



UNIVERSIDADE ESTADUAL DE SANTA CRUZ
PRO-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL
EM CIÊNCIA E TECNOLOGIA

LUIS AUGUSTO FRANCO LÓPEZ

RECONSTRUÇÃO DE ROTAS METABÓLICAS DE BACTÉRIAS DO GÊNERO *Klebsiella*
SPP. COMO FERRAMENTA NO APRIMORAMENTO DO CONTROLE BIOLÓGICO DA
MOSCA *Ceratitis capitata* (MEDFLY)

ILHÉUS-BA
2020

LUIS AUGUSTO FRANCO LÓPEZ

**RECONSTRUÇÃO DE ROTAS METABÓLICAS DE
BACTÉRIAS DO GÊNERO *Klebsiella* SPP. COMO
FERRAMENTA NO APRIMORAMENTO DO CONTROLE
BIOLÓGICO DA MOSCA *Ceratitis capitata* (MEDFLY)**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional em Ciência e Tecnologia da Universidade Estadual de Santa Cruz, como parte das exigências para obtenção do título de Mestre em Modelagem Computacional em Ciência e Tecnologia.

Orientador: Prof. Dr. César Alberto Bravo Pariente

ILHÉUS-BA
2020

L864

Lópes, Luis Augusto Franco.

Reconstrução de rotas metabólicas de bactérias do gênero *Klebsiella* spp. como ferramenta no aprimoramento do controle biológico da mosca *Ceratitis capitata* (medfly) / Luis Augusto Franco Lópes. – Ilhéus, BA: UESC, 2020.

51 f.: il.

Orientador: César Alberto Bravo Pariente.

Dissertação (Mestrado) – Universidade Estadual Santa Cruz. Programa de Pós-Graduação em Modelagem Computacional em Ciência e Tecnologia.

Inclui referências e apêndices.

1. Biotecnologia. 2. Mosca-do-mediterrâneo. 3. Insetos nocivos – Controle biológico. 4. *Klebsiella*. 5. Modelos matemáticos. 6. Simbiose. I. Título.

CDD 660.6

LUIS AUGUSTO FRANCO LÓPEZ

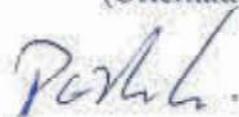
RECONSTRUÇÃO DE ROTAS METABÓLICAS DE
BACTÉRIAS DO GÊNERO *Klebsiella* SPP. COMO
FERRAMENTA NO APRIMORAMENTO DO CONTROLE
BIOLÓGICO DA MOSCA *Ceratitis capitata* (MEDFLY)

Ilhéus-BA, 10/02/2020

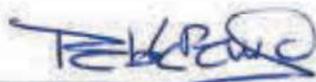
Comissão Examinadora



Prof. Dr. César Alberto Bravo Pariente
UESC
(Orientador)



Prof. Dr. Paulo Eduardo Ambrósio
UESC



Prof. Dr. Pablo Ivan Pereira Ramos
FIOCRUZ

Dedico este trabalho a Deus, aos meus pais
e ao meu irmão.

Agradecimentos

- A Deus, pela sabedoria e iluminação para realizar esta dissertação.
- Ao Programa das Organizações dos Estados Americanos (OEA), pela oportunidade de participar das bolsas para estudos de Pós-Graduação nas universidades do Brasil.
- À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro.
- Ao Programa de Pós-Graduação em Modelagem Computacional em Ciência e Tecnologia de Universidade Estadual de Santa Cruz, pela formação e apoio nesta etapa de discente.
- Ao meu Orientador Dr. César Alberto Bravo Pariente, por aceitar-me como seu aluno, pelas suas ideias, paciência, dedicação e apoio no caminho da formação no Mestrado.
- Aos professores e discentes do PPGMC e aos meus amigos da turma 2018, pelo apoio e sugestões nas disciplinas e na realização da dissertação. Por sempre falar devagar, nos primeiros meses que entendia pouco o português.
- Ao pessoal do NBCGIB, pelo espaço e apoio ao nosso trabalho dissertação.
- Aos membros da banca examinadora por aceitar o convite e contribuir para o enriquecimento do trabalho.
- Aos meus pais, Luis Fco. e Julieta , e ao meu irmão, José, pelo amor e apoio constantes apesar da distância.
- À minha namorada, Gabriela, eu agradeço pelo amor, pela dedicação, pela compreensão, pelo incentivo e pela solidariedade.
- À família de amigos hispanofalantes que me adotou na chegada no Brasil.
- Ao amigo da turma Frederico Chaves pelo apoio e pelas inúmeras contribuições.

"Let me tell you the secret that has led me to my goal. My strength lies solely in my tenacity."

Louis Pasteur

Reconstrução de rotas metabólicas de bactérias do gênero *Klebsiella* spp. como ferramenta no aprimoramento do controle biológico da mosca *Ceratitis capitata* (medfly)

Resumo

Os cientistas têm desenvolvido vários métodos no controle de pragas, como o controle químico com inseticidas e como o controle biológico com a técnica do inseto estéril (SIT, pelas suas siglas em inglês *Sterile Insect Technique*) baseada na produção de machos inférteis produzidos por irradiação e liberados posteriormente no meio ambiente. Atualmente, a técnica SIT é utilizada no controle de população de *Ceratitis capitata* e essa técnica é complementada com pesquisa da sua microbiota. Baseada nessa abordagem é utilizada a identificação dos mecanismos moleculares nas bactérias que formam essa microbiota da mosca utilizando a técnica de reconstrução de rotas metabólicas dessas bactérias, para avaliar possível relação simbiótica entre hospedeiro e bactéria. Neste processo simbiótico entre a mosca e as bactérias a metabolização do nitrogênio é realizado pela bactéria posto que a mosca não é capaz de metabolizar o nitrogênio. Baseado nessa informação nesta dissertação nós focamos na utilização da técnica de reconstrução das rotas metabólicas utilizando uma técnica não usual baseada na teoria de grafos para identificar rotas envolvidas na metabolização do nitrogênio e identificar esses mecanismos moleculares entre *C. capitata* e *Klebsiella* a partir dos dados sequenciamento das bactérias presentes na microbiota da mosca. As sequências foram processadas pela metodologia proposta nesta dissertação obtendo dez rotas metabólicas, seis rotas com L-Glutamato como produto, três com Amônia como produto e uma rota metabólica que não foi utilizada nesta pesquisa posto que não produz produtos metabólicos finais. Os principais produtos metabólicos utilizados pela mosca são o L-Glutamato e Amônia esses produtos são necessários na síntese de novas proteínas no processo de metamorfose da mosca. Concluímos desde o ponto de vista da metabolização do nitrogênio não há diferença entre as cepas. Com base nessas informações, analisamos e avaliamos a análises das rotas metabólicas com EM. As considerações acima discutidas sugerem que podemos identificar possíveis rotas metabólicas envolvidas no processo de simbiose entre *Klebsiella* e *C. capitata* utilizando a metodologia proposta de reconstrução de rotas metabólicas baseada na teoria de grafos, que pode ser utilizada para melhorar a produção de machos inférteis de *C. capitata* para a liberação e controle populacional.

Palavras-chave: *Ceratitis capitata*, *Klebsiella*, Reconstrução rotas metabólicas e Simbiose.

Reconstruction of metabolic pathways of *Klebsiella* spp. bacteria as a tool improve the biologic control of medfly fly (*Ceratitis capitata*)

Abstract

Nowadays, scientists develop multiple methods to control pests, like chemical control as the insecticides and biological control as the Sterile Insect Technique (SIT) to release infertile males produced by irradiation. This technique is complemented with research in *Ceratitis capitata* microbiota, this approach is based on the identification of molecular mechanisms using the reconstruction of metabolic pathways to improve the SIT technique. We focus on the reconstruction of metabolic pathways involved in nitrogen metabolization because the bacteria are in charge of metabolizing the nitrogen compounds and make available for the medfly. We propose a pipeline based on graph theory to process the genome sequences of the three *Klebsiella* bacteria, reconstruct metabolic pathways and identify molecular mechanisms. The result of processing the biological data of the three strains of *Klebsiella* were ten pathways involved in the symbiosis process between *Klebsiella* and medfly. The results of the pathways analyses present six pathways where the L-Glutamate was a final product, three pathways with ammonia product and one pathway was discarded because it was an internal pathway and don't generate a metabolic product. Those pathways corresponding to L-Glutamate and ammonia metabolic products are part of nitrogen metabolism in *Klebsiella*, which shows that this metabolic process is one of the molecular mechanisms involved in symbiosis process between *Klebsiella* and *C. capitata*. We conclude based on the nitrogen metabolization that doesn't exist a difference in the three strains of *Klebsiella*. Finally, by the results of the sequences analysis, in the medfly's metamorphose the L-Glutamate and Ammonia are used in the synthesis of new proteins and those are products of nitrogen metabolization. The final consideration suggest that the pipeline propose can be used as first step to identify molecular mechanisms that can improve the production of industrial *C. capitata*.

Keywords: *Ceratitis capitata*, *Klebsiella oxytoca*, Computational modeling, Stoichiometric modeling, Symbiosis, Elementary modes.

Lista de figuras

Figura 1 – Esquema das bases técnicas da dissertação	3
Figura 2 – Dogma central da biologia molecular	4
Figura 3 – Controle de pragas utilizando a técnica do inseto estéril	9
Figura 4 – Taxa metabólica da metamorfose de <i>Ceratitis capitata</i>	11
Figura 5 – Taxa metabólica da metamorfose de <i>Ceratitis capitata</i>	12
Figura 6 – Controle de pragas utilizando a técnica do inseto estéril na mosca do mediterrâneo	13
Figura 7 – Exemplificação das rotas metabólicas	15
Figura 8 – Representação de rotas metabólicas a partir do grafo	16
Figura 9 – Análises das rotas metabólicas	18
Figura 10 – Pipeline proposto para realizar a reconstrução de rotas metabólicas .	19
Figura 11 – Grafo da metabolização de nitrogênio nas bactérias <i>Klebsiella</i>	21
Figura 12 – Modelagem estequiométrico	22
Figura 13 – Subgrafo de reações químicas envolvidas na metabolização do nitrogênio nas bactérias <i>Klebsiella</i>	29
Figura 14 – Matriz estequiométrica das rotas metabólicas envolvidas na metabolização de nitrogênio nas bactérias <i>Klebsiella</i>	30
Figura 15 – Representação matricial dos EM obtidos pelas análises das rotas metabólicas envolvidas na metabolização de nitrogênio nas bactérias <i>Klebsiella</i>	31
Figura 16 – Elementary Modes gerados pelas análises das rotas metabólicas nas bactérias <i>Klebsiella</i> que geram L-Glutamato	33
Figura 17 – Elementary Modes gerados pelas análises das rotas metabólicas nas bactérias <i>Klebsiella</i> . que geram amônia e reações internas	34
Figura 18 – Script para o uso da ferramenta SPAdes	43
Figura 19 – Grafo de pesquisa e sua matriz de adjacência	44
Figura 20 – Potências 2, 3 e 4 da matriz de adjacência	45
Figura 21 – Potências 5, 6 e 7 da matriz de adjacência	46
Figura 22 – Subgrafo exemplo do problema utilizado na aplicação de DNA computing	47
Figura 23 – Codificação do grafo problema de acordo a técnica de DNA computing	48
Figura 24 – Reação de ligação	49
Figura 25 – Produto de reação de ligação	50
Figura 26 – Electroforeses do PCR <i>in silico</i>	51

Lista de tabelas

Tabela 1 – Recursos computacionais utilizados	17
Tabela 2 – Contigs obtidos para cada cepa das bactérias de gênero <i>Klebsiella</i> spp.	24
Tabela 3 – Genes identificados a partir dos contigs da bactéria <i>Klebsiella pneumoniae</i> - cepa 1	26
Tabela 4 – Genes identificados a partir dos contigs das bactérias <i>Klebsiella oxytoca</i> - cepa 2 e <i>Klebsiella oxytoca</i> - cepa 3	27
Tabela 5 – Metabolitos envolvidos no subgrafo	28

Lista de abreviaturas e siglas

BLAST	Basic Local Alignment Search Tool
CACAU	Centro de Armazenamento de dados e Computação Avançada da UESC
CNA	CellNetAnalyzer
DNA	Ácido Desoxirribonucléico
EBI	European Bioinformatics Institute
EBML	European Molecular Biology Laboratory
EM	Elementary Modes
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	National Center of Biotechnological Information
pb	Pares de bases
PPGMC	Programa de Pós-Graduação em Modelagem Computacional em Ciência e Tecnologia
RNA	Ácido Ribonucléico
UESC	Universidade Estadual de Santa Cruz

Sumário

1 – Introdução	1
1.1 Problematização	1
1.2 Objetivos	2
1.2.1 Objetivo Geral	2
1.2.2 Objetivos Específicos	2
1.3 Motivação	2
1.4 Bases Técnicas da dissertação	3
1.4.1 Base da Biologia	4
1.4.2 Base da Ciência da Computação	5
1.5 Organização do trabalho	6
1.5.1 Base da Ciência da Computação	6
1.6 Organização do trabalho	7
2 – Revisão de Literatura	8
2.1 Controle de pragas	8
2.2 Técnica do inseto estéril	9
2.3 <i>Ceratitis capitata</i> , Mosca do mediterrâneo	10
2.4 <i>Klebsiella</i>	13
2.5 Reconstrução de rotas metabólicas	14
2.6 Rotas metabólicas	14
2.6.1 Modelagem estequiométrica	15
2.6.2 Teoria dos grafos	15
2.7 Recursos computacionais utilizados no processamento de dados	16
2.8 Elementary Modes	17
3 – Reconstrução de rotas metabólicas de <i>Klebsiella</i>	19
3.1 Origem dos dados biológicos	19
3.2 Montagem dos contigs	20
3.3 Identificação dos genes a partir dos contigs gerados	20
3.4 Identificação das rotas metabólicas a partir dos genes encontrados	21
3.5 Modelagem estequiométrica das rotas metabólicas	21
3.6 Análise das rotas metabólicas utilizando CNA	22
4 – Resultados e discussão: execução do processo de reconstrução de rotas metabólicas	23

4.1	Montagem dos contigs a partir dos dados de sequenciamento das bactérias <i>Klebsiella</i>	24
4.2	Identificação dos genes envolvidos na metabolização do nitrogênio a partir dos contigs gerados	25
4.3	Identificação das rotas metabólicas da metabolização de nitrogênio a partir dos genes encontrados	28
4.4	Modelagem estequiométrica das rotas metabólicas das bactérias <i>Klebsiella</i>	29
4.5	Análise das rotas metabólicas das cepas de <i>Klebsiella</i> sp. utilizando CNA	31
5	– Considerações finais	36
5.1	Conclusões	36
5.2	Trabalhos Futuros	37
	Referências	38
	Apêndices	42
	APÊNDICE A – Script utilizado para rodar a ferramenta SPAdes @CACAU . . .	43
	APÊNDICE B – Procurar caminhos em grafos de acordo as potências da sua matriz de adjacência	44
	APÊNDICE C – Aplicação de DNA computing	47
	C.1 Uso de DNA computing	49

1 Introdução

Durante os últimos anos, pesquisas do controle biológico de pragas de insetos têm sido intensificadas, especificamente na mosca do mediterrâneo *Ceratitis capitata*. Uma das técnicas do controle biológico utilizada na mosca é a Técnica do Inseto Estéril (SIT, pelas suas siglas em inglês *Sterile Insect Technique*). A técnica SIT é complementada por pesquisas na identificação de conjunto de microrganismos chamado microbiota dentro da mosca. De acordo com (AHARON et al., 2013) as bactérias do gênero *Klebsiella* são organismos simbiotes da *C. capitata* uma vez que contém mecanismos moleculares relacionados com grupos metabólicos que coexistem e complementam aos grupos metabólicos do hospedeiro (mosca). Esses mecanismos moleculares são necessários no ciclo de vida da mosca, como por exemplo, a metabolização do nitrogênio que é usado pelas moscas no processo digestivo (AMI et al., 2009).

