está presente em todos os genomas exceto nos patovares *centellae*, *viticola* e *vitiscarnosae*. Os metabólitos xenoamicin A e xenoamicin B estão presentes somente no patovar *vitiswoodrowii*.



**Figura 11.** Heatmap dos metabólitos secundários preditos pelo AntiSMASH para os genomas dos patovares de *Xanthomonas campestris*. No eixo x encontram-se os metabólitos preditos, e no eixo Y os patovares de *X. campestris*. A escala do heatmap representa a presença e ausência do metabólito no genoma do patovar. Os metabólitos representados em vermelho indicam a ausência do mesmo no genoma, já a cor rosa clara indica a presença do metabólito no genoma.

## 5.8 Predição dos efetores TAL

A predição dos efetores TAL nos genomas de *X. campestris*, por meio do software TargetALE, demonstrou que somente o patovar *X. campestris* pv. *campestris*, de acesso AM920689.1, apresentou TALEs em seu genoma. Sendo dois TALEs preditos para esse patovar, o primeiro TALE (TALES1) com início no nucleotídeo 2595972 e fim no nucleotídeo de número 2599383; o segundo TALE (TALES2) predito tem início no nucleotídeo de número 2572623 e fim no nucleotídeo de número 2576157 do genoma. Para os demais genomas não foi possível prever TALEs. Além disso, foi possível prever quais os genes de *Brassica oleracea* são alvos dos TALEs preditos (Tabela 4). Sendo o TALEs1 o único que houve predição de genes alvos.

**Tabela 4.** Relação dos genes alvos de TALEs preditos pelo *software* TargeTALE.

TALE	Gene Alvo	Função da proteína codificada pelo gene alvo	Código de Acesso (Uniprot)	Organismo alvo
TALES1	Não caracterizado	Ligação de íons cálcio	A0A0D3A649	<i>Brassica oleracea</i> var. <i>oleracea</i>
TALES1	Não caracterizado	Não caracterizada	A0A0D3BTC5	<i>Brassica oleracea</i> var. <i>oleracea</i>
TALES1	Não caracterizado	Não caracterizada	A0A0D3CE74	<i>Brassica oleracea</i> var. <i>oleracea</i>
TALES1	106343852	Protease específica da ubiquitina dependente de tiol	A0A0D3CGL3	<i>Brassica oleracea</i> var. <i>oleracea</i>
TALES1	Não caracterizado	Não caracterizada	A0A0D3E5T7	<i>Brassica oleracea</i> var. <i>oleracea</i>
TALES1	Não caracterizado	Atividade de ribonuclease híbrida RNA-DNA	A0A0D3A8Q7	<i>Brassica oleracea</i> var. <i>oleracea</i>
TALES1	106332693	Regulação positiva do ritmo circadiano	A0A0D3B5Z2	<i>Brassica oleracea</i> var. <i>oleracea</i>
TALES1	Não caracterizado	Não caracterizada	A0A0D3C3Y2	<i>Brassica oleracea</i> var. <i>oleracea</i>

## 5.9 Análise dos genes de biossíntese de xantana

Os genes relacionados à biossíntese de xantana em *X. campestris* resultantes da revisão da literatura estão listados na tabela 5, com seus respectivos operons, funções, códigos de acesso e referência.

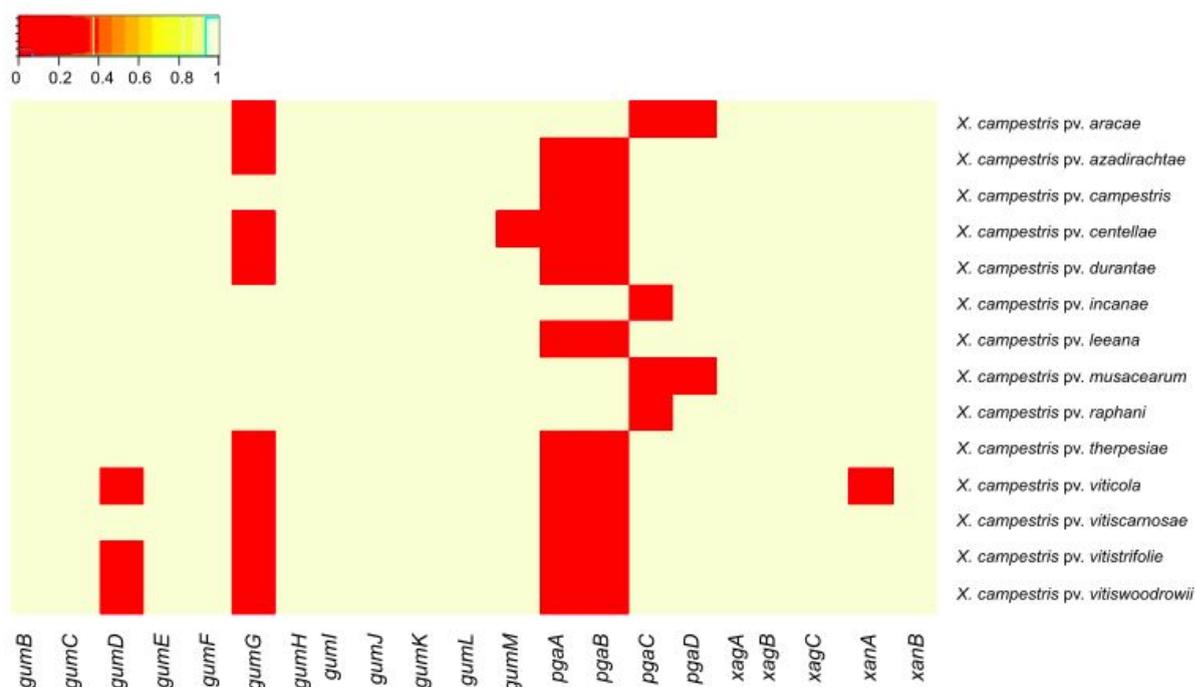
**Tabela 5.** Relação dos genes da rota de biossíntese de xantana em *Xanthomonas campestris*. Seus respectivos operons, genes, funções e códigos de acesso do banco de dados Uniprot.

Operon/Gene	Nome do gene	Função da Proteína	Código de Acesso (Uniprot)	Referência
	<i>gumB</i>	Proteína de membrana associada ao transporte de polissacarídeo	Q93A84	(Vojnov et al., 2001)
	<i>gumC</i>	Biossíntese de lipopolissacarídeo	Q8P7Z8	(Vojnov et al., 2001)
	<i>gumD</i>	Enzima catalisadora da produção de glucosil-pirofosforil- <i>u</i> ndecaprenol	Q56770	(Vojnov et al., 2001)

<b>Operon <i>gum</i></b>	<i>gumE</i>	Biossíntese de lipopolissacarídeo	Q8P800	(Vojnov et al., 2001)
	<i>gumF</i>	Atividade de transferase, transferência de grupos acila, exceto grupos amino-acil	Q8P801	(Vojnov et al., 2001)
	<i>gumG</i>	Atividade de transferase, transferência de grupos acila, exceto grupos amino-acil	Q8P802	(Vojnov et al., 2001)
	<i>gumH</i>	Atividade de transferase, transferência de grupos glicosil	Q56774	(Vojnov et al., 2001)
	<i>gumI</i>	Enzima que catalisa a formação do polissacarídeo xantana	Q56775	(Vojnov et al., 2001)
	<i>gumJ</i>	Biossíntese de lipopolissacarídeo	Q8P805	(Vojnov et al., 2001)
	<i>gumK</i>	Enzima glicolípídeo 2-beta-glucuronosiltransferase	Q8GCH2	(Vojnov et al., 2001)
	<i>gumL</i>	Atividade de transferase	Q3BRV5	(Vojnov et al., 2001)
	<i>gumM</i>	Atividade transferase, transferindo grupos hexosil	Q8P807	(Vojnov et al., 2001)
<b>Operon <i>pga</i></b>	<i>pgaA</i>	Atividade transportadora transmembranar de poli-beta-1,6-N-acetil-D-glucosamina	A0A3Q9H6J7	(Wang et al., 204)
	<i>pgaB</i>	Atividade da hidrolase, atuando nas ligações carbono-nitrogênio (mas não peptídicas)	A0A3Q9H2Y2	(Wang et al., 204)
	<i>pgaC</i>	Formação de biofilme e aderência celular	A0A2H1T835	(Wang et al., 204)
	<i>pgaD</i>	Formação de biofilme e aderência celular	A0A1T1S0G6	(Wang et al., 204)
<b>Operon <i>xag</i></b>	<i>xagA</i>	Formação de biofilme e aderência celular	B0RVF3	(An et al., 2019)
	<i>xagB</i>	Atividade transferase, transferindo grupos glicosil	B0RVF2	(An et al., 2019)

	<i>xagC</i>	Formação de biofilme e aderência celular	B0RVF1	(An et al., 2019)
<b>Gene</b>	<i>xanA</i>	Atividade de fosfohexose mutase	B0RVK5	(Katzen et al., 1996)
<b>Gene</b>	<i>xanB</i>	Biossíntese de xantana	B0RVK6	(Katzen et al., 1996)

O alinhamento da sequência dos genes da tabela 5 com os genomas de cada patovar está representado na figura 12, na qual é possível visualizar os genes *gumB*, *gumC*, *gumE*, *gumF*, *gumH*, *gumI*, *gumJ*, *gumK*, *gumL*, *xagA*, *xagB*, *xagC* e *xanB*, os quais estão presentes em todos os genomas dos patovares. O gene *gumD* está presente em todos os genomas exceto nos genomas de *viticola*, *vitistrifoliae* e *vitiswoodrowii*. O gene *gumG* está presente nos genomas de *campestris*, *incanae*, *leeana*, *musacearum* e *raphani*. O gene *gumM* está presente em todos os genomas, exceto no patovar *centellae*. Os genes *pgaA* e *pgaB* estão presentes somente nos genomas dos patovares *aracae*, *incanae*, *musacearum* e *raphani*. O gene *pgaC* está presente em todos os genomas, exceto nos genomas de *aracae*, *incanae*, *musacearum* e *raphani*. Já o gene *pgaD* está presente em todos os genomas, exceto nos patovares *aracae* e *musacearum*. Por fim, o gene *xanA* está presente em todos os patovares, exceto no genoma de *viticola*.



**Figura 12.** Heatmap dos genes de biossíntese de xantana presentes nos genomas dos patovares de *Xanthomonas campestris*. No eixo x encontram-se os genes relacionados com a biossíntese da xantana, e no eixo Y os patovares de *X. campestris*. A escala do heatmap representa a presença e ausência do gene no genoma do patovar. Os genes representados em vermelho indicam a ausência do mesmo no genoma, já a cor rosa clara indica a presença do gene no genoma.

### 5.10 Análise dos genes de virulência

Os genes de virulência de *X. campestris* resultantes da revisão da literatura estão listados na tabela 6, com suas respectivas funções, códigos de acesso e referência.

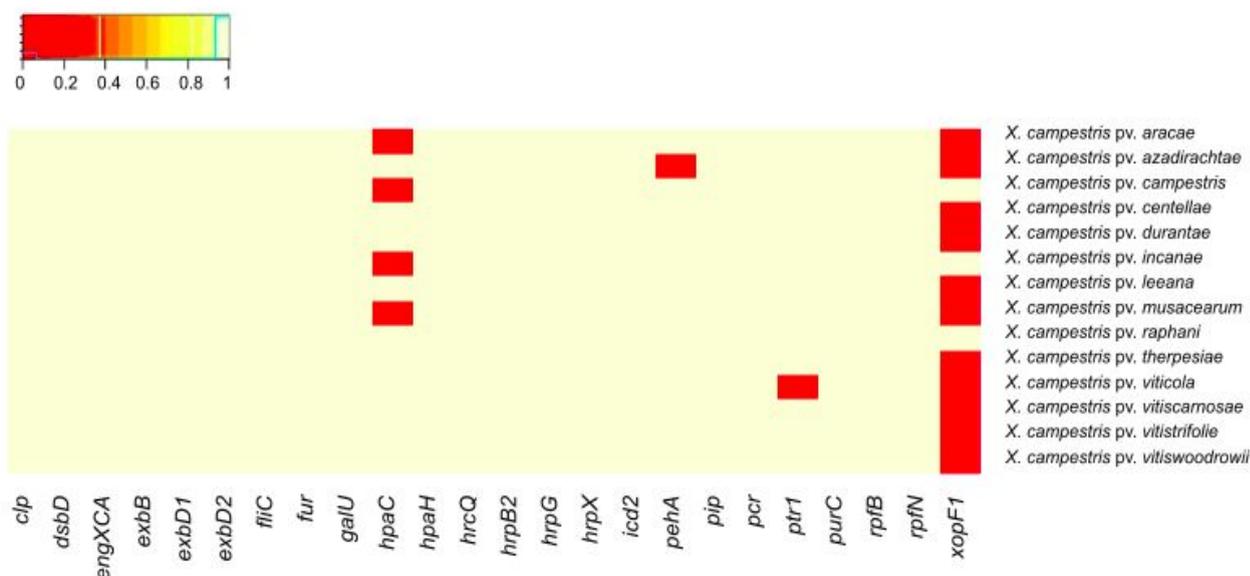
**Tabela 6.** Genes de virulência presentes nos genomas de *X. campestris*.

