

UNIVERSIDADE FEDERAL DE PELOTAS
CDTec – Centro de Desenvolvimento Tecnológico
Graduação Bacharelado em Biotecnologia

Trabalho de Conclusão de Curso



**Pangenoma de *Xanthomonas campestris* como Ferramenta para
Caracterização de Genes com Potencial Biotecnológico**

Amanda Munari Guimarães

Pelotas, 2017

Amanda Munari Guimarães

**Pangenoma de *Xanthomonas campestris* como Ferramenta para
Caracterização de Genes com Potencial Biotecnológico**

Trabalho de Conclusão de Curso
apresentado ao Centro de
Desenvolvimento Tecnológico da
Universidade Federal de Pelotas, como
requisito parcial à obtenção do título de
Bacharel em Biotecnologia.

Orientador: Luciano da Silva Pinto

Pelotas, RS, 2017

Amanda Munari Guimarães

**Pangenoma de *Xanthomonas campestris* como Ferramenta para
Caracterização de Genes com Potencial Biotecnológico**

**Trabalho de Conclusão de Curso aprovado, como requisito parcial, para
obtenção do grau de Bacharel Biotecnologia, Centro de Desenvolvimento
Tecnológico, Universidade Federal de Pelotas.**

Data da Defesa: 27 de Novembro de 2017

Banca examinadora:

**Prof. Dr. Luciano da Silva Pinto (Orientador)
Doutor em Biotecnologia pela Universidade Federal de Pelotas**

**Msc. Frederico Schmitt Kremer
Mestre em Biotecnologia pela Universidade Federal de Pelotas**

**Prof. Dr. Ismail Teodoro de Souza Junior
Doutor em Fitotecnia pela Universidade Federal do Rio Grande do Sul**

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

G963p Guimarães, Amanda Munari

Pangenoma de *Xanthomonas campestris* como ferramenta para caracterização de genes com potencial biotecnológico / Amanda Munari Guimarães ; Luciano da Silva Pinto, orientador ; Frederico Schmitt Kremer, coorientador. — Pelotas, 2017.

66 f.

Trabalho de Conclusão de Curso (Bacharelado em Biotecnologia) — Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2017.

1. Genômica comparativa. 2. Fitopatógeno. 3. Genoma núcleo. 4. Xantana. 5. Proteínas efectoras. I. Pinto, Luciano da Silva, orient. II. Kremer, Frederico Schmitt, coorient. III. Título.

CDD : 576.162

Elaborada por Ubirajara Buddin Cruz CRB: 10/901

Dedico este trabalho aos meus pais, que sempre apoiaram minhas escolhas

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.

Aos professores da graduação do curso de Biotecnologia que foram sempre amáveis, severos quando necessário, amigos e grandes fomentadores do conhecimento científico.

Agradeço todos os integrantes do grupo BioProLab que me acompanharam nos 4 anos de graduação e sempre me motivaram, deram apoio, foram amigos e, principalmente, me ensinaram muito sobre a biotecnologia e sobre trabalhar em grupo. À Luciana, minha orientadora da extensão que me fez refletir muito sobre o papel do pesquisador na sociedade, e a importância da divulgação correta da ciência. Em especial, o Luciano meu orientador de iniciação científica por 4 anos que depositou confiança em mim, aprimorar conhecimentos e, acima de tudo, me incentivou a ser uma pesquisadora independente. Meu “co-orientador” Frederico, sem o qual esse trabalho não seria possível. Tem minha eterna gratidão pela amizade, ensinamentos, troca de conhecimentos científicos e por me mostrar quão maravilhosa é a bioinformática, tenho certeza que ele foi o responsável por despertar a admiração e amor que tenho por ela.

Aos colegas da graduação, em especial à Gabrielle Navarro, a Gabriana Timm e Jessica Dorneles que sempre estiveram ao meu lado. Sem elas essa etapa da minha vida teria tido menos graça, menos palhaçadas, menos comemorações e com certeza os trabalhos e seminários não teriam sido os mesmos, já que 100% deles fizemos juntas. Essa amizade quero levar comigo e guardo as melhores recordações em meu coração.

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 pelas noites mal dormidas, pelas madrugadas nos plantões, pelos abraços nos momentos

mais críticos, pelas comemorações nos momentos de alegria e por sempre apoiaram minhas escolhas, fazendo o possível, e às vezes o impossível, para eu conquistar meus sonhos. Por terem me ensinado a ser responsável, ter caráter, ser um 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.

Aos meus avós, tios, tias e primos. Ao Walesson e a Leane, graças a eles tive contato com a cultura mineira, e acima de tudo foram extremamente carinhosos e preocupados comigo, sem nunca medirem esforços para me ajudar. Ao Thiago Perdigão, que me acompanhou esses 4 anos em Pelotas sendo meu companheiro, meu guardião e, principalmente, amigo, o qual me fez dar boas risadas, permitiu que eu amadurecesse e exercitasse meu pensamento crítico, além de me ensinar sobre música, economia, filosofia, literatura e sempre instigando meus estudos. Jamais esquecerei tudo aquilo que aprendi com ele.

Resumo

GUIMARÃES, Amanda Munari. **Pangenoma de *Xanthomonas campestris* como Ferramenta para Caracterização de Genes com Potencial Biotecnológico**. 2017. Trabalho de conclusão de curso - Graduação de Biotecnologia, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2017.

Xanthomonas campestris é uma espécie de bactéria fitopatogênica gram-negativa, a qual possui cerca de 140 patovares que acometem diversos hospedeiros de importância econômica como a banana, uva e pimentão. A partir de 14 genomas de *X. campestris* fez-se análises do pangenoma a fim de prospectar genes de possível interesse biotecnológico. Através da análise de pangenômica estrutural foi possível determinar o tipo do pangenoma, sendo esse do tipo fechado. Tendo os genomas re-annotados, a busca por grupos ortólogos gerou 3480 grupos pertencentes ao core genoma, além de determinar o tamanho de 1.2 Mb para as regiões genômicas conservadas entre as cepas. Na análise metabólica foram encontradas rotas de processos celulares, de respostas à alteração do ambiente e de processamento de informações genéticas. Por meio da análise de genes da biossíntese de xantana, identificou-se que, somente os patovares *viticola*, *thespesiae*, *vitiswoodrowii*, *vitistrifoliae*, *vitiscarnosae*, *durantae*, *azadirachtae* e *leeana* apresentam potencial para a produção de xantana. Em relação às proteínas efetoras, observou-se 171 grupos de proteínas preditas presentes nos 14 genomas, entretanto não foram identificadas proteínas efetoras quando realizado o Blastp dos grupos. A falta de acurácia na identificação de proteínas efetoras sugere que o *software* utilizado não apresenta calibragem para o gênero *Xanthomonas*. Em relação aos fatores de virulência *dsbD*, *engXCA*, *exbB*, *exbD1*, *exbD2*, *fliC*, *fur*, *galU*, *hrpX*, *icd2*, *pip*, *prc*, *purC*, *rpfB* e *rpfG* foram encontrados em todos patovares, já os genes *rpfN*, *hpaH*, *hrpB2*, *xopF1*, *hrcQ*, *rpfC*, *rpfF*, *pehA*, *hpaC* e *hrpG* não estão presentes em todos os genomas.

Palavras-chaves: Genômica Comparativa, Fitopatógeno, Genoma Núcleo, Xantana, Proteínas Efetoras.

Abstract

GUIMARÃES, Amanda Munari. **Pangenome of *Xanthomonas campestris* as a Tool for Characterization Biotechnological Potential Genes.** 2017. Graduation thesis - Graduação de Biotecnologia, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2017.

Xanthomonas campestris is gram-negative phytopathogenic bacteria species which has about 140 pathovars that affect several hosts of economic importance such as bananas, grapes and peppers. Pangenome analyzes were performed for 14 genomes of *X. campestris* in order to prospect genes of possible biotechnological interest. Through the analysis of structural pangenomics it was possible to determine the type of pan-genome, being of the closed type. Through the re-annotated genome was possible search for orthologous groups this generated 3480 groups belonging to the core genome and determined the size of 1.2 Mb for the conserved genomic regions between the strains. Metabolic analysis showed the presence of routes of cellular processes, responses to alterations in the environment and processing of genetic information. Xanthan biosynthesis analysis demonstrated only *viticola*, *thespesiae*, *vitiswoodrowii*, *vitistrifoliae*, *vitiscarnosae*, *durantae*, *azadirachtae* and *leeana* pathovars with potential for the production of xanthan. We observed 171 groups of effector proteins present in all 14 genomes, but no effector proteins were identified when Blastp was performed. Lack of accuracy in the identification of effector proteins suggests that the software used doesn't present calibration for the genus *Xanthomonas*. Virulence factors *dsbD*, *engXCA*, *exbB*, *exbD1*, *exbD2*, *fliC*, *fur*, *galU*, *hrpX*, *icd2*, *pip*, *prc*, *purC*, *rpfB* and *rpfG* were found in all strains, on the other hand *rpfN*, *hpaH*, *hrpB2*, *xopF1*, *hrcQ*, *rpfC*, *rpfF*, *pehA*, *hpaC* and *hrpG* genes weren't found in some pathovars.

Keyword: Comparative Genomics, Phytopathogen, Xanthan, Core Genome, Effective Proteins.

Lista de Figuras

| | |
|--|----|
| Figura 1. Sintomas causados pelo gênero <i>Xanthomonas</i> | 17 |
| Figura 2. Esquema representativo dos sistema de secreção de proteínas em <i>Xanthomonas spp</i> | 19 |
| Figura 3. Esquema da rota da biossíntese de xantana em <i>Xanthomonas</i> | 22 |
| Figura 4. Diagrama de Venn ilustrando as categorias de um pangenoma..... | 25 |
| Figura 5. Resumo da metodologia utilizada na análise do pangenoma de <i>Xanthomonas campestris</i> | 28 |
| Figura 6. Diagrama dos grupos ortólogos do pangenoma de <i>X. campestris</i> | 35 |
| Figura 7. Curva descrevendo o tamanho do pangenoma de <i>X. campestris</i> | 36 |
| Figura 8. Funções preditas para o <i>core</i> genoma de <i>X. campestris</i> | 37 |
| Figura 9. Rotas metabólicas preditas para <i>core</i> genoma de <i>X. campestris</i> | 38 |
| Figura 10. Rotas metabólicas preditas para cada patovar de <i>X. campestris</i> em comparação com a análise do <i>core</i> genoma..... | 39 |
| Figura 11. <i>Heatmap</i> dos genes da rota de biossíntese de xantana nos genomas de <i>X. campestris</i> | 40 |
| Figura 12. Diagrama das proteínas efetoras preditas para o <i>core</i> genoma de <i>X. campestris</i> | 41 |
| Figura 13. <i>Heatmap</i> dos genes de virulência presentes nos patovares de <i>X. campestris</i> | 42 |

Lista de Tabelas

| | |
|---|----|
| Tabela 1. Patovares de <i>X. campestris</i> analisados, seus respectivos hospedeiros, códigos de acesso e referência..... | 29 |
| Tabela 2. Relação dos genes da rota de biossíntese de xantana analisados, seus respectivos e códigos de acesso..... | 33 |
| Tabela 3. Genes de virulência presentes em <i>X. campestris</i> | 43 |

Lista de Abreviaturas e Siglas

DSF - fator de sinal difusível do inglês *diffusible signal factor*

EPS - polissacarídeos extracelulares do inglês *extracellular polysaccharides*

FAA - fasta *of amino acid*

FFN - fasta *feature nucleotides*

FNA- fasta *of nucleic acid*

GBK - genbank

GFF - gene *feature format*

Hrp - resposta hipersensível de patogenicidade do inglês *hypersensitive pathogenicity response*

LPS - lipopolissacarídeos

MAMPs/PAMPs - padrões moleculares associados aos patógenos do inglês *pathogen-associated molecular patterns*

Mb - megabase

NCBI - centro nacional de informação biotecnológica do inglês *national center for biotechnology information*

NGS - sequenciamento de nova geração do inglês *next generation sequencing*

ORF - quadro aberto para leitura do inglês *open read frame*

PTI - imunidade desencadeada por PAMPs do inglês *PAMPs triggered immunity*

pv - patovar

T1SS - sistema de secreção do tipo 1 do inglês *type 1 secretion system*

T2SS - sistema de secreção do tipo 2 do inglês *type 2 secretion system*

T3SS - sistema de secreção do tipo 3 do inglês *type 3 secretion system*

T4SS - sistema de secreção do tipo 4 do inglês *type 4 secretion system*

T5SS - sistema de secreção do tipo 5 do inglês *type 5 secretion system*

T6SS - sistema de secreção do tipo 6 do inglês *type 6 secretion system*

TBL - *table*

Sumário

| | |
|--|----|
| 1. Introdução | 15 |
| 2. Revisão Bibliográfica | 17 |
| 2.1 Gênero <i>Xanthomonas</i> | 17 |
| 2.2 <i>X. campestris</i> | 20 |
| 2.3 Genômica..... | 22 |
| 2.4 Genômica de Microrganismos Procariotos..... | 23 |
| 2.5 Abordagem Pangenômica..... | 24 |
| 3. Objetivos | 27 |
| 3.1 Objetivos gerais..... | 27 |
| 3.2 Objetivos específicos..... | 27 |
| 4. Material e Métodos | 28 |
| 4.1 Resumo da abordagem utilizada..... | 28 |
| 4.2 Obtenção dos genomas de <i>X. campestris</i> | 29 |
| 4.3 Re-anotação estrutural de cada gene..... | 30 |
| 4.4 Análise dos grupos ortólogos..... | 31 |
| 4.5 Análise pangenômica estrutural..... | 32 |
| 4.6 Anotação funcional dos genes..... | 32 |
| 4.7 Análise Metabólica..... | 32 |
| 4.7.1 Rotas Metabólicas..... | 32 |
| 4.7.2 Biossíntese de xantana..... | 33 |
| 4.8 Análise de proteínas efetoras..... | 34 |
| 4.9 Análise dos fatores de virulência..... | 34 |
| 5. Resultados | 35 |
| 5.1 Análise dos grupos ortólogos..... | 35 |
| 5.2 Análise pangenômica estrutural..... | 36 |
| 5.3 Anotação funcional dos genes..... | 36 |
| 5.4 Análise metabólica..... | 38 |
| 5.4.1 Rotas Metabólicas..... | 38 |
| 5.4.2 Biossíntese de xantana..... | 39 |

| | |
|---|-----------|
| 5.5 Análise de proteínas efetoras..... | 40 |
| 5.6 Análise de fatores de virulência..... | 41 |
| 6. Discussão..... | 45 |
| 7. Conclusão..... | 51 |
| 8. Referências..... | 52 |
| 9. Material complementar | 60 |

1. Introdução

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 um mecanismo conhecido como Sistemas de Secreção, podendo ser do tipo I,II,III,IV,V e VI, sendo esses responsáveis por secretar proteínas para o meio extracelular, além de secretar proteínas no interior das células do hospedeiro. Esse gênero 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 mais de 140 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, a resposta hipersensível e de patogenicidade, e a goma xantana, a qual auxilia na sobrevivência e fixação da bactéria no tecido do hospedeiro. Esse polímero é comercialmente utilizada na indústria devido suas propriedades de emulsificação, estabilização, viscosidade, solubilidade e espessante.