Neste contexto, uma pesquisa fundamentada na reconstrução de rotas metabólicas a partir de genomas das bactérias *Klebsiella* pode contribuir para identificar mecanismos moleculares de *Klebsiella* como possíveis alvos da engenharia metabólica na linha de produção da mosca pela técnica SIT, que possa aprimorar o controle biológico da mosca. No entanto, utilizar ferramentas de bioinformática para processar dados biológicos. Nesta dissertação são descritas as etapas envolvidas na reconstrução de rotas metabólicas a partir de genomas de *Klebsiella* utilizando conceitos da área de biologia e ciência da computação. O objetivo desta pesquisa utilizar uma técnica de reconstrução de rotas metabólicas a partir de genomas para identificar mecanismos moleculares envolvidos na metabolização do nitrogênio das bactérias que a mosca poderia utilizar no processo digestivo e ciclo de vida.

1.1 Problematização

Para complementar as pesquisas do efeito da microbiota na mosca, é imprescindível identificar os mecanismos moleculares que possam apresentar como a bactéria pode melhorar o processo digestivo e desenvolvimento da mosca produzida industrialmente. No entanto, essa identificação é tardia, posto que, sem dados de sequenciamento os cientistas têm que utilizar métodos de laboratório que acrescentam o tempo de pesquisa. Diante disso, o problema que segue é: como identificar computacionalmente mecanismos moleculares das bactérias presentes na microbiota da mosca utilizando dados de sequenciamento?

1.2 Objetivos

Nesta seção apresentamos os objetivos da dissertação, tanto em termos gerais quanto específicos. Com o auxílio dos objetivos propostos, foram obtidos os resultados esperados dessa dissertação.

1.2.1 Objetivo Geral

1. Desenvolver uma metodologia para realizar a reconstrução de rotas metabólicas a partir dos dados de sequenciamento de bactérias do gênero *Klebsiella* presentes na microbiota da mosca.

1.2.2 Objetivos Específicos

1. Realizar a montagem dos genomas das três cepas de *Klebsiella* obtidas em colaboração de uma pesquisa realizada pela Universidad del Valle de Guatemala.
2. Aplicar a técnica de reconstrução de rotas metabólicas de acordo com (PAPIN et al., 2003); a partir dos genomas das três bactérias do gênero *Klebsiella*.
3. Propor o uso de uma técnica de reconstrução de rotas metabólicas baseada na teoria dos grafos.
4. Avaliar biologicamente os resultados obtidos do processo computacional.

1.3 Motivação

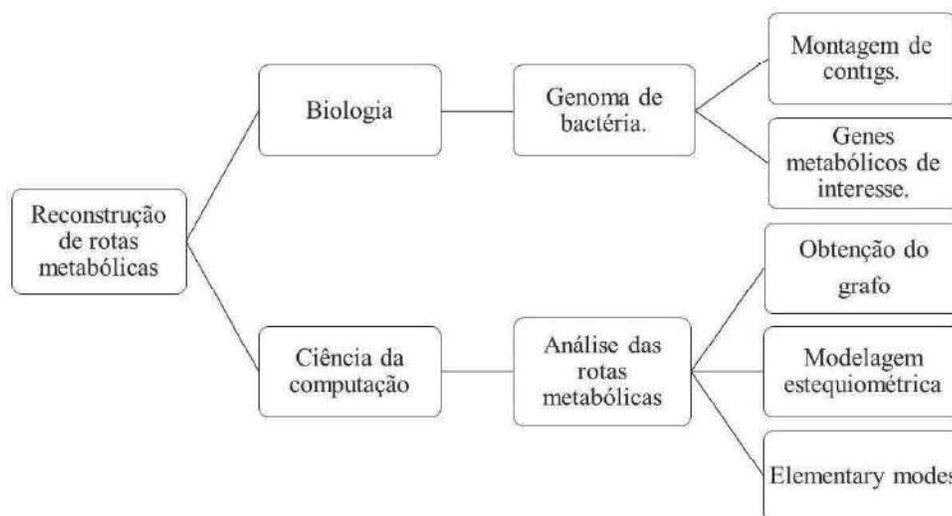
A utilização de modelos matemáticos e computacionais como ferramentas podem facilitar a análise de sistemas biológicos como no caso da identificação dos mecanismos moleculares entre uma bactéria e a mosca. A integração das áreas da biologia e ciência da computação permite processar dados biológicos e obter soluções para problemas biológicos com auxílio de ferramentas computacionais. Nesta dissertação tendo natureza multidisciplinar, é proposto um processo de 5 etapas que para processar resultados de sequenciamento de genomas e gerar o produto que é a reconstrução de rotas metabólicas, onde foram identificadas moléculas envolvidas na relação de simbiose com a mosca. A modelagem matemática de uma rota metabólica é de forma matricial, além disso, foi utilizada uma técnica não usual para a análise baseada na teoria dos grafos avaliando os caminhos em grafos dirigidos. Essa técnica de modelagem de rotas metabólicas, gera todas as possíveis rotas dentro das bactérias e isso permite identificar os mecanismos moleculares realizados pela bactéria como possíveis alvos na engenharia metabólica na produção de *C. capitata* pela técnica SIT.

A ideia de procurar mecanismos moleculares utilizando a reconstrução de rotas metabólicas foi obtida do trabalho de (PINZÓN et al., 2010) onde ele trabalha com a utilização de reconstrução de rotas metabólicas como procedimento na caracterização de interações planta – patógeno. A principal motivação desse estudo, foi trabalhar analogamente com a possível relação que existe entre bactérias *Klebsiella* e hospedeiro *C. capitata*. A aplicação desta abordagem vai permitir identificar mecanismos moleculares envolvidos na metabolização do nitrogênio realizado pelas bactérias utilizando a técnica de reconstrução de rotas metabólicas. Esses mecanismos são importantes na linha de produção da mosca pela técnica SIT. Os mecanismos são a produção de metabolitos como L-glutamato que é usado nas sínteses de novas proteínas no ciclo de metamorfose da mosca. Além disso, a *Klebsiella* tem a capacidade de deslocar a bactéria *Pseudomonas* spp. ocupando maior nicho ecológico dentro da mosca.

1.4 Bases Técnicas da dissertação

A identificação de mecanismos moleculares utilizando a reconstrução das rotas metabólicas das bactérias *Klebsiella* é um projeto multidisciplinar, pois utiliza de conceitos da área da biologia e ciências da computação. A Figura 1 mostra os conceitos básicos nos quais está embasado essa dissertação, trabalhando com as duas áreas de interesse, para obter uma solução a um problema biológico com auxílio de ferramentas computacionais.

Figura 1 – Esquema das bases técnicas da dissertação

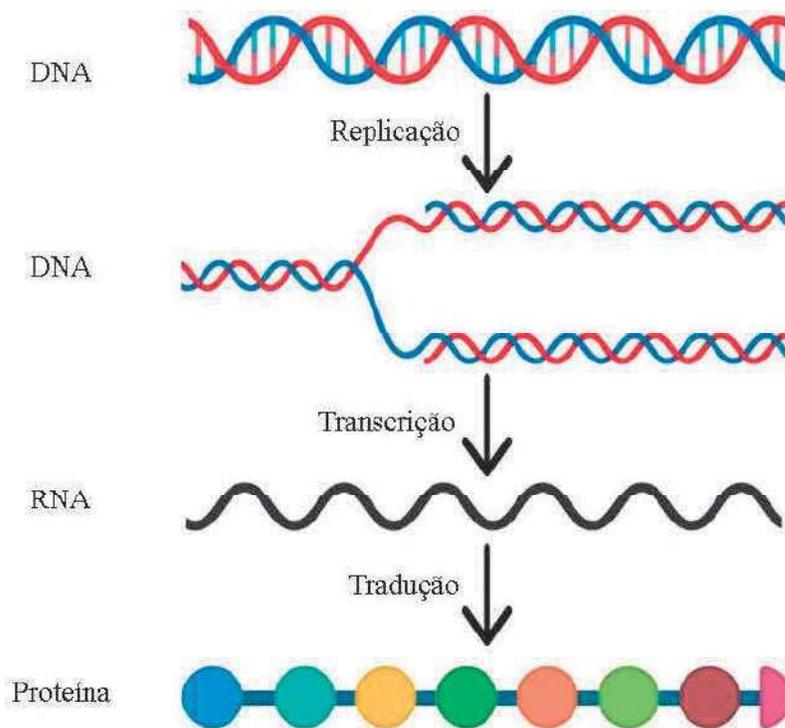


Fonte: Autoria própria

1.4.1 Base da Biologia

A biologia é uma área com múltiplos ramos que permitem o estudo de organismos e a vida. Os organismos são compostos por múltiplas moléculas biológicas que interagem entre si gerando sistemas complexos. As principais moléculas biológicas conhecidas são o DNA, RNA e proteínas. A interação dessas moléculas é chamada de "Dogma Central da Biologia Molecular"(Figura 2). O processo apresenta o fluxo de informação genética iniciando com uma molécula de DNA na etapa de replicação. Nesta etapa são geradas duas cópias idênticas de DNA a partir da molécula inicial. O DNA logo sofre um processo de transcrição onde é gerado o RNA. Essa informação genética na forma de RNA é levada até citoplasma onde os ribossomos realizam a síntese de proteínas pela etapa de tradução. As proteínas são moléculas biológicas compostas pelos aminoácidos. Estas moléculas são responsáveis pela maioria dos processos metabólicos dentro das células, apresentando funções enzimáticas e de transporte (CLARK, 2009).

Figura 2 – Dogma central da biologia molecular



Fonte: Autoria própria

As proteínas reagem com outras biomoléculas e substâncias químicas obtendo metabólitos (produtos do metabolismo). As reações entre biomoléculas ajudam nos processos celulares dos organismos. As séries destas reações químicas são chamadas rotas metabólicas. As rotas metabólicas ocorrem com auxílio de múltiplas moléculas como enzimas, cofatores, substrato e produto. Nesta dissertação, o foco foi as rotas me-

tabólicas envolvidas na metabolização de nitrogênio nas bactérias *Klebsiella*. Dentro das reações que compõem o sistema de metabolização de nitrogênio podem-se encontrar reações de fixação, assimilação e regulação de nitrogênio. Além disso, existem enzimas que interagem com metabólitos intermediários para acelerar as reações químicas. Identificação de mecanismos moleculares como as rotas metabólicas é um problema biológico porque precisa de grandes quantidades de informação. Foi apresentado um processo de 5 etapas que permite identificar as rotas metabólicas a partir de resultados de sequenciamento e identificar mecanismos moleculares das bactérias que possam ser utilizados na produção de *C. capita*.

1.4.2 Base da Ciência da Computação

As rotas metabólicas precisam de abordagens onde toda a informação possa ser organizada. Um exemplo é a descrição matemática das funções sistemáticas por elas fornecidas. As rotas metabólicas são complexas, mas possuem componentes, como os valores estequiométricos, que têm informação sobre cada reação. Dessa forma, a modelagem estequiométrica de uma rota metabólica permite identificar a topologia da rota. Os valores estequiométricos são os principais componentes da modelagem, sendo esses valores a quantidade molar de substrato e produto. As rotas metabólicas podem-se representar na forma matricial, neste caso, os valores contidos na matriz são os dados estequiométricos de cada reação que acontece dentro do sistema de metabolismo. A matriz estequiométrica (S) contém informação de número de metabólitos (m) e o número de reações (n) (PAPIN et al., 2003).

As rotas metabólicas podem ser representadas por conceitos básicos de teoria dos grafos. De acordo com os resultados obtidos nas análises da parte de biologia, a identificação de genes permite determinar a quantidade de moléculas presentes no sistema. No estudo, foi utilizado o enfoque de um grafo dirigido, onde os metabólitos são representados pelos vértices do grafo. Além disso, as reações que acontecem dentro do sistema de metabolização de nitrogênio são representadas pelas arestas. Com a representação na forma de grafo podem-se aplicar algoritmos que permitam identificar os caminhos dentro do sistema (OLIVEIRA; BAILEY, 2001). Sendo assim, para identificar as reações e os metabólitos envolvidos foi utilizado o enfoque da análise de Elementary modes (EM). Os EM são todos os possíveis caminhos que possam existir dentro do sistema de metabolização. Para fazer a análise é necessário fixar vértices de intercâmbio (entrada e saída), permitindo assim, obter caminhos dentro do grafo que representa mecanismos moleculares que possam solucionar o problema biológico proposto utilizando ferramentas computacionais.

1.5 Organização do trabalho

O trabalho apresentado é dividido em cinco capítulos: O capítulo 1, mostra a relevância do estudo, descrevendo a introdução, a problematização, os objetivos, a motivação e as bases teóricas da dissertação. No capítulo 2 é apresentada a revisão de literatura sobre conceitos necessários no desenvolvimento da dissertação. O capítulo 3 descreve a metodologia empregada, detalhando cada etapa do processo de reconstrução de rotas metabólicas, além disso, são listadas as ferramentas computacionais utilizadas. No capítulo 4 há a aplicação do processo computacional nos resultados dos sequenciamentos fornecidos para as cepas das bactérias *Klebsiella*. Por fim, o capítulo 5 apresenta a discussão e as conclusões da interpretação dos mecanismos moleculares identificados nas bactérias *Klebsiella* que podem ser utilizados como alvos em engenharia metabólica na produção da mosca pela técnica SIT da dissertação. Além disso, são apresentados os trabalhos futuros que poderão ser desenvolvidos.

1.5.1 Base da Ciência da Computação

As rotas metabólicas precisam de abordagens onde toda a informação possa ser organizada. Um exemplo é a descrição matemática das funções sistemáticas por elas fornecidas. As rotas metabólicas são complexas, mas possuem componentes, como os valores estequiométricos, que têm informação sobre cada reação. Dessa forma, a modelagem estequiométrica de uma rota metabólica permite identificar a topologia da rota. Os valores estequiométricos são os principais componentes da modelagem, sendo esses valores a quantidade molar de substrato e produto. As rotas metabólicas podem-se representar na forma matricial, neste caso, os valores contidos na matriz são os dados estequiométricos de cada reação que acontece dentro do sistema de metabolismo. A matriz estequiométrica (S) contém informação de número de metabólitos (m) e o número de reações (n) (PAPIN et al., 2003).