Gene	Função	Código de acesso (Uniprot)	Referência
<i>clp</i>	Regulador de transcrição global de fatores de virulência	CLP_XANCP	(Lo et al., 2019)

<b><i>dsbD</i></b>	Relacionado com a biogênese do citocromo c-type e relação com a virulência	B0RN48_XANCB	(Chen et al., 2017)
<b><i>engXCA</i></b>	Endonuclease extracelular	GUNA_XANCP	(He & Zhang, 2008)
<b><i>exbB</i></b>	Sistema de aquisição de ferro férrico dependente de TonB	EXBB_XANCP	(Wiggerich & Puhler, 2000)
<b><i>exbD1</i></b>	Sistema de aquisição de ferro férrico dependente de TonB	EXBD1_XANCB	(Wiggerich & Puhler, 2000)
<b><i>exbD2</i></b>	Sistema de aquisição de ferro férrico dependente de TonB	EXBD2_XANCB	(Wiggerich & Puhler, 2000)
<b><i>fliC</i></b>	flagelina	Q3BU10_XANC5	(Yu et al., 2014)
<b><i>fur</i></b>	Metabolismo do ferro e funções associadas à virulência, além do fator de transcrição Fur-like	Q8PAL3_XANCP	(Liu et al., 2016)
<b><i>galU</i></b>	Desempenha um papel na ligação bacteriana, motilidade celular, síntese de ePS e LPS, patogenicidade e tolerância ao estresse	Q8P8Q1_XANCP	(Liao et al., 2014)
<b><i>hpaC</i></b>	Transloca proteínas efectoras	Q3BYK8_XANC5	(Hausner & Buttner, 2014)
<b><i>hpaH</i></b>	Desempenha o papel de glicosil hidrolases	Q3BYJ1_XANC5	(Hausner et al., 2017)
<b><i>hrcQ</i></b>	Proteína secretada pelo Sistema de Secreção do Tipo III	Q8PB96_XANCP	(Hausner & Buttner, 2014)
<b><i>hrpB2</i></b>	Transportador de prótons	A0A1B0UHJ8_XANCA	(Hausner & Buttner, 2014)
<b><i>hrpG</i></b>	Regulação de Genes e patogenicidade	Q3BW18_XANC5	(Rashid et al., 2016)
<b><i>hrpX</i></b>	Sensor fosforilado de kinase	Q8P9A6_XANCP	(Rashid et al., 2016)
<b><i>icd2</i></b>	Codifica uma desidrogenase de isocitrato funcional	B0RWU3_XANCB	(Chiang et al., 2017)
<b><i>pehA</i></b>	Gene de poligalacturonase	A0A1Y6HAX6_9XANT	
<b><i>pip</i></b>	Gene de virulência proline iminopeptidase	Q8PC98_XANCP	(Santos et al., 2017)
<b><i>prc</i></b>	Protease específica Tail	Q8P591_XANCP	(Liao et al., 2016)
<b><i>ptr1</i></b>	Putativa pteridina redutase	B0RNR9_XANCB	(Dow et al., 1990)
<b><i>purC</i></b>	Gene de virulência	PUR7_XANCP	(Yuan et al., 2013)

<b><i>rpfB</i></b>	Regulador de patogenicidade	Q3BUB1_XANC5	(Noh et al., 2014)
<b><i>rpfN</i></b>	Causou a repressão da produção de enzimas e polissacarídeos	B0RRR6_XANCB	(Noh et al., 2014)
<b><i>xopF1</i></b>	Proteína Efetora Tipo III	G0CIQ6_XANCA	(Roux et al., 2015)

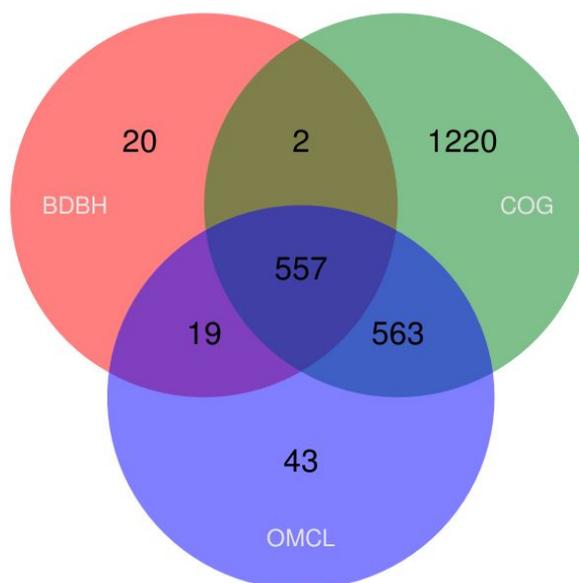
O alinhamento da sequência dos genes da tabela 6 com os genomas de cada patovar está representado na figura 13, na qual é possível visualizar os genes *clp*, *dsbD*, *engXCA*, *exbB*, *exbD1*, *exbD2*, *flic*, *fur*, *galU*, *hpaH*, *hrcQ*, *hrpB2*, *hrpG*, *hrpX*, *icd2*, *pip*, *pcr*, *purC*, *rpfB* e *rpfN*, os quais estão presentes em todos os genomas dos patovares. O gene *hpaC* está presente em todos os genomas, exceto nos genomas dos patovares *aracae*, *campestris*, *incanae* e *musacearum*. O gene *pehA* está presente em todos os genomas, exceto no patovar *azadirachtae*. Já o gene *ptr1* está presente em todos os genomas, exceto no patovar *viticola*. Por fim, o gene *xopF1* está presente somente nos genomas dos patovares *campestris*, *incanae* e *raphani*.



**Figura 13.** Heatmap dos genes de virulência presentes nos genomas dos patovares de *Xanthomonas campestris*. No eixo x encontram-se os genes de virulência, e no eixo Y os patovares de *X. campestris*. A escala do heatmap representa a presença e ausência do gene no genoma do patovar. Os genes representados em vermelho indicam a ausência dele no genoma, já a cor rosa clara indica a presença do gene no genoma.

### 5.11 Predição de proteínas efetoras

A predição de proteínas efetoras nos genomas de *X. campestris* resultou num total de 557 clusters de proteínas consenso entre os três algoritmos: BDBH, OrthoMCL e COG (Figura 14). Desses, 172 clusters estão presentes nos 14 genomas analisados, 120 clusters em pelo menos 2 patovares e não foi encontrado nenhum cluster 40 exclusivo a um único patovar (Anexo D).



**Figura 14.** Diagrama de venn das proteínas efetoras preditas para o *core* genoma de *Xanthomonas campestris*. Em rosa estão representados os clusters de proteínas preditos pelo algoritmo do BDBH. Em verde são os clusters preditos pelo algoritmo do COG. Em roxo os clusters preditos pelo OrthoMCL. Sendo o valor 557 os clusters consensos dos três programas.

### 5.12 Análise dos genes de DSF

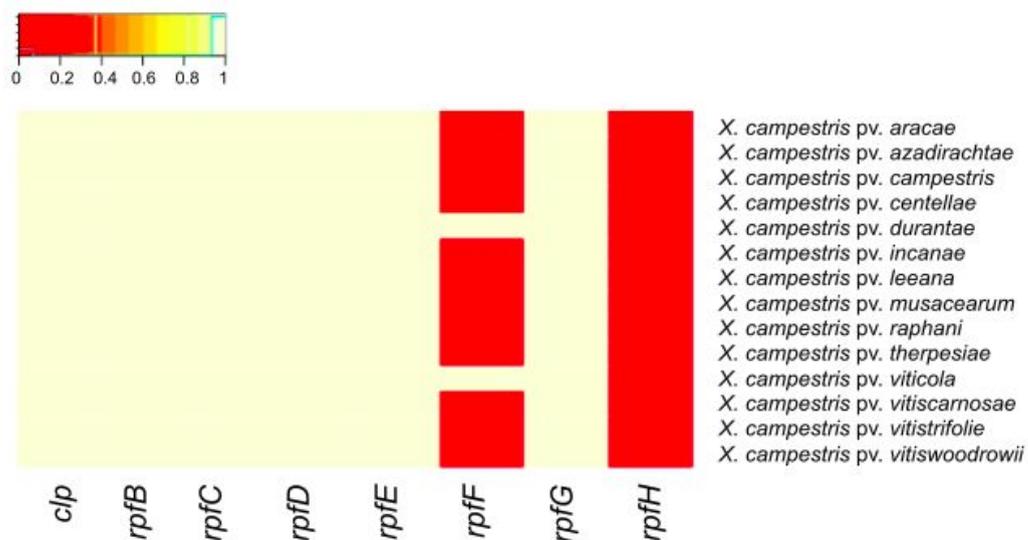
Os genes relacionados ao DSF em *X. campestris* resultantes da revisão da literatura estão listados na tabela 7, com suas respectivas funções, códigos de acesso e referência.

**Tabela 7.** Relação dos genes de DSF em *Xanthomonas campestris*.

Gene	Função	Código de Acesso (Uniprot)	Referência
<i>clp</i>	Regulador de transcrição, ativando ou reprimindo a expressão de um grande conjunto de genes na via de DSF	CLP_XANCP	(Liao <i>et al.</i> , 2016)
<i>rpfB</i>		O05174_XANCA	(Gudesblast <i>et al.</i> , 2009)
	Membro do sistema		

<b><i>rpfC</i></b>	regulador RpfG/RpfC, necessário para virulência e formação de biofilmes. Envolvido na detecção e resposta ao DSF	RPFC_XANC8	(O'Connell et al., 2013)
<b><i>rpfD</i></b>	Não caracterizada	Q9L4E0	(Cho et al., 2013)
<b><i>rpfE</i></b>	Não caracterizada	Q9L4D4	(Cho et al., 2013)
<b><i>rpfF</i></b>	Não caractrizada	Q7CLS3	(Cho et al., 2013)
<b><i>rpfG</i></b>	O RpfG ativado degrada o di-GMP cíclico em GMP, levando à ativação do gene <i>clp</i>	Q4UU85	(Rott et al., 2013)
<b><i>rpfH</i></b>	Não caracterizada	B0RSR0	(Rott et al., 2013)

O alinhamento da sequência dos genes da tabela X com os genomas de cada patovar está representado na figura 15, na qual é possível visualizar os genes *clp*, *rpfB*, *rpfC*, *rpfD*, *rpfE* e *rpfG*, os quais estão presentes em todos os 14 genomas. Já o gene *rpfF* está somente nos genomas dos patovares *centellae* e *viticola*. Por fim, o gene *rpfH* não está presente em nenhum genoma.



**Figura 15.** Heatmap dos genes de DSF presentes nos genomas dos patovares de *Xanthomonas campestris*. No eixo x encontram-se os genes envolvidos no DSF, e no eixo Y os patovares de *X. campestris*. A escala do heatmap representa a presença e ausência do gene no genoma do patovar. Os genes representados em vermelho indicam a ausência do mesmo no genoma, já a cor rosa clara indica a presença do gene no genoma.

## 6 DISCUSSÃO

Utilizando o *software* GET\_HOMOLOGUES foi possível obter o *core*, o *soft*, o *shell* e o *cloud* genoma para os patovares de *X. campestris*, conforme referido acima. Os resultados vão de encontro com o esperado, sendo os resultados obtidos dependentes, portanto, dos parâmetros considerados. O método utilizado demonstrou que o *core* genoma (Figura 6), equivalente a aproximadamente 70%. Cerca de 6% dos *clusters* gênicos corresponde ao *cloud* genoma, ou seja, a *clusters* de genes presentes em somente um genoma. Essa distribuição de pangenoma vai de encontro com os resultados obtidos por Policarpo, (2015) para sorotipos de *Streptococcus pneumoniae*.

A análise da classificação do pangenoma de *X. campestris* utilizando o programa PanGP revelou que este é do tipo fechado, uma vez que o gráfico gerado (Figura 8) demonstrou que a adição de mais genomas à presente análise não contribuiria para a aquisição de novos genes (Tettelin *et al.*, 2005; Muzzi *et al.*, 2007). Em outras palavras, para 14 genomas utilizados na análise do pangenoma, até o genoma 4 houve um aumento no número de grupos ortólogos no *core* genoma, a partir do genoma 5 houve um decréscimo, ou seja, a presença de mais genomas não contribui para a determinação do genoma núcleo de *X. campestris*.

Na comparação de identidade de nucleotídeo, por meio do cálculo de ANI, foi possível identificar três clusterizações distintas (Figura 9). Sendo os patovares *campestris*, *raphani* e *incanae* pertencentes a um cluster. Esse resultado vai de encontro com os achados genômicos de Roux *et al.* (2015). As demais clusterizações não foram reportadas na literatura, fato esse destaca a necessidade da mineração dos dados genômicos depositados a fim de melhor caracterizar os patovares pertencentes à espécie *X. campestris*.