As ferramentas da genômica moderna possibilitam a caracterização dos organismos, de forma a permitir o entendimento de sua genética e fisiologia, eventos evolutivos, bioquímica, desenvolvimento de novos compostos biotecnológicos e patogênese de doenças. Nesse contexto, o sequenciamento de nova geração (NGS, *next generation sequencing*) permitiu um aumento expressivo no volume de dados genômicos em bancos públicos, como o GenBank. Atualmente, existem 65 genomas completos de *X. campestris* depositados no GenBank. Considerando esse volume

de informações, estudos comparativos possibilitam compreender a plasticidade destes genomas e suas características únicas.

Em vista disso, a análise do pangenoma busca compreender e comparar diferentes isolados de uma mesma espécie ou gênero com a finalidade de analisar aspectos evolutivos, adaptativos e interações entre patógeno-hospedeiro. Nesse sentido, o presente trabalho, através de abordagens *in silico*, objetivou analisar o pangenoma e genoma núcleo da espécie *X. campestris* a fim de identificar genes e rotas metabólicas que possam caracterizar os patovares, além de identificar possíveis produtos de aplicação biotecnológica.

2. Revisão Bibliográfica

2.1 Gênero *Xanthomonas*

Originada do grego, a palavra *Xanthomonas* significa “xanthos” = amarelo e “monas” = unidade. O gênero de bactérias *Xanthomonas* está distribuído em 1470 famílias e abarca em torno de 27 espécies gram-negativas fitopatogênicas, algumas delas se dividem em diferentes patovares, conforme a especificidade de hospedeiro, sintomas, doença e modo de infecção. Os membros do gênero causam doenças em pelo menos 124 espécies monocotiledôneas e 268 espécies de dicotiledôneas, incluindo *X. citri*, *X. oryzae* e *X. campestris* pv. *campestris*, as quais infectam respectivamente citros (laranja, limão e lima), arroz e crucíferas (brócolis e couve-flor) (ROYER *et al.*, 2013; RODRIGUEZ-R *et al.*, 2012). Em geral, esses fitopatógenos causam uma variedade de sintomas, incluindo necrose, cancos, manchas, podridão, hipertrofia e hiperplasia, além de afetar uma variedade de partes da planta, incluindo folhas, hastes, caules, flores, brotos, sementes e frutas (Figura 1) (JACQUES *et al.*, 2016).

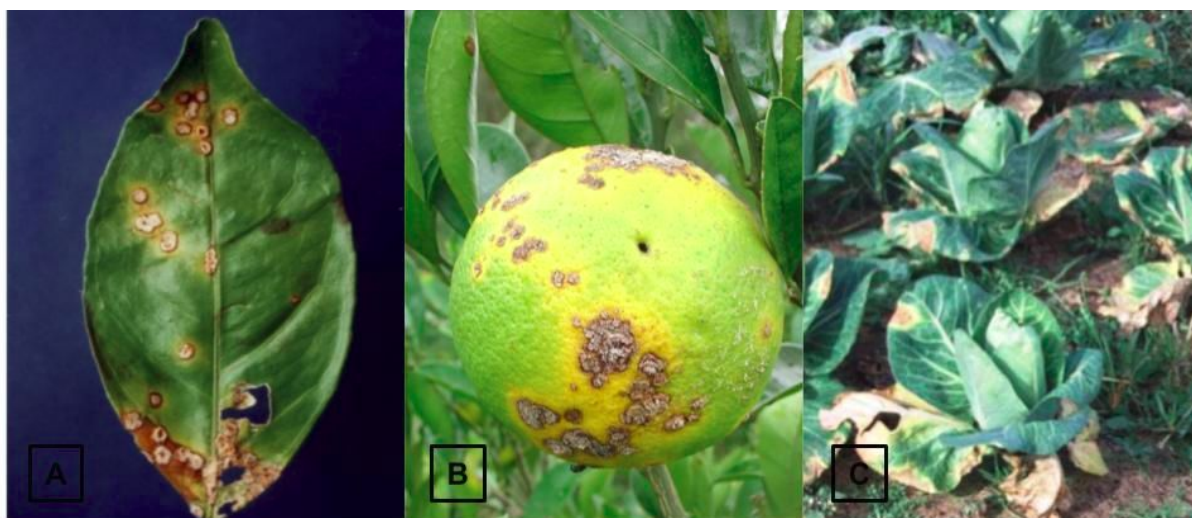


Figura 1. Sintomas causados pelo gênero *Xanthomonas*. (A) Cancro cítrico causado por *X. citri* em folhas de laranja. (B) Cancro cítrico causado pela *X. citri* subsp. *citri* no fruto da laranja. (C) Podridão negra causada pela *X. campestris* pv. *campestris* em couve. Adaptado de OLIVEIRA, et al., (2008); VICENTE & HOLUB, (2013).

As doenças causadas pelo gênero têm, normalmente, como foco inicial de contaminação as sementes, embora práticas agrícolas como a poda, irrigação, transporte de plantas contaminadas, solo contaminado e insetos sejam um meio comum de disseminação. O processo de infecção ocorre, primeiramente, nos tecidos foliares, onde através de feridas e abertura dos estômatos e hidatódios a bactéria adentra o parênquima mesofílico. A primeira interação molecular entre o patógeno e o hospedeiro ocorre enquanto a bactéria ainda está nas superfícies dos tecidos, no qual através de diversos sensores há o reconhecimento do ambiente como favorável à infecção ou não (PARKINSON *et al.*, 2007; RYAN *et al.*, 2009)

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; THIEME; BONAS, 2006). Esse reconhecimento acarreta numa resposta imune basal desencadeada por PAMPs (PTI – *PAMPs Triggered Immunity*) e que, geralmente, freiam a ação do patógeno. No entanto, as espécies do gênero *Xanthomonas* desenvolveram mecanismos de evadir essa defesa do hospedeiro através de 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 uma 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 (CIANCIOOTTO; 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.

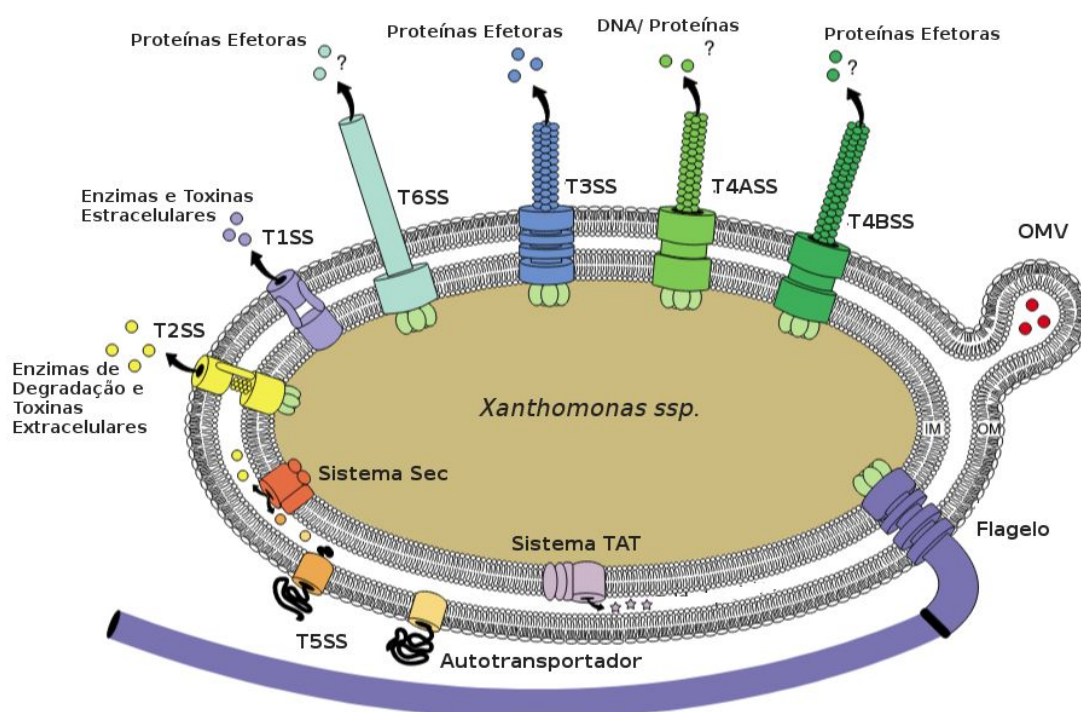


Figura 2. Esquema representativo dos sistemas de secreção de proteínas em *Xanthomonas* spp. Os sistemas T2S e T5S dependem do sistema Sec ou TAT para o transporte de proteínas através da membrana interna. Os sistemas T3S, T4S e T6S estão associados às estruturas de pilus extracelulares para translocar 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 BUTTNER & BONAS, 2009.

2.2 *X. campestris*

A espécie *X. campestris* é a mais complexa dentro do gênero, uma vez que apresenta mais de 140 patovares devido às inúmeras plantas hospedeiras que podem infectar e ao espectro de doenças. Poucas espécies apresentam essa variedade de hospedeiros como *X. campestris* (MINGZHI; ZILING; YONGQUAN, 2007; SCHATSCHNEIDER *et al.*, 2017).

Dentre os patovares de *X. campestris*, este trabalho analisou 14, no entanto somente três deles demonstraram apresentar bibliografia que descrevesse sintomas, doença, processos de infecção e características da patogenia. Dentre esses, está *X. campestris* pv. *musacearum*, o qual é responsável por causar a murcha bacteriana da banana (falsa banana). O processo de infecção ocorre, primeiramente, nos tecidos florais, entretanto o processo ainda não é bem compreendido (TRIPATHI, L. *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 (NAKATO; MAHUKU; COUTINHO, 2017). O primeiro relato da doença foi em 1968 na Etiópia, atualmente a doença está confirmada em 21 distritos de Uganda, no norte de Kivu, República Democrática do Congo, Ruanda, Tanzânia, Burundi e 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.

Em segundo, *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; EVERETT; ROBERTS, 2006).

O processo de patogenicidade 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, celulases, 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).

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 goma xantana, sendo esse um heteropolissacarídeo sintetizado por meio de processo fermentativo aeróbico, que utiliza a glicose ou sacarose como principal substrato. A síntese da goma xantana envolve algumas rotas metabólicas e genes com *xanA*, *xanB* e o *operon gum* (Figura 3). Do ponto de vista comercial, a goma xantana é o polissacarídeo microbiano mais importante, uma vez que suas propriedades de emulsificação, estabilização, solubilidade, espessamento e viscosidade podem ser empregadas em diferentes setores como na indústria alimentícia, farmacêutica e de cosméticos (STEFFENS *et al.*, 2016; TAO *et al.*, 2012).

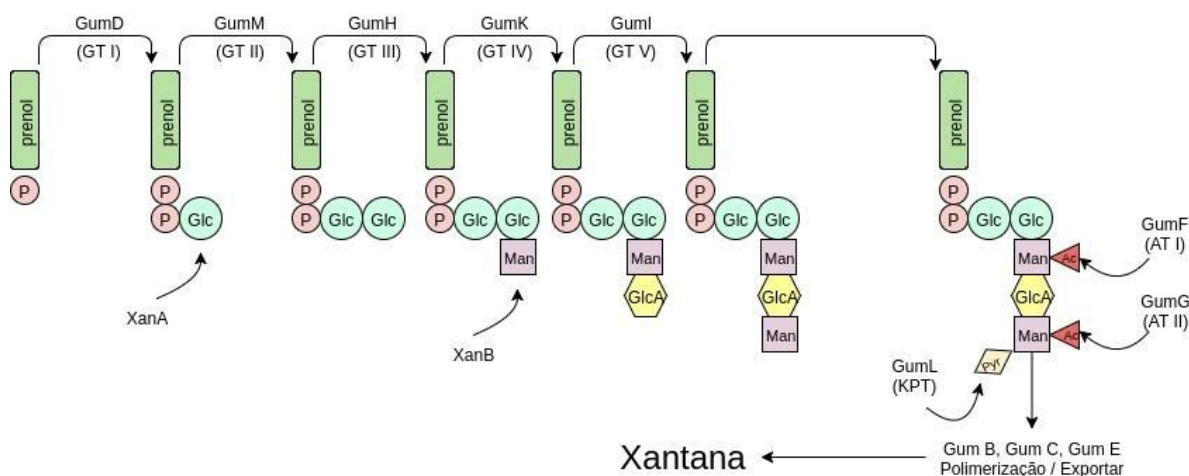


Figura 3. Esquema da rota da biossíntese de xantana em *Xanthomonas campestris*. Os componentes das ligações lipídicas são representados por: Glc, glucose; Man, manose; GlcA, ácido glucurônico; Ac, grupo acetilo; Pyr, grupo pirúvico. A designação de cada proteína é seguido pela sua função proposta da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

2.3 Genômica

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 contém 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 (STAMATOYANNOPOULOS, 2004).

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 sub áreas, 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, também, análise funcional de genes de organismos em ambientes específicos. Por fim, a epigenômica busca entender as modificações epigenéticas no material genético (MOROZOVA; MARRA, 2008; WOLD; MYERS, 2007).

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 biológicos. 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 (FIELDS, 2007).