As rotas metabólicas podem ser representadas por conceitos básicos de teoria dos grafos. De acordo com os resultados obtidos nas análises da parte de biologia, a identificação de genes permite determinar a quantidade de moléculas presentes no sistema. No estudo, foi utilizado o enfoque de um grafo dirigido, onde os metabólitos são representados pelos vértices do grafo. Além disso, as reações que acontecem dentro do sistema de metabolização de nitrogênio são representadas pelas arestas. Com a representação na forma de grafo podem-se aplicar algoritmos que permitam identificar os caminhos dentro do sistema (OLIVEIRA; BAILEY, 2001). Sendo assim, para identificar as reações e os metabólitos envolvidos foi utilizado o enfoque da análise de Elementary modes (EM). Os EM são todos os possíveis caminhos que possam existir dentro do sistema de metabolização. Para fazer a análise é necessário fixar vértices de intercâmbio (entrada e saída), permitindo assim, obter caminhos dentro do grafo

que representa mecanismos moleculares que possam solucionar o problema biológico proposto utilizando ferramentas computacionais.

1.6 Organização do trabalho

O trabalho apresentando é dividido em cinco capítulos: O capítulo 1, mostra a relevância do estudo, descrevendo a introdução, a problematização, os objetivos, a motivação e as bases teóricas da dissertação. No capítulo 2 é apresentada a revisão de literatura sobre conceitos necessários no desenvolvimento da dissertação. O capítulo 3 descreve a metodologia empregada, detalhando cada etapa do processo de reconstrução de rotas metabólicas, além disso, são listadas as ferramentas computacionais utilizadas. No capítulo 4 há a aplicação do processo computacional nos resultados dos sequenciamentos fornecidos para as cepas das bactérias *Klebsiella* e sua interpretação. Por fim, o capítulo 5 as conclusões da interpretação dos mecanismos moleculares identificados nas bactérias *Klebsiella* que podem ser utilizados como alvos em engenharia metabólica na produção da mosca pela técnica SIT da dissertação. Além disso, são apresentados os trabalhos futuros que poderão ser desenvolvidos para acrescentar o conhecimento nesta área.

2 Revisão de Literatura

No presente capítulo foram apresentados os conceitos teóricos necessários no desenvolvimento da dissertação. São descritos os modelos matemáticos e as ferramentas computacionais utilizadas na técnica proposta de reconstrução das rotas metabólicas das bactérias do gênero *Klebsiella*. Ainda neste capítulo, são apresentados os conceitos básicos da biologia como a abordagem do controle de pragas e as bases biológicas dos componentes da dissertação. Além disso, apresentamos os conceitos básicos da aplicação computacional a um conjunto de dados biológicos.

2.1 Controle de pragas

Os insetos são uma das classes de seres vivos mais diversas no mundo, pois têm a capacidades de adaptar-se a distintas condições de ambiente. Os insetos têm um papel na ecologia de ecossistemas, mas algumas espécies são consideradas pragas. As pragas de insetos são principalmente observadas na agricultura e as consequências de suas ações têm efeito na redução de produtividade de safras. Como resultado, os insetos são alvos de técnicas de controle de pragas para reduzir o tamanho das populações e evitar perdas nas safras (OLIVEIRA et al., 2014). O desenvolvimento do controle de pragas é focado em análises custo-benefício, abordando o impacto nos produtores, sociedade e meio ambiente. Dependendo da praga e do cultivo, o controle pode ser de diferente tipo, por exemplo:

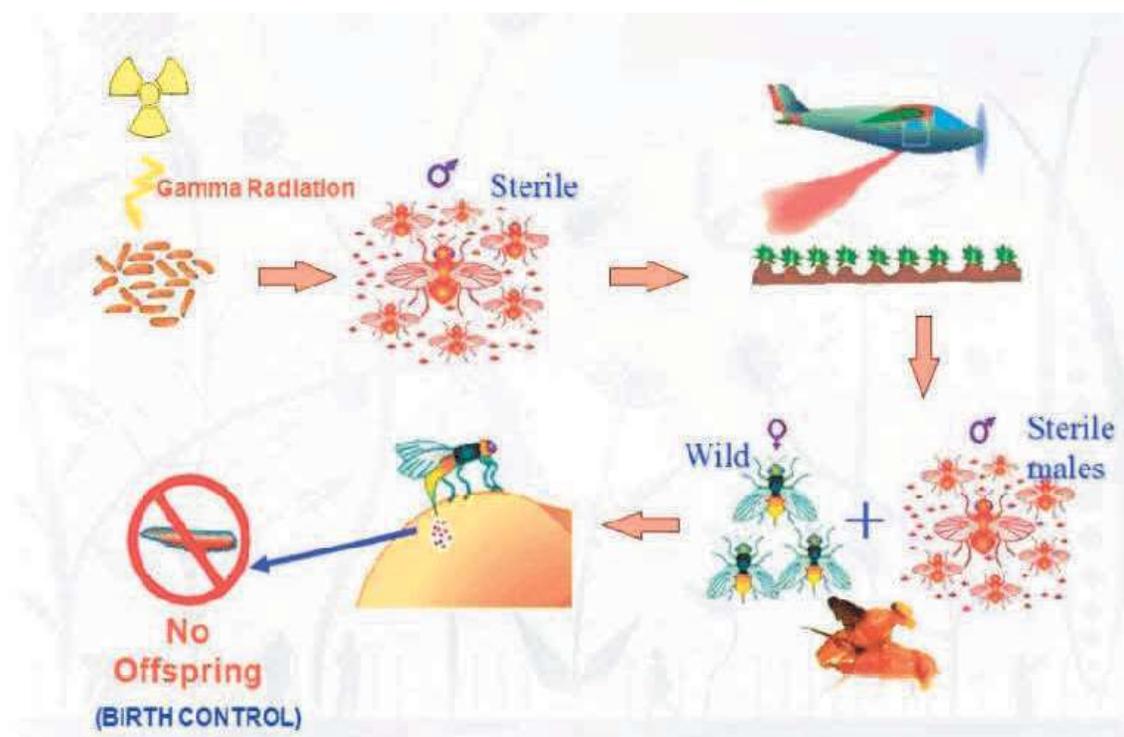
- Controle biológico
- Controle mecânico
- Controle cultural
- Modificação do meio ambiente
- Manipulação genética
- Controle químico
- Métodos regulatórios

Cada controle tem diferentes alvos que podem variar entre hospedeiro e seu parasita. Neste trabalho, nos focamos na técnica do inseto estéril no controle biológico da *C. capitata* e seu bactéria simbiote *Klebsiella*.

2.2 Técnica do inseto estéril

Nesta dissertação, o foco foi a microbiota que compõem os machos estéreis de *Ceratitis capitata* obtidos pela técnica SIT realizada por Moscamed, Guatemala. A técnica do inseto estéril conhecida como "SIT" (pelas suas siglas em inglês "Sterile Insect Technique") é uma técnica de esterilização de insetos para controlar a população de um inseto em particular. Essa é uma técnica amigável ao meio ambiente, comparada aos inseticidas e pesticidas usados pelos agricultores. Ela é baseada na irradiação dos insetos pelos raios X e γ – *gama*, causando efeito sobre o espermatozoide do macho de *C. capitata*, que torna inviável no momento da fecundação do óvulo (Fig. 3) (LEFTWICH et al., 2014).

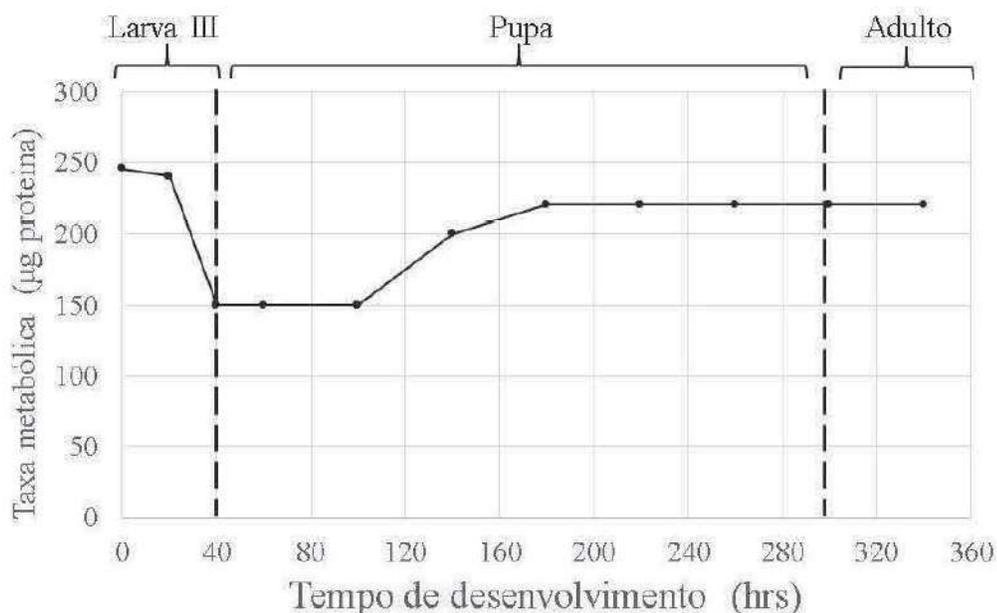
Figura 3 – Controle de pragas utilizando a técnica do inseto estéril



Fonte: (DEMIRBAS-UZEL, 2018)

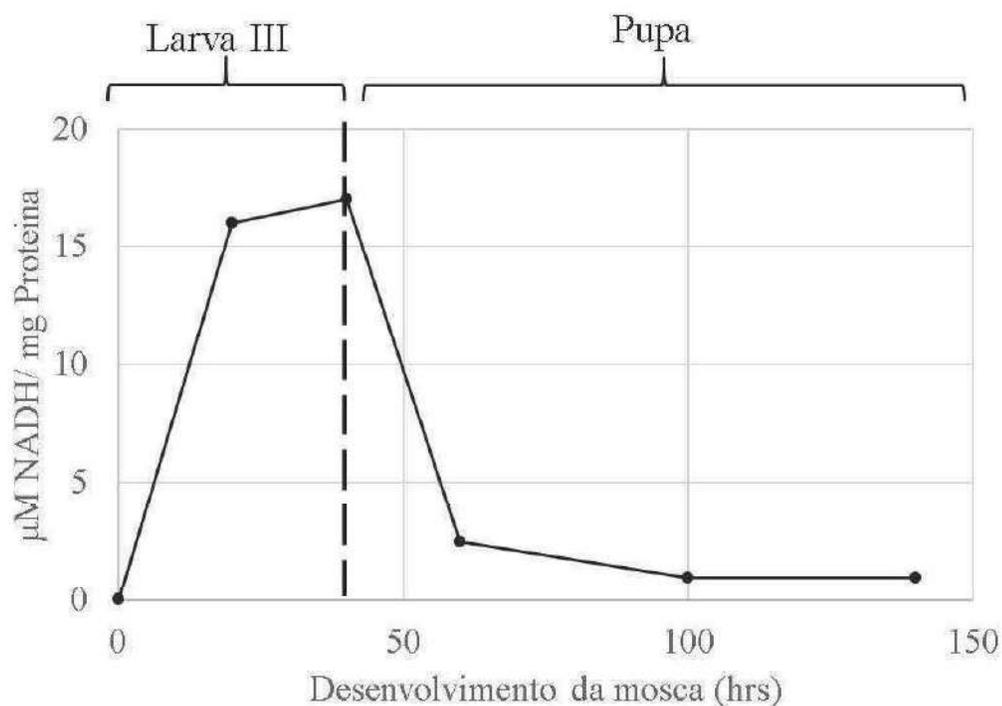
2.3 *Ceratitis capitata*, Mosca do mediterrâneo

A mosca do Mediterrâneo *Ceratitis capitata* (Diptera: Tephritidae) é uma das maiores pragas conhecidas. O inseto tem capacidade de adaptação em qualquer meio e, por consequência, ele pode estabelecer-se na maioria dos ambientes do mundo como Guatemala, que forneceu dados das pesquisas realizadas para nosso trabalho. Além disso, é um inseto que pode se alimentar de qualquer tipo de comida e tem múltiplas plantas hospedeiras (RICALDE et al., 2012). Durante o processo da metamorfose da *Ceratitis capitata*, ocorre um fluxo da energia dentro da mosca, intercâmbio interno e externo, onde o balanço energético é negativo ($E_{in} < E_{out}$). Por consequência, a energia presente no processo vai para fora do sistema e energia não entra. Para suportar esse processo, a larva da mosca usa reservas de energia como: lipídeos, glucose e proteínas. Dentro da metamorfose ocorre processo bioquímicos como a digestão de proteínas, produzindo aminoácidos livres que são usados como moléculas para realizar a síntese de novas proteínas. A taxa metabólica é diferente nas etapas da metamorfose, a Figura 4 apresenta a representação de taxa metabólica vs tempo de desenvolvimento. No estágio de larva III a quantidade é aproximadamente 250 ug de proteína, nesta etapa está presente a proteína de cutícula da larva. Porém as 40 hrs os valores de proteína presentes decrescem devido à digestão de proteínas gerando aminoácidos livres no meio. Aproximadamente às 90 hrs a quantidade de proteínas aumenta, uma vez que o processo de sínteses de proteínas de cutícula de larva é iniciado. Essa proteína está presente até fase de adulto de *C. capitata*.

Figura 4 – Taxa metabólica da metamorfose de *Ceratitis capitata*

Fonte: Adaptada de (NESTEL et al., 2003).

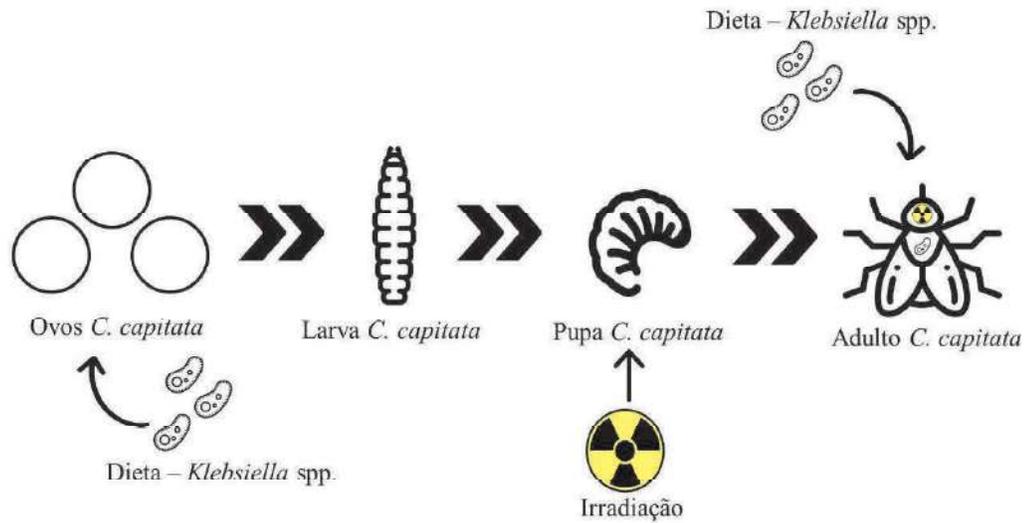
A taxa metabólica está associada com a atividade enzimática da enzima glutamato desidrogenase (GDH). A enzima tem a capacidade de degradar o glutamato, gerando como produto a amônia e outros intermediários metabólicos. A atividade enzimática tem relação com a concentração de amônia no meio. Quando a concentração é maior ou igual a 1 mM a produção de amônia é incrementada, mas quando as concentrações são menores que 1 mM a produção de Glutamato é maior. Isso ocorre, pois, a enzima GDH tem baixa afinidade pelo amônio. Na Figura 5 é apresentada a atividade da enzima glutamato desidrogenase, às 40 hrs de desenvolvimento da mosca a enzima têm seu ponto máximo de atividade, isso indica que tem maior produção de amônio. Além disso, nas 40 hrs a enzima e o substrato têm a maior interação, logo a concentração de substrato decresce. A produção de L-Glutamato também é importante na metamorfose da *Ceratitis capitata* na produção de novas proteínas para os seguintes processos bioquímicos até chegar ao estágio de adulto.