Por fim, podemos considerar que GET\_HOMOLOGUES é um método eficiente para a obtenção de um pangenoma para espécies bacterianas. Os resultados obtidos podem contribuir para outros estudos, como por exemplo, para análises de funções de genes presentes no *core*, análise filogenética, bem como para comparação de patovares e suas características de patogenicidade e virulência, sendo este último caso analisado neste trabalho.

O rápido desenvolvimento da genômica trouxe uma mudança de paradigma para a pesquisa de patogênese bacteriana. A análise genômica comparativa foi iniciada estudando organismos filogeneticamente distantes; por outro lado, a comparação intra-espécies de várias patovares está se tornando uma das novas tendências para o estudo de patógenos bacterianos (Whittam & Bumbaugh, 2002). Além disso, anotações completas usando tecnologias genômicas comparativas e funcionais são essenciais para obter uma compreensão sistemática da biologia das doenças causadas pelos patovares de *X. campestris*, bem como a elucidação de rotas metabólicas conservadas. As informações obtidas a partir dessa compreensão podem revelar, sistematicamente, as variações genéticas e organização do genoma entre diferentes patovares bacterianos. Bem como, fornece informações sobre a natureza molecular da virulência e especificidade ao hospedeiro. Nesse contexto, a genética molecular extensiva de *X. campestris* identificou os principais determinantes da patogenicidade e fatores de virulência, dentre os quais se destacam os exopolissacarídeos (EPS), sistemas de translocação de proteínas e proteínas efetoras (Qian *et al.*, 2005).

A comparação das predições de rotas metabólicas para os genomas de *X. campestris*, presente na Figura 10 baseou-se no banco de dados KEGG *pathways* usando a ferramenta BlastKOALA. Os resultados apontam para uma diferença nos números de predições para rotas relacionadas a processamento de informação genética. Além disso, majoritariamente as predições foram referentes as rotas de metabolismo celular, sendo essa predição também distribuída diferencialmente entre os patovares. Esse fato pode indicar que nessas duas rotas é onde se

encontram a maioria dos genes exclusivos de cada isolado, bem como genes associados a patogenicidade e especificidade ao hospedeiro. Fato esse, também observado por Meng et al. (2017), através da análise pangenômica de *Streptococcus mutans*.

Os genomas de *X. campestris* foram analisados usando o *software* AntiSMASH para identificar clusters de genes e vias metabólicas associadas à biossíntese de metabólitos secundário. Baseado nos resultados gerados, foi possível identificar nos genomas analisados *clusters* de genes responsáveis pela biossíntese de xantoferrina, xantomonadina, fosfonato, lassopeptídeo, pseudopironina A e B, bactericina, bananamides, cichopectin, NRPs e xenoamicina A e B. A xantoferrina é sideróforo importante para o crescimento celular em condições com baixa concentração de ferro, sendo um fator relevante para a virulência de *X. campestris* (Pandey et al., 2016). Xantomonadina é produzida pela maioria das bactérias do gênero *Xanthomonas*, ela é responsável pelo pigmento amarelo das colônias além de conferir proteção ao DNA dos raios ultravioletas, fato esse que permite a habilidade de sobreviver no ambiente (He et al., 2011). O fosfonato e a bactericina desempenham um papel importante na inibição do crescimento de outros microrganismos pelas bactérias do gênero *Xanthomonas*, através da sua atividade bacteriostática e bactericida (Montesinos & Vilardell, 2001). Os lassopeptídeos são uma classe de peptídeos modificados que apresentam função antimicrobial e antiviral (Cheung-Lee et al., 2019). A função da pseudopironina A e B em *Xanthomonas* ainda não foi elucidada na literatura. As bananamides ainda não foram elucidadas em *Xanthomonas*, em contrapartida em *Pseudomonas* elas demonstraram desempenhar atividade antimicrobial (Omoboye et al., 2019). Por fim, a xenoamicina é um composto com atividade antiparasitária, no entanto não há a caracterização desse composto em *Xanthomonas* (Zhou et al., 2015).

A análise de efetores TAL demonstrou que somente o patovar *X. campestris* pv. *Campestris* apresentou esse efetor em seu genoma. Além disso, foi possível

predizer, pra esse patovar, os respectivos alvos no hospeiro, nos quais os efetores TAL possivelmente atuariam. Esse fato demonstra a adaptação do patovar a diferentes ambientes e genótipos de hospedeiro. No entanto para os outros patovares não foi possível identificar efetores TAL, tão pouco predizer seus alvos. Esse fato pode estar relacionado à ausência das sequências codificantes de TALEs nos genomas, isso se dá, pois, essas sequências têm como característica a repetição em *tandem* de aminoácidos e alguns montadores de genomas têm em seu algoritmo a função de excluir regiões repetitivas geradas pelo sequenciamento.

*X. campestris* possui como polissacarídeo extracelular a xantana. A xantana é um polímero de unidades repetidas de pentassacarídeos com a estrutura manose-( $\beta$ -1,4)-glucurônica ácido-( $\beta$ -1,2)-manose-( $\beta$ -1,3)-celobiose. Os resultados da análise dos genes presentes nos genomas dos patovares de *X. campestris*, referentes à biossíntese da xantana nos revelaram que os genes *gumBCEFHIJKLM* e os genes *xagA*, *xagB*, *xagC*, e *xanB* estão presentes em todos os patovares, inferindo que todas as cepas são capazes de produzir goma xantana, além de produzir, provavelmente, com a mesma eficiência, uma vez que Katzen *et al.* (1996) e Vojno *et al.* (2001) demonstram que os genes do operon *gum* estão envolvidos na biossíntese de xantana. Além disso análises transcricionais revelaram que a região dos genes *gum* é expressa, principalmente, como um operon de 16kb, tendo o promotor localizado a montante de *gumB*.

Em relação aos genes *xagABC*, Tao *et al.*, (2010) demonstraram que eles são importantes na regulação da produção, e na ausência desses, a biossíntese de xantana apresentou menor rendimento. A ausência do gene *gumD* nos patovares *X. campestris* pv. *viticola*, *X. campestris* pv. *vitistrifolie* e em *X. campestris* pv. *vitiswoodrowii* pode indicar uma deficiência no processo de polimerização, uma vez que a biossíntese da xantana ocorre na membrana celular, onde o polímero é construído a partir de glicosiltransferases, e os genes envolvidos nesse processo de construção polimérica são os genes *gumD*, *gumM*, *gumH*, *gumK* e *gumI* (Ielpi *et al.*, 1993; Stankowski *et al.*, 1993). Além desses genes, a biossíntese da xantana

depende dos precursores de açúcar de nucleotídeo UDP-glucose, de GDP-manose e UDP-glucuronato.

Os genes *xanA* e *xanB* codificam as enzimas que catalisa a maioria dos passos dentro das vias de síntese de UDP-glucose e GDP-manose (Lin et al., 1995). Com isso, tem-se que a ausência do gene *xanA* no patovar *X. campestris* pv. *viticola* sugere uma deficiência no processo de catálise nas vias de síntese de UDP-glucose e GDP-manose. Os genes *pgaABCD* foram reportados por Wang; Preston & Romeo (2004) como fundamentais para a atividade de formação de biofilme da xantana. Além disso, análises funcionais demonstraram que a expressão conjunta dos três genes *gumD*, *xagB* e *pgaC* são importantes para a formação de biofilme em *X. campestris*, sendo a formação de biofilme importante na patogenicidade, uma vez que está relacionado com a colonização da bactéria no hospedeiro. Em vista disso, os patovares que apresentam em seus genomas os três genes são *X. campestris* pv. *azadirachtae*, *X. campestris* pv. *campestris*, *X. campestris* pv. *centellae*, *X. campestris* pv. *durantae*, *X. campestris* pv. *leena* e *X. campestris* pv. *thespesiae*, isso indica que, possivelmente, esses patovares formam a estrutura de biofilme.

A virulência de *X. campestris* depende de vários fatores, incluindo a síntese de enzimas extracelulares e a síntese do polissacarídeo extracelular xantano. A síntese desses diversos fatores por *X. campestris* são positivamente controlados por diversos genes. Nossos resultados da análise de genes de virulência demonstraram que todos patovares apresentam os genes *dsbD*, o qual atua como uma redutase no espaço periplásmico e, também, participa na biossíntese dos citocromos-c. Chen, et al. (2017) demonstrou que a ausência desse gene afeta a respiração celular de *X. campestris* por perda de síntese dos citocromos-c, logo houve crescimento mais lento, utilizando carboidratos pobres, como galactose, xilose, maltose, sacarose e glicose. Conseqüentemente, houve a diminuição da atividade enzimática, e uma menor produção de EPS.

Os genes *engXCA*, *exbB*, *exbD1* e *exbD2* também, estão presentes em todos patovares. Sabe-se que o gene *engXCA* codifica para uma endoglucanase extracelular ENGXCA, sendo ela importante no processo de patogenia (Gouch et al., 1990). Já *exbB*, *exbD1* e *exbD2* foram descritos por Wiggerich & Pühler (2000) como codificadores de proteína auxiliares que formam o sistema TomB, o qual faz a incorporação do ferro férrico envolvido no mecanismo de resposta hipersensitiva e de patogenicidade. Esse mecanismo é caracterizado pela rápida necrose do tecido do hospedeiro, sintoma, esse, característico de infecções por *X. campestris*. O gene *fliC*, presente em todos patovares de *X. campestris*, codifica para flagelina, a qual é essencial para a formação do flagelo, estrutura que além de conferir motilidade à bactéria, também é um fator de virulência (Yu et al., 2014). O *galU* é um gene que desempenha a função de aderência da célula bacteriana, motilidade e síntese de LPS (Liao et al., 2014). Já os genes *hrpX*, *hrpB2* e *hrpG* fazem parte do cluster *hrp* e regulam fatores de virulência como o T3SS (Rashid et al., 2016).

Chiang et al., (2017) retratou o gene *icd2* em *X. campestris*, demonstrando que mutantes com ausência do mesmo diminuíram sua virulência, no entanto ainda é desconhecido o papel desse gene na virulência de *Xanthomonas*. O *hpaC* se mostrou ausente nos patovares *aracae*, *campestris*, *incanae* e *musacearum*. A ausência de *hpaC* nesse patovares pode indicar uma composição de proteínas do T3SS diferente dos outros patovares. Uma vez que, juntamente com *hrpB2*, *HrcQ* e *HpaH*, o gene *hpaC* codificam proteínas que fazem parte do T3SS (Hausner & Buttner, 2014). Os 14 patovares apresentam, também, em seus genomas os genes *pehA*, *prc*, *purC* e *pip*, os quais desempenham função essencial no processo de virulência. O *pehA* regula a expressão dos genes *clp* e *rpfF* (Hsiao et al., 2008). Yuan et al. (2013) e Santos et al. (2017) relataram que a ausência de *purC* e *pip* diminuiu a virulência, interrompendo a biossíntese de purina a qual é essencial nos processos fisiológicos de *X. campestris* pv. *campestris*, respectivamente. Já o *prc* é um gene altamente conservado em *Xanthomonas*, sendo ele um modulador periplasmático das respostas de estímulo do envelope celular. Além disso, esse gene se comportou da mesma forma que *purC* e *pip* quando inativado,

demonstrando que sua ausência diminui a virulência, possivelmente pela produção de exopolissacarídeos, diminuição na motilidade dependente do flagelo, causando dificuldade em formar biofilme, assim como menor regulação da transcrição de genes codificadores de proteases em *X. campestris pv. campestris* (Liao et al., 2016). Somente os patovares *campestris*, *incanae* e *raphani* apresentaram em seu genoma o gene *xopF1*, em vista disso os demais patovares, por não possuírem esse gene, podem apresentar uma composição de proteínas de T3SS diferente, visto que segundo Roux et al. (2015) *xopF1* faz parte do T3SS. Dow et al. (1990) relatou que *prt1* codifica para proteases importantes na patogênese de *X. campestris pv. campestris* e mutantes com ausência desse gene apresentaram em seus hospedeiros redução nos sintomas de podridão negra. O patovar *viticola* foi o único a não apresentar esse gene em seu genoma., uma vez que desempenha o papel de proteína efetora.

As proteínas efetoras são um conjunto de proteínas secretadas por bactérias patogênicas nos hospedeiros, por meio dos sistemas de secreção. Elas são importantes na invasão dos tecidos do hospedeiro, na supressão do sistema imune, além de auxiliar a sobrevivência do patógeno. Sendo, portanto, relacionadas à virulência da bactéria. O repertório de proteínas efetoras varia de espécie para espécie, assim como pode haver exclusivas aos patovares. Nesse sentido, a predição de proteínas efetoras para os genomas analisados de *X. campestris* permitiu identificar, conforme o Anexo D, um montante de 172 *clusters* de proteínas, as quais representam aquelas compartilhadas por todos os patovares. O alinhamento desses *clusters* revelou que nenhuma das proteínas preditas eram de fato efetoras, esse fato sugere que o *software* EffectiveDB não foi capaz de prever proteínas efetoras para *X. campestris*, uma vez que Collingro et al. (2017), através da mesma ferramenta, predisse essas proteínas para *Chlamydia trachomatis*. Reforçando, então, a necessidade do desenvolvimento de *softwares* adequados para a predição de proteínas efetoras em *Xanthomonas* a fim de gerar uma identificação mais acurada.