2.4 Genômica de microrganismos procariotos

Os microrganismos são os seres vivos mais bem adaptados, numerosos e antigos do planeta 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. As abordagens da bioinformática e da biologia molecular permitem a prospecção *in silico* de informações, a partir de dado 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 *Genome* do NCBI (www.ncbi.nlm.nih.gov) constam aproximadamente 15.800 entradas de Procariotos. Já no banco GOLD (www.genomesonline.org), atualmente, estão registrados cerca de 68.600 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, produtos biotecnológicos e aspectos evolutivos.

2.5 Abordagem pangenômica

Uma das abordagens da genômica é a análise do pangenoma, a qual foi descrita a primeira vez por Tettelin *et al* em 2005 através do estudo do pangenoma de *Streptococcus agalactiae*. Por pangenômica compreende-se a comparação entre

diferentes isolados de uma mesma espécie ou 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 4) (MUZZI; DONATI, 2011).

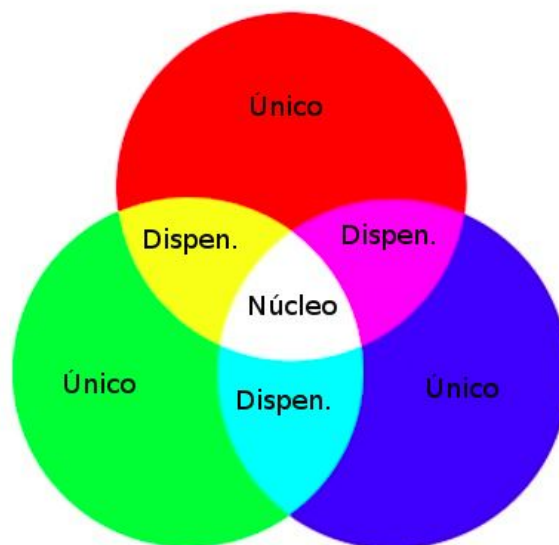


Figura 4. Diagrama de Venn ilustrando as categorias de um pangenoma. A região em branco representa os que são compartilhadas por todos os isolados. As regiões em amarelo, rosa e verde menta representam o genoma acessório ou dispensável, o qual compreende os genes compartilhado entre dois ou mais isolados, mas não por todos. As cores verdes, azul e vermelho 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 menor 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; MEDINI *et al.*, 2005).

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. Por fim, os genes espécie-específicos são aqueles presente em somente um isolado, os quais desempenham funções adaptativas, estão relacionados a virulência e patogenicidade em organismos patogênicos, em organismos não patogênicos esses genes podem estar relacionados a rotas metabólicas (DONATI *et al.*, 2010).

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 dos seus genes. A primeira é denominada de pangenoma aberto, o que significa que a probabilidade de um novo genoma sequenciado fornecer 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. A segunda denomina-se pangenoma fechado, a qual 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).

3. Objetivos

3.1 Objetivos gerais

Esse estudo teve como principal objetivo a análise *in silico* do pangenoma da *X. campestris*, tendo em vista a identificação de genes e rotas metabólicas diferencialmente distribuídos em cada patovar (genomas acessórios), e compartilhados por todas as cepas (genoma núcleo), visando a identificação de potenciais produtos biotecnológicos.

3.2 Objetivos específicos

- Obter os genomas de *X. campestris* no genbank;
- Extrair e indexar os seus respectivos proteomas;
- Identificar grupos ortólogos;
- Mapear os genes de rotas metabólicas;
- Analisar genes da rota de biossíntese de Xantana;
- Mapear as funções dos genes;
- Analisar fatores de virulência;
- Identificar proteínas efetoras.

4. Material e Métodos

4.1 Resumo da abordagem utilizada

O resumo das abordagens utilizadas na seleção, filtragem e re-anotação estrutural dos genomas utilizados, assim como as abordagens de pangenômica estrutural, análise metabólica, análise de proteínas efetoras, das rotas de biossíntese de xantana, dos grupos ortólogos, anotação funcional e proteínas efetoras estão descritas na Figura 5.

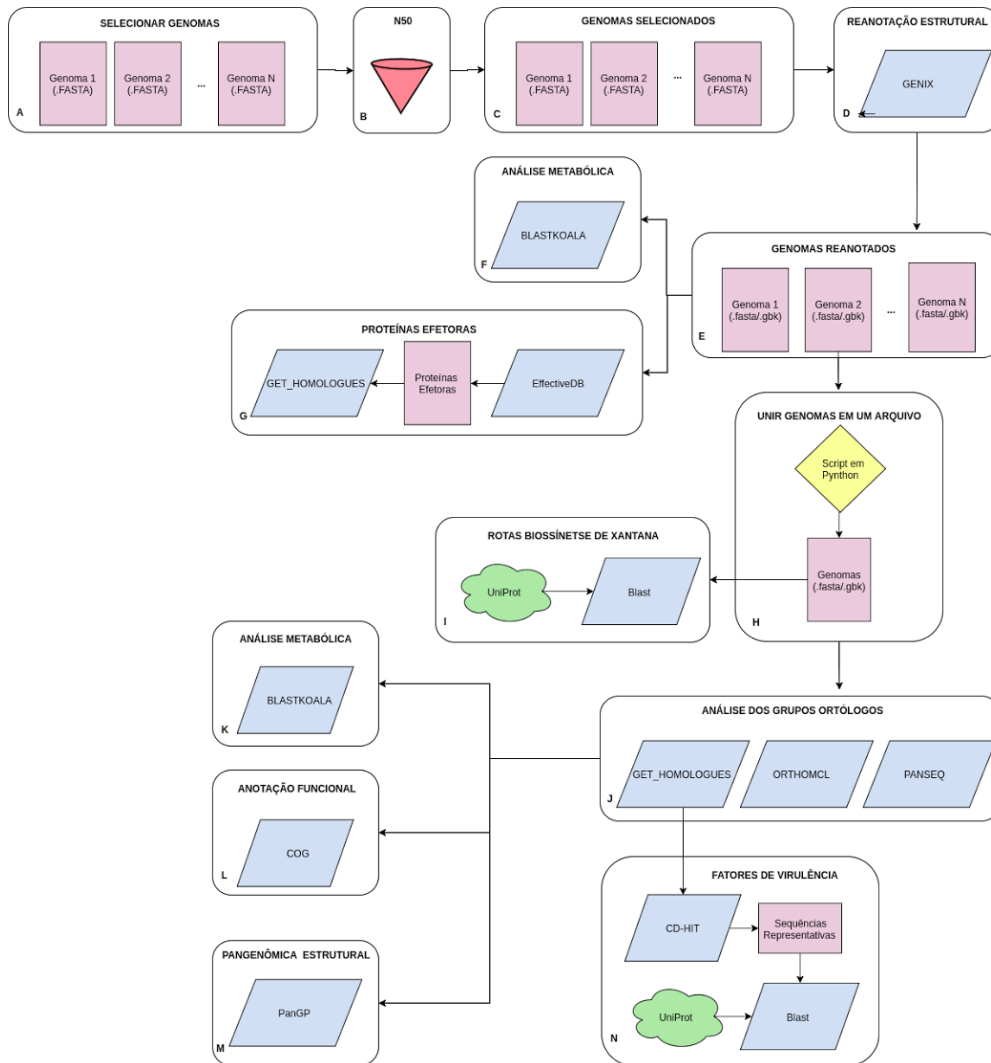


Figura 5. Resumo da metodologia utilizada na análise do pangenoma de *Xanthomonas campestris*. (A) Utilizando o banco de dados GenBank do NCBI foram selecionados 63 genomas de *X. campestris*, sendo 7 genomas completos e 56 genomas rascunhos; (B) Utilizando a métrica N50, os 63 genomas foram filtrados em 14 genomas representativos; (C) Foram selecionados 14 genomas representativos. (D) Re-anotação estrutural dos genes dos 14 genomas, utilizando a ferramenta GENIX; (E) Genomas re-anotados no formato .fasta e .gbk; (F) Os arquivos .fasta dos genomas re-anotados de cada patovar foram submetidos na plataforma online do BlastKoala; (G) Para cada patovar utilizou-se o *software* EffectiveDB para buscar por proteínas efetoras nos genomas, em seguida através do GET_HOMOLOGUES fez-se a análise de clusters dessas proteínas efetoras; (H)

Os 14 genomas reanotados foram unidos em um único arquivo por meio de um *script*; (I) Análise de rotas de biossíntese de xantana utilizando as ferramentas UniProt e Blast; (J) Análise dos grupos ortólogos utilizando três abordagens: GET_HOMOLOGUES, PanSeq e OrthoMCL; (K) Análise Metabólica do *core* genoma gerado pela análise do GET_HOMOLOGUES; (L) Anotação Funcional dos genes do *core* genoma utilizando o COG; (M) Para determinar o tipo de pangenoma para *X. campestris* utilizou-se a ferramenta PanGP; (N) Busca por genes relacionados a fatores de virulência, primeiro utilizando o CD-HIT para identificar genes redundantes do *core* genoma, gerando um arquivo com as sequências representativas de cada gene, em seguida foi feito o alinhamento com BLAST dos genes do *core* genoma com os genes de virulência, reportados na bibliografia.

4.2 Obtenção dos genomas de *Xanthomonas campestris*

O primeiro passo foi obter, através do banco de dados GenBank presente na plataforma do NCBI, 63 genomas de *X. campestris*, sendo 7 genomas completos e 56 rascunhos, desse 63 foram selecionados 14 genomas, um referente a cada patovar, com base na métrica N50. Os patovares analisados foram: *musacearum*, *incanae*, *viticola*, *raphani*, *X. campestris*, *arecae*, *centellae*, *thespesiae*, *leeana*, *vitiswoodrowii*, *vitiscarnosae*, *durantae*, *azadirachtae* e *vitistrifoliae* (Tabela 1).

Tabela 1. Patovares de *X. campestris* analisados, seus respectivos hospedeiros, códigos de acesso e referência.

| Hospedeiro | Patovar | Código de Acesso (GenBank) | Referência |
|--------------------------------------|--|----------------------------|--|
| Falsa banana | <i>X. campestris</i> pv. <i>musacearum</i> | AGFQ00000000.1 | (ERIC MAINA MAGEMBE et al., 2014) |
| Informação não disponível | <i>X. campestris</i> pv. <i>incanae</i> | ATNN01000001.1 | (ROUX et al., 2015) |
| Uva (<i>Vitis vinifera</i> L.) | <i>X. campestris</i> pv. <i>viticola</i> | CBZT010000001.1 | (MIDHA; PATIL, 2014) |
| Rabanete (<i>Raphanus sativus</i>) | <i>X. campestris</i> pv. <i>raphani</i> | CP002789.1 | (BOGDANOVE et al., 2011) |
| Crucíferas | <i>X. campestris</i> strain 17 | CP011256.1 | (BAO, W., LEUNG, F.C.C. AND JIANG, 2015) |
| Informação não disponível | <i>X. campestris</i> pv. <i>arecae</i> | JSBP01000001.1 | (SAPP, M. AND STUDHOLME, 2014) |
| Informação não disponível | <i>X. campestris</i> pv. <i>centellae</i> | LOJR01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |
| Informação não disponível | <i>X. campestris</i> pv. <i>thespesiae</i> | LOJU01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |

| | | | |
|---------------------------|--|----------------|---|
| Informação não disponível | <i>X. campestris</i> pv. <i>vitiswoodrowii</i> | LOKG01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |
| Informação não disponível | <i>X. campestris</i> pv. <i>vitistrifoliae</i> | LOKH01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |
| Informação não disponível | <i>X. campestris</i> pv. <i>vitiscarnosae</i> | LOKI01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |
| Informação não disponível | <i>X. campestris</i> pv. <i>durantae</i> | LOKP01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |
| Informação não disponível | <i>X. campestris</i> pv. <i>azadirachtae</i> | LOKS01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |
| Informação não disponível | <i>X. campestris</i> pv. <i>leeana</i> | LOJY01000001.1 | (BANSAL, K., MIDHA, S. AND PATIL, 2015) |

4.3 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; CAMACHO *et al.* 2009), 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).

4.4 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.

4.5 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 (KANEHISA; SATO; MORISHIMA, 2016). Esse *software* utiliza como arquivos de entrada no formato .FASTA, além disso para utilizá-lo são necessárias outras ferramentas bioinformáticas como PGAP (ZHAO *et al.*, 2012), OrthoMCL (FISCHER *et al.*, 2011) and PanOCT (FOUTS *et al.*, 2012).

4.6 Anotação funcional dos genes

Quanto à anotação funcional dos genes, a ferramenta escolhida foi o Cluster of Ortholog Groups (COG) (TATUSOV *et al.*, 2000), a qual foi executada através de um *script*.

4.7 Análise Metabólica

4.7.1 Rotas Metabólicas

Foram realizadas duas análises metabólicas, uma com o *core* genoma resultante da análise com a ferramenta GET_HOMOLOGUES, e outra com cada patovar desse estudo de forma individual. Nos dois casos a ferramenta de bioinformática utilizada foi o BlastKoala (KANEHISA; SATO; MORISHIMA, 2016). Para a análise metabólica do *core* genoma, um *script* foi feito para unir todos os genes do *core* genoma em um único arquivo .fasta. O arquivo gerado foi, então, submetido para as análises do BlastKoala. Já para cada patovar, os arquivos dos genomas re-annotados pelo GENIX no formato .faa foram submetido, um a um, na plataforma *online* do programa. Para os critérios “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, e o banco de dados de genes “*species_prokaryotes*”. Com os campos selecionados cada genoma de cada patovar foi submetido.

4.7.2 Biossíntese de Xantana

Os genes relacionados à biossíntese de xantana utilizados para essa análise foram aqueles reportados por (DARRASSE *et al.*, 2013). Sendo esses e seus respectivos *operons* e códigos de acesso representados na Tabela 2. Após selecionar os genes, fez-se o *download* das sequências através do banco de dados UniProt. Por fim, através da ferramenta Blastp foi feita a busca por similaridades entre os genes de biossíntese de xantana descritos na bibliografia e os genomas dos patovares.