Figura 5 – Taxa metabólica da metamorfose de *Ceratitis capitata*

Fonte: Adaptada de (NESTEL et al., 2003).

Esta espécie de praga tem impacto econômico para os agricultores pois as frutas e vegetais são usadas por ela como lugar para realizar a oviposição. Logo, suas larvas se alimentam das frutas, facilitando o apodrecimento (MINAS et al., 2011). Para realizar o controle biológico dessa praga, é utilizada a técnica SIT, que consiste em um processo desde ovos até adultos (Fig 6). A irradiação favorece que organismos como as bactérias *Klebsiella* consigam realizar o processo simbiótico. Isso ocorre pois a radiação elimina a microbiota original do macho, que passa a ser alimentado com dieta constituída de *Klebsiella* para realizar a recolonização do trato digestivo e assegurar sua competência para após ser liberado (GUERFALI et al., 2018).

Figura 6 – Controle de pragas utilizando a técnica do inseto estéril na mosca do mediterrâneo



Fonte: Autoria Própria

2.4 *Klebsiella*

Foi identificada na microbiota da mosca a presença de bactérias que pertencem à família Enterobacteriaceae e dentre elas a que possui maior presença é a do gênero *Klebsiella* (BEHAR et al., 2005). As bactérias deste gênero são Gram-negativa, além de ser um patógeno oportunista que afeta o hospedeiro através de um sistema imunológico comprometido ou enfraquecido. A bactéria *Klebsiella oxytoca*, é uma bactéria simbiótica presente no desenvolvimento da mosca, do ovo até a fase adulta. Esta bactéria é importante na fixação do nitrogênio que é usado pelas moscas, além de desempenhar um papel na escolha do par selvagem para copular. De acordo com (AHARON et al., 2013), a microbiota está associada ao trato digestivo e tem um papel importante no “fitness” do desenvolvimento da mosca. Além disso, (NESTEL et al., 2003) descreveram que moléculas, tais como aminoácidos e proteínas, que as bactérias podem produzir contribuem no desenvolvimento da mosca até a fase adulta. Essas moléculas foram nosso alvo para identificar mecanismos moleculares de produção.

2.5 Reconstrução de rotas metabólicas

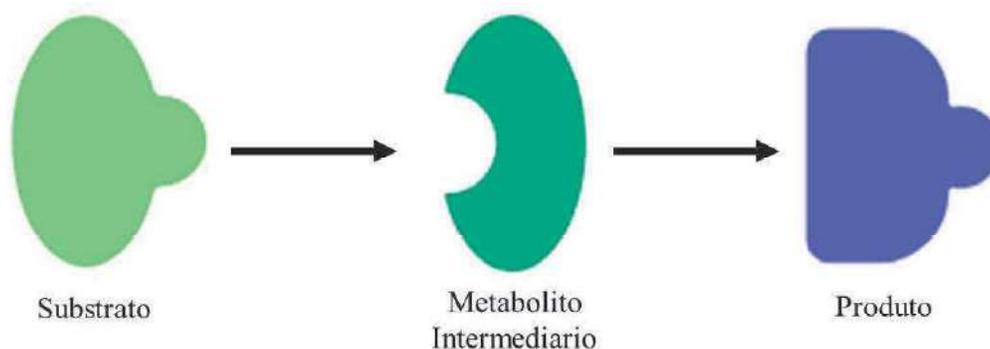
O processo de reconstrução destas rotas metabólicas possui uma série de etapas utilizando múltiplas ferramentas computacionais e modelos matemáticos que fornecem estruturas que permitem o estudo, evolução, organização e princípios desses sistemas celulares (TENAZINHA; VINGA, 2011). No contexto da biologia computacional, uma rota metabólica descreve funções sistemáticas que ocorrem através de metabólitos (componentes celulares envolvidos no metabolismo), onde esses componentes são conectados para gerar um determinado produto (FAUST et al., 2011). Cada reação em uma rota metabólica pode conter enzimas, cofatores ou outros componentes que facilitam o metabolismo (FEIST et al., 2008). Neste trabalho descrevemos as etapas envolvidas na reconstrução de rotas metabólicas com os dados do sequenciamento das três cepas de *Klebsiella* com um enfoque multidisciplinar utilizando conceitos da área de biologia e ciência da computação. Os dados são processados pela ferramenta computacional que realiza a montagem de *contigs* (sequências contínuas de sobreposição). Os *contigs* contêm sequências de genes que codificam metabólitos envolvidos na metabolização do nitrogênio nas bactérias. Esses genes são identificados em bases de dados gerando o alinhamento com sequências parecidas e determinando a presença/ausência dos genes nos *contigs* das bactérias. Portanto, os genes procurados são aqueles que apresentam participação na metabolização de nitrogênio nas bactérias importantes para simbiose com a mosca.

2.6 Rotas metabólicas

As bactérias do gênero *Klebsiella* apresentam rotas metabólicas que estão envolvidas na simbiose com a mosca do mediterrâneo. Rotas metabólicas são um conjunto de reações químicas que consistem em um conjunto de compostos inicial usados como substrato, que com auxílio de enzimas ou outras moléculas é transformada em um produto (Fig. 7). O produto resultante é o substrato da reação subsequente, até que a estrutura química final é alcançada pelo metabólito (COX, 2007). As rotas metabólicas podem ser reconstruídas por processamentos computacionais de dados biológicos a partir de dados de genomas como é proposto em nossa metodologia, Apêndice B e Apêndice C (TOYA et al., 2011). Essa abordagem de reconstrução de rotas metabólicas permite identificar mecanismos moleculares de interesse na área de pesquisa.

A reconstrução de rotas metabólicas tem múltiplas abordagens e podem ser representadas matematicamente por dois conceitos da área de ciência da computação, como utilizando uma técnica não usual baseada na teoria dos grafos ou modelagem estequiométrica.

Figura 7 – Exemplificação das rotas metabólicas



Fonte: Autoria Própria

2.6.1 Modelagem estequiométrica

A estequiometria das rotas metabólicas de fixação de nitrogênio nas bactérias descreve as arquiteturas e topologias das rotas presentes. Além disso, os coeficientes estequiométricos são taxas molares onde os substratos são convertidos em produtos nas reações químicas (SCHILLING et al., 2000). As rotas metabólicas da *Klebsiella* são analisadas pelo modelo estequiométrico da reação. Geralmente, isso é analisado de acordo com um estado estável hipotético, em que o metabolismo não muda com relação ao tempo e todas as concentrações intermediárias são constantes (TOYA et al., 2011). Depois de determinar esses coeficientes de metabólitos de interesse, determina-se uma matriz de fluxo. A matriz (S), que contém $m \times n$ elementos, onde m (linhas) corresponde ao número total de metabólitos e n (colunas) ao total de reações que ocorrem no sistema (SCHILLING et al., 2000). A matriz S , o vector de fluxo v (rotas) e o estado estável das rotas metabólicas podem ser relacionados pela seguinte equação:

$$S * v = 0 \quad (1)$$

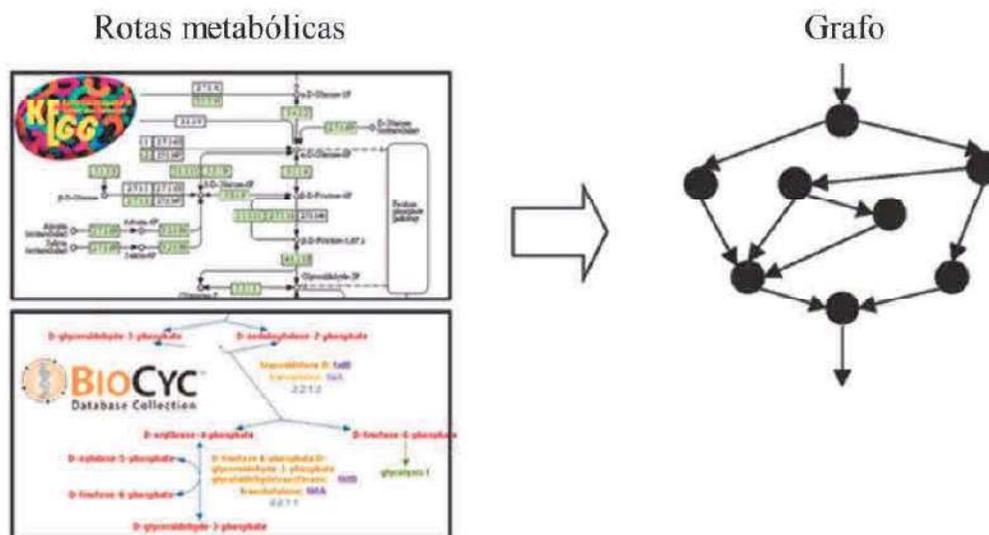
Na modelagem estequiométrica os sinais dos coeficientes representam o consumo do metabolito (-) e a produção (+), além disso, os vértices de intermediários indicam o caminho do vértice inicial até vértice final da rota metabólica.

2.6.2 Teoria dos grafos

A partir das bases de dados biológicas conhecidas é possível obter informação de rotas metabólicas. As rotas metabólicas podem ser representadas pelos grafos dirigidos (Fig. 8), onde o grafo (G) é definido por (V, E) , sendo V um conjunto de vértices representando os metabólitos envolvidos nas reações químicas, e E é um conjunto de arestas representando as conexões entre vértices (reações que acontecem nas rotas metabólicas).

A representação na forma de grafos permite identificar caminhos entre vértices fixados utilizando algoritmos de caminho curto ou fluxo máximo (PAVLOPOULOS et al., 2011).

Figura 8 – Representação de rotas metabólicas a partir do grafo



Fonte: Adaptada de: (TOYA et al., 2011)

2.7 Recursos computacionais utilizados no processamento de dados

Os recursos utilizados são listados na Tabela 1, onde é apresentado o resumo das ferramentas utilizadas na dissertação. Além disso, foram utilizadas 3 bases de dados:

- Centro Nacional de Informação Biotecnológica (NCBI, pelas suas siglas em inglês, <<https://www.ncbi.nlm.nih.gov/>>).
- Laboratório Europeu de Biologia Molecular com a parceria do Instituto Europeu de Bioinformática (EBML-EBI, pelas suas siglas em inglês, <<https://www.ebi.ac.uk/>>).
- Enciclopédia Kyoto de Genes e Genomas (KEGG, pelas suas siglas em inglês, <<https://www.kegg.jp/>>).

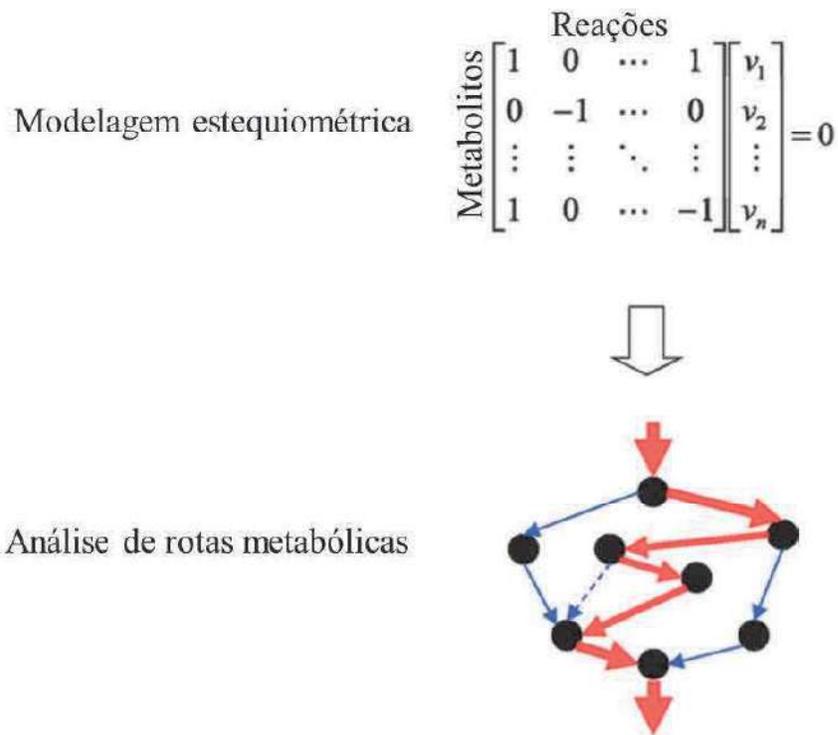
Tabela 1 – Recursos computacionais utilizados na dissertação (Fonte: Autoria Própria).

Nome	Versão	Utilidade	Referência
SPAdes	3.13.0	Montagem dos contigs	(BANKEVICH et al., 2012)
BLAST	2.8.1	Identificação genes de interesse	(ALTSCHUL et al., 1990)
KEGG	2019	Grafo rotas metabólicas / Modelagem estequiométrica	(OGATA et al., 1999)
CellNetAnalyzer	2019.1	Análise de Elementary Modes	(KLAMT et al., 2007)
MATLAB R2014b	8.4	Interfase CellNetAnalyzer	www.mathworks.com

2.8 Elementary Modes

A análise dos fluxos metabólicos (MFA - pela sua sigla em inglês "Metabolic Flux Analysis") é um método utilizado para estimar o fluxo metabólico intracelular das vias metabólicas que não podem ser observadas diretamente (Fig. 9). A estimativa dos fluxos ajuda a elucidar os mecanismos do metabolismo nos níveis de transcrição e tradução. Tal explicação auxilia no entendimento dos sistemas biológicos de interesse. Com o modelo estequiométrico, os fluxos metabólicos intracelulares podem ser calculados usando a técnica de Elementary Modes (EM). A análise pelos EM facilita identificação de todos os possíveis caminhos que existem desde vértices de entrada até vértices saída. Além disso, os EM são conhecidos por ser um conjunto de enzimas e metabolitos que podem reagir no estado estável como a técnica MFA descreve. ([SCHILLING et al., 1999](#)).