Por fim, foi possível realizar a identificação de genes diferencialmente distribuídos entre os patovares relacionados à biossíntese dos fatores DSF. Os resultados, conforme a Figura 15, demonstram que todos os patovares apresentam a capacidade de produzir sinal DSF, uma vez que eles possuem o gene *c/p*, o qual é responsável pela regulação da transcrição do conjunto de genes da via de biossíntese de DSF.

## 7 DISCUSSÃO GERAL E PERSPECTIVAS

A análise do pangenoma de *X. campestris* possibilitou identificar grupos ortólogos, além de estabelecer o tamanho do *core*, do *soft*, do *shell* e do *cloud* genoma, tendo sido possível identificar as características do pangenoma dessa espécie, bem como calcular a identidade média de nucleotídeos dos genomas dos patovares analisados. Os resultados dessa análise podem servir de base para estudos futuros de filogenia, bem como para trabalhos que visam melhor elucidar o papel real dos genes de virulência diferencialmente distribuídos. Bem como estudos similares ao descrito por Patel et al. (2016), no qual utilizou-se da abordagem pangenômica aliada a análise de SNPs e de filogenia para gerar árvores de relação entre isolados. Esse tipo de caracterização permite demonstrar fatores filogenéticos específicos para a aquisição de fatores de virulência localizados em ilhas de patogenicidade, além de possibilitar a identificação de marcadores de virulência pra identificação de isolados.

Com a anotação metabólica, predição de metabólitos secundários e análise dos genes de virulência nos genomas de *X. campestris*, foi possível identificar as rotas metabólica, metabólitos secundários e genes de virulência diferencialmente distribuídos entre os 14 genomas analisados. Em trabalhos futuros será possível identificar alternativas para controle do patógeno, uma vez que já foi demonstrado que é possível bloquear enzimaticamente a proteína RpfF, a qual é responsável pela sinalização da biossíntese do fator DSF. Esse bloqueio foi demonstrado por Ionecu et al. (2013) em *Xylella fastidiosa*, nesse estudo foi possível elucidar que a ação enzimática de bloqueio da proteína RpfF acarretou a perda de motilidade, diminuição do crescimento celular e diminuição da virulência. Além disso, será possível, futuramente, clonar e expressar de forma heteróloga os genes de proteínas dos metabólitos secundários apresentados com atividade antimicrobiana e antiviral. Esses metabólitos são de particular interesse biotecnológico, principalmente em estudos farmacêuticos e de controle biológico.

## 8 CONCLUSÃO GERAL

Através das abordagens de bioinformática utilizadas, foi possível identificar as características do pangenoma de *Xanthomonas campestris*. Bem como anotar rotas metabólicas e predizer produtos do metabolismo secundário de cada patovar. Por fim, identificamos genes que codificam para fatores de virulência nos diferentes patovares de *X. campestris*.

## 9 REFERÊNCIAS

- ALLEN, E.E. & BANFIELD, J.F. (2005). Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology*. 3(6). p. 489–498.
- ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. 215(3). p. 403–410.
- AN, S.-Q., POTNIS, N., DOW, M., VORHÖLTER, F.-J., HE, Y.-Q., BECKER, A., TEPPER, D., LI, Y., WANG, N., BLERIS, L. & TANG, J.-L. (2019). Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiology Reviews*.
- ASLAM, S.N., NEWMAN, M.A., ERBS, G., MORRISSEY, K.L., CHINCHILLA, D., BOLLER, T., JENSEN, T.T., DE CASTRO, C., IERANO, T., MOLINARO, A., JACKSON, R.W., KNIGHT, M.R. & COOPER, R.M. (2008). Bacterial Polysaccharides Suppress Induced Innate Immunity by Calcium Chelation. *Current Biology*. 18(14). p. 1078–1083.
- BARBER, C.E., TANG, J.L., FENG, J.X., PAN, M.Q., WILSON, T.J.G., SLATER, H., DOW, J.M., WILLIAMS, P. & DANIELS, M.J. (1997). A novel regulatory system required for pathogenicity of *Xanthomonas campestris* is mediated by a small diffusible signal molecule. *Molecular Microbiology*. 24(3). p. 555–566.
- BEZUIDT, O.K., PIERNEEF, R., GOMRI, A.M., ADESIOYE, F., MAKHALANYANE, T.P., KHARROUB, K. & COWAN, D.A. (2016). The *Geobacillus* pan-genome: Implications for the evolution of the genus. *Frontiers in Microbiology*.
- BLIN, K., SHAW, S., STEINKE, K., VILLEBRO, R., ZIEMERT, N., LEE, S.Y., MEDEMA, M.H. & WEBER, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic acids research*. 47(W1). p. W81–W87.
- BOCHTLER, M. (2012). Structural basis of the TAL effector-DNA interaction. *Biological Chemistry*. 393(10). p. 1055–1066.
- BOGDANOVE, A.J., KOEBNIK, R., LU, H., FURUTANI, A., ANGIUOLI, S.V., PATIL, P.B., VANSLUYS, M.A., RYAN, R.P., MEYER, D.F., HAN, S.W., APARNA, G., RAJARAM, M., DELCHER, A. L., PHILLIPPY, A. M., PUIU, D., SCHATZ, M. C., SHUNWAY, M., SOMMER, D. D., TRAPNELL, C. P., ET ALL (2011). Two new complete genome sequences offer insight in to host and tissue specificity of plant pathogenic *Xanthomonas* sp. *Journal of Bacteriology*. 193(19). p. 5450–5464.
- BÜTTNER, D. & BONAS, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews*. 34(2). p. 107–133.
- CHANG, K.W., WENG, S.F. & TSENG, Y.H. (2001). UDP-glucose dehydrogenase gene of *Xanthomonas campestris* is required for virulence. *Biochemical and Biophysical Research Communications*. 287(2). p. 550–555.
- CHEN, L., WANG, M., HUANG, L., ZHANG, Z., LIU, F. & LU, G. (2017). XC\_0531 encodes a c-type cytochrome biogenesis protein and is required for pathogenesis in *Xanthomonas campestris* pv. *campestris*. *BMC Microbiology*. 17(1). p. 142.

- CHEUNG-LEE, W.L., PARRY, M.E., CARTAGENA, A.J., DARST, S.A. & JAMES LINK, A. (2019). Discovery and structure of the antimicrobial lasso peptide citrocin. *Journal of Biological Chemistry*. 294(17). p. 6822–6830.
- CHIANG, Y.-C., LIAO, C.-T., DU, S.-C. & HSIAO, Y.-M. (2017). Functional characterization and transcriptional analysis of *icd2* gene encoding an isocitrate dehydrogenase of *Xanthomonas campestris* pv. *campestris*. *Archives of Microbiology*. 199(6). p. 917–929.
- CHO, J.H., YOON, J.M., LEE, S.W., NOH, Y.H. & CHA, J.S. (2013). *Xanthomonas oryzae* pv. *oryzae* RpfE regulates virulence and carbon source utilization without change of the DSF production. *Plant Pathology Journal*. 29(4). p. 364–373.
- CIANCOTTO, N.P. & WHITE, R.C. (2017). Expanding role of type II secretion in bacterial pathogenesis and beyond. *Infection and Immunity*. 85(5).
- COLLINGRO, A., KÖSTLBACHER, S., MUSSMANN, M., STEPANAUSKAS, R., HALLAM, S.J. & HORN, M. (2017). Unexpected genomic features in widespread intracellular bacteria: Evidence for motility of marine chlamydiae. *ISME Journal*. 11(10). p. 2334–2344.
- CONESA, A. & GÖTZ, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*.
- CONTRERAS-MOREIRA, B. & VINUESA, P. (2013). GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology*. 79(24). p. 7696–7701.
- COUGER, M.B., HANAFY, R.A., MITACEK, R.M. BUDD, C, FRENCH, D.P., HOFF, W.D., ELSHAHED, M.S. & YOUSSEF, N.H. (2005). The draft genome of *Xanthomonas* sp. Strain Mitacek01 Expands the pangenome of a Genus of Plant Pathogens. *Genome announcements*. 3(6).
- DANILEVICZ, M.F., TAY FERNANDEZ, C.G., MARSH, J.I., BAYER, P.E. & EDWARDS, D. Plant pangenomics: approaches, applications and advancements. *Current Opinion in Plant Biology*.
- DENG, Y., WU, J., TAO, F. & ZHANG, L.H. (2011). Listening to a new language: DSF-based quorum sensing in gram-negative bacteria. *Chemical Reviews*. 111(1). p. 160–179.
- DOW, J.M., CLARKE R., B., MILLIGAN, D.E., TANG, J.L. & DANIELS, M.J. (1990). Extracellular proteases from *Xanthomonas campestris* pv. *campestris*, the black rot pathogen. *Applied and Environmental Microbiology*. 56(10). p. 2994–2998.
- DOW, J.M. & DANIELS, M.J. (1994). Pathogenicity determinants and global regulation of pathogenicity of *Xanthomonas campestris* pv. *campestris*. *Current Topics in Microbiology and Immunology*. 192. p. 29–41.
- DOW, J.M., CROSSMAN, L., FINDLAY, K., HE, Y.Q., FENG, J.X. & TANG, J.L. (2003). Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proceedings of the National*

- Academy of Sciences of the United States of America*. 100(19). p. 10995–11000.
- DOW, M. (2008). Diversification of the function of cell-to-cell signaling in regulation of virulence within plant pathogenic xanthomonads. *Science Signaling*. 1(21).
- EBERHARDT, R.Y., HAFT, D.H., PUNTA, M., MARTIN, M., O'DONOVAN, C. & BATEMAN, A. (2012). AntiFam: A tool to help identify spurious ORFs in protein annotation. *Database*.
- EICHINGER, V., NUSSBAUMER, T., PLATZER, A., JEHL, M.A., ARNOLD, R. & RATTEI, T. (2016). EffectiveDB - Updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Research*. 44(D1). p. D669–D674.
- FOUTS, D.E., BRINKAC, L., BECK, E., INMAN, J. & SUTTON, G. (2012). PanOCT: Automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Research*. 40(22).
- GLASNER, J.D., MARQUEZ-VILLAVICENCIO, M., KIM, H.S., JAHN, C.E., MA, B., BIEHL, B.S., RISSMAN, A.I., MOLE, B., YI, X., YANG, C.H., DANGL, J.L., GRANT, S.R., PERNA, N.T. & CHARKOWSKI, A.O. (2008). Niche-specificity and the variable fraction of the *Pectobacterium* pan-genome. *Molecular Plant-Microbe Interactions*. 21(12). p. 1549–1560.
- GONÇALVES, E.R. & ROSATO, Y.B. (2002). Phylogenetic analysis of *Xanthomonas* species based upon 16S-23S rDNA intergenic spacer sequences. *International Journal of Systematic and Evolutionary Microbiology*. 52(2). p. 355–361.
- GOUGH, C.L., DOW, J.M., KEEN, J., HENRISSAT, B. & DANIELS, M.J. (1990). Nucleotide sequence of the engXCA gene encoding the major endoglucanase of *Xanthomonas campestris* pv. *campestris*. *Gene*. 89(1). p. 53–59.
- GRIFFITHS-JONES, S., BATEMAN, A., MARSHALL, M., KHANNA, A. & EDDY, S.R. (2003). Rfam: An RNA family database. *Nucleic Acids Research*. 31(1). p. 439–441.
- GUDESBLAT, G.E., TORRES, P.S. & VOJNOV, A.A. (2009). *Xanthomonas campestris* overcomes *Arabidopsis* stomatal innate immunity through a DSF cell-to-cell signal-regulated virulence factor. *Plant Physiology*. 149(2). p. 1017–1027.
- GÜRLEBECK, D., THIEME, F. & BONAS, U. (2006). Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant. *Journal of Plant Physiology*. 163(3). p. 233–255.
- HANDELSMAN, J. (2004). Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*. 68(4). p. 669–685.
- HAUSNER, J., HARTMANN, N., JORDAN, M. & BÜTTNER, D. (2017). The predicted lytic transglycosylase HpaH from *Xanthomonas campestris* pv. *vesicatoria* associates with the type III secretion system and promotes effector protein translocation. *Infection and Immunity*. 85(2).