Tabela 2. Relação dos genes da rota de biossíntese de xantana analisados, seus respectivos *operons* e códigos de acesso.

| Operon / Gene | Gene Name | Código (UniProt) |
|-------------------|-------------|------------------|
| Operon <i>gum</i> | <i>gumB</i> | Q93A84 |
| | <i>gumC</i> | Q8P7Z8 |
| | <i>gumD</i> | Q56770 |
| | <i>gumE</i> | Q8P800 |
| | <i>gumF</i> | Q8P801 |
| | <i>gumG</i> | Q8P802 |
| | <i>gumH</i> | Q56774 |
| | <i>gumI</i> | Q56775 |
| | <i>gumJ</i> | Q8P805 |
| | <i>gumK</i> | Q8GCH2 |
| | <i>gumL</i> | Q3BRV5 |
| | <i>gumM</i> | Q8P807 |
| Operon <i>pga</i> | <i>pgaA</i> | A0A0E2YPP2 |

| | | |
|-------------------|-------------|------------|
| | <i>pgaB</i> | A0A1T1S0K0 |
| | <i>pgaC</i> | A0A1T1S0K0 |
| | <i>pgaD</i> | A0A1T1S0G6 |
| Operon <i>xag</i> | <i>xagA</i> | B0RVF3 |
| | <i>xagB</i> | B0RVF2 |
| | <i>xagC</i> | B0RVF1 |
| Gene | <i>xanA</i> | B0RVK5 |
| Gene | <i>xanB</i> | B0RVK6 |

4.8 Análise de proteínas efetoras

A análise de proteínas efetoras foi feita 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 fazem parte do *core* genoma, quais são do genoma acessório e quais são específicas a cada patovar.

4.9 Análise dos fatores virulência

Para análise dos fatores de virulência, a primeira etapa foi realizar uma revisão bibliográfica em busca de genes relacionados à virulência da bactéria. Após selecionar os genes, fez-se o *download* das sequências através do banco de dados UniProt. Em seguida, utilizando os resultados dos clusters de proteínas gerados pelo GET_HOMOLOGUES foi feito um *script* para determinar a sequência representativa de cada cluster utilizando a ferramenta CD-HIT. Por fim, através da ferramenta Blastp foi feito a busca por similaridades entre os genes de virulência descritos na bibliografia e as sequências representativas.

5. Resultados

5.1 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* genome compreende o tamanho de 1.2 Mb.

Através OrthoMCL foram preditos 5466 grupos ortólogos, sendo 2734 considerados grupos que estão presentes em todos patovares e 3113 grupos presentes em pelo menos 95% dos patovares.

A terceira análise, através do GET_HOMOLOGUES, demonstrou um número de 3480 grupos ortólogos consensos para os diferentes algoritmos utilizados pela ferramenta, sendo 6.288 identificados através do próprio OrthoMCL (re-implementação), 10.575 a partir do COG e 3.765 por meio do DBDH (Figura 6).

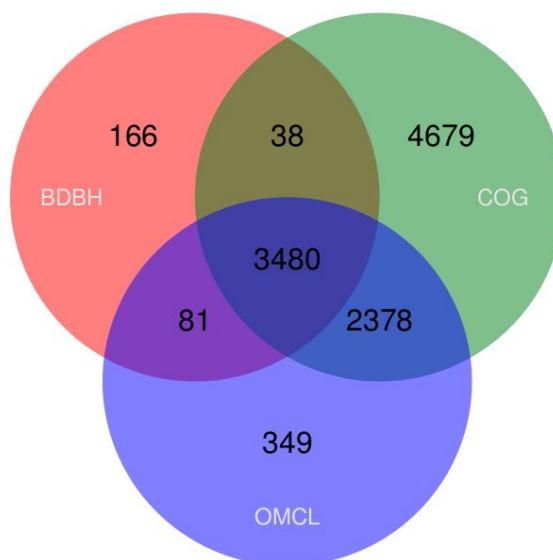


Figura 6. Diagrama dos grupos ortólogos do pangenoma de *X. campestris*. Em rosa a predição dos grupos ortólogos pelo algoritmo BDBH, em verde os grupos ortólogos preditos pelo COG, em roxo os grupos ortólogos gerados pelo OrthoMCL.

5.2 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 4100 grupos, já para o *core* genoma ~1750 a 3200 grupos. Sobre a característica do pangenoma, tem-se que esse estudo é do tipo pangenoma fechado (Figura 7).

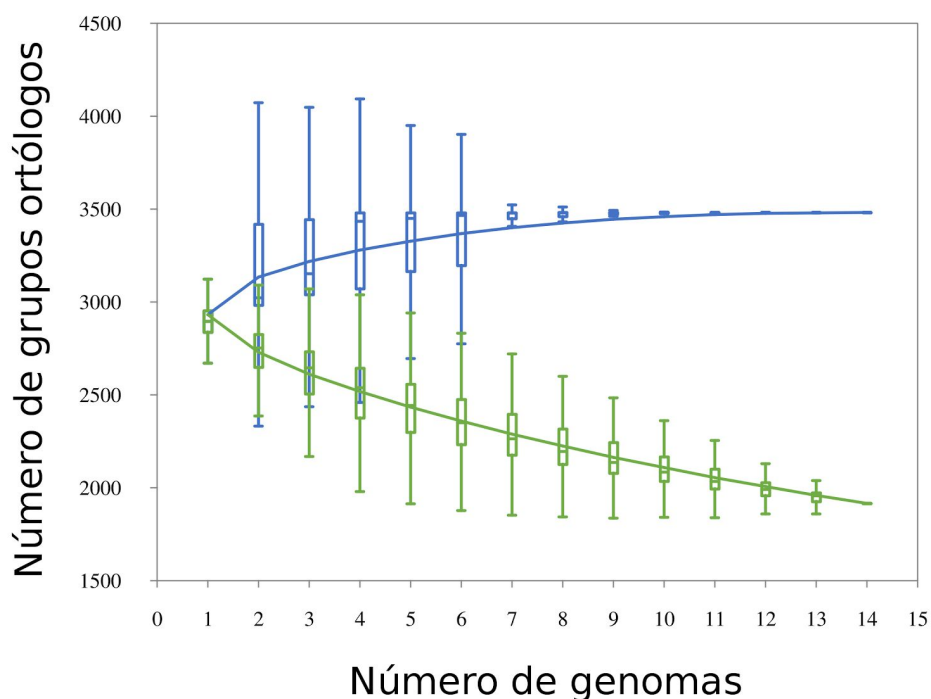


Figura 7. Curva descrevendo o tamanho do pangenoma de *X. campestris*. No eixo x está o número genomas utilizados, no eixo y o número de grupos ortólogos. Os dados em azul representam o pangenoma e os dados em verde representa o *core* genoma.

5.3 Anotação funcional dos genes

A partir dos 3480 *clusters* de genes do *core* genoma preditos pelo GET_HOMOLOGUES (Figura 8), a anotação funcional dos genes revelou nenhum hit encontrado (1963 grupos); previsão de funções gerais (359 grupos); mecanismos de transdução de sinal (292 grupos); transcrição (283 grupos); parede celular, membrana e biogênese do envelope (276 grupos); transporte de aminoácidos e metabolismo (244 grupos); transporte de carboidratos e metabolismo (227 grupos);

translocação, estrutura ribossomal e biogênese (227 grupos); com funções desconhecidas (223 grupos); conservação e produção de energia (212 grupos); transporte de íons inorgânicos e metabolismo (211 grupos); modificações pós tradução, rotação de proteínas, chaperonas (182 grupos); transporte de lipídios e metabolismo (179 grupos); transporte de coenzimas e metabolismo (176 grupos); replicação, recombinação e reparo (163 grupos); modificações celulares (152 grupos), tráfego intracelular, secreção e transporte vesicular (122 grupos); mecanismos de defesa (118 grupos), biossíntese de metabólitos secundários, transporte e catabolismo (101 grupos); transporte de nucleotídeos e metabolismo (74 grupos), mobiloma: profagos, transposons (73 grupos), estruturas extracelulares (41 grupos), controle do ciclo celular, divisão celular e particionamento cromossômico (39 grupos); citoesqueleto (2 grupos); processamento e modificação de RNA (1 grupos); estrutura e dinâmica da cromatina (1 grupos).



Figura 8. Funções previstas para o core genoma de *X. campestris*.

5.4 Análise metabólica

5.4.1 Rotas metabólicas

Realizamos a análise metabólica do *core* genoma, a qual revelou rotas de processos celulares, processamento de informações do ambiente (resposta à alterações ambientais), processamento de informações genéticas (tradução, transcrição e replicação), metabolismo (rotas de catabolismo e anabolismo), fisiologia (manutenção do homeostase) (Figura 9). Entretanto, um número considerável de proteínas teve como *hit* entradas no banco de dados que não possuem nenhum processo metabólico vinculado, assim como algumas das proteínas analisadas não possuem *hits*.

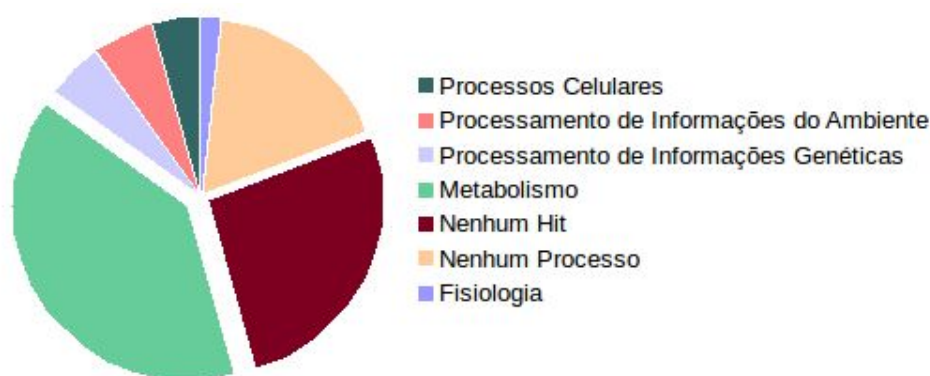


Figura 9. Rotas metabólicas previstas para o *core* genoma de *X. campestris*. As rotas metabólicas previstas pelo *software* BlastKoala para o *core* genoma foram relacionadas a processos celulares, processamento de informações do ambiente, processamento de informações genéticas, metabolismo, nenhum hit, nenhum processo e fisiologia.

A análise metabólica de cada patovar demonstrou regularidade quando a presença de rotas de processos celulares, fisiologia, metabolismo, processamento de informações ambientais, processamento de informações genéticas (Figura 10). No entanto as rotas que não foram encontrados *hits* apresentam diferença entre os patovares.

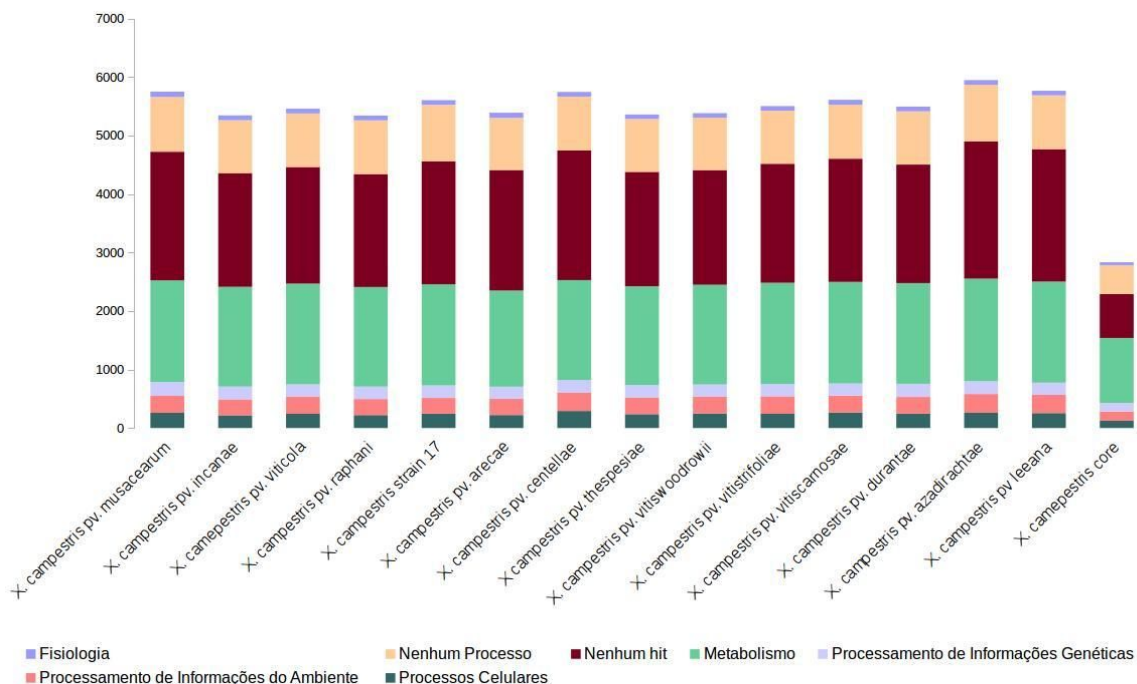


Figura 10. Gráfico das rotas metabólicas previstas pelo software BlastKoala para cada patovar de *X. campestris* em comparação com a análise do core genoma. As rotas metabólicas previstas pelo software BlastKoala foram relacionadas a processos celulares, processamento de informações do ambiente, processamento de informações genéticas, metabolismo, nenhum hot, nenhum processo e fisiologia.

5.4.2 Biossíntese de xantana

Quanto aos genes da rota de biossíntese de xantana, foram encontrados em todos os patovares os 12 genes *gumBCDEFGHIJKLM*, *xagA*, *xagB*, *xagC*, *xanA* e *xanB*. Entretanto, os genes *pgaA*, *pgaB* e *pgaC* estão presente em todos patovares exceto em *incanae* e *raphani*. Já o gene *pgaD* não é encontrado em *musacearum*, *incanae*, *raphani*, *X. campestris* strain 17, *arecae* e *centellae* (Figura 11).