Figura 9 – Análises das rotas metabólicas

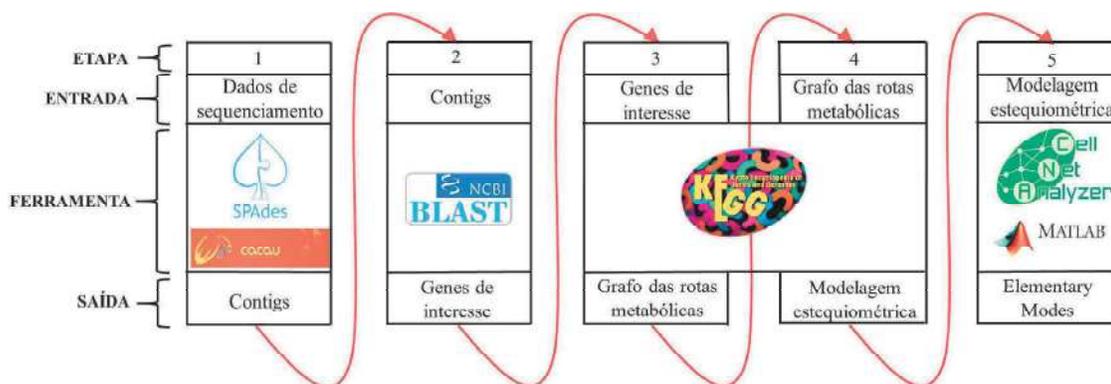


Fonte: (TOYA et al., 2011)

3 Reconstrução de rotas metabólicas de *Klebsiella*

Neste capítulo é descrito com suas 5 etapas o pipeline gerado para realizar a reconstrução de rotas metabólicas. A Figura 10 destaca os dados de entrada e saída para cada etapa do processo. A ideia deste estudo é aplicar o processamento a um conjunto de dados biológicos (dados de sequenciamento) das bactérias do gênero *Klebsiella*, com o auxílio das ferramentas mencionadas no diagrama. O resultado desejado foi a reconstrução de rotas metabólicas a partir do genoma destas bactérias identificadas na microbiota da *Ceratitis capitata*. Essas rotas logo foram comparadas para identificar possíveis mecanismos moleculares como possíveis alvos na engenharia metabólica na produção de *C. capitata* pela técnica SIT.

Figura 10 – Pipeline proposto para realizar a reconstrução de rotas metabólicas



Fonte: Autoria própria

3.1 Origem dos dados biológicos

Os dados biológicos foram fornecidos por um grupo de pesquisa da Universidad del Valle de Guatemala em conjunto com MOSCAMED. O nome da pesquisa foi “Caracterización genética de cepas de *Klebsiella* spp. aislada de ambientes asociados a *Ceratitis capitata* (Weidemann), por medio de secuenciación de su genoma, extracción de plásmidos y Rep-PCR”. O objetivo do estudo foi realizar o sequenciamento dos genomas das três bactérias *Klebsiella* spp. potencialmente simbióticas com *C. capitata*. O objeto da pesquisa foram as bactérias isoladas da microbiota da mosca. A metodologia foi baseada em 1) Provas bioquímicas; 2) Isolamento de DNA genômico; 3) Análises dos genomas e seu sequenciamento realizado. O sequenciamento foi realizado no equipamento Illumina MiniSeq (FIGUEROA; NAVARRETE, 2018). Os resultados

do sequenciamento foram 5445479 leituras para o isolado -1 de *Klebsiella pneumoniae*, 3817864 leituras para o isolado-2 de *Klebsiella oxytoca* e 4141211 leituras para o isolado-3 de *Klebsiella oxytoca*. Os dados biológicos do sequenciamento de cada genoma foram 2 arquivos em formato FASTAQ correspondentes às leituras *forward* (R1) e *reverse* (R2) fornecendo os dados para nossa pesquisa.

3.2 Montagem dos contigs

Para iniciar o processamento na etapa 1, foi utilizado o supercomputador Centro de Armazenamento de dados e Computação Avançada da UESC (CACAU, pelas suas siglas no português) com as seguintes características: Bull novascale com 20 nós R422-E1, 2 processadores Intel®Xeon® E5430 @ 2.66GHz QuadCore, 16 GB Memória, 1 placa infiniband, 1 placa gigabit ethernet.

Os resultados do sequenciamento das bactérias *Klebsiella* foram processados pela ferramenta de montagem de sequências SPAdes (BANKEVICH et al., 2012). O script da utilização de SPAdes encontra-se no Apêndice A rodado no CACAU com os parâmetros:

- -o: output onde a informação foi armazenada.
- -careful: reduz o número de discordâncias e espaços dentro das sequências.
- -pe1-1 e -pe1-2: leituras de sequenciamento.
- -m: limite de memória em Gb, nosso caso foi 15 Gb.

O procedimento para realizar a montagem foi gerar alinhamentos entre as *reads* (trechos de DNA produto do sequenciamento). O alinhamento procurou fragmentos de sequências em comum entre os nodes gerando sequências contínuas de sobreposição de regiões denominadas *contigs* (BASTIDE; MCCOMBIE, 2007).

3.3 Identificação dos genes a partir dos contigs gerados

As sequências dos genes envolvidos na metabolização do nitrogênio foram obtidas do Centro Nacional de Informação Biotecnológica dos Estados Unidos (NCBI, pelas suas siglas em inglês) e Laboratório Europeu da Biologia Molecular com a parceria do Instituto Europeu de Bioinformática (EBML-EBI, pelas suas siglas em inglês). Na etapa 2 foi realizado o alinhamento dos contigs obtidos na etapa 2 contra as sequências dos genes de interesse envolvidos nos mecanismos moleculares da metabolização do nitrogênio nas bactérias. A identificação desses genes foi realizada pelo BLAST de nucleotídeos (ALTSCHUL et al., 1990). O parâmetro para realizar o alinhamento foi “Highly similar sequences” e parâmetros padrão na pontuação do alinhamento.

3.4 Identificação das rotas metabólicas a partir dos genes encontrados

Na etapa 3, o processamento dos genes relacionados com a produção dos metabólitos envolvidos na metabolização do nitrogênio nas bactérias foi realizado com a base de dados KEGG (OGATA et al., 1999). Os dados de entrada no KEGG foram os nomes desses genes identificados pelo alinhamento. Logo, os genes foram procurados nas rotas metabólicas obtidas pela ferramenta KEGG, obtendo no final um grafo (Figura. 11) para cada genoma com informação geral de todo o metabolismo das bactérias. No entanto, foi extraído um subgrafo só com a informação relacionada aos genes identificados na metabolização do nitrogênio para cada cepa das bactérias *Klebsiella* spp.

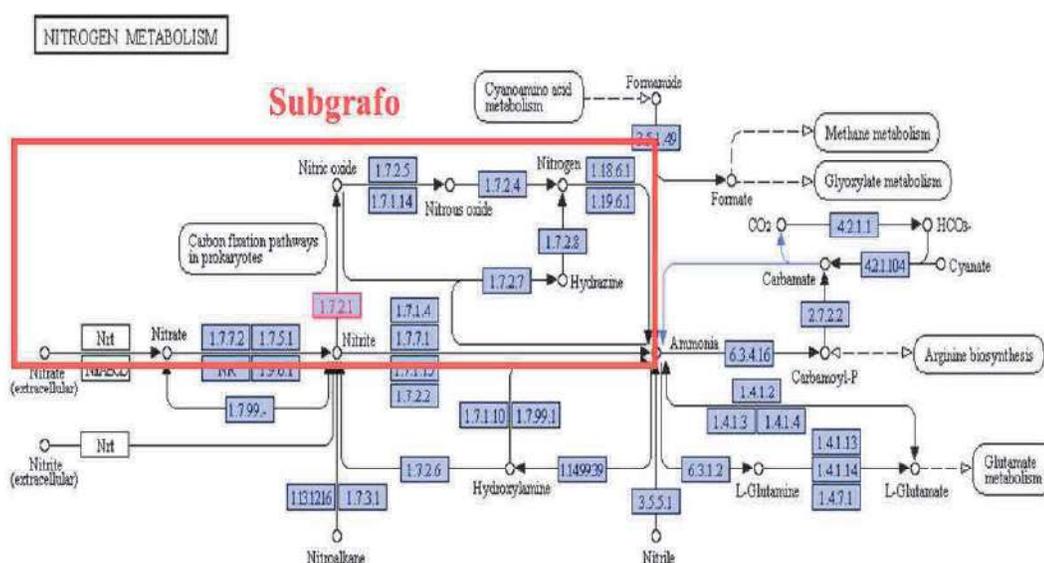


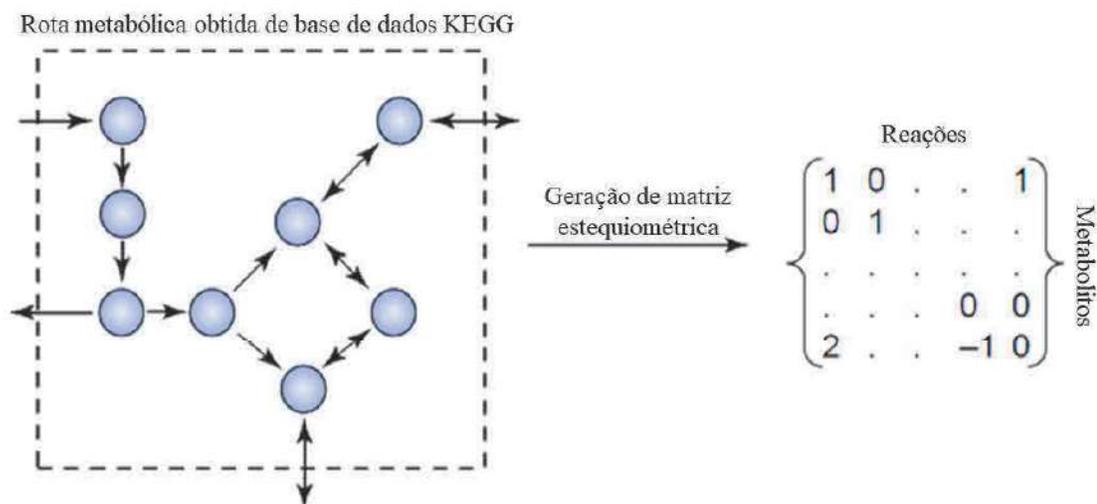
Figura 11 – Grafo da metabolização de nitrogênio nas bactérias *Klebsiella*

Fonte: KEGG

3.5 Modelagem estequiométrica das rotas metabólicas

Considerando os grafos obtidos no passo anterior, na etapa 4 para cada cepa, foram realizados os modelamentos estequiométricos. As rotas metabólicas obtidas na etapa anterior foram representadas matematicamente pela matriz estequiométrica (S). A matriz contém os coeficientes estequiométricos fornecidos pela base de dados KEGG relacionados aos metabolitos envolvidos nos mecanismos moleculares das reações químicas da metabolização do nitrogênio nas bactérias. Na matriz S o número de reações presentes são colunas enquanto os metabolitos (substrato, produtos e cofatores) estão em linhas (Fig 12) (PAPIN et al., 2003).

Figura 12 – Modelagem estequiométrica



Fonte: (PAPIN et al., 2003)

3.6 Análise das rotas metabólicas utilizando CNA

Na etapa 5, as análises das rotas metabólicas foram realizadas com os dados obtidos dos grafos das reações para cada cepa de *Klebsiella* spp., onde pode-se extrair a informação dos valores estequiométricos. As análises foram obtidas pelo processamento na ferramenta CellNetAnalyzer (KLAMT et al., 2007) para cada cepa das bactérias, e os resultados foram os *Elementary mode* (EM, pelas suas siglas em inglês). As análises dos EM calcula todas as soluções de possíveis caminhos de metabolitos de intercâmbio de entrada até metabolitos de saída na rota metabólica (TRINH et al., 2009). Para o modelo foram considerados 10 metabolitos, além de identificar 12 reações (caminhos internos), 2 produtos de entrada e 2 produtos de saída para fazer a análise dos EM para cada conjunto de dados de cada cepa das bactérias.

4 Resultados e discussão: execução do processo de reconstrução de rotas metabólicas

No presente capítulo são desenvolvidos os pontos referentes à discussão dos resultados obtidos, além disso, é descrito o processo necessário na identificação dos mecanismos moleculares das bactérias *Klebsiella* potencialmente alvos na engenharia metabólica na produção de *C. capitata* pela técnica SIT. A abordagem trabalhada foi aplicada aos dados biológicos realizando as análises das possíveis rotas metabólicas. Os seguintes tópicos apresentam os resultados das 5 etapas aplicadas para os dados das três bactérias.

Desde as primeiras aplicações dos sequenciamentos dos genomas, as áreas de biologia e ciência da computação têm evoluído conjuntamente no processamento dos dados biológicos, produzindo cada vez em maior escala e com poder computacional necessário no processamento desses dados. Por conseguinte, a modelagem de sistemas biológicos e o desenvolvimento de ferramentas computacionais que permitissem resolver questões biológicas têm ganhado notoriedade. O controle biológico de pragas é alvo dos principais problemas biológicos cuja solução pode ser complementada através da aplicação de processamento computacional de dados biológicos. Em virtude da complexidade do problema, múltiplas técnicas são utilizadas para aprimorar o controle da praga como controle químico e controle biológico principalmente das pragas de insetos. A mosca *Ceratitidis capitata* é considerada uma das pragas mais prejudiciais pelas perdas que ela gera. Por consequência múltiplas pesquisas trabalham no aprimoramento do controle biológico da mosca utilizando a técnica SIT.

O controle biológico da mosca *C. capitata* usando a técnica SIT gera adultos machos inférteis. Esses machos são liberados em áreas definidas onde podem copular com fêmeas selvagens, sem gerar descendência, fazendo com que a população da mosca no meio ambiente seja reduzida (KLASSEN et al., 2005). Por isso, a técnica SIT é aprimorada constantemente e uma das abordagens de aprimoramento é a adição de bactérias na dieta da *C. capitata* após ser irradiada. Essas bactérias são conhecidas por ser potencialmente simbióticas com a mosca do mediterrâneo. As bactérias utilizadas, do gênero *Klebsiella*, possuem mecanismos moleculares que a mosca precisa para seu desenvolvimento no ciclo de produção pela técnica SIT (AHARON et al., 2013).

O objetivo da dissertação foi identificar mecanismos moleculares com auxílio da técnica de reconstrução de rotas metabólicas utilizando resultados de sequenciamento

dos genomas de três bactérias *Klebsiella*. O intuito desses mecanismos moleculares é ser utilizados como alvos da engenharia genética e metabólica para poder aprimorar processos bioquímicos, digestivos e de imunidade da *C. capitata* produzida pela técnica SIT. Portanto, para cumprir esse objetivo os três genomas de *Klebsiella* foram processados pela metodologia proposta (Fig. 10) e as ferramentas computacionais apresentadas na Tabela 1.

4.1 Montagem dos contigs a partir dos dados de sequenciamento das bactérias *Klebsiella*

Na etapa 1 do processo, a ferramenta SPAdes gerou a montagem dos *contigs* das três cepas das bactérias do gênero *Klebsiella*. Os genomas destas bactérias têm aproximadamente comprimento de 6 Mbp e os resultados são apresentados na Tabela 2 em relação ao número de *contigs* gerados após realizar o processamento de montagem de *contigs* com a ferramenta SPAdes. Neste caso todos os *contigs* gerados são indispensáveis por causa das sequências de DNA contidas em cada um deles. Os genes de interesse que tem participação nos mecanismos moleculares da metabolização do nitrogênio nas bactérias podem estar presentes em um ou dois *contigs*. Sendo assim, é necessário conservar todos os *contigs* gerados para fazer o alinhamento e identificar os genes. Na tabela também são apresentados os *contigs* que tem maior e menor comprimento de acordo à cepa de bactéria onde foram extraídos. Os *contigs* com maior comprimento podem ser melhores para essa dissertação, mas é importante avaliar todos os *contigs* e procurar sequências de genes conhecidos.

Tabela 2 – Contigs obtidos para cada cepa das bactérias de gênero *Klebsiella* spp. Fonte: (Autoria Própria).