- HAUSNER, J. & BÜTTNER, D. (2014). The YscU/FlhB homologue HrcU from *Xanthomonas* controls type III secretion and translocation of early and late substrates. *Microbiology (United Kingdom)*. 160(PART 3). p. 576–588.
- HE, Y.W., WANG, C., ZHOU, L., SONG, H., DOW, J.M. & ZHANG, L.H. (2006). Dual signaling functions of the hybrid sensor kinase RpfC of *Xanthomonas campestris* involve either phosphorelay or receiver domain-protein interaction. *Journal of Biological Chemistry*. 281(44). p. 33414–33421.
- HE, Y.W. & ZHANG, L.H. (2008). Quorum sensing and virulence regulation in *Xanthomonas campestris*. *FEMS Microbiology Reviews*. 32(5). p. 842–857.
- HE, Y.W., WU, J., ZHOU, L., YANG, F., HE, Y.Q., JIANG, B. LE, BAI, L., XU, Y., DENG, Z., TANG, J.L. & ZHANG, L.H. (2011). *Xanthomonas campestris* diffusible factor is 3-hydroxybenzoic acid and is associated with xanthomonadin biosynthesis, cell viability, antioxidant activity, and systemic invasion. *Molecular Plant-Microbe Interactions*. 24(8). p. 948–957.
- HSIAO, Y.M., ZHENG, M.H., HU, R.M., YANG, T.C. & TSENG, Y.H. (2008). Regulation of the *pehA* gene encoding the major polygalacturonase of *Xanthomonas campestris* by Clp and RpfF. *Microbiology*. 154(3). p. 705–713.
- HYATT, D., CHEN, G.L., LOCASCIO, P.F., LAND, M.L., LARIMER, F.W. & HAUSER, L.J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 11.
- IELPI, L., COUSO, R.O. & DANKERT, M.A. (1993). Sequential assembly and polymerization of the polyprenol-linked pentasaccharide repeating unit of the xanthan polysaccharide in *Xanthomonas campestris*. *Journal of Bacteriology*. 175(9). p. 2490–2500.
- JACQUES, M.-A., ARLAT, M., BOULANGER, A., BOUREAU, T., CARRÈRE, S., CESBRON, S., JACQUES, M.-A., ARLAT, M., BOULANGER, A., BOUREAU, T., CARRÈRE, S., CESBRON, S., CHEN, N.W.G., COCIANCICH, S., DARRASSE, A., DENANCÉ, N., FISCHER-LE SAUX, M., GAGNEVIN, L., KOEBNIK, R., LAUBER, E., NOËL, L.D., PIERETTI, I., PORTIER, P., PRUVOST, O., RIEUX, A., ET AL. (2016). Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annual Review of Phytopathology*. 54(1). p. 163–187.
- JANSEN, R.C. & NAP, J.P. (2001). Genetical genomics: The added value from segregation. *Trends in Genetics*. 17(7). p. 388–391.
- KANEHISA, M., GOTO, S., SATO, Y., FURUMICHI, M. & TANABE, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*. 40(D1).
- KANEHISA, M., SATO, Y. & MORISHIMA, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*. 428(4). p. 726–731.
- KATZEN, F., FERREIRO, D.U., ODDO, C.G., IELMINI, M.V., BECKER, A., PÜHLER, A. & IELPI, L. (1998). *Xanthomonas campestris* pv. *campestris* gum mutants: Effects on

- xanthan biosynthesis and plant virulence. *Journal of Bacteriology*. 180(7). p. 1607–1617.
- KATZEN, F., BECKER, A., ZORREGUIETA, A., PÜHLER, A. & IELPI, L. (1996). Promoter analysis of the *Xanthomonas campestris* pv. *campestris* gum operon directing biosynthesis of the xanthan polysaccharide. *Journal of Bacteriology*. 178(14). p. 4313–4318.
- KÖNIG, S., ROMOTH, L. & STANKE, M. (2018). Comparative genome annotation. In *Methods in Molecular Biology*. Humana Press Inc., p. 189–212.
- KREMER, F.S., ESLABÃO, M.R., DELLAGOSTIN, O.A. & DA PINTO, L.S. (2016). Genix: A new online automated pipeline for bacterial genome annotation. *FEMS Microbiology Letters*. 363(23).
- LAGESEN, K., HALLIN, P., RØDLAND, E.A., STÆRFELDT, H.H., ROGNES, T. & USSERY, D.W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*. 35(9). p. 3100–3108.
- LAING, C., BUCHANAN, C., TABOADA, E.N., ZHANG, Y., KROPINSKI, A., VILLEGAS, A., THOMAS, J.E. & GANNON, V.P.J. (2010). Pan-genome sequence analysis using Panseq: An online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics*. 11.
- LASLETT, D. & CANBACK, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*. 32(1). p. 11–16.
- LENG, M., LU, Z.J., QIN, Z.S., QI, Y.H., LU, G.T. & TANG, J.L. (2019). Flp, a Fis-like protein, contributes to the regulation of type III secretion and virulence processes in the phytopathogen *Xanthomonas campestris* pv. *campestris*. *Molecular Plant Pathology*. 20(8). p. 1119–1133.
- LI, J. & BURMEISTER, M. (2005). Genetical genomics: combining genetics with gene expression analysis. *Human Molecular Genetics*. 14(suppl\_2). p. R163–R169.
- LI, W. & GODZIK, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 22(13). p. 1658–1659.
- LIAO, C.T., DU, S.C., LO, H.H. & HSIAO, Y.M. (2014). The galU gene of *Xanthomonas campestris* pv. *campestris* is involved in bacterial attachment, cell motility, polysaccharide synthesis, virulence, and tolerance to various stresses. *Archives of microbiology*. 196(10). p. 729–738.
- LIAO, C.T., LIU, Y.F., CHIANG, Y.C., LO, H.H., DU, S.C., HSU, P.C. & HSIAO, Y.M. (2016). Functional characterization and transcriptome analysis reveal multiple roles for prc in the pathogenicity of the black rot pathogen *Xanthomonas campestris* pv. *campestris*. *Research in Microbiology*. 167(4). p. 299–312.
- LIMA, W.C., PAQUOLA, A.C.M., VARANI, A.M., VAN SLUYS, M.A. & MENCK, C.F.M. (2008). Laterally transferred genomic islands in Xanthomonadales related to pathogenicity and primary metabolism. *FEMS Microbiology Letters*. 281(1). p. 87–97.

- LIU, H., DONG, C., ZHAO, T., HAN, J., WANG, T., WEN, X. & HUANG, Q. (2016). Functional analysis of the ferric uptake regulator gene *fur* in *Xanthomonas vesicatoria*. *PLoS ONE*. 11(2).
- LO, H.H., LIAO, C.T., LI, C.E., CHIANG, Y.C. & HSIAO, Y.M. (2019). The *clpX* gene plays an important role in bacterial attachment, stress tolerance, and virulence in *Xanthomonas campestris* pv. *campestris*. *Archives of Microbiology*.
- LOCKHART, D.J. & WINZELER, E.A. (2000). Genomics, gene expression and DNA arrays. *Nature*. 405(6788). p. 827–836.
- LONJON, F., LOHOU, D., CAZALÉ, A.C., BÜTTNER, D., RIBEIRO, B.G., PÉANNE, C., GENIN, S. & VAILLEAU, F. (2017). HpaB-Dependent Secretion of Type III Effectors in the Plant Pathogens *Ralstonia solanacearum* and *Xanthomonas campestris* pv. *vesicatoria*. *Scientific Reports*. 7(1).
- LÓPEZ, N.I., HAEDO, A.S. & MÉNDEZ, B.S. (1999). Evaluation of *Xanthomonas campestris* survival in a soil microcosm system. *International Microbiology*. 2(2). p. 111–114.
- LOWE, T.M. & EDDY, S.R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*. 25(5). p. 955–964.
- LUCACIU, R., PELIKAN, C., GERNER, S.M., ZIOUTIS, C., KÖSTLBACHER, S., MARX, H., HERBOLD, C.W., SCHMIDT, H. & RATTEI, T. (2019). A Bioinformatics Guide to Plant Microbiome Analysis. *Frontiers in Plant Science*. 10.
- MACHADO, K.C.T., FORTUIN, S., TOMAZELLA, G.G., FONSECA, A.F., WARREN, R.M., WIKER, H.G., DE SOUZA, S.J. & DE SOUZA, G.A. (2019). On the impact of the pangenome and annotation discrepancies while building protein sequence databases for bacteria proteogenomics. *Frontiers in Microbiology*. 10(JUN).
- MANSFIELD, J., GENIN, S., MAGORI, S., CITOVSKY, V., SRIARIYANUM, M., RONALD, P., DOW, M., VERDIER, V., BEER, S. V., MACHADO, M.A., TOTH, I., SALMOND, G. & FOSTER, G.D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*. 13(6). p. 614–629.
- MARSCHALL, T., MARZ, M., ABEEL, T., DIJKSTRA, L., DUTILH, B., KERSEY, P., WIGARD, P., NOVAK, A., PATEN, B., PORUBSKY, D., RIVAL, E., ALKAN, C., BAAIJENS, J., BAKKER, P.I.W. DE, BOEVA, V., CHIAROMONTE, F., CHIKHI, R., FRANCESCA, D., CIJVAT, R., ET AL. (2016). The Era of Pan-Genomics: Current Status, Promises and Challenges. In *bioRxiv*. p. 043430.
- MAXIMIANO, M.R., OLIVEIRA-NETO, O.B., FRANCO, O.L. & MEHTA, A. (2017). Validation of an in vitro system for studies of pathogenicity mechanisms in *Xanthomonas campestris*. *FEMS Microbiology Letters*. 364(22).
- MENG, P., LU, C., ZHANG, Q., LIN, J. & CHEN, F. (2017). Exploring the Genomic Diversity and Cariogenic Differences of *Streptococcus mutans* Strains Through Pan-Genome and Comparative Genome Analysis. *Current Microbiology*. 74(10). p. 1200–1209.

- McINERNEY, M.J., STRUCHTEMEYER, C.G., SIEBER, J., MOUTTAKI, H., STAMS, A.J.M., SCHINK, B., ROHLIN, L. & GUNSALUS, R.P. (2008). Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. In *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc., p. 58–72.
- MONTESINOS, E. & VILARDELL, P. (2001). Effect of bactericides, phosphonates and nutrient amendments on blast of dormant flower buds of pear: A field evaluation for disease control. *European Journal of Plant Pathology*. 107(8). p. 787–794.
- MOROZOVA, O. & MARRA, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*. 92(5). p. 255–264.
- MUZZI, A. & DONATI, C. (2011). Population genetics and evolution of the pan-genome of *Streptococcus pneumoniae*. *International Journal of Medical Microbiology*. 301(8). p. 619–622.
- NAWROCKI, E.P. & EDDY, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 29(22). p. 2933–2935.
- NOH, Y.H., KIM, S.Y., HAN, J.W., SEO, Y.S. & CHA, J.S. (2014). Expression of colSR genes increased in the rpf mutants of *Xanthomonas oryzae* pv. *oryzae* KACC10859. *Plant Pathology Journal*. 30(3). p. 304–309.
- O'CONNELL, A., AN, S.Q., MCCARTHY, Y., SCHULTE, F., NIEHAUS, K., HE, Y.Q., TANG, J.L., RYAN, R.P. & DOW, J.M. (2013). Proteomics analysis of the regulatory role of Rpf/DSF cell-to-cell signaling system in the virulence of *Xanthomonas campestris*. *Molecular Plant-Microbe Interactions*. 26(10). p. 1131–1137.
- OCIMATI, W., WERE, E., GROOT, J.C.J., TITTONELL, P., NAKATO, G.V. & BLOMME, G. (2018). Risks posed by intercrops and weeds as alternative hosts to *Xanthomonas campestris* pv. *musacearum* in Banana Fields. *Frontiers in Plant Science*. 9.
- OMOBOYE, O.O., GEUDENS, N., DUBAN, M., CHEVALIER, M., FLAHAUT, C., MARTINS, J.C., LECLÈRE, V., ONI, F.E. & HÖFTE, M. (2019). *Pseudomonas* sp. COW3 produces new bananamide-type cyclic lipopeptides with antimicrobial activity against *Pythium myriotylum* and *Pyricularia oryzae*. *Molecules*. 24(22).
- PANDEY, S.S., PATNANA, P.K., RAI, R. & CHATTERJEE, S. (2017). Xanthoferrin, the  $\alpha$ -hydroxycarboxylate-type siderophore of *Xanthomonas campestris* pv. *campestris*, is required for optimum virulence and growth inside cabbage. *Molecular plant pathology*. 18(7). p. 949–962.
- PATEL, I.R., GANGIREDLA, J., LACHER, D.W., MAMMEL, M.K., JACKSON, S.A., LAMPEL, K.A. & ELKINS, C.A. (2016). FDA Escherichia coli Identification (FDA-ECID) microarray: A pangenome molecular toolbox for serotyping, virulence profiling, molecular epidemiology, and phylogeny. *Applied and Environmental Microbiology*. 82(11). p. 3384–3394.
- PEDROSO DE OLIVEIRA, R., UENO, B., BUENO, W., OTTO, S., KOLLER, C. & GOMES DA ROCHA, P.S. Cancro cítrico: epidemiologia e controle. Available at: [www.cpact.embrapa.br](http://www.cpact.embrapa.br).