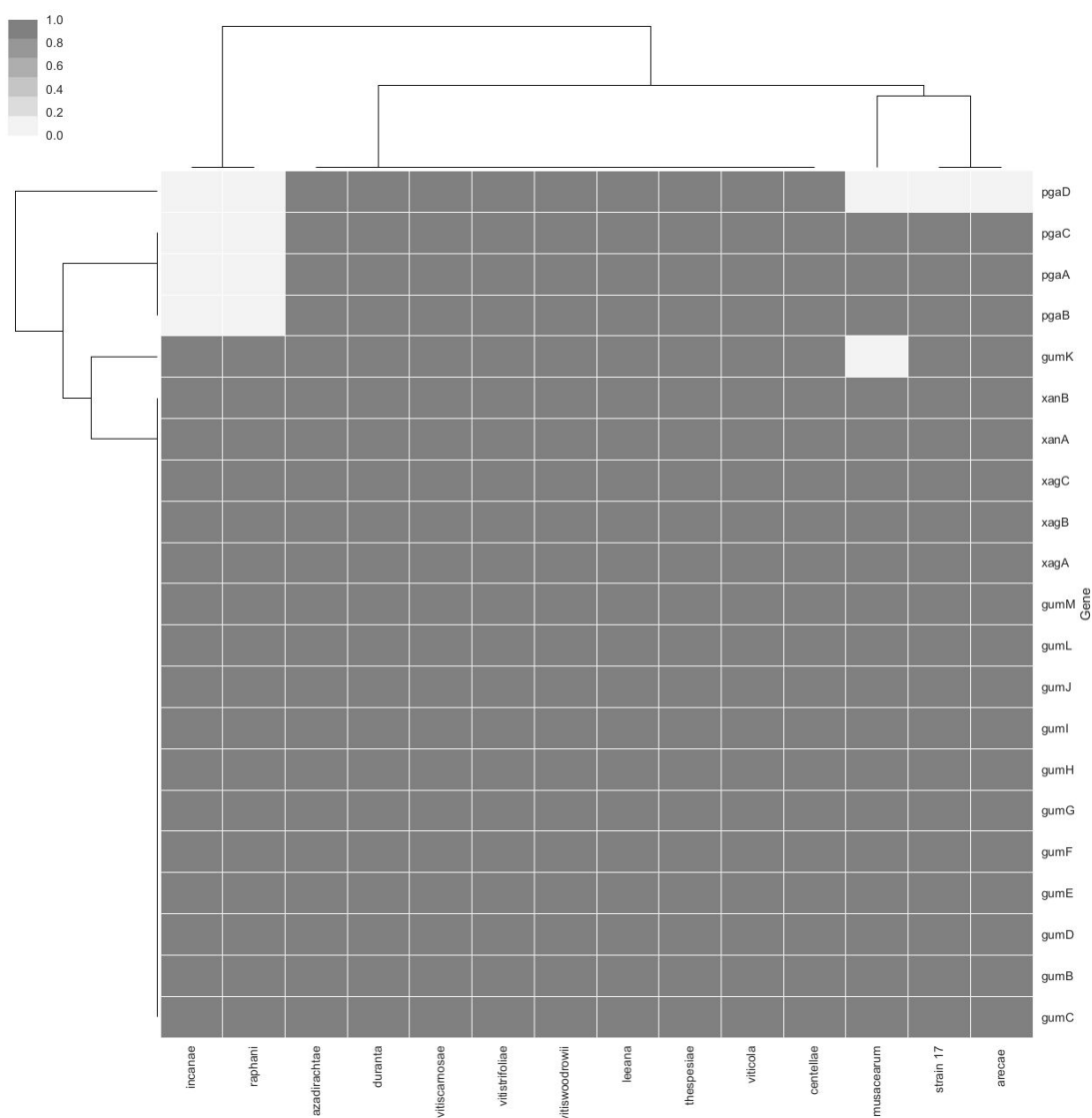


Figura 11. Heatmap dos genes da rota de biossíntese de xantana nos genomas de *X. campestris*. Na horizontal estão indicados os 14 patovares analisados. Na vertical os genes de biossíntese de xantana. Os quadrados pretos representam a presença do gene no genoma, e em branco a ausência do gene.

5.5 Análise de proteínas efetoras

A busca por proteínas efetoras nos genomas de *X. campestris* resultou num total de 557 clusters de proteínas consenso entre os três algoritmos: BDBH, OrthoMCL e COG. Desses, 171 clusters estão presente nos 14 genomas analisados, 120 clusters em pelo menos 2 patovares e não foi encontrado nenhum cluster

exclusivo a um único patovar (Material Complementar 4). No entanto, nenhuma proteína efetora foi predita para *X. campestris*.

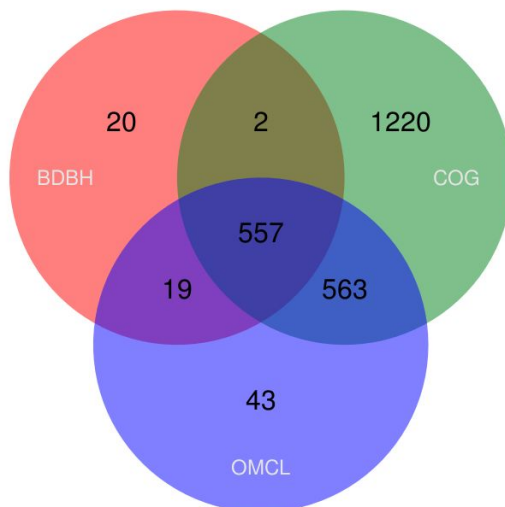


Figura 12. Diagrama das proteínas efetoras preditas para o core genoma de *X. campestris*. Em rosa estão representados os *clusters* de proteínas preditos pelo algoritmo do BDBH. Em verde são os *clusters* preditos pelo algoritmo do COG. Em roxo os *clusters* preditos pelo OrthoMCL. Sendo o valor 557 os *clusters* consensos dos três programas.

5.6 Análise dos fatores virulência

Em relação aos genes de virulência analisados (Tabela 3). Os genes encontrados em todos os patovares são: *dsbD*, *engXCA*, *exbB*, *exbD1*, *exbD2*, *fliC*, *fur*, *galU*, *hrpX*, *icd2*, *pip*, *prc*, *purC*, *rpfB*, *rpfG* e *ptr1*. Entretanto, *clp* foi encontrado em todos patovares, exceto no *thespesiae*, assim como *rpfN* não foi encontrado em *azadirachtae*, além de *hpaH* e *hrpB2* não presentes em *X. campestris strain 17*. Os genes *xopF1* e *hrcQ* só está presente no patovar *incanae* e *raphani*. O *rpfC* e *rpfF* estão presentes em *incanae*, *vitistrifoliae*, *vitiswoodrowii*, *centellae*, *raphani* e *azadirachtae*. O *pehA* nos patovares *incanae*, *vitistrifoliae*, *vitiscarnosae*, *raphani* e em *X. campestris strain 17*. Já o gene *hpaC* está presente em todos os patovares à exceção de *incanae*, *musacearum*, *raphani*, *arecae* e *X. campestris strain 17*. Por fim, o gene *hrpG* não se encontra nos patovares *incanae* e *raphani*.

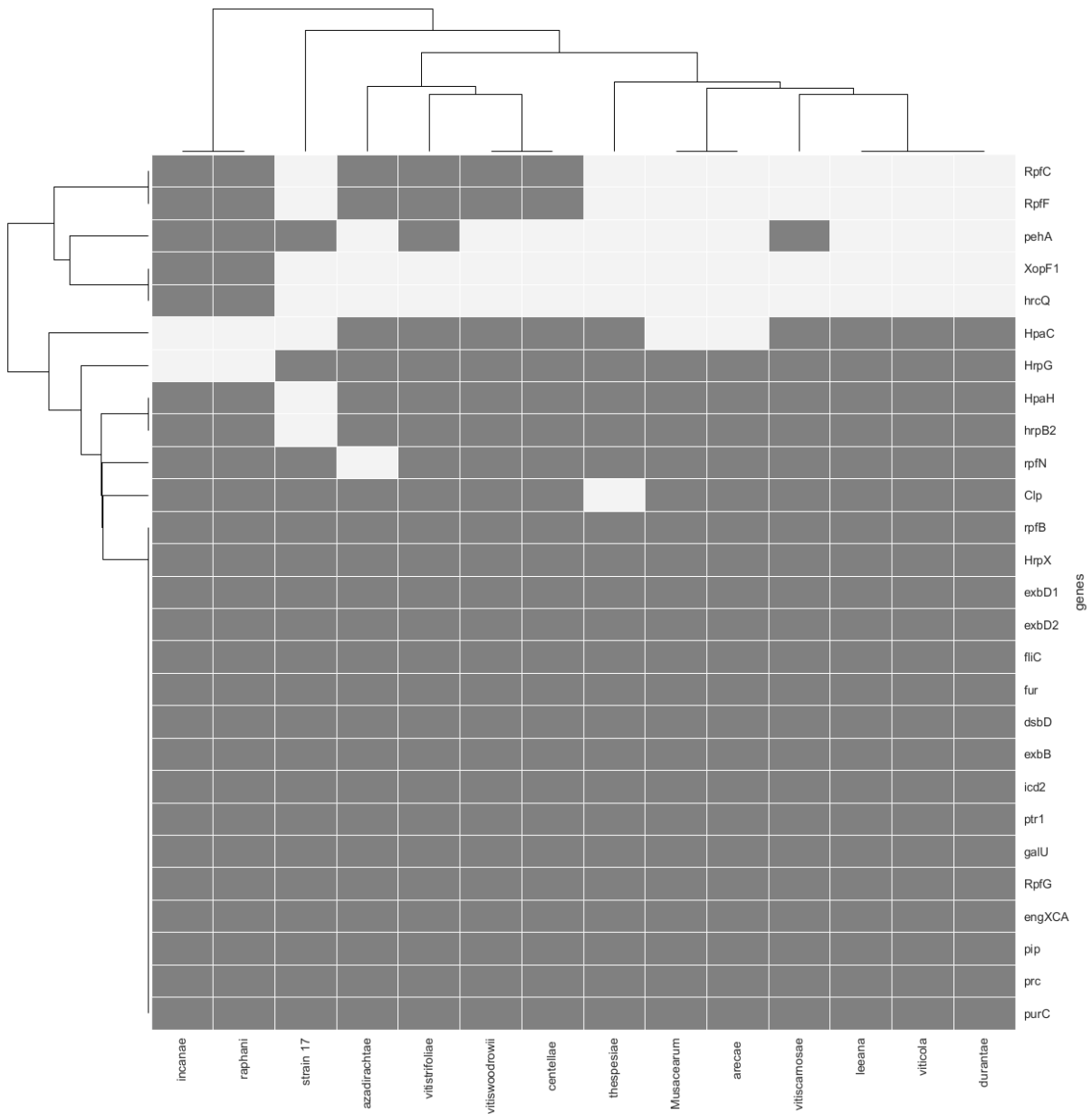


Figura 13. Heatmap dos genes de virulência presentes nos patovares de *X. campestris*. Na horizontal estão indicados os 14 patovares analisados. Na vertical os genes dos fatores de virulência. Os quadrados pretos representam a presença do gene no genoma, e em branco a ausência do gene.

Tabela 3. Genes de virulência presentes em *X. campestris*.

| Gene | Função | Organismo | Código de Acesso (UniProt) | Referência |
|---------------|--|--------------------------------------|-----------------------------------|--|
| <i>clp</i> | Regulador de transcrição global que regula fatores de virulência | <i>X. campestris pv. campestris</i> | CLP_XANCP | (LIAO <i>et al.</i> , 2016) |
| <i>dsbD</i> | Relacionado com a biogênese do citocromo c-type e relação com a virulência | <i>X. campestris pv. campestris</i> | B0RN48_XANCB | (CHEN <i>et al.</i> , 2017) |
| <i>engXCA</i> | Endonuclease extracelular | <i>X. campestris pv. campestris</i> | GUNA_XANCP | (HE; ZHANG, 2008) |
| <i>exbB</i> | Sistema de aquisição de ferro férrico dependente de TonB | <i>X. campestris pv. campestris</i> | EXBB_XANCP | (WIGGERICH; PÜHLER, 2000) |
| <i>exbD1</i> | Sistema de aquisição de ferro férrico dependente de TonB | <i>X. campestris pv. campestris</i> | EXBD1_XANCB | (WIGGERICH; PÜHLER, 2000) |
| <i>exbD2</i> | Sistema de aquisição de ferro férrico dependente de TonB | <i>X. campestris pv. campestris</i> | EXBD2_XANCB | (WIGGERICH; PÜHLER, 2000) |
| <i>fliC</i> | flagelina | <i>X. campestris pv. vesicatoria</i> | Q3BU10_XANC5 | (YU <i>et al.</i> , 2014) |
| <i>fur</i> | Metabolismo do ferro e funções associadas à virulência além do fator de transcrição Fur-like | <i>X. campestris pv. campestris</i> | Q8PAL3_XANCP | (PANDEY <i>et al.</i> , 2016) |
| <i>galU</i> | Desempenha um papel na ligação bacteriana, motilidade celular, síntese de ePS e LPS, patogenicidade e tolerância ao estresse | <i>X. campestris pv. campestris</i> | Q8P8Q1_XANCP | (LIAO <i>et al.</i> , 2014) |
| <i>hpaC</i> | Transloca proteínas efetoras | <i>X. campestris pv. vesicatoria</i> | Q3BYK8_XANC5 | (HAUSNER; BUTTNER, 2014) |
| <i>hpaH</i> | Transloca proteínas efetoras | <i>X. campestris pv. vesicatoria</i> | Q3BYJ1_XANC5 | (HAUSNER, Jens <i>et al.</i> , 2017) |
| <i>hrcQ</i> | Proteína do Sistema de Secreção do Tipo III | <i>X. campestris pv. vesicatoria</i> | Q8PB96_XANCP | (LORENZ; HAUSNER, Jens; BÜTTNER, 2012) |
| <i>hrpB2</i> | Transportador de prótons | <i>X. campestris pv. campestris</i> | A0A1B0UHJ8_XANCA | (HAUSNER; BUTTNER, 2014) |
| <i>hrpG</i> | Regulação de Genes e patogenicidade | <i>X. campestris pv. vesicatoria</i> | Q3BW18_XANC5 | (RASHID; IKAWA; TSUGE, 2016) |
| <i>hrpX</i> | Sensor fosforilado de kinase | <i>X. campestris pv. campestris</i> | Q8P9A6_XANCP | (RASHID; IKAWA; TSUGE, 2016) |
| <i>icd2</i> | Codifica uma desidrogenase de isocitrato funcional | <i>X. campestris pv. campestris</i> | B0RWU3_XANCB | (CHIANG <i>et al.</i> , 2017) |