	<i>K. pneumoniae</i> - cepa 1	<i>K. oxytoca</i> - cepa 2	<i>K. oxytoca</i> - cepa 3
Maior contig (pb)	311,754	522,307	845,485
Menor contig (pb)	78	78	78
No. total de contigs	958	233	331

O ponto de partida desta dissertação de dissertação foram os dados fornecidos pela pesquisa realizada na Guatemala. Esses dados foram obtidos na forma de 2 arquivos de sequenciamento R1 e R2 para cada um dos genomas de *Klebsiella*. Os documentos R1 e R2 na etapa 1 foram utilizados para realizar a montagem dos contigs. Esses contigs (Tabela 2) são fragmentos de DNA com os quais pode-se reconstruir sequências completas de genes ou genomas. Dentro dos contigs obtidos das três bactérias *Klebsiella* é possível identificar diferenças na quantidade de contigs gerados. A diferença deve-se ao fato de ser três bactérias diferentes. A principal diferença é a natureza dessas bactérias,

a *K. pneumoniae* – cepa 1 foi utilizada como microrganismo de referência, a *K. oxytoca* – cepa 2 foi isolada da microbiota da *C. capitata* produzida na indústria e por último a *K. oxytoca* – cepa 3 foi isolada da microbiota da mosca silvestre.

A montagem dos contigs das cepas 1 e 3 apresentam maior número de contigs obtidos em comparação com os contigs obtidos da cepa 2. Por essa razão foram procuradas sequências alvo como por exemplo fatores de virulência que podem estar presentes nestas cepas e podam ser um critério de diferença; esses fatores foram o gene *shv* e *rmpa*. O gene *rmpa* está presente nas três cepas de bactérias, mas o gene *shv* esta ausente na cepa 2. Mesmo que a diferença possa estar relacionada com a presença/ausência de determinadas sequências genéticas. De acordo com (LIAO et al., 2011) as bactérias do gênero *Klebsiella* contém resistências aos antibióticos, porém, o 20% das cepas deste gênero produz β -lactamases. Foi identificada que a cepa 2 não contém a sequência genética relacionada com o gene *shv*, associado com β -lactamase sulfidrílo. Esse gene codifica resistência aos antibióticos β -lactâmicos. Essa ausência foi identificada pela técnica de bioengenharia reversa posto que é possível uma modificação genética na cepa utilizada na indústria. Baseado nessa informação nos propomos que para a indústria que utiliza a *K. oxytoca* – cepa 2 na dieta da mosca *C. capitata* é importante não liberar moscas com bactérias resistentes aos antibióticos β -lactâmicos pelo perigo que representa. As bactérias podem realizar a transferência de material genético, neste caso esse gene, e criar bactérias resistentes perigosas no meio ambiente gerando resistências aos antibióticos.

4.2 Identificação dos genes envolvidos na metabolização do nitrogênio a partir dos contigs gerados

O processamento dos *contigs* obtidos na etapa 2 foi realizado com o BLAST, identificando regiões específicas similares aos genes envolvidos nos mecanismos moleculares da metabolização de nitrogênio nas bactérias. A informação desses genes foi obtida das bases de dados NBI, EBI e KEGG. Os genes identificados (Tabela 3 e 4) têm importância na simbiose entre as bactérias *Klebsiella*. e seu hospedeiro *Ceratitis capitata*, devido ao fato de que são genes envolvidos na metabolização de nitrogênio na bactéria com impacto no processo digestivo da mosca. Esses genes têm a função de fixação, assimilação e regulação do nitrogênio e os demais genes tem importância na síntese de enzimas envolvidas no metabolismo de nitrogênio. Na Tabela 3 são apresentados os genes identificados na bactéria *Klebsiella pneumoniae* - cepa 1.

Tabela 3 – Genes identificados a partir dos contigs da bactéria *Klebsiella pneumoniae* - cepa 1 (Fonte: Autoria Própria).

Nome do gene	Função	Comprimento (pb)	No. de acesso
<i>nifU</i>	Fixação de nitrogênio	387	KPN_02861
<i>hpxB</i>	Assimilação de nitrogênio	738	MF417539.1
<i>allC</i>	Assimilação de nitrogênio	2516	2273639_02901
<i>ureG</i>	Assimilação de nitrogênio	618	29381131
<i>glnA</i>	Síntese de glutamina sintetase	928	LC011557.1
<i>gltB</i>	Síntese de glutamato sintetase	7930	AY035435.1
<i>gluD</i>	Síntese de glutamato desidrogenase	1382	29379502
<i>ntrA</i>	Regulação de nitrogênio	1935	X03147.1
<i>ntrB</i>	Regulação de nitrogênio	1408	X03146.1
<i>ntrC</i>	Regulação de nitrogênio	1575	X02617.1
<i>narI</i>	Síntese nitrato reductase	681	KONIH1_14050
<i>narH</i>	Síntese nitrato reductase	1545	KONIH1_14060
<i>narZ</i>	Síntese nitrato reductase	3741	KONIH1_14070
<i>norBC</i>	Síntese oxido nitroso reductase	2734	KONIH1_22945
<i>nirK</i>	Síntese nitrito reductase	2874	AAA25099.1

Na Tabela 4, são listados os genes identificados para as bactérias *Klebsiella oxytoca* - cepa 2 e *Klebsiella oxytoca* - cepa 3. Esses genes estão envolvidos nos mecanismos moleculares da metabolização de nitrogênio nas bactérias *Klebsiella oxytoca*. A diferença dos genes da *Klebsiella pneumoniae* e as cepas de *Klebsiella oxytoca* são os genes envolvidos na fixação do nitrogênio. A cepa 1 precisa só de 1 gene, mas a cepa 2 e 3 precisam de 3 genes para realizar a função de fixação do nitrogênio nas bactérias.

Tabela 4 – Genes identificados a partir dos contigs das bactérias *Klebsiella oxytoca* - cepa 2 e *Klebsiella oxytoca* - cepa 3 (Fonte: Autoria Própria).

Nome do gene	Função	Comprimento (pb)	No. de acesso
<i>nifH</i>	Fixação de nitrogênio	263	KJ940124.1
<i>nifD</i>	Fixação de nitrogênio	1523	Y00316.1
<i>nifK</i>	Fixação de nitrogênio	1557	CAA29588.1
<i>allS</i>	Assimilação de nitrogênio	738	MF417539.1
<i>allC</i>	Assimilação de nitrogênio	2516	2273639_02901
<i>ureG</i>	Assimilação de nitrogênio	618	29381131
<i>glnA</i>	Síntese de glutamina sintetase	928	LC011557.1
<i>gltB</i>	Síntese de glutamato sintetase	7930	AY035435.1
<i>gluD</i>	Síntese de glutamato desidrogenase	1382	29379502
<i>ntrA</i>	Regulação de nitrogênio	1935	X03147.1
<i>ntrB</i>	Regulação de nitrogênio	1408	X03146.1
<i>ntrC</i>	Regulação de nitrogênio	1575	X02617.1
<i>narI</i>	Síntese nitrato reductase	681	KONIH1_14050
<i>narH</i>	Síntese nitrato reductase	1545	KONIH1_14060
<i>narZ</i>	Síntese nitrato reductase	3741	KONIH1_14070
<i>norBC</i>	Síntese oxido nitroso reductase	2734	KONIH1_22945
<i>nirK</i>	Síntese nitrito reductase	2874	AAA25099.1

Por outro lado, encontraram-se na literatura informação de genes envolvidos na metabolização do nitrogênio na bactéria. Esses genes foram identificados na etapa 2 com ajuda da ferramenta BLAST. O processamento procurou regiões similares entre os contigs e genes obtidos das bases de dados do NCBI, EBI e KEGG. Dessa forma foram encontrados 15 genes (Tabela 3) para a cepa *K. pneumoniae* – cepa 1 e 17 genes para as cepas de *K. oxytoca* - cepas 2 e 3 - (Tabela 4). Esses genes têm relação potencialmente simbiótica com a mosca pois estão envolvidos no metabolismo do nitrogênio, especificamente na fixação, assimilação e regulação do nitrogênio. Nos genes identificados, existe uma diferença entre os genes envolvidos na fixação do nitrogênio devido ao fato de se tratar de espécies distintas de bactérias. Entre as espécies, a maioria os genes são conservados, porém, pela diversidade genética os genes podem mudar de sequências, mas não de função, como neste caso. Esse mecanismo molecular da metabolização do nitrogênio é importante, uma vez que a maior parte da dieta dos insetos é pobre em nitrogênio (JR., 1985; WALDBAUER; FRIENDMAN, 1991). Dessa forma, a mosca precisa da metabolização do nitrogênio para obter compostos necessários para seu desenvolvimento no ciclo de vida.

4.3 Identificação das rotas metabólicas da metabolização de nitrogênio a partir dos genes encontrados

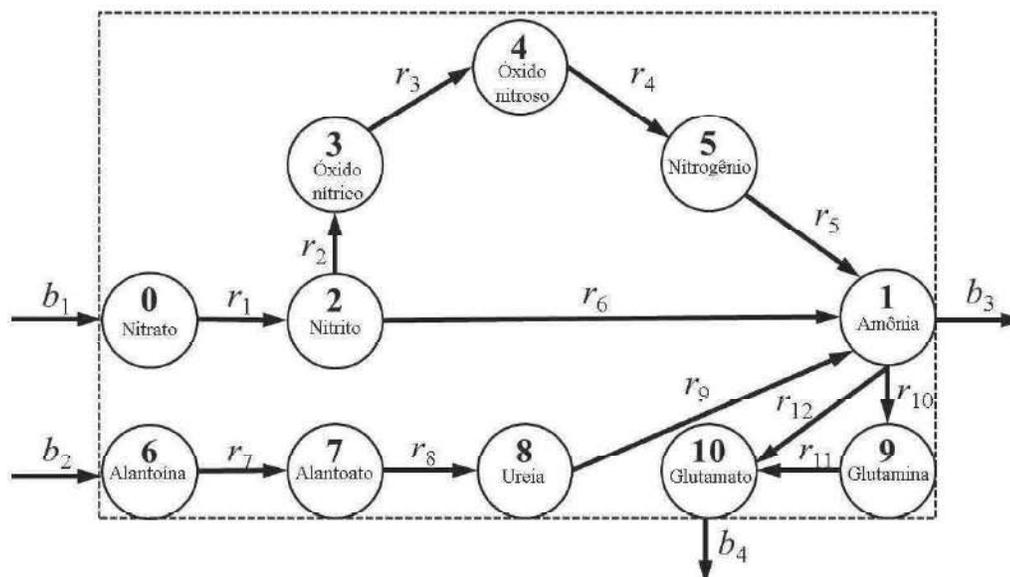
Na etapa 3, foram identificadas as possíveis rotas metabólicas pela base de dados KEGG (OGATA et al., 1999). Cada gene encontrado pelo BLAST codifica enzimas ou unidades de proteínas que tem função no metabolismo das bactérias *Klebsiella*. Os compostos envolvidos nas rotas metabólicas encontradas (Tabela 5), são 11 metabolitos, sendo cada um representado pelo número de vértice correspondente no subgrafo representado na Fig 13. A numeração apresentada vai de 0 até 10, em que representam as funções que os metabolitos encontrados desempenham na metabolização do nitrogênio nas bactérias, onde cada metabolito envolvido é um composto com nitrogênio na estrutura química. O intuito da metabolização do nitrogênio é obter compostos químicos mais simples para a utilização da mosca.

Tabela 5 – Metabolitos envolvidos no subgrafo (Fonte: Autoria Própria).

Vértice	Metabolito
0	Nitrato
1	Amônia
2	Nitrito
3	Óxido nítrico
4	Oxido nitroso
5	Nitrogênio
6	S-Alantoína
7	Alantoato
8	Ureia
9	L-Glutamina
10	L-Glutamato

Na Fig 13 é apresentado o subgrafo, levando em consideração cada um dos metabolitos encontrados. O subgrafo é a representação das rotas metabólicas envolvidas na metabolização de nitrogênio nas bactérias *Klebsiella*. Esse subgrafo foi projetado de acordo com a teoria dos grafos, representando cada metabolito como vértices e cada reação, que ocorre na metabolização do nitrogênio nas bactérias, como arestas conectando os vértices. Além disso, foram fixados os metabólitos de entrada e saída, esses são representados por b_i , onde b_1 (Nitrato) e b_2 (S-Alantoína) são metabólitos de entrada e b_3 (Amônia) e b_4 (L-Glutamato) são os metabólitos de saída.

Figura 13 – Subgrafo de reações químicas envolvidas na metabolização do nitrogênio nas bactérias *Klebsiella*



Fonte: Autoria própria

4.4 Modelagem estequiométrica das rotas metabólicas das bactérias *Klebsiella*

Na etapa 4 do processo foi gerada a matriz estequiométrica (S) (Fig 14), na qual possui 16 colunas que correspondem ao número de reações químicas no grafo. As primeiras 12 colunas são as representações das reações que acontecem dentro da metabolização de nitrogênio nas bactérias e as últimas 4 colunas são as reações de entrada e saída dos metabolitos fixados. As 11 linhas são os metabolitos envolvidos na via de metabolização de nitrogênio nas bactérias do gênero *Klebsiella*. Os valores na matriz são coeficientes estequiométricos fornecidos pela base de dados KEGG para cada reação. Os coeficientes que têm o sinal negativo (-) indicam que o metabolito foi consumido na reação e os coeficientes com sinal positivo (+) são os produtos gerados pela reação que aconteceu.

Figura 14 – Matriz estequiométrica das rotas metabólicas envolvidas na metabolização de nitrogênio nas bactérias *Klebsiella*.

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	b_1	b_2	b_3	b_4	
Vértice																	
0	-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Nitrato
1	0	0	0	0	1	1	0	0	1	-1	0	-1	0	0	-1	0	Amônia
2	1	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	Nitrito
3	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	Óxido nítrico
4	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	Oxido nitroso
5	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	Nitrogênio
6	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	1	0	S-Alantoin
7	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	Alantoato
8	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	Ureia
9	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	L-Glutamina
10	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	-1	L-Glutamato

Fonte: Autoria própria

Os metabolitos e reações envolvidas na metabolização do nitrogênio foram representados com o subgrafo obtido do KEGG, sendo utilizado na etapa 4 para construir a modelagem estequiométrica, complementada com a informação dos coeficientes estequiométricos fornecidos pela base de dados. O modelo foi baseado na equação 1; para trabalhar baseado no estado estável das rotas metabólicas. A matriz presente na Fig 14 mostra os coeficientes estequiométricos envolvidos nas reações da metabolização do nitrogênio nas bactérias *Klebsiella*. A modelagem estequiométrica gerou as representações na forma matricial das reações que acontecem dentro do sistema da metabolização do nitrogênio nas bactérias. A utilização dos conceitos matemáticos aplicados aos dados biológicos foram necessários no momento de analisar as rotas metabólicas, permitindo supor as condições das rotas metabólicas de acordo com o estado estável das concentrações dos metabolitos.

4.5 Análise das rotas metabólicas das cepas de *Klebsiella* sp. utilizando CNA

As análises dos EM são realizadas pela ferramenta CellNetAnalyzer que permite a visualização do grafo gerado (Fig 13). O modelo foi elaborado com o formato de fluxos (caminhos) entre vértices, inserindo os coeficientes estequiométricos de cada reação de acordo com a codificação do grafo. Os EM obtidos são apresentados na representação matricial, onde as colunas são o número de EM encontrados e as filas são as reações envolvidas nos EM (Fig 15).