- POLICARPO, A.D. (2015). Análise do pangenoma de *Streptococcus pneumoniae* e comparação de genomas dos serótipos 1 e 3.
- QIAN, W., JIA, Y., REN, S.X., HE, Y.Q., FENG, J.X., LU, L.F., SUN, Q., YING, G., TANG, D.J., TANG, H., WU, W., HAO, P., WANG, L., JIANG, B. LE, ZENG, S., GU, W.Y., LU, G., RONG, L., TIAN, Y., ET AL. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Research*. 15(6). p. 757–767.
- RASHID, M.M., IKAWA, Y. & TSUGE, S. (2016). GamR, the LysR-type galactose metabolism regulator, regulates hrp gene expression via transcriptional activation of two key hrp regulators, HrpG and HrpX, in *Xanthomonas oryzae* pv. *oryzae*. *Applied and Environmental Microbiology*. 82(13). p. 3947–3958.
- RIGANO, L.A., SICILIANO, F., ENRIQUE, R., SENDÍN, L., FILIPPONE, P., TORRES, P.S., QÜESTA, J., DOW, J.M., CASTAGNARO, A.P., VOJNOV, A.A. & MARANO, M.R. (2007). Biofilm formation, epiphytic fitness, and canker development in *Xanthomonas axonopodis* pv. *citri*. *Molecular Plant-Microbe Interactions*. 20(10). p. 1222–1230.
- RODRIGUEZ-R, L.M., GRAJALES, A., ARRIETA-ORTIZ, M.L., SALAZAR, C., RESTREPO, S. & BERNAL, A. (2012). Genomes-based phylogeny of the genus *Xanthomonas*. *BMC Microbiology*. 12.
- ROTT, P., FLEITES, L.A., MENSI, I., SHEPPARD, L., DAUGROIS, J.H., DOW, J.M. & GABRIEL, D.W. (2013). The RpfCG two-component system negatively regulates the colonization of sugar cane stalks by *Xanthomonas albilineans*. *Microbiology (United Kingdom)*. 159(PART 6). p. 1149–1159.
- ROUX, B., BOLOT, S., GUY, E., DENANCÉ, N., LAUTIER, M., JARDINAUD, M.F., FISCHER-LE SAUX, M., PORTIER, P., JACQUES, M.A., GAGNEVIN, L., PRUVOST, O., LAUBER, E., ARLAT, M., CARRÈRE, S., KOEBNIK, R. & NOËL, L.D. (2015). Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. *BMC Genomics*. 16(1).
- ROYER, M., KOEBNIK, R., MARGUERETTAZ, M., BARBE, V., ROBIN, G.P., BRIN, C., CARRERE, S., GOMEZ, C., HÜGELLAND, M., VÖLLER, G.H., NOËLL, J., PIERETTI, I., RAUSCH, S., VERDIER, V., POUSSIER, S., ROTT, P., SÜSSMUTH, R.D. & COCIANCICH, S. (2013). Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. *BMC Genomics*. 14(1). p. 658.
- RUKAYADI, Y., SUWANTO, A., TJAHJONO, B. & HARLING, R. (2000). Survival and epiphytic fitness of a nonpathogenic mutant of *Xanthomonas campestris* pv. *glycines*. *Applied and Environmental Microbiology*. 66(3). p. 1183–1189.
- RYAN, R.P., KOEBNIK, R., SZUREK, B., BOUREAU, T., BERNAL, A., BOGDANOVE, A. & DOW, J.M. (2009). Passing GO (gene ontology) in plant pathogen biology: A report from the *Xanthomonas* genomics conference. In *Cellular Microbiology*. p. 1689–1696.

- RYAN, R.P., VORHÖLTER, F.J., POTNIS, N., JONES, J.B., VAN SLUYS, M.A., BOGDANOVA, A.J. & DOW, J.M. (2011). Pathogenomics of *Xanthomonas*: Understanding bacterium-plant interactions. *Nature Reviews Microbiology*. 9(5). p. 344–355.
- SANGER, F. & COULSON, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*. 94(3).
- SANTOS, C., MAXIMIANO, M.R., RIBEIRO, D.G., OLIVEIRA-NETO, O.B., MURAD, A.M., FRANCO, O.L. & MEHTA, A. (2017a). Differential accumulation of *Xanthomonas campestris* pv. *campestris* proteins during the interaction with the host plant: Contributions of an in vivo system. *Proteomics*. 17(12).
- SCHATSCHNEIDER, S., SCHNEIDER, J., BLOM, J., LÉTISSE, F., NIEHAUS, K., GOESMANN, A. & VORHÖLTER, F.J. (2017). Systems and synthetic biology perspective of the versatile plant-pathogenic and polysaccharide-producing bacterium *Xanthomonas campestris*. *Microbiology (United Kingdom)*. 163(8). p. 1117–1144.
- SCHEIBNER, F., HARTMANN, N., HAUSNER, J., LORENZ, C., HOFFMEISTER, A.K. & BÜTTNER, D. (2018). The type III secretion chaperone HpaB controls the translocation of effector and non-effector proteins from *Xanthomonas campestris* pv. *vesicatoria*. *Molecular Plant-Microbe Interactions*. 31(1). p. 61–74.
- SCHOLZE, H. & BOCH, J. (2010). TAL effector-DNA specificity. *Virulence*. 1(5). p. 428–432.
- SILVA, A.M.F., DE MENEZES, E.F., DE SOUZA, E.B., DE MELO, N.F. & MARIANO, R. DE L.R. (2012). Sobrevivência de *Xanthomonas campestris* pv. *viticola* em tecido infectado de videira. *Revista Brasileira de Fruticultura*. 34(3). p. 757–765.
- STANKOWSKI, J.D., MUELLER, B.E. & ZELLER, S.G. (1993). Location of a second O-acetyl group in xanthan gum by the reductive-cleavage method. *Carbohydrate research*. 241. p. 321–6.
- STOODLEY, P., SAUER, K., DAVIES, D.G. & COSTERTON, J.W. (2002). Biofilms as Complex Differentiated Communities. *Annual Review of Microbiology*. 56(1). p. 187–209.
- STUDHOLME, D.J., KEMEN, E., MACLEAN, D., SCHORNACK, S., ARITUA, V., THWAITES, R., GRANT, M., SMITH, J. & JONES, J.D.G. (2010). Genome-wide sequencing data reveals virulence factors implicated in banana *Xanthomonas* wilt. *FEMS Microbiology Letters*. 310(2). p. 182–192.
- SUTHERLAND, I.W. (2001). Biofilm exopolysaccharides: A strong and sticky framework. *Microbiology*. 147(1). p. 3–9.
- TAO, F., SWARUP, S. & ZHANG, L.H. (2010). Quorum sensing modulation of a putative glycosyltransferase gene cluster essential for *Xanthomonas campestris* biofilm formation. *Environmental Microbiology*. 12(12). p. 3159–3170.
- TETTELIN, H., MASIGNANI, V., CIESLEWICZ, M.J., DONATI, C., MEDINI, D., WARD, N.L., ANGIUOLI, S. V., CRABTREE, J., JONES, A.L., DURKIN, A.S., DEBOY, R.T., DAVIDSEN, T.M., MORA, M., SCARSELLI, M., MARGARIT Y ROS, I., PETERSON, J.D., HAUSER, C.R.,

- SUNDARAM, J.P., NELSON, W.C., ET AL. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome.” *Proceedings of the National Academy of Sciences of the United States of America*. 102(39). p. 13950–13955.
- TORRES, P.S., MALAMUD, F., RIGANO, L.A., RUSSO, D.M., MARANO, M.R., CASTAGNARO, A.P., ZORREGUIETA, A., BOUARAB, K., DOW, J.M. & VOJNOV, A.A. (2007). Controlled synthesis of the DSF cell-cell signal is required for biofilm formation and virulence in *Xanthomonas campestris*. *Environmental Microbiology*. 9(8). p. 2101–2109.
- TRINDADE, L.C. DA, MARQUES, E., LOPES, D.B. & FERREIRA, M.Á. DA S.V. (2007). Development of a molecular method for detection and identification of *Xanthomonas campestris* pv. *viticola*. *Summa Phytopathologica*. 33(1). p. 16–23.
- TRIPATHI, L., MWAKA, H., TRIPATHI, J.N. & TUSHEMERIRWE, W.K. (2010). Expression of sweet pepper Hrap gene in banana enhances resistance to *Xanthomonas campestris* pv. *musacearum*. *Molecular Plant Pathology*. 11(6). p. 721–731.
- UDAONDO, Z., MOLINA, L., SEGURA, A., DUQUE, E. & RAMOS, J.L. (2016). Analysis of the core genome and pangenome of *Pseudomonas putida*. *Environmental Microbiology*. 18(10). p. 3268–3283.
- VICENTE, J.G., EVERETT, B. & ROBERTS, S.J. (2006). Identification of isolates that cause a leaf spot disease of brassicas as *Xanthomonas campestris* pv. *raphani* and pathogenic and genetic comparison with related pathovars. *Phytopathology*. 96(7). p. 735–745.
- VICENTE, J.G. & HOLUB, E.B. (2013). *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Molecular Plant Pathology*. 14(1). p. 2–18.
- VOJNOV, A.A., SLATER, H., DANIELS, M.J. & DOW, J.M. (2001). Expression of the gum operon directing xanthan biosynthesis in *Xanthomonas campestris* and its regulation in planta. *Molecular Plant-Microbe Interactions*. 14(6). p. 768–774.
- VORHÖLTER, F.J., SCHNEIKER, S., GOESMANN, A., KRAUSE, L., BEKEL, T., KAISER, O., LINKE, B., PATSCHKOWSKI, T., RÜCKERT, C., SCHMID, J., SIDHU, V.K., SIEBER, V., TAUCH, A., WATT, S.A., WEISSHAAR, B., BECKER, A., NIEHAUS, K. & PÜHLER, A. (2008). The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *Journal of Biotechnology*. 134(1–2). p. 33–45.
- WANG, L.H., HE, Y., GAO, Y., WU, J.E., DONG, Y.H., HE, C., WANG, S.X., WENG, L.X., XU, J.L., TAY, L., FANG, R.X. & ZHANG, L.H. (2004). A bacterial cell-cell communication signal with cross-kingdom structural analogues. *Molecular Microbiology*. 51(3). p. 903–912.
- WANG, X., PRESTON, J.K. & ROMEO, T. (2004). The pgaABCD Locus of *Escherichia coli* Promotes the Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation. *Journal of Bacteriology*. 186(9).p. 2724–2734. Wang, X., Preston, J.K. & Romeo, T. (2004). The pgaABCD Locus of *Escherichia coli* Promotes the

- Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation. *Journal of Bacteriology*. 186(9).p. 2724-2734.
- WASUKIRA, A., TAYEBWA, J., THWAITES, R., PASZKIEWICZ, K., ARITUA, V., KUBIRIBA, J., SMITH, J., GRANT, M. & STUDHOLME, D.J. (2012). Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pathovar *musacearum*. *Genes*. 3(3). p. 361–377.
- WHITTAM, T.S. & BUMBAUGH, A.C. (2002). Inferences from whole-genome sequences of bacterial pathogens. *Current opinion in genetics & development*. 12(6). p. 719–25.
- WIGGERICH, H.G. & PÜHLER, A. (2000). The *exbD2* gene as well as the iron-uptake genes *tonB*, *exbB* and *exbD1* of *Xanthomonas campestris* *pv. campestris* are essential for the induction of a hypersensitive response on pepper (*Capsicum annuum*). *Microbiology*. 146(5). p. 1053–1060.
- WOYKE, T., TIGHE, D., MAVROMATIS, K., CLUM, A., COPELAND, A., SCHACKWITZ, W., LAPIDUS, A., WU, D., MCCUTCHEON, J.P., McDONALD, B.R., MORAN, N.A., BRISTOW, J. & CHENG, J.-F. (2010). One Bacterial Cell, One Complete Genome N. Ahmed (ed.). *PLoS ONE*. 5(4). p. E10314.
- YAN, X., TAO, J., LUO, H.L., TAN, L.T., RONG, W., LI, H.P. & HE, C.Z. (2019). A type III effector *XopLXcc8004* is vital for *Xanthomonas campestris* pathovar *campestris* to regulate plant immunity. *Research in Microbiology*. 170(3). p. 138–146.
- YU, C., CHEN, H., TIAN, F., LEACH, J.E. & HE, C. (2014). Differentially-expressed genes in rice infected by *Xanthomonas oryzae* *pv. oryzae* relative to a flagellin-deficient mutant reveal potential functions of flagellin in host–pathogen interactions. *Rice*. 7(1). p. 1–11.
- YUN, M.H., TORRES, P.S., EL OIRDI, M., RIGANO, L.A., GONZALEZ-LAMOTHE, R., MARANO, M.R., CASTAGNARO, A.P., DANKERT, M.A., BOUARAB, K. & VOJNOV, A.A. (2006). Xanthan induces plant susceptibility by suppressing callose deposition. *Plant Physiology*. 141(1). p. 178–187.
- YUAN, Z., WANG, L., SUN, S., WU, Y. & QIAN, W. (2013). Genetic and proteomic analyses of a *Xanthomonas campestris* *pv. campestris* purc mutant deficient in purine biosynthesis and virulence. *Journal of Genetics and Genomics*. 40(9). p. 473–487.
- ZHANG, C., LV, M., YIN, W., DONG, T., CHANG, C., MIAO, Y., JIA, Y. & DENG, Y. (2019). *Xanthomonas campestris* promotes diffusible signal factor biosynthesis and pathogenicity by utilizing glucose and sucrose from host plants. *Molecular Plant-Microbe Interactions*. 32(2). p. 157–166.
- ZHANG, H., WEI, J., QIAN, W. & DENG, C. (2020). Analysis of HrpG regulons and HrpG-interacting proteins by ChIP-seq and affinity proteomics in *Xanthomonas campestris*. *Molecular Plant Pathology*. p. MPP.12903.
- ZHANG, Y., WEI, C., JIANG, W., WANG, L., LI, C., WANG, Y., DOW, J.M. & SUN, W. (2013). The HD-GYP domain protein RpfG of *Xanthomonas oryzae* *pv.*