| | | | | |
|--------------|---|---|------------------|----------------------------------|
| <i>pehA</i> | Gene de poligalacturonase | <i>X. campestris</i> pv. <i>campestris</i> | A0A1Y6HAX6_9XANT | (HSIAO <i>et al.</i> , 2008) |
| <i>pip</i> | Gene de virulência proline iminopeptidase | <i>X. campestris</i> pv. <i>campestris</i> | Q8PC98_XANCP | (SANTOS <i>et al.</i> , 2017) |
| <i>prc</i> | Protease específica Tail | <i>X. campestris</i> pv. <i>campestris</i> | Q8P591_XANCP | (LIAO <i>et al.</i> , 2016) |
| <i>ptr1</i> | Putativa pteridina redutase | <i>X. campestris</i> pv. <i>campestris</i> | B0RNR9_XANCB | (DOW <i>et al.</i> , 1990) |
| <i>purC</i> | Gene de virulência | <i>X. campestris</i> pv. <i>campestris</i> | PUR7_XANCP | (YUAN <i>et al.</i> , 2013) |
| <i>rpfB</i> | Regulador de patogenicidade | <i>X. campestris</i> pv. <i>campestris</i> | Q3BUB1_XANC5 | (NOH <i>et al.</i> , 2014) |
| <i>rpfC</i> | Proteína sensorial / reguladora | <i>X. campestris</i> pv. <i>campestris</i> | RPFC_XANC8 | (NOH <i>et al.</i> , 2014) |
| <i>rpfF</i> | Enoyl-CoA hidratase / proteína da família isomerase RpfF | <i>X. campestris</i> pv. <i>campestris</i> | Q3BUB2_XANC5 | (HSIAO <i>et al.</i> , 2008) |
| <i>rpfG</i> | Regulador de resposta de fosfodiesterase di-GMP cíclico | <i>X. campestris</i> pv. <i>campestris</i> | RPFG_XANC8 | (NOH <i>et al.</i> , 2014) |
| <i>rpfN</i> | Causou a repressão da produção de enzimas e polissacarídeos | <i>X. campestris</i> pv. <i>campestris</i> | B0RRR6_XANCB | (NOH <i>et al.</i> , 2014) |
| <i>xopF1</i> | Proteína Efetora Tip0III | <i>X. campestris</i> pv. <i>campestris</i> | G0CIQ6_XANCA | (ROUX <i>et al.</i> , 2015) |

6. Discussão

X. campestris é considerada a espécie mais complexa dentro do gênero *Xanthomonas*, pois apresenta cerca de 140 patovares, os quais possuem alta especificidade ao hospedeiros, sendo, assim, divergentes quanto aos mecanismos de patogenicidade (QIAN et al., 2005). Além disso, *X. campestris* produz goma xantana, a qual é amplamente empregada nas indústrias alimentícias, farmacêuticas e de cosméticos devido suas características de viscosidade, emulsão, solubilidade, espessamento e estabilidade (HARDING et al., 1987). Análises pangenômicas permitem a identificação e o entendimento de aspectos evolutivos, adaptativos e interações entre patógeno-hospedeiro, por meio da comparação entre diferentes isolados de uma mesma espécie. Além disso, permite a identificação de genes de interesse biotecnológico, tais como genes envolvidos na biossíntese de xantana.

Desta forma, o presente trabalho focou-se em analisar, através de abordagens *in silico*, o pangenoma de *X. campestris* a fim de buscar por genes e proteínas que possam ser utilizados como possíveis produtos biotecnológicos. Dentre eles, estão os genes envolvidos na biossíntese de xantana tais como os 12 genes *gumBCDEFGHIJKLM*, os quais são expressos em operon (KATZEN et al., 1996; VOJNO et al., 2001). A síntese de xantana ocorre na membrana celular, onde unidades de glicose-glicose-manose-glucuronato-manose são construídas a partir de glicosiltransferases codificadas pelos genes D, M, H, K e I (IELPI et al., 1993; STANKOWSKI et al., 1993). Além desses genes, a biossíntese 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). Os genes *xagABC* foram reportados por Tao; Swaruo & Zhang (2010) como codificadores um sistema de glicosiltransferase conservado associado à biossíntese de xantana, ademais, demonstraram que a eliminação desses genes resultou em diminuição da produção do polissacarídeo. Da mesma forma, os genes *pgaABCD* foram reportados por Wang; Preston & Romeo (2004) como fundamentais para a biossíntese e funcionalidade do polissacarídeo.

O gene de virulência *dsbD* 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, diminuindo a atividade enzimática, como consequência menor produção de EPS, o qual é um importante fator de virulência. Gouch, *et. al* (1990) elucidou o gene *engXCA*, o qual codifica para uma endoglucanase extracelular ENGXCA, sendo ela importante no processo de patogenia. Os genes *exbB*, *exbD1* e *exbD2* foram descritos por Wiggerich & Pühler (2000) para *X. campestris* como codificadores de proteína auxiliares que formam o sistema TomB, o qual faz a incorporação do ferro férrico e é de extrema importância para o mecanismo de resposta hipersensitiva e de patogenicidade, sendo essa resposta caracterizada pela rápida necrose do tecido do hospedeiro, sintoma, esse, característico de infecções por *X. campestris*. O gene que *fliC* codifica para flagelina foi estudado por Yu *et al.* (2014), suas análises demonstraram que ele está associado à patogenicidade de *X. oryzae pv oryzae*. 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; IKAWA; TSUGE, 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* faz parte do T3SS juntamente com *hrpB2*, *HrcQ* e *HpaH*, os quais codificam proteínas que translocam outras proteínas efetoras (HAUSNER & BUTTNER, 2014). O *pehA* regula a expressão dos genes *clp* e *rpfF* (HSIAO *et al.*, 2008). Os genes *rpf* segundo Noh *et al.* (2014) são fatores de regulação de patogenicidade e são responsáveis pelo fator de sinal de difusão (DSF). Segundo Roux *et al.* (2015) *xopF1* faz parte do T3SS, uma vez que desempenha o papel de proteína efetora. 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. oryzae pv. oryzae* e *X. campestris pv. campestris*, respectivamente. Já o *prc* é 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). Dow *et al.* (1990) relatou que *prt1* codifica para proteases importantes na patogenia de *X. campestris* pv. *campestris* e mutantes com ausência desse gene apresentaram em seus hospedeiros redução nos sintomas de podridão negra.

A análise das características do pangenoma dos 14 genomas analisados, revelou que este é do tipo fechado, uma vez que o gráfico gerado pelo PanGP (Figura 9) demonstrou que quanto mais genomas fossem adicionados, menos grupos ortólogos seriam identificados. Essa situação se dá, possivelmente, pois o *core* genoma representa grupos ortólogos em comum entre todos os genomas em questão, como cada patovar de *X. campestris* apresenta alta especificidade ao hospedeiro, logo possuem genes exclusivos a cada um e a adição de genomas não contribuirá para a determinação do *core* genoma. A evidência para essa situação se dá com os números de grupos ortólogos demonstrados. 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 muitos genomas não contribuem para a determinação do genoma núcleo de *X. campestris*.

Os 2734 grupos ortólogos identificados em todos patovares pelo OrthoMCL diferem do número dos mesmos identificados pelo GET_HOMOLOGUES, sendo este algoritmo responsável pela identificação de 3480 grupos consensos para o genoma núcleo. Essa divergência de predição do OrthoMCL em si e quando o mesmo estava integrado ao GET_HOMOLOGUES demonstram a importância da comparação entre algoritmos para gerar grupos ortólogos mais consistentes.

A anotação funcional dos grupos ortólogos do *core* genoma é essencial para que se possa determinar quais processos celulares e metabólicos são compartilhados entre os patovares. Foram encontrados, majoritariamente, grupos

com nenhuma função relacionada, sugerindo que os genes em questão não possuem similaridade com o genes do banco de dados do COG, ou que a função do gene ainda não tenha sido elucidada. A mesma análises do genoma acessório deve ser feita, futuramente, a fim de determinar quais são as funções dos genes que não são compartilhados entre os patovares, determinando, assim, quais características são exclusivas, qual a consequência e investigar possíveis aplicações para esses genes.

A análise metabólica do *core* genoma revelou grande parte da predição foi relacionada a vias de metabolismo, o que significa que a maior parte dos genes presentes no genoma núcleo estão relacionados à rotas de catabolismo e anabolismo, por exemplo. Kim *et al.* (2017) relatou o mesmo resultado através do estudo do pangenoma de *Bacillus*, demonstrando que, majoritariamente, os grupos ortólogos do genoma núcleo estão relacionados à rotas de metabolismo celular. Assim como Meng *et al.* (2017) demonstrou por meio do pangenoma de *Streptococcus mutans* a presença, predominantemente, de rotas ligadas ao metabolismo, as quais representaram 35,6% dos grupos ortólogos do *core* genoma. As predições que não apresentaram hit, também, estão em grande quantidade. Após esse, a predição para nenhum processo metabólica representa uma considerável fração do *core* genoma. Em relação ao estudo de cada patovar quanto a caracterização de rotas metabólicas, houve regularidade quando a presença de rotas de processos celulares, fisiologia, metabolismo, processamento de informações ambientais, processamento de informações genéticas, significando, provavelmente, que são processo comuns aos patovares. Em contrapartida, as predições que resultaram em nenhum hit ou nenhum processo metabólico mostram haver diferença entre os patovares, isso pode estar associado a genes exclusivos de cada isolado, genes relacionados à patogenicidade e especificidade ao hospedeiros.

Os 12 genes *gumBCDEFGHIJKLM* e os genes *xagA*, *xagB*, *xagC*, *xanA* e *xanB* presentes em todos os patovares inferem que todas as cepas são capazes de produzir goma xantana, além de produzir, provavelmente, com a mesma eficiência, uma vez que Tao; Swarup & Zhang (2010) demonstraram que *xagABC* são importantes na regulação da produção, e na ausência desses, a biossíntese de

xantana apresentou menor rendimento. Entretanto, os genes *pgaABC* estão ausentes nos patovares *raphani* e *incanae*, em vista disso, a funcionalidade e o processo de biossíntese da xantana estariam prejudicados. Da mesma forma, a ausência de *pgaD* em *musacearum*, *incanae*, *raphani*, *X. campestris* strain 17, *arecae* e *centellae*, estaria interferindo na biossíntese e qualidade da xantana. Segundo essa análise, somente os patovares *viticola*, *thespesiae*, *vitiswoodrowii*, *vitistrifoliae*, *vitiscarnosae*, *durantae*, *azadirachtae* e *leeana* apresentam potencial para a produção de xantana.

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 podem haver exclusivas aos patovares. Nesse sentido, a busca por essas proteínas nos genomas analisados de *X. campestris* permitiu identificar em 95% do genomas 172 clusters de proteínas efetoras, no entanto quando foi feito o Blastp dessas proteínas preditas, não houveram proteínas efetoras preditas pelo EffectiveDB, sugerindo que o *software* não apresenta calibragem para o gênero *Xanthomonas*, pois Collingro *et al.* 2017 através da mesma ferramenta conseguiu prever proteínas efetoras da *Chlamydia trachomatis*. Sendo assim, há a necessidade de realizar uma nova análise utilizando outras ferramentas de predição de proteínas efetoras, a fim de gerar uma identificação mais acurada.

A ausência dos fatores de virulência codificados pelos genes *rpfCFN*, provavelmente, acarretaria na falta de regulação de fatores de virulência nos patovares *incanae*, *vitistrifoliae*, *vitiswoodrowii*, *centellae*, *raphani* e *azadirachtae*, uma vez que demonstraram não possuir esses genes. Já a ausência de *hpaCH* em *incanae*, *musacearum*, *raphani*, *arecae* e *X. campestris* strain 17, pode significar uma má funcionalidade do T3SS, pois esses dois genes fazem parte do sistema de secreção auxiliando na translocação de proteínas efetoras. Da mesma forma *hrcQ*, *hrpG* e *xopF1* ausentes em *raphani* e *incanae* influenciam na efetividade do T3SS. Assim como, *hrpB2* ausente em cepa 17 de *X. campestris* acarretaria em uma

deficiência do T3SS. Por fim, *pehA* ausente em *incanae*, *vitistrifoliae*, *vitiscarnosae*, *raphani* e em *X. campestris strain 17* representa uma perda de regulação dos genes *rpf* e *clp*, significando, provavel, deficiência no T3SS.

As próximas etapas consistem em identificar, de forma mais detalhada, os processos metabólicos. Assim como, será necessário utilizar outras abordagens para a identificação proteínas efetoras em *X. campestris*. Para os genes de virulência, análises experimentais serão necessárias para determinar se a ausência de determinados genes estão relacionadas à especificidade ao hospedeiro e doença. Além disso, será necessário utilizar de métodos experimentais para determinar a qualidade e produção da goma xantana nos patovares.

7. Conclusão

A partir da abordagem utilizada foi possível identificar rotas metabólicas para os patovares de *X. campestris*, assim como para o genoma núcleo. A busca por genes da rota de biossíntese de xantana revelou que somente os patovares *viticola*, *thespesiae*, *vitiswoodrowii*, *vitistrifoliae*, *vitiscarnosae*, *durantae*, *azadirachtae* e *leena* apresentam potencial para a produção de xantana. Proteínas efetoras não foram preditas, e a presença de genes de virulência apresentam variações entre os patovares analisados.