Figura 15 – Representação matricial dos EM obtidos pelas análises das rotas metabólicas envolvidas na metabolização de nitrogênio nas bactérias *Klebsiella*.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	Vértice inicial	Aresta	Vértice Final
P =	0	1	0	1	1	1	1	1	0	0	0	r_1	2
	0	0	0	0	1	1	0	1	0	0	2	r_2	3
	0	0	0	0	1	1	0	1	0	0	3	r_3	4
	0	0	0	0	1	1	0	1	0	0	4	r_4	5
	0	0	0	0	1	1	0	1	0	0	5	r_5	1
	0	1	0	1	0	0	1	0	0	0	2	r_6	1
	1	0	1	0	0	0	0	0	1	0	6	r_7	7
	1	0	1	0	0	0	0	0	1	0	7	r_8	8
	1	0	1	0	0	0	0	0	1	0	8	r_9	1
	1	1	0	0	0	1	0	0	0	1	1	r_{10}	9
	1	1	0	0	0	1	0	0	0	1	9	r_{11}	10
	0	0	1	1	1	0	0	0	0	1	1	r_{12}	10
	0	1	0	1	1	1	1	1	0	0	-	b_1	0
	1	0	1	0	0	0	0	0	1	0	-	b_2	6
	0	0	0	0	0	0	1	1	1	0	-	b_3	1
	1	1	1	1	1	1	0	0	0	0	-	b_4	10

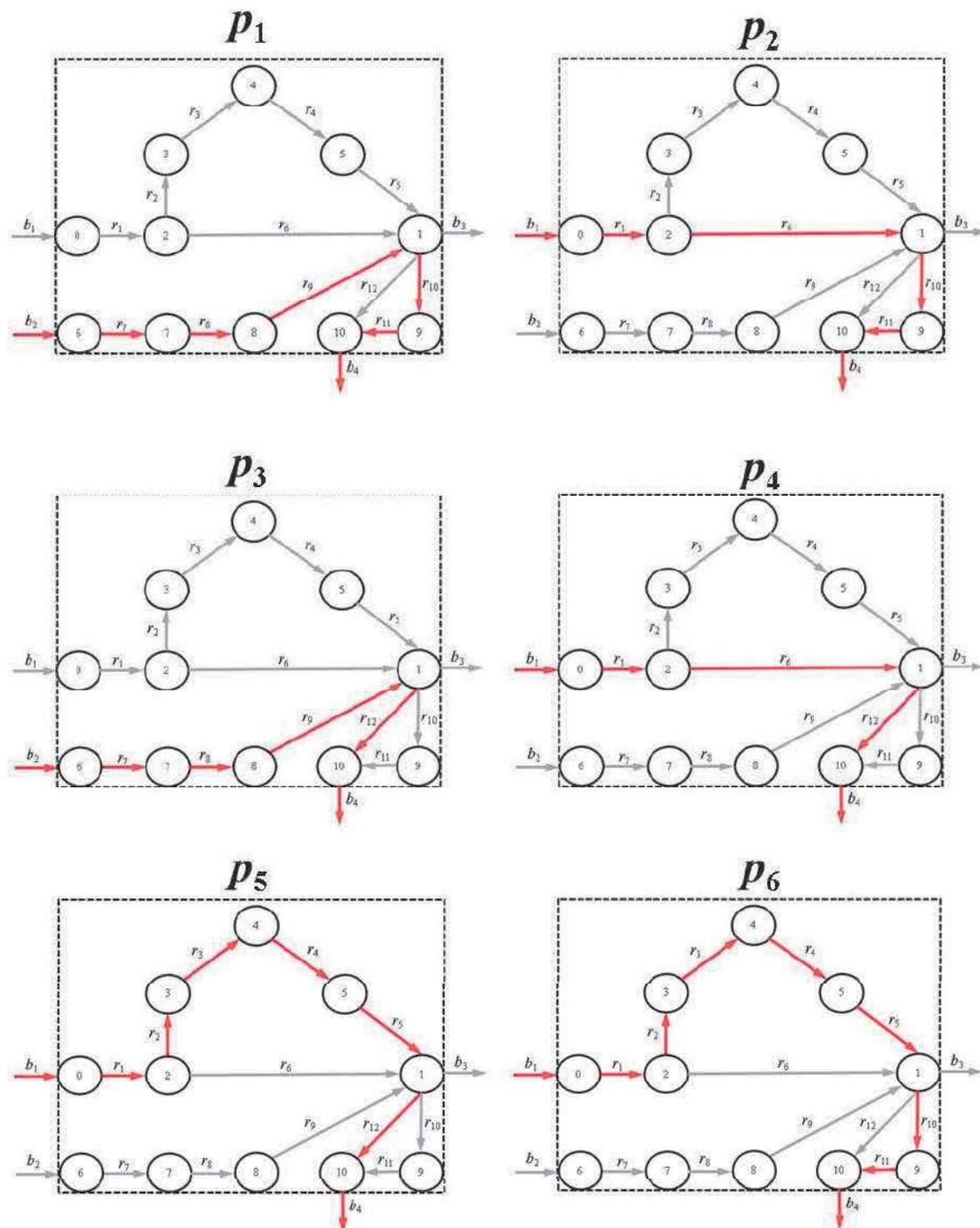
Fonte: Autoria própria

De acordo com o subgrafo obtido na Fig 13 na etapa 5, foi utilizada a análise de rotas metabólicas pelos “Elementary Modes”. Esta técnica permite identificar todos os possíveis caminhos que existem no sistema definido pelo usuário (TRINH et al., 2009), esses EM foram obtidos pela ferramenta CellNetAnalyzer, que tem uma interface gráfica no MATLAB. Dos 10 EM obtidos, 9 foram úteis para essa dissertação, pois segundo (SCHILLING et al., 2000) são classificados como tipo I, ou seja, são rotas funcionais onde o intercâmbio de fluxo é ativo, indo desde um vértice com um metabolito de entrada até qualquer vértice de saída fixado. Além disso o EM No. 10 foi descartado porque é de tipo III, o que indica que forma ciclos internos (fluxo inativo) na rota metabólica (MAARLEVELD et al., 2013).

As análises das rotas metabólicas geraram resultados iguais nas três cepas de *Klebsiella*, identificando os mecanismos moleculares dessas bactérias e sua importância no ciclo de vida da mosca *C. capitata*. No ciclo de vida da mosca o processo da metamorfose precisa do mecanismo molecular da metabolização de nitrogênio. A primeira abordagem foi focada nos metabolitos de saída fixados nas análises dos EM. A amônia é um produto de excreção da metabolização de nitrogênio na bactéria, mas também é um reagente na formação de glutamina dentro do ciclo da reutilização de amônia. Logo a glutamina reage com a enzima glutamato sintetase e produz L – Glutamato. Os resultados bioquímicos destas reações são relacionados de acordo com a atividade da enzima glutamato desidrogenase, onde o substrato desta enzima é o L-Glutamato, e o produto é a amônia. A atividade enzimática (Fig 5) apresenta uma queda aproximadamente às 40 hrs, o que indica que a enzima não tem substrato para reagir, não conseguindo produzir amônia. Porém o ciclo pode acontecer em baixas concentrações de amônia, produzindo glutamina e voltar a gerar L-Glutamato. O L-Glutamato é uma molécula de interesse nas simbioses com a mosca, pois é um aminoácido utilizado para realizar a síntese de novas proteínas nos estágios da metamorfose da mosca.

Nos EM são apresentados os 10 EM obtidos das análises de rotas metabólicas das bactérias *Klebsiella* (Fig 16 e 17). Os EM foram codificados como p_i . Sugerido por (PAPIN et al., 2003) foram classificados em 2 grupos. O primeiro grupo apresenta os resultados obtidos para as rotas onde o produto final foi o L-Glutamato, composto importante na relação simbiótica entre a bactéria e a mosca (Fig 16).

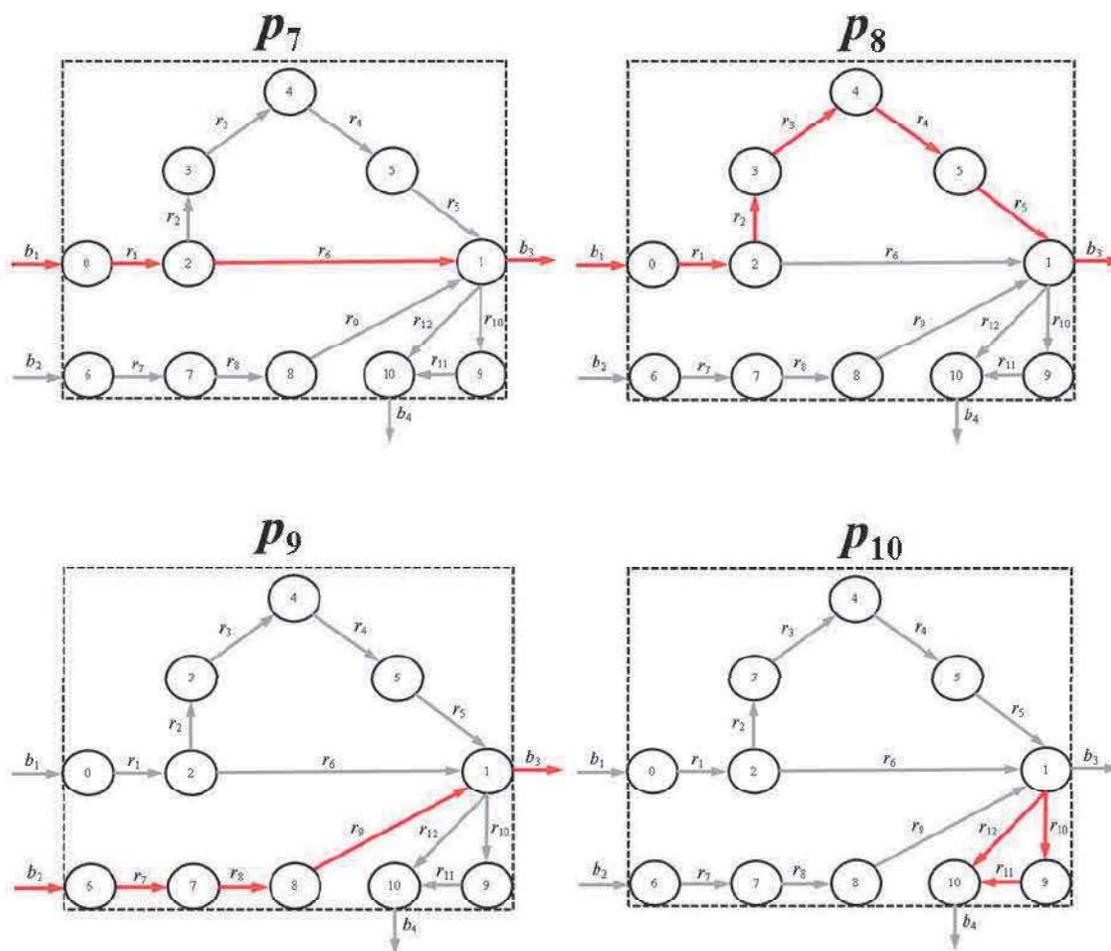
Figura 16 – Elementary Modes gerados pelas análises das rotas metabólicas nas bactérias *Klebsiella* que geram L-Glutamato



Fonte: Autoria própria

O segundo grupo (Fig 17) são representados os EM restantes. Os EM $p_7 - p_9$ mostram as possíveis rotas onde o produto final é a amônia. O p_{10} representa um caminho interno que não produz intercambio de metabolitos no sistema, e não inicia em um vértice de entrada, além disso, não gera um metabolito de saída.

Figura 17 – Elementary Modes gerados pelas análises das rotas metabólicas nas bactérias *Klebsiella*. que geram amônia e reações internas



Fonte: Autoria própria

No processo da metamorfose a *C. capitata* sofre processos bioquímicos necessários para seu desenvolvimento. A segunda abordagem foi avaliar o fluxo de energia necessário para completar a metamorfose desde a fase de ovo até a adulta da mosca. Nestas etapas a mosca precisa de energia em múltiplas formas para realizar a transformação. O foco desta análise foram as proteínas como uma fonte principal dessa energia, pois são essas as moléculas que sofrem digestão na metamorfose, gerando aminoácidos livres. Esses aminoácidos são usados principalmente nas sínteses das novas proteínas

nos diferentes estágios da mosca. Segundo (NESTEL et al., 2003), essas proteínas são de cutícula de larva e pupa, e a taxa metabólica (Fig 4) delas apresentam variações durante o tempo de desenvolvimento da mosca. As variações ocorrem porque no processo de digestão, a concentração de proteína decresce, isso gera aminoácidos livres no meio. Após aproximadamente 60 hrs, esses mesmos aminoácidos são utilizados nas sínteses de proteínas no estágio seguinte da mosca.

Após a identificação do mecanismo molecular da metabolização do nitrogênio nas bactérias *Klebsiella* e sua importância no desenvolvimento da mosca, é importante adicionar que essas bactérias também têm a capacidade de concorrer com outras bactérias. Na microbiota da mosca estão presentes múltiplas enterobactérias e *Pseudomonas* spp. Na figura são apresentadas as porções de bactérias presentes em diferentes estágios do desenvolvimento da mosca. De acordo com os resultados foi proposto o mecanismo molecular de resistência/imunidade entre *Klebsiella* e *Pseudomonas* spp. como alvo também da engenharia metabólica. A *Klebsiella* acaba deslocando a *Pseudomonas* da microbiota da mosca. Isso tem importância na imunidade/resistência a *Pseudomonas* spp. pois essa bactéria é um patógeno da mosca, gerando efeito negativo na porcentagem da eclosão dos ovos, e reduzindo a porcentagem de ovos exitosos na eclosão (YISHAI et al., 2018).

5 Considerações finais

Baseado nos resultados no presente capítulo apresentamos as principais conclusões referentes ao processo proposto de reconstrução de rotas metabólicas e as propostas para trabalhos futuros no modelo experimental apresentado.

5.1 Conclusões

Analisando os resultados obtidos conclui-se:

- Os resultados da reconstrução de rotas metabólicas permitiram comparar as rotas envolvidas na metabolização do nitrogênio nas três cepas de bactéria. A comparação foi feita a partir dos metabolitos e as reações presentes. No entanto entre as bactérias *Klebsiella* não existe diferença entre as rotas metabólicas. Por consequência, foram obtidos dados de energia e termodinâmica que permitissem diferenciar as cepas. Esses dados foram relacionados com a metamorfose da mosca, uma vez que nessa fase a mosca utiliza múltiplas fontes de energia para completar o processo. Dessa forma, foi estudada a utilização de proteínas como fonte de energia no processo de metamorfose. Neste processo ocorre a digestão de proteínas gerando aminoácidos livres como o L-Glutamato.
- O mecanismo molecular identificado da metabolização do nitrogênio é um possível alvo na engenharia metabólica e genética das bactérias *Klebsiella*. O processo molecular pode ser modificado para melhorar a produção de metabolitos de interesse como o L-Glutamato, sendo este composto um metabolito importante nas sínteses das proteínas da larva e pupa na metamorfose da mosca. Estas bactérias modificadas pela engenharia metabólica e genética podem ser adicionadas na dieta da *C. capitata* produzida pela técnica SIT.
- O mecanismo molecular identificado, de imunidade/resistência à *Pseudomonas* spp., está relacionado com a concorrência com as bactérias *Klebsiella*. A *Pseudomonas* spp. é uma bactéria patogena no ciclo de vida da mosca *C. capitata* e sua presença na microbiota faz que o porcentagem da eclosão de ovos seja menor (80%) ao porcentagem de eclosão de ovos atingido pelas moscas com as bactérias *Klebsiella* na microbiota (90%). Foi proposto que o mecanismo de imunidade/resistência à *Pseudomonas* spp. é devido a genes de proteínas que faz a *Klebsiella* deslocar a *Pseudomonas* spp.
- Os resultados de reconstrução de rotas metabólicas mostram a identificação dos mecanismos de metabolização do nitrogênio e de resistência/imunidade a *Pseudo-*

monas. Esses mecanismos são realizados pela bactéria *Klebsiella* e produz produtos metabólicos de importância como o L-Glutamato e Amônia utilizados no processo de metamorfoses da mosca. Os resultados apresentados concordam com pesquisas realizadas anteriormente na área de reconstrução de rotas metabólicas de *K. pneumoniae* que mostram a importância das bactérias *Klebsiella* spp. para seu uso na biotecnologia de acordo com a capacidade de metabolizar produtos nitrogenados. Esses produtos são utilizados pela mosca e podem melhorar a produção desta logo da aplicação da técnica SIT.