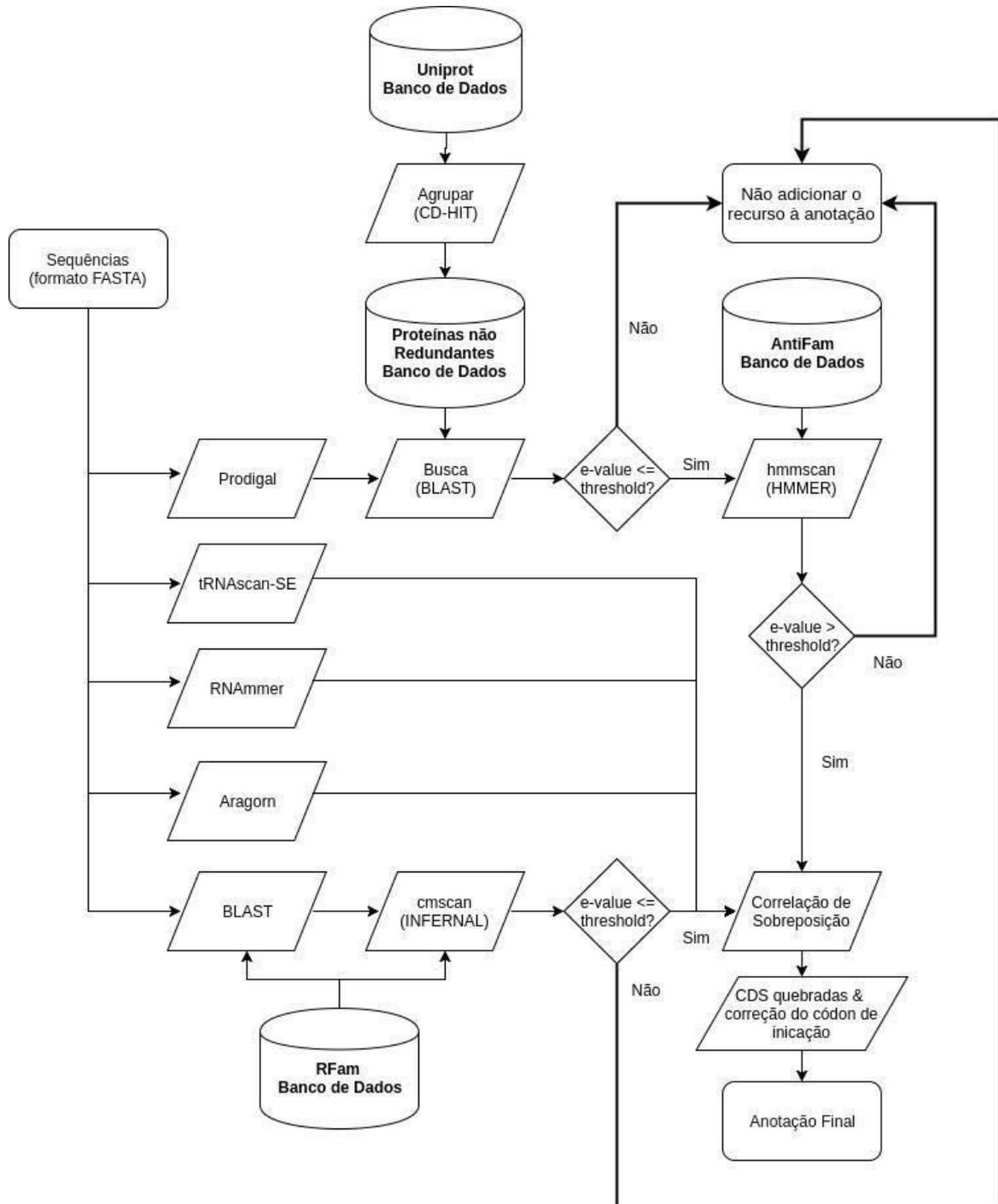
oryzicola regulates synthesis of extracellular polysaccharides that contribute to biofilm formation and virulence on rice. *PLoS one*. 8(3). p. E59428.

ZHAO, Y., JIA, X., YANG, J., LING, Y., ZHANG, Z., YU, J., WU, J. & XIAO, J. (2014). PanGP: A tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics*. 30(9). p. 1297–1299.

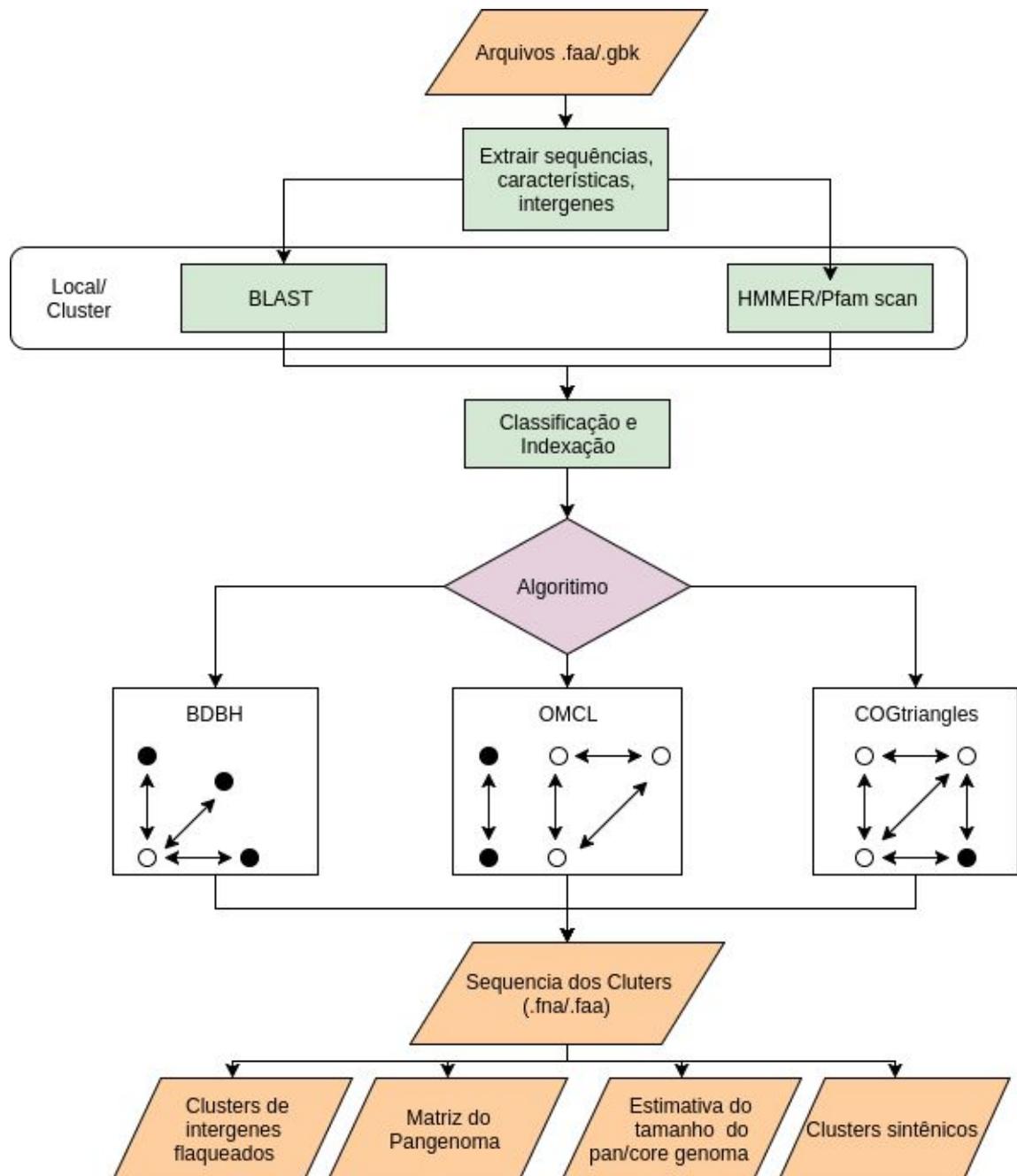
ZHAO, Y., WU, J., YANG, J., SUN, S., XIAO, J. & YU, J. (2012). PGAP: Pan-genomes analysis pipeline. *Bioinformatics*. 28(3). p. 416–418.

ZHOU, L., WANG, X.-Y., SUN, S., YANG, L.-C., JIANG, B.-L. & HE, Y.-W. (2015). Identification and characterization of naturally occurring DSF-family quorum sensing signal turnover system in the phytopathogen *Xanthomonas*. *Environmental Microbiology*. 17(11). p. 4646–4658.

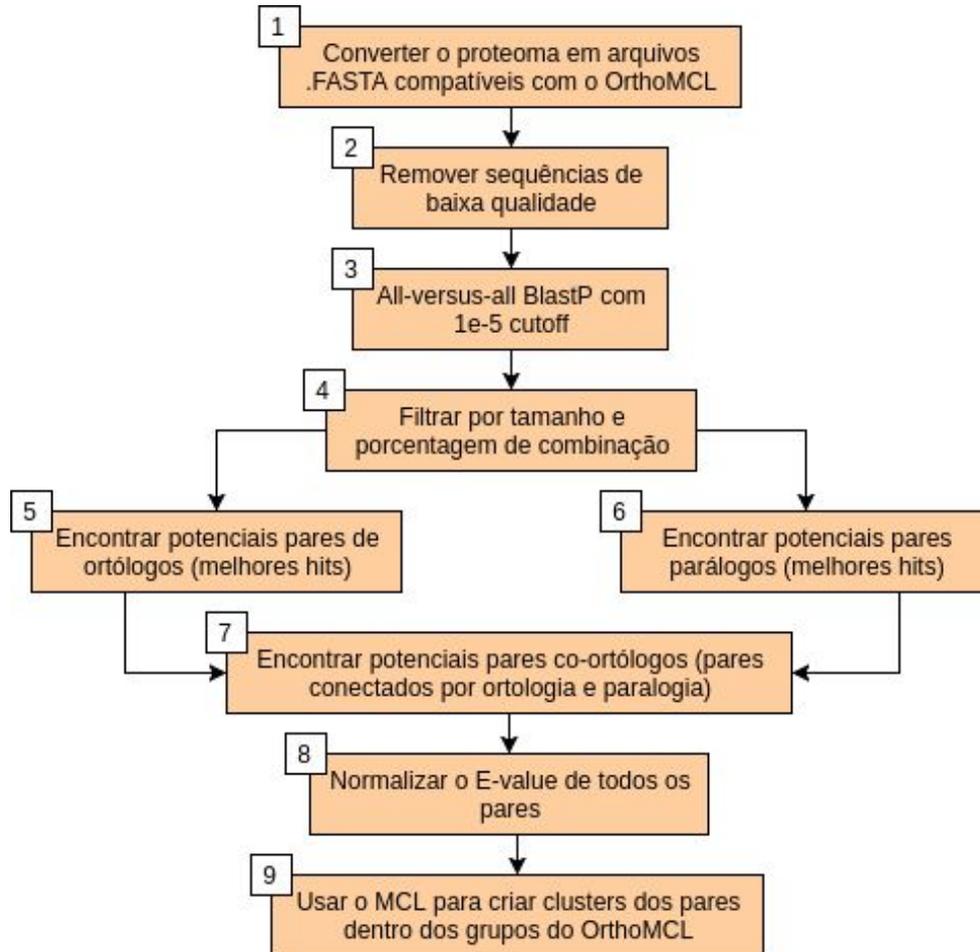
## 10 ANEXOS



**Anexo A.** Pipeline do *software* GENIX. Adaptado de KREMER *et al.*, 2016.



**Anexo B.** Visão geral do algoritmo do *software* GET\_HOMOLOGUES. Adaptado de (CONTRERAS-MOREIRA; VINUESA, 2013).



**Anexo C.** Visão geral do algoritmo do software OrthoMCL. (1) Os proteomas devem estar no formato .fasta cumprindo os requisitos simples; (2) Os arquivos dos proteomas são filtrados baseado no tamanho e na porcentagem de *stop codons*; (3) Através do BlastP os proteomas são comparados entre si; (4) Para cada par gerado há uma porcentagem de similaridade; (5) Selecionar os melhores hits a partir de todos os pares de proteínas ortólogos; (6) Selecionar os melhores hits a partir de todos os pares de proteínas parálogos; (7) Encontrar pares de proteínas que estão conectados através da paralogia e ortologia; (8) Ranquear os grupos ortólogos através do E-value; (9) Criação dos clusters de ortólogos para os grupos preditos. Adaptado de Fischer *et al.*, 2011.