8. Referências

- ALLEN, E. E.; BANFIELD, J. F. Community genomics in microbial ecology and evolution. **Nature Reviews Microbiology**, 2005. v. 3, n. 6, p. 489–498.
- ALTSCHUL, S.F. *et al.* Basic local alignment search tool. **Journal of molecular biology**, 1990 v. 215(3), pp.403–10.
- BANSAL, K., MIDHA, S. AND PATIL, P. B. No Title. **Unpublished**, 2015.
- BAO, W., LEUNG, F.C.C. AND JIANG, J. Complete Genome Sequence of *Xanthomonas campestris* Strain 17, a Lignin-degrading Bacterium Isolated from Forest Soil. **Unpublished**, 2015.
- BOGDANOVA, A. J. *et al.* Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. **Journal of Bacteriology**, 2011. v. 193, n. 19, p. 5450–5464.
- CAMACHO, C. *et al.* BLAST+: architecture and applications. **BMC bioinformatics**, 2009 v. 10(1), p.421.
- CHANG, J. Wake of the flood: ascribing functions to the wave of type III effector proteins of phytopathogenic bacteria. **Current Opinion in Microbiology**, 2004. v. 7, n. 1, p. 11–18.
- CHEN, L. *et al.* XC_0531 encodes a c-type cytochrome biogenesis protein and is required for pathogenesis in *Xanthomonas campestris* pv. *campestris*. **BMC Microbiology**, 2017 v. 17 p. 142.
- CHIANG, Y.-C. *et al.* Functional characterization and transcriptional analysis of *icd2* gene encoding an isocitrate dehydrogenase of *Xanthomonas campestris* pv. *campestris*. **Archives of microbiology**, 2017. v. 199, n. 6, p. 917–929.
- CIANCIOFFO, N. P.; WHITE, R. C. Expanding Role of Type II Secretion in Bacterial Pathogenesis and Beyond. **Infection and Immunity**, 2017. v. 85, n. 5, p. e00014-17.
- COLLINGRO, A. *et al.* Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. **The ISME Journal**, 2017. v. 11, n. 10, p. 2334–2344.
- CONTRERAS-MOREIRA, B.; VINUESA, P. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. **Applied and Environmental Microbiology**, 2013 v. 79 p. 7696 –7701.

- DARRASSE, A. *et al.* Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. **BMC Genomics**, 2013. v. 14, n. 1, p. 761.
- DELCHER, A. L.; SALZBERG, S. L.; PHILLIPPY, A. M. Using MUMmer to identify similar regions in large sequence sets. Hoboken, NJ, USA: **Current protocols in bioinformatics**, 2003. v. Chapter 10, p. Unit 10.3.
- DOW, J. M. *et al.* Extracellular proteases from *Xanthomonas campestris* pv. *campestris*, the black rot pathogen. **Applied and environmental microbiology**, 1990. v. 56, n. 10, p. 2994–8.
- EBERHARDT, R.Y. *et al.* AntiFam: a tool to help identify spurious ORFs in protein annotation. **Database : the journal of biological databases and curation**, 2012, p.bas003.
- EICHINGER, V. *et al.* EffectiveDB--updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. **Nucleic acids research**, 2016. v. 44, n. D1, p. D669-74.
- EDDY, S.R. Accelerated Profile HMM Searches. W. R. Pearson. ed. **PLoS computational biology**, 2011 v. 7(10), p.e1002195.
- ERIC MAINA MAGEMBE, B. *et al.* A genomics approach to the development of molecular diagnostic tools for banana *Xanthomonas* wilt, gray leaf. 2014.
- FIELDS, S. MOLECULAR BIOLOGY: Site-Seeing by Sequencing. **Science**, 2007. v. 316, n. 5830, p. 1441–1442.
- FISCHER, S. *et al.* Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. **Current Protocols in Bioinformatics**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.
- FU, L. *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics**, 2012. v. 28, n. 23, p. 3150–3152.
- GERLACH, R.G. & HENSEL, M. Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. **Int. J. of Med. Microbiology**, 2007 v. 297:401-15.
- GLASNER, J. D. *et al.* Niche-Specificity and the Variable Fraction of the *Pectobacterium* Pan-Genome. **Molecular Plant-Microbe Interactions**, 2008. v. 21, n. 12, p. 1549–1560.
- GOUGH, C.L. DOW, J. M. *et al.* Nucleotide sequence of the engXCA gene encoding the major endoglucanase of *Xanthomonas campestris* pv. *campestris*. **Gene**, 1990. v. 89, n. 1, p. 53–59.

GRIFFITHS-JONES, S. *et al.* Rfam: an RNA family database. **Nucleic acids research**, 2003. v. 31(1), pp.439–41.

GÜRLEBECK, D.; THIEME, F.; BONAS, U. Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant. **Journal of Plant Physiology**, 2006. v. 163, n. 3, p. 233–255.

HANDELSMAN, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. **Microbiology and Molecular Biology Reviews**, 2004. v. 68, n. 4, p. 669–685.

HAUSNER, J. *et al.* 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**, 2017. v. 85, n. 2, p. e00788-16.

HAUSNER, J.; BUTTNER, D. The YscU/FlhB homologue HrcU from *Xanthomonas* controls type III secretion and translocation of early and late substrates. **Microbiology**, 2014. v. 160, n. Pt_3, p. 576–588.

HE, Y.-W.; ZHANG, L.-H. Quorum sensing and virulence regulation in *Xanthomonas campestris*. **FEMS Microbiology Reviews**, 2008. v. 32, n. 5, p. 842–857.

HSIAO, Y.-M. *et al.* Regulation of the *pehA* gene encoding the major polygalacturonase of *Xanthomonas campestris* by Clp and RpfF. **Microbiology (Reading, England)**, 2008. v. 154, n. Pt 3, p. 705–13.

HYATT, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC bioinformatics**, 2010. v. 11, p.119.

IELPI, L., COUCO, R.O., DARKERT, M.A. Sequential assembly and polymerization of the polyprenol-linked pentasaccharide repeating unit of the xanthan polysaccharide in *Xanthomonas campestris*. **J. Bacteriol**, 1993. v. 175, p. 2490–2500.

JACQUES, M.-A. *et al.* Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. **Annual Review of Phytopathology**, 2016. v. 54, n. 1, p. 163–187.

JANSEN, R. C. ; NAP, J.P. Genetical genomics: the added value from segregation. **Trends in Genetics**, 2001. v. 17, n. 7, p. 388–391.

KANEHISA, M.; SATO, Y.; MORISHIMA, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. **Journal of molecular biology**, 2016. v. 428, n. 4, p. 726–731.

KATZEN, F. *et al.* *Xanthomonas campestris* pv. *campestris* gum mutants: effects on xanthan biosynthesis and plant virulence. **Journal of bacteriology**, 1998. v. 180, n. 7, p. 1607–17.

KIM, N. H.; CHOI, H. W.; HWANG, B. K. *Xanthomonas campestris* pv. *vesicatoria* Effector AvrBsT Induces Cell Death in Pepper, but Suppresses Defense Responses in Tomato. **Molecular Plant-Microbe Interactions**, 2010. v. 23, n. 8, p. 1069–1082.

KIM, Y. *et al.* Pan-genome analysis of *Bacillus* for microbiome profiling. **Scientific reports**, 2017. v. 7, n. 1, p. 10984.

KREMER, F. S.; ESLABÃO, M. R.; DELLAGOSTIN, O. A., PINTO, L. S. Genix: a new online automated pipeline for bacterial genome annotation. **FEMS Microbiologist Letters**, 2016. v. 363 (23).

KRISTENSEN, D. M. *et al.* A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. **Bioinformatics (Oxford, England)**, 2010. v. 26, n. 12, p. 1481–7.

LAGESEN, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. **Nucleic acids research**, 2007 v. 35(9), pp.3100–8.

LAINING, C. *et al.* Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. **BMC Bioinformatics**, 2010. v. 11, n. 1, p. 461.

LASLETT, D. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. **Nucleic Acids Research**, 2004 v. 32(1), pp.11–16.

LI, W. & GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics (Oxford, England)**, 2006. v. 22(13), pp.1658–9.

LIAO, C.-T. *et al.* 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**, 2014. v. 196, n. 10, p. 729–738.

LIAO, C.-T. *et al.* Functional characterization and transcriptome analysis reveal multiple roles for *prc* in the pathogenicity of the black rot pathogen *Xanthomonas campestris* s pv. *campestris*. **Research in Microbiology**, 2016. v. 167, n. 4, p. 299–312.

LIN, C.S., LIN, N.T., YANG, B.Y., WENG, S.F., TSENG, Y.H. Nucleotide sequence and expression of UDP-glucose dehydrogenase gene required for the synthesis of xanthan gum in *Xanthomonas campestris*. **Biochem. Biophys. Res. Commun**, 1995. v. 207, p. 223–230.

- LIU, L. *et al.* The type III effector AvrXccB in *Xanthomonas campestris* pv. *campestris* targets putative methyltransferases and suppresses innate immunity in Arabidopsis. **Molecular Plant Pathology**, 2017. v. 18, n. 6, p. 768–782.
- LOCKHART, D. J.; WINZELER, E. A. Genomics, gene expression and DNA arrays. **Nature**, 2000. v. 405, n. 6788, p. 827–836.
- LONJON, F. *et al.* HpaB-Dependent Secretion of Type III Effectors in the Plant Pathogens *Ralstonia solanacearum* and *Xanthomonas campestris* pv. *vesicatoria*. **Scientific Reports**, 2017. v. 7, n. 1, p. 4879.
- LORENZ, C.; HAUSNER, J.; BÜTTNER, D. HrcQ Provides a Docking Site for Early and Late Type III Secretion Substrates from *Xanthomonas*. **PLoS ONE**, 2012. v. 7, n. 11, p. e51063.
- LOWE, T.M. & EDDY, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic acids research**, 1997 v. 25(5), pp.955–64.
- MARSCHALL, T. *et al.* Computational Pan-Genomics: Status, Promises and Challenges. **BioRXIV**, 2016. p. 43430.
- MCINERNEY, M. J. *et al.* Physiology, Ecology, Phylogeny, and Genomics of Microorganisms Capable of Syntrophic Metabolism. **Annals of the New York Academy of Sciences**, 2008. v. 1125, n. 1, p. 58–72.
- MENG, P. *et al.* Exploring the Genomic Diversity and Cariogenic Differences of *Streptococcus mutans* Strains Through Pan-Genome and Comparative Genome Analysis. **Current Microbiology**, 2017. v. 74, n. 10, p. 1200–1209.
- MIDHA, S.; PATIL, P. B. Genomic insights into the evolutionary origin of *Xanthomonas axonopodis* pv. *citri* and its ecological relatives. **Applied and Environmental Microbiology**, 2014. v. 80, n. 20, p. 6266–6279.
- MINGZHI, L. I.; ZILING, S. U. N.; YONGQUAN, L. I. *Xanthomonas campestris*. **Ratio**, 2007. v. 15, n. 30370939, p. 639–642.
- MOROZOVA, O.; MARRA, A. M. Applications of next-generation sequencing technologies in functional genomics. **Genomics**, 2008. v. 92, n. 5, p. 255–264.
- MUZZI, A.; DONATI, C. Population genetics and evolution of the pan-genome of *Streptococcus pneumoniae*. **International Journal of Medical Microbiology**, 2011. v. 301, n. 8, p. 619–622.
- NAKATO, V.; MAHUKU, G.; COUTINHO, T. *Xanthomonas campestris* pv. *musacearum* : a major constraint to banana, plantain and enset production in Central and East Africa over the past decade. **Molecular Plant Pathology**, 2017.

- NAWROCKI, E.P., KOLBE, D.L. & EDDY, S.R. Infernal 1.0: inference of RNA alignments. **Bioinformatics (Oxford, England)**, 2009 v. 25(10), p.1335–7.
- NOH, Y.-H. *et al.* Expression of colSR Genes Increased in the rpf Mutants of *Xanthomonas oryzae pv. oryzae* KACC10859. **The Plant Pathology Journal**, 2014. v. 30, n. 3, p. 304–309.
- OLIVEIRA, R. P *et al.* **Cancro cítrico: epidemiologia e controle**. Pelotas: Embrapa Clima Temperado, 2008. v. 234.
- PARKINSON, N. *et al.* Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. **International Journal of Systematic and Evolutionary Microbiology**, 2007. v. 57,p. 2881–2887.
- QIAN, W. *et al.* Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris pv. campestris*. **Genome research**, 2005. v. 15, n. 6, p. 757–67.
- RASHID, M. M.; IKAWA, Y.; TSUGE, S. 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**, 2016. v. 82, n. 13, p. 3947–3958.
- RODEN, J. A. *et al.* A genetic screen to isolate type III effectors translocated into pepper cells during *Xanthomonas* infection. **Proceedings of the National Academy of Sciences of the United States of America**, 2004. v. 101, n. 47, p. 16624–9.
- RODRIGUEZ-R, L. M. *et al.* Genomes-based phylogeny of the genus *Xanthomonas*. **BMC Microbiology**, 2012. v. 12, n. 1, p. 43.
- ROYER, M. *et al.* Genome mining reveals the genus *Xanthomonas* to be a promising reservoir for new bioactive non-ribosomally synthesized peptides. **BMC genomics**, 2013. v. 14, p. 658.
- ROUX, B. *et al.* Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. **BMC genomics**, 2015. v. 16, n. 1, p. 975.
- RYAN, R. P. *et al.* Passing GO (gene ontology) in plant pathogen biology: a report from the *Xanthomonas* Genomics Conference. **Cellular Microbiology**, 2009. v. 11, n. 12, p. 1689–1696.
- RYAN, R. P. *et al.* Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. **Nature Reviews Microbiology**, 2011. v. 9, n. 5, p. 344–355.

SCHATTSCHNEIDER, S. *et al.* Systems and synthetic biology perspective of the versatile plant-pathogenic and polysaccharide-producing bacterium *Xanthomonas campestris*. **Microbiology**, 2017. v. 163, n. 8, p. 1117–1144.

SANTOS, C. *et al.* Differential accumulation of *Xanthomonas campestris* pv. *campestris* proteins during the interaction with the host plant: Contributions of an in vivo system. **PROTEOMICS**, 2017. v. 17, n. 12, p. 1700086.

SAPP, M. AND STUDHOLME, D. J. Genome sequencing of *Xanthomonas vasicola* and closely related strains. **Unpublished**, 2014.

SCHEIBNER, F. *et al.* The type III secretion chaperone HpaB controls the translocation of effector and non-effector proteins from *Xanthomonas campestris* pv. *vesicatoria*. **Molecular Plant-Microbe Interactions**, 2017. p. MPMI-06-17-0138-R.

SILVA, A. M. F. *et al.* Sobrevivência de *Xanthomonas campestris* pv. *viticola* em tecido infectado de videira. **Revista Brasileira de Fruticultura**, 2012. v. 34, n. 3, p. 757–765.

STAMATOYANNOPOULOS, J. A. The genomics of gene expression. **Genomics**, 2004. v. 84, n. 3, p. 449–457.

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. **Carbohydr. Res**, 1993. v. 241, p. 321–326.

STEFFENS, T. *et al.* The influence of a modified lipopolysaccharide O-antigen on the biosynthesis of xanthan in *Xanthomonas campestris* pv. *campestris* B100. **BMC Microbiology**, 2016, v. 16:93.

TAO, F.; SWARUP, S.; ZHANG, L.-H. Quorum sensing modulation of a putative glycosyltransferase gene cluster essential for *Xanthomonas campestris* biofilm formation. **Environmental Microbiology**, 2010. v. 12, n. 12, p. 3159–3170.

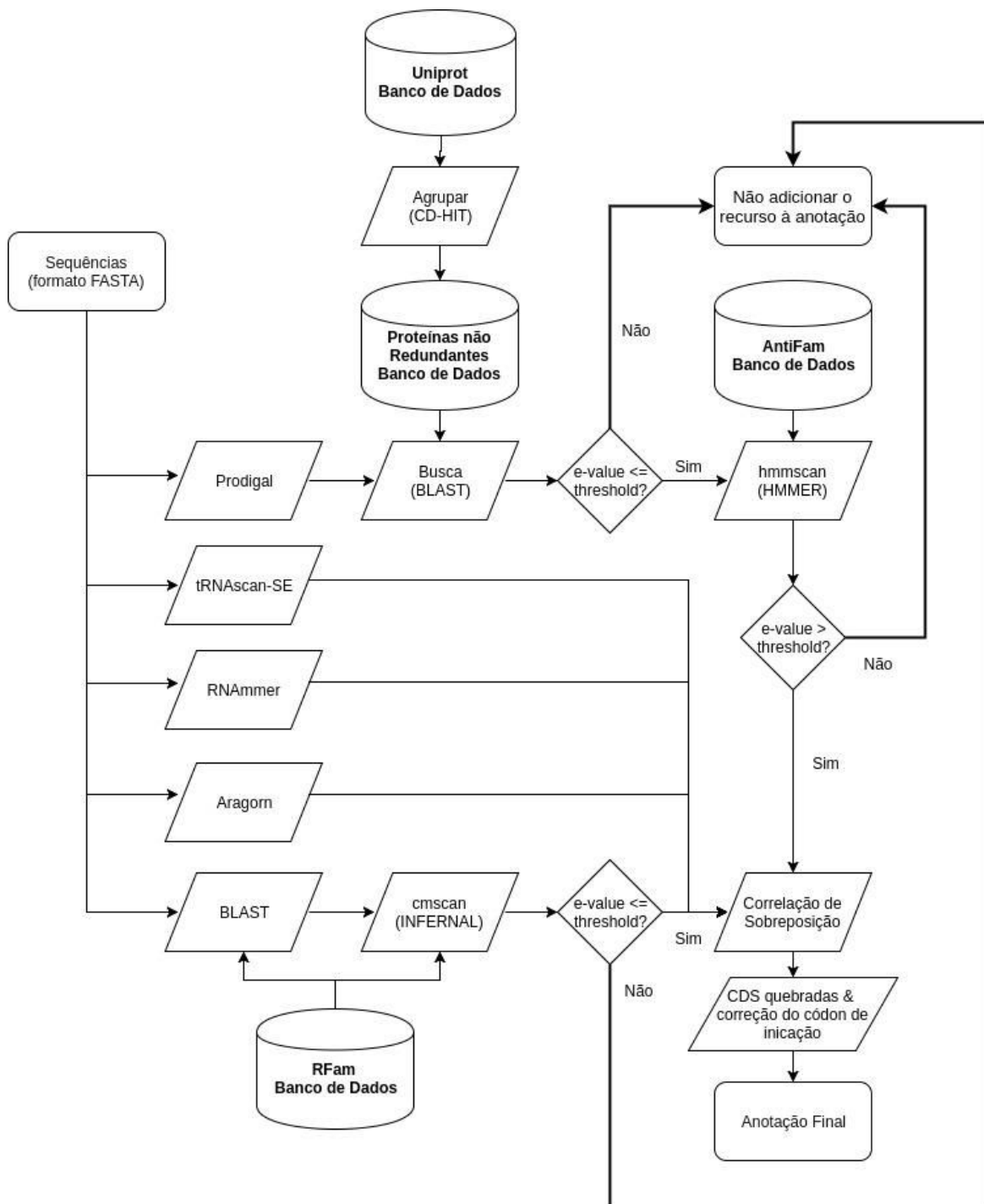
TAO, F. *et al.* Genome sequence of *Xanthomonas campestris* JX, an industrially productive strain for Xanthan gum. **Journal of bacteriology**, 2012. v. 194, n. 17, p. 4755–6.

TATUSOV, R. L. *et al.* The COG database: a tool for genome-scale analysis of protein functions and evolution. **Nucleic Acids Research**, 2000. v. 28, n. 1, p. 33–36.

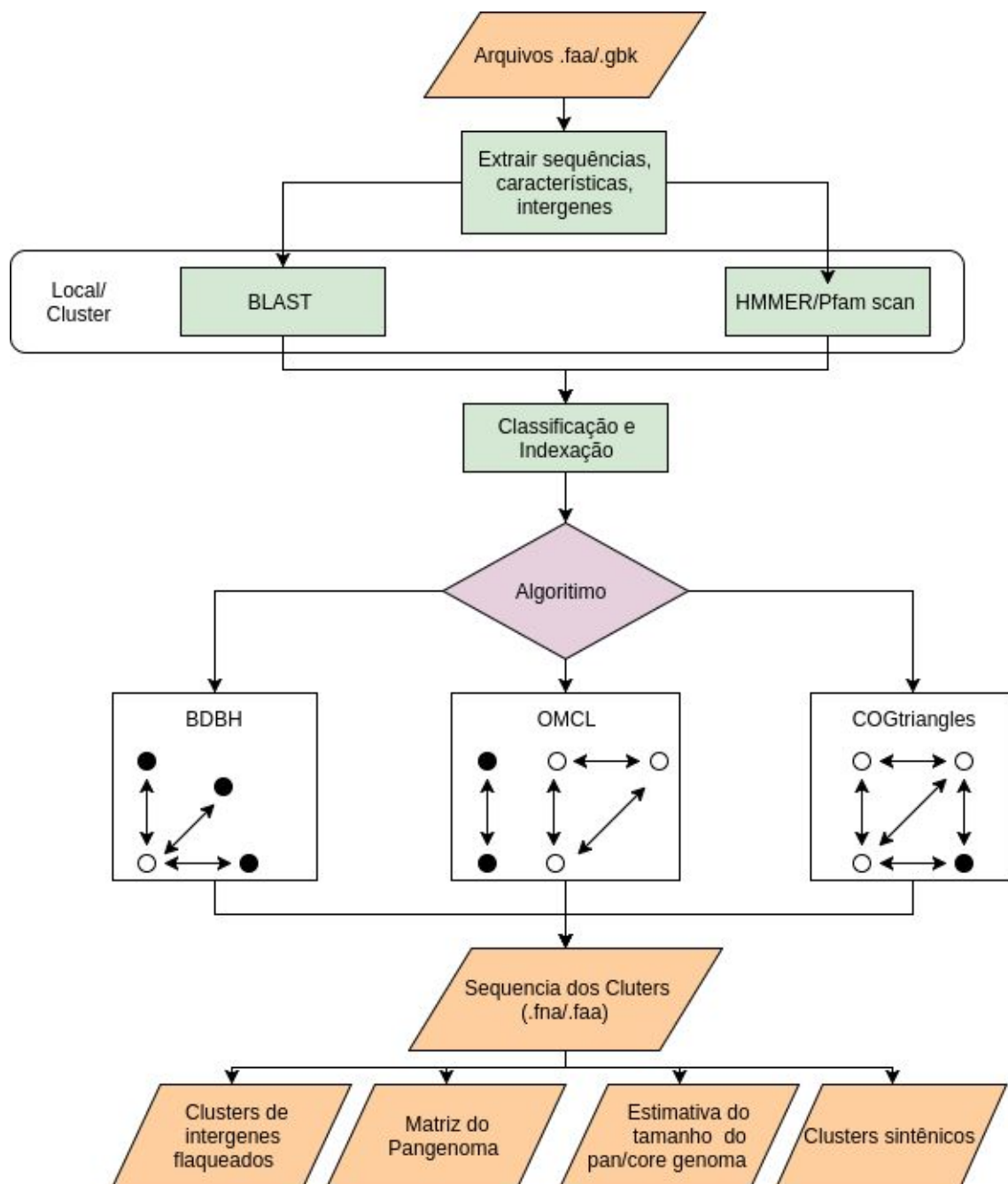
TRINDADE, L. C. Da *et al.* Development of a molecular method for detection and identification of *Xanthomonas campestris* pv. *viticola*. **Summa Phytopathologica**, 2007. v. 33, n. 1, p. 16–23.

- TRIPATHI, L. *et al.* Expression of sweet pepper Hrap gene in banana enhances resistance to *Xanthomonas campestris* pv. *musacearum*. **Molecular Plant Pathology**, 2010. v. 11, n. 6, p. 721–731.
- UDAONDO, Z. *et al.* Analysis of the core genome and pangenome of *Pseudomonas putida*. **Environmental Microbiology**, 2016. v. 18, n. 10, p. 3268–3283.
- VICENTE, J. G.; EVERETT, B.; ROBERTS, S. J. Identification of Isolates that Cause a Leaf Spot Disease of Brassicas as *Xanthomonas campestris* pv. *raphani* and Pathogenic and Genetic Comparison with Related Pathovars. **Bacteriology**, 2006. v. 96, n. 7.
- VICENTE, J. G.; HOLUB, E. B. *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**, 2013. v. 14, n. 1, p. 2–18.
- VOJNOV, A.A., Slater, H., Daniels, M.J., Dow, J.M. Expression of the gum operon directing xanthan biosynthesis in *Xanthomonas campestris* and its regulation in planta. **Mol. Plant Microbe Interact**, 2001. v. 14, p. 768–774.
- WANG, X.; PRESTON, J. F.; ROMEO, T. The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. **Journal of bacteriology**, 2004. v. 186, n. 9, p. 2724–34.
- WASUKIRA, A. *et al.* Genome-Wide Sequencing Reveals Two Major Sub-Lineages in the Genetically Monomorphic Pathogen *Xanthomonas campestris* Pathovar *musacearum*. **Genes**, 2012. v. 3, n. 4, p. 361–37.
- WIGGERICH, H.-G.; PÜHLER, A. 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**, 2000. v. 146, n. 5, p. 1053–1060.
- WOLD, B.; MYERS, R. M. Sequence census methods for functional genomics. **Nature Methods**, 2007. v. 5, n. 1, p. 19–21.
- WOYKE, T. *et al.* One Bacterial Cell, One Complete Genome. **One Bacterial Cell One Complete Genome. PLoS ONE**, 2010. v. 5, n. 4.
- YU, C. *et al.* 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 (New York, N.Y.)**, 2014. v. 7, n. 1, p. 20.
- YUAN, Z. *et al.* Genetic and proteomic analyses of a *Xanthomonas campestris* pv. *campestris* *purC* mutant deficient in purine biosynthesis and virulence. **Journal of genetics and genomics**, 2013. v. 40, n. 9, p. 473–87.

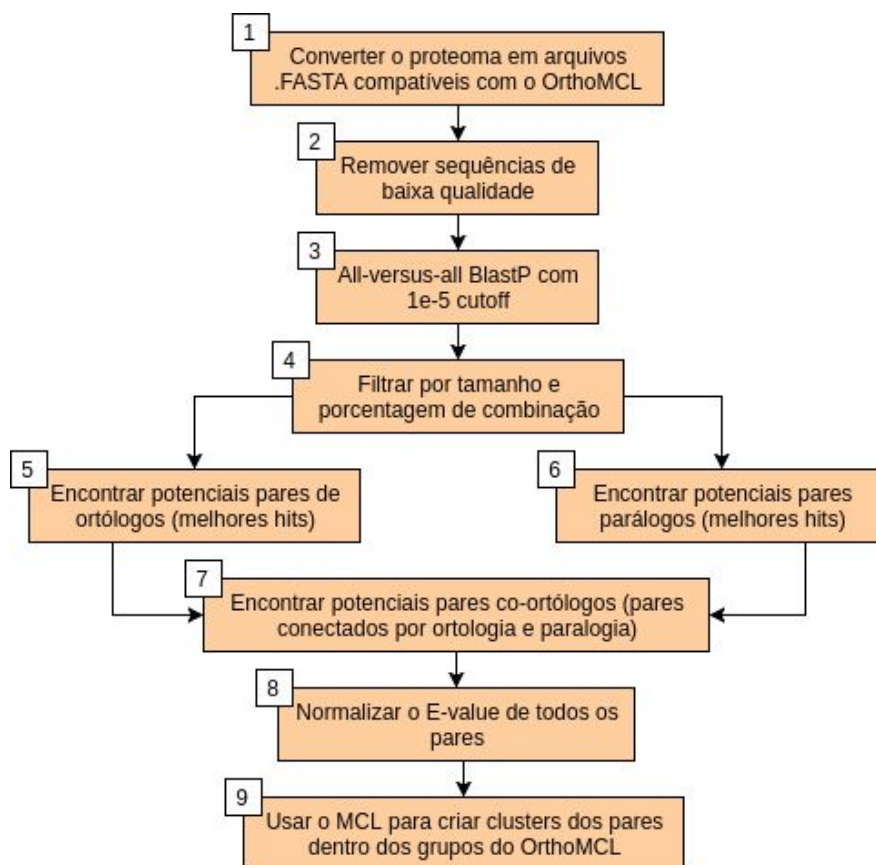
9. Material complementar



Material complementar 1. Pipeline do software GENIX. Adaptado de KREMER *et al.*, 2016.



Material complementar 2. Visão geral do algoritmo do software GET_HOMOLOGUES. Adaptado de (CONTRERAS-MOREIRA; VINUESA, 2013).



Material complementar 3. 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 ortólogas; (6) Selecionar os melhores hits a partir de todos os pares de proteínas parálogas; (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.

Material Complementar 4. Proteínas efetoras presentes em todos os patovares de *X. campestris* analisados.

| Código do Cluster | Produto |
|--------------------------------|--|
| 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 guanyltransferase |
| 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 FigG |
| Cluster 2388_2315019-2315427-1 | Flagellar basal-body rod protein FigC |
| 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 | Lacl 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 methyltransferase |
| 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-dehydroquinase synthase |
| Cluster2551_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 HsIV |
| 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 | Acyltransferase |
| 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 |