**Anexo D.** Proteínas efectoras presentes em todos os patovares de *Xanthomonas campestris*.

<b>Código do Cluster</b>	<b>Produto</b>
Cluster 2085_5018-7463-1	DNA gyrase subunit B
Cluster 2086_7530-8355-1	Membrane protein
Cluster 2087_12277-12700-1	Biopolymer transport ExbD1 protein
Cluster 2088_14804-15575-1	Pyridoxine 5'-phosphate synthase
Cluster 2102_194590-195973-1	Cation transporter
Cluster 2106_246163-246943-1	Exodeoxyribonuclease III
Cluster 2107_247824-248439-1	4-phosphopantetheinyl transferase
Cluster 2108_261357-261594-1	50S ribosomal protein L28
Cluster 2115_290008-290755-1	Ketosynthase
Cluster 2127_364886-365738-1	Shikimate dehydrogenase (NADP(+))
Cluster 2128_366520-368656-1	ATP-dependent DNA helicase DinG
Cluster 2134_408232-409324-1	Catalase-related peroxidase
Cluster 2136_414009-414372-1	Uncharacterized protein
Cluster 2137_416168-416948-1	Histidine biosynthesis protein HisIE
Cluster 2139_424407-424689-1	Uncharacterized protein TetR family transcriptional regulator
Cluster 2144_433141-433789-1	
Cluster 2145_444465-445323-1	Catalase
Cluster 2153_491436-493041-1	Cytochrome c oxidase subunit 1
Cluster 2155_493824-494700-1	MFS transporter
Cluster 2157_512756-513104-1	Uncharacterized protein
Cluster 2162_543340-543637-1	Peptidoglycan-binding protein LysM
Cluster 2164_556212-557163-1	Cell division protein FtsX
Cluster 2165_558936-559812-1	RNA polymerase sigma factor RpoH
Cluster 2167_569185-570562-1	Multidrug transporter MatE
Cluster 2169_589885-591175-1	Polymerase
Cluster 2170_591159-591984-1	Lipopolysaccharide biosynthesis protein
Cluster 2171_595186-596320-1	GTP cyclohydrolase-2
Cluster 2180_630258-630426-1	Uncharacterized protein
Cluster 2183_640651-641518-1	Dehydrogenase
Cluster 2334_1909473-1910037-a	Extradiol ring-cleavage dioxygenase
Cluster 2186_692957-693452-1	Uncharacterized protein
Cluster 2188_722986-723787-1	ATP synthase subunit a
Cluster 2192_763573-764056-1	Diguanylate cyclase
Cluster 2196_799338-800742-1	Mannose-1-phosphate guanylyltransferase
Cluster 2198_823359-826194--1	Valine--tRNA ligase
Cluster 2205_918982-919804-1	Ribosomal RNA small subunit methyltransferase I
Cluster 2207_922818-923877-1	Ribosomal RNA small subunit methyltransferase H
Cluster 2209_939061-939973-1	UDP-3-O-acyl-N-acetylglucosamine deacetylase
Cluster 2213_982906-984118-1	S-adenosylmethionine synthase
Cluster 2215_996648-997530-1	ABC transporter permease
Cluster 2220_1054483-1054915-1	Nucleoside diphosphate kinase regulator
Cluster 2221_1057954-1058518-1	Alkyl hydroperoxide reductase subunit C
Cluster 2222_1062260-1063259-1	Exodeoxyribonuclease IX
Cluster 2224_1066396-1068745-1	Uncharacterized protein
Cluster 2227_1074507-1075593-1	Branched-chain-amino-acid aminotransferase
Cluster 2230_1116785-1117253-1	30S ribosomal protein S7
Cluster 2231_1136593-1137985-1	Phospho-2-dehydro-3-deoxyheptonate aldolase
Cluster 2232_1149708-1151748-1	2,4-dienoyl-CoA reductase
Cluster 2235_1182600-1184127-1	Exopolyphosphatase
Cluster 2085_5018-7463-1	Glutaredoxin
Cluster 2238_1193370-1194132-1	Membrane protein
Cluster 2245_1273659-1274802-1	Membrane protein
Cluster 2257_1354304-1356410-1	Catalase
Cluster 2259_1359313-1362241-1	Glycine dehydrogenase
Cluster 2265_1468143-1470390-1	Catalase-peroxidase
Cluster 2283_1555877-1556738-1	Regulatory protein
Cluster 2284_1562374-1563328-1	GTPase Era

Cluster 2289	1597654-1597840-1	Uncharacterized protein
Cluster 2302	1659004-1661461-1	Outer membrane protein assembly factor BamA
Cluster 2304	1670107-1670935-1	30S ribosomal protein S2
Cluster 2307	1700978-1702172-1	Multidrug transporter
Cluster 2308	1702178-1703759-1	Multidrug resistance protein B
Cluster 2309	1713074-1713275-1	Uncharacterized protein
Cluster 2313	1721517-1723071-1	Fumarate hydratase class I
Cluster 2321	1816875-1818015-1	5-methyltetrahydrofolate--homocysteine methyltransferase
Cluster 2323	1826770-1827520-1	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB
Cluster 2332	1901917-1902478-1	Poly(Hydroxyalcanoate) granule associated protein
Cluster 2334	1909473-1910037-a	TonB-like protein
Cluster 2344	1961175-1962177-1	Uncharacterized protein
Cluster 2347	2012178-2012994-1	Phosphomethylpyrimidine kinase
Cluster 2348	2015904-2016480-1	Glycine cleavage system regulatory protein
Cluster 2350	2051623-2052991-1	Poly(A) polymerase I
Cluster 2353	2077755-2078523-1	Methyltransferase
Cluster 2356	2131186-2131885-1	Enolase-phosphatase E1
Cluster 2359	2159687-2161427-1	SsDNA exonuclease RecJ
Cluster 2360	2167742-2169260-1	Lysine--tRNA ligase
Cluster 2361	2169398-2170535-1	Cation tolerance protein CutA
Cluster 2362	2176820-2179589-1	Aconitate hydratase
Cluster 2373	2249725-2250493-1	RNA polymerase sigma factor FlIA
Cluster 2374	2269308-2269647-1	Flagellar motor switch protein FlIN
Cluster 2379	2294406-2295516-1	Aminotransferase
Cluster 2380	2301266-2302595-1	Flagellar hook-associated protein 2
Cluster 2382	2304333-2305539-1	Flagellar hook protein FlgL
Cluster 2385	2310473-2311259-1	Flagellar basal-body rod protein FlgG
Cluster 2388	2315019-2315427-1	Flagellar basal-body rod protein FlgC
Cluster 2390	2335414-2335633-1	Translation initiation factor IF-1
Cluster 2392	2337680-2338649-1	Thioredoxin reductase
Cluster 2394	2353302-2354157-1	Membrane protein
Cluster 2409	2487362-2488781-1	Uncharacterized protein
Cluster 2410	2489265-2490075-1	3-deoxy-manno-octulosonate cytidyltransferase
Cluster 2417	2517864-2521365-1	Acriflavin resistance protein
Cluster 2418	2535592-2536765-1	PHA synthase subunit
Cluster 2420	2544914-2545430-1	tRNA-specific adenosine deaminase
Cluster 2432	2634164-2635184-1	Acetylornithine carbamoyltransferase
Cluster 2433	2636164-2637598-1	Cysteine--tRNA ligase
Cluster 2437	2654868-2655558-1	Phytoene synthase
Cluster 2438	2656364-2657084-1	Ubiquinone biosynthesis O-methyltransferase
Cluster 2441	2672174-2672735-1	Adenine phosphoribosyltransferase
Cluster 443	2698524-2699613-1	Ribonuclease D
Cluster 2448	2731599-2732964-1	Succinate-semialdehyde dehydrogenase
Cluster 2457	2784887-2785883-1	LacI family transcriptional regulator
Cluster 2460	2807449-2807803-1	Preprotein translocase, YajC subunit
Cluster 2461	2824771-2825278-1	Phosphopantetheine adenyltransferase
Cluster 2464	2849260-2849554-1	Uncharacterized protein
Cluster 2477	2910423-2911473-1	GDP-mannose:glycolipid 4-beta-D-mannosyltransferase
Cluster 2479	2916218-2917673-1	Polysaccharide biosynthesis protein GumD
Cluster 2480	2917915-2919340-1	Polysaccharide biosynthesis protein GumC
Cluster 2485	2964109-2966551-1	Type IV secretion system energizing component VirB4
Cluster 2487	2975716-2977738-1	UvrABC system protein B
Cluster 2488	2994785-2995289-1	N5-carboxyaminoimidazole ribonucleotide mutase
Cluster 2489	3000615-3000876-1	30S ribosomal protein S15
Cluster 2490	3015672-3017907-1	NADH-quinone oxidoreductase
Cluster 2491	3017903-3019238-1	NADH-quinone oxidoreductase subunit F
Cluster 2492	3034903-3035791-1	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
Cluster 2502	3115410-3115806-1	Membrane protein
Cluster 2503	3134770-3135187-1	Ribosomal silencing factor RsfS
Cluster 2505	3136139-3137168-1	DNA polymerase III subunit delta
Cluster 2518	3237130-3237646-1	Membrane protein

Cluster 2520_3239781-3240891-1	Chemotaxis response regulator protein-glutamate methylesterase
Cluster 2523 3249877-3250738-1	Pirin
Cluster 2526 3260989-3261997-1	Xanthine dehydrogenase
Cluster 2528 3274340-3275339-1	Octaprenyl-diphosphate synthase
Cluster 2532 3290219-3290759-1	DNA-3-methyladenine glycosylase
Cluster 2534 3294107-3295238-1	Twitching motility protein PilT
Cluster 2536 3297408-3298269-1	Pyrroline-5-carboxylate reductase
Cluster 2539 3309984-3312303-1	TonB-dependent receptor
Cluster 2540 3316148-3317123-1	FAD:protein FMN transferase
Cluster 2544 3340794-3341667-1	Nucleotide-binding protein KWM 0119655
Cluster 2550 3382447-3383593-1	3-dehydroquinate synthase
Cluster 2551 3402162-3403545-1	Serine dehydratase
Cluster 2552 3407714-3408746-1	Homoserine acetyltransferase
Cluster 2553 3421667-3422276-1	TetR family transcriptional regulator
Cluster 2557 3464509-3464725-1	Uncharacterized protein
Cluster 2559 3503584-3504484-1	Coenzyme PQQ synthesis protein B
Cluster 2565 3543645-3545553-1	Probable potassium transport system protein kup
Cluster 2566 3547907-3548567-1	Uncharacterized protein
Cluster 2574 3636613-3637609-1	Pseudouridine synthase
Cluster 2576 3642524-3643400-1	Succinate--CoA ligase [ADP-forming] subunit alpha
Cluster 2596 3798025-3798481-1	Uncharacterized protein
Cluster 2597 3798477-3799449-1	ATPase
Cluster 2598 3799469-3800489-1	ATPase AAA
Cluster 2601 3814809-3815190-1	Endoribonuclease
Cluster 2604 3824962-3827134-1	(P)ppGpp synthetase
Cluster 2605 3832469-3834902-1	Thymidine phosphorylase
Cluster 2607 3849193-3850483-1	Glutamate-1-semialdehyde 2,1-aminomutase
Cluster 2614 3898773-3899775-1	Ketol-acid reductoisomerase (NADP(+))
Cluster 2615 3899820-3901542-1	Acetolactate synthase
Cluster 2616 3902973-3904536-1	2-isopropylmalate synthase
Cluster 2620 3913919-3914573-1	TetR family transcriptional regulator
Cluster 2623 3922556-3922754-1	Uncharacterized protein
Cluster 2629 3971932-3972271-1	Uncharacterized protein
Cluster 2633 3983364-3984141-1	Transcriptional regulator
Cluster 2638 4054034-4055048-1	Lipoyl synthase
Cluster 2644 4103655-4104207-1	ATP-dependent protease subunit HslV
Cluster 2648 4125268-4127206-1	Glucans biosynthesis glucosyltransferase H
Cluster 2650 4156486-4157416-1	Magnesium transporter
Cluster 2659 4210145-4212833-1	Pyruvate dehydrogenase E1 component
Cluster 2679 4359669-4360158-1	MarR family transcriptional regulator
Cluster 2680 4361248-4362769-1	Dihydroorotate dehydrogenase
Cluster 2681 4369344-4370415-1	2-oxoisovalerate dehydrogenase
Cluster 2683 4396176-4397934-1	Malto-oligosyltrehalose trehalohydrolase
Cluster 2685 4413918-4414260-1	Uncharacterized protein
Cluster 2686 4445494-4446355-1	Transcriptional regulator, IclR family
Cluster 2691_4495203-4496235-1	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
Cluster 2692 4503891-4504284-1	Uncharacterized protein
Cluster 2700 4571349-4572960-1	Methylcrotonoyl-CoA carboxylase
Cluster 2705 4607144-4607774-1	Acytransferase
Cluster 2710 4656663-4657275-1	Phosphatidylethanolamine-binding protein
Cluster 2717 4703552-4704374-1	2-keto-4-pentenoate hydratase
Cluster 2728 4837678-4838281-1	GTP cyclohydrolase 1