5.2 Trabalhos Futuros

Baseados na identificação de mecanismos moleculares pela reconstrução de rotas metabólicas, sugerem-se os trabalhos a seguir.

- Obter dados de concentrações de metabolitos que permitam complementar a técnica de reconstrução de rotas metabólicas e identificar os fluxos metabólicos mais prováveis.
- Obter e processar dados termodinâmicos das rotas metabólicas e do processo de metamorfose da mosca do mediterrâneo, para avaliar espontaneidade das reações que podem ocorrer.
- Realizar bioengenharia reversa para identificar outras sequências que possam-se relacionar com a diferença de contigs gerados entre cepas. Além disso, adicionar técnicas tipo CRISPR-Cas9 no processo de edição genética de genomas e comparar resultados.
- Identificar os mecanismos moleculares envolvidos com outros possíveis genes de imunidade e resistência às bactérias potencialmente simbióticas.
- Realizar engenharia metabólica e genética nos mecanismos moleculares identificados, bem como realizar a experimentação no laboratório para validar os dados gerados in silico pelas análises computacionais.

Referências

ADLEMAN, L. Molecular Computation of Solutions to Combinatorial Problems. **Science**, v. 266, n. 5187, p. 1021–1024, 1994.

AHARON, Y.; PASTERNAK, Z.; YOSEF, M. B.; BEHAR, A.; LAUZON, C.; YUVAL, B.; JURKEVITCH, E. Phylogenetic, metabolic, and taxonomic diversities shape mediterranean fruit fly microbiotas during ontogeny. **Applied and Environmental Microbiology**, v. 79, n. 1, p. 303–313, 2013. ISSN 00992240.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 1990. ISSN 00222836.

AMI, E. B.; YUVAL, B.; JURKEVITCH, E. Manipulation of the microbiota of mass-reared Mediterranean fruit flies *Ceratitis capitata* (Diptera : Tephritidae) improves sterile male sexual performance. **The ISME Journal**, Nature Publishing Group, v. 4, n. 1, p. 28–37, 2009. ISSN 1751-7362. Disponível em: <<http://dx.doi.org/10.1038/ismej.2009.82>>.

BANKEVICH, A.; KULIKOV, A. S.; PRJIBELSKI, A. D.; TESLER, G.; VYAHHI, N.; SIROTKIN, A. V.; PHAM, S.; DVORKIN, M.; PEVZNER, P. A.; BANKEVICH, A.; NIKOLENKO, S. I.; PYSHKIN, A. V.; NURK, S.; GUREVICH, A. A.; ANTIPOV, D.; ALEKSEYEV, M. A.; LESIN, V. M. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, 2012. ISSN 1066-5277.

BASTIDE, M.; MCCOMBIE, W. Assembling genomic DNA sequences with PHRAP. **Bioinformatics**, v. 11, n. 4, p. 1–15, 2007.

BEHAR, A.; YUVAL, B.; JURKEVITCH, E. Enterobacteria-mediated nitrogen fixation in natural populations of the fruit fly *Ceratitis capitata*. **Molecular Ecology**, v. 14, n. 9, p. 2637–2643, 2005. ISSN 09621083.

CLARK, D. **Molecular Biology**. 1 ed.. ed. [S.l.]: Elsevier, 2009. 803 p. ISBN 9780123785893.

COX, M. M. **PRINCIPLES OF BIOCHEMISTRY**. 5th. ed. Wisconsin, USA: W. H. Freeman and Company, 2007. 1294 p. ISBN 9780716771081.

DEMIRBAS-UZEL, G. **Improving Sterile Insect Technique for tsetse flies through research on their symbiont and pathogens** *Improving Sterile Insect Technique for tsetse flies through research on their symbiont and pathogens ausgeführt zum Zwecke der Erlangung des akademischen*. 257 p. Tese (Doutorado) — Technische Universität Wien, 2018.

FAUST, K.; CROES, D.; HELDEN, J. V. Prediction of metabolic pathways from genome-scale metabolic networks. **BioSystems**, Elsevier Ireland Ltd, v. 105, n. 2, p. 109–121, 2011. ISSN 0303-2647. Disponível em: <<http://dx.doi.org/10.1016/j.biosystems.2011.05.004>>.

FEIST, A. M.; HERRGÅRD, M. J.; THIELE, I.; REED, J. L.; PALSSON, B. Ø. Reconstruction of biochemical networks in microorganisms. **Nature reviews. Microbiology**, v. 7, p. 129–143, 2008.

FIGUEROA, N.; NAVARRETE, C. **Caracterización genética de cepas de Klebsiella spp. aislada de ambientes asociados a Ceratitis capitata (Weidemann), por medio de secuenciación de su genoma, extracción de plásmidos y Rep-PCR**. 77 p. Tese (Doutorado) — Universidad del Valle de Guatemala, 2018.

GUERFALI, M. M.; DJOBBI, W.; CHARAABI, K.; HAMDEN, H.; FADHL, S.; MARZOUKI, W.; DHAOUEDI, F.; CHEVRIER, C. Evaluation of *Providencia rettgeri* pathogenicity against laboratory Mediterranean fruit fly strain (*Ceratitidis capitata*). **PloS one**, v. 13, n. 5, p. 1–18, 2018.

JR., F. S. Food utilization by insects: interpretation of observed differences between dry weight and energy efficiencies. **Entomologia Experimentalis et Applicata**, v. 39, n. 1, p. 47–60, 1985.

KLAMT, S.; SAEZ-RODRIGUEZ, J.; GILLES, E. D. Structural and functional analysis of cellular network with CellNetAnalyzer. **BMC systems biology**, v. 2, n. 1, p. 13, 2007. ISSN 17520509. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19040715>>.

KLASSEN, W.; CURTIS, C.; DYCK, V.; HENDRICH, J.; ROBINSON, A. **Sterile Insect Technique Principles and Practice in Area-Wide Integrated Pest Management**. [S.l.]: Springer, 2005. 783 p. ISBN 9781402040504.

LEFTWICH, P. T.; KOUKIDOU, M.; REMPOULAKIS, P.; GONG, H.-f.; ZACHAROPOULOU, A.; FU, G.; CHAPMAN, T.; ECONOMOPOULOS, A.; VONTAS, J.; ALPHEY, L.; LIMITED, O.; DRIVE, I.; PARK, M.; OX, O.; ALPHEY, L. Genetic elimination of field-cage populations of Mediterranean fruit flies. **Proceedings Biological Sciences**, v. 281, n. 1792, p. 9, 2014.

LIAO, Y.-c.; HUANG, T.-w.; CHEN, F.-c.; CHARUSANTI, P.; HONG, J. S. J.; CHANG, H.-y.; TSAI, S.-f.; PALSSON, B. O.; HSIUNG, C. A. An Experimentally Validated Genome-Scale Metabolic Reconstruction of *Klebsiella pneumoniae* MGH 78578 , i YL1228 . **Journal of Bacteriology**, v. 193, n. 7, p. 1710–1717, 2011.

MAARLEVELD, T. R.; KHANDELWAL, R. A.; OLIVIER, B. G.; TEUSINK, B.; BRUGGEMAN, F. J. Basic concepts and principles of stoichiometric modeling of metabolic networks. **Biotechnology Journal**, v. 8, n. 9, p. 997–1008, 2013. ISSN 18606768.

MINAS, R. S. D.; DOLINSKI, C.; CARVALHO, S.; SOUZA, R. M. D. BIOLOGICAL CONTROL OF FRUIT FLY *Ceratitidis capitata* USING ENTOMOPATHOGENIC NEMATODES IN LABORATORY. **Scientia Agraria**, v. 12, n. 2, p. 115–119, 2011.

NESTEL, A. D.; TOLMASKY, D.; RABOSSI, A.; LUIS, A.; NESTEL, D.; TOLMASKY, D.; RABOSSI, A. Lipid , Carbohydrates and Protein Patterns During Metamorphosis of the Mediterranean Fruit Fly , *Ceratitidis capitata* (Diptera : Tephritidae) Published By : Entomological Society of America Lipid , Carbohydrates and Protein Patterns During Metamorphosis o. **Entomological Society of America**, v. 96, n. 3, p. 237–244, 2003.

OGATA, H.; GOTO, S.; SATO, K.; FUJIBUCHI, W.; BONO, H.; KANEHISA, M. KEGG: Kyoto encyclopedia of genes and genomes. **Nucleic Acids Research**, v. 27, n. 1, p. 29–34, 1999. ISSN 03051048.

OLIVEIRA, C. M.; AUAD, A. M.; MENDES, S. M.; FRIZZAS, M. R. Crop losses and the economic impact of insect pests on Brazilian agriculture. **Crop Protection**, Elsevier Ltd, v. 56, p. 50–54, 2014. ISSN 0261-2194. Disponível em: <<http://dx.doi.org/10.1016/j.cropro.2013.10.022>>.

OLIVEIRA, J. S.; BAILEY, C. G. An Algebraic-combinatorial Model for the Identification and Mapping of Biochemical Pathways. p. 1163–1196, 2001.

PAPIN, J. A.; PRICE, N. D.; WIBACK, S. J.; FELL, D. A.; PALSSON, B. O. Metabolic pathways in the post-genome era. **Trends in Biochemical Sciences**, v. 28, n. 5, p. 250–258, 2003. ISSN 09680004.

PAVLOPOULOS, G. A.; SECRIER, M.; MOSCHOPOULOS, C. N.; SOLDATOS, T. G.; KOSSIDA, S.; AERTS, J.; SCHNEIDER, R.; BAGOS, P. G. Using graph theory to analyze biological networks. **BioData Mining**, BioMed Central Ltd, v. 4, n. 1, p. 10, 2011. ISSN 1756-0381. Disponível em: <<http://www.biodatamining.org/content/4/1/10>>.

PINZÓN, A.; RODRIGUEZ-R, L. M.; BERNAL, A.; RESTREPO, S. Targeted metabolic reconstruction : a novel approach for the characterization of plant- pathogen interactions. **Briefings in Bioinformatics**, v. 12, n. 2, p. 151–162, 2010.

RICALDE, M. P.; NAVA, D. E.; LOECK, A. E.; DONATTI, M. G. Temperature-dependent development and survival of Brazilian populations of the Mediterranean fruit fly , *Ceratitis capitata* , from tropical , subtropical and temperate regions. **Journal of Insect Science**, v. 12, n. 33, p. 1–11, 2012.

SCHILLING, C. H.; LETSCHER, D.; PALSSON, B. O. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. **Journal of Theoretical Biology**, v. 203, n. 3, p. 229–248, 2000. ISSN 00225193.

SCHILLING, C. H.; SCHUSTER, S.; PALSSON, B. O. Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-genomic Era. **Biotechnol. Prog.**, v. 15, n. 3, p. 296–303, 1999. ISSN 1520-6033.

SETUBAL, J.; MEIDANIS, J. **Introduction to Computational Molecular Biology**. [S.l.]: PWS PUBLISHING COMPANY, 1997. 277 p. ISBN 0534952623.

TENAZINHA, N.; VINGA, S. A Survey on Methods for Modeling and Analyzing Integrated Biological Networks. **IEEE CS**, v. 8, n. 4, p. 943–958, 2011.

TOYA, Y.; KONO, N.; ARAKAWA, K.; TOMITA, M. Metabolic Flux Analysis and Visualization †. **Journal of Proteoma Research**, v. 10, n. 8, p. 3313–3323, 2011.

TRINH, C. T.; WLASCHIN, A.; SRIENC, F. Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Characterizing Cellular Metabolism. **Applied Microbiology**, v. 81, n. 5, p. 813–826, 2009. ISSN 15378276. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909134/pdf/nihms176852.pdf>>.

WALDBAUER, G. P.; FRIENDMAN, S. R. o. E. Self-Selection of Optimal Diets by Insects. **Annual Review of Entomology**, v. 36, n. 1, p. 43–63, 1991.

YISHAI, D.; BEN-YOSEF, M.; YUVAL, B.; JURKEVITCH, E. The host fruit amplifies mutualistic interactions between *Ceratitis capitata* larvae and associated bacteria. **bioRxiv**, v. 1, n. 1, p. 1–29, 2018.

Apêndices

APÊNDICE A – Script utilizado para rodar a ferramenta SPAdes @CACAU

No script utilizado para rodar a ferramenta SPAdes (Fig 18), é definido o nome do trabalho, o tempo de permanência na fila, a quantidade de nós e cores usadas, e a saída. Além disso, é definido o endereço do programa no CACAU. Os Elementos R1 e R2 são os dados de entrada do programa para fazer a montagem das sequências.

```
#!/bin/bash

#SBATCH -J SpdCepa2 #nomedojob
#SBATCH --partition long #fila
#SBATCH --nodes 1 #quantidade de nós
#SBATCH --ntasks 8 #quantidade de cores
#SBATCH -o log_spades_cep2.log #log de saída do slurm

source /usr/local/vars/spades-3.13.0_vars.sh

FASTQ_DIRECTORY=/home_nfs/augfranco/2019-08-03-Cepa2/data
R1=Galaxy3-AG02-Klebsiella-oxytoca-planta-KOP_S2_L001_R1_001.fastq
R2=Galaxy8-AG02-Klebsiella-oxytoca-planta-KOP_S2_L001_R2_001.fastq

cd /home_nfs/augfranco/2019-08-03-Cepa2/assemblies/spades

spades.py -o spades_cep2 --careful --pe1-1 $FASTQ_DIRECTORY/$R1 --pe1-2 $FASTQ_DIRECTORY/$R2 -t 8 -m 15
```

Figura 18 – Script para o uso da ferramenta SPAdes

Fonte: Autoria Própria

APÊNDICE B – Procurar caminhos em grafos de acordo as potências da sua matriz de adjacência

No presente apêndice apresentamos outra possível abordagem para solucionar o problema de interesse. Foi proposto trabalhar com um grafo e sua matriz de adjacência (Fig. 19). O subgrafo obtido é apresentado na figura 19A, com 11 vértices, sendo o grafo $G = (V, E)$. A matriz de adjacência, de ordem $|V| * |V|$, foi apresentada na figura 19B, onde $|V|$ é o número de vértices.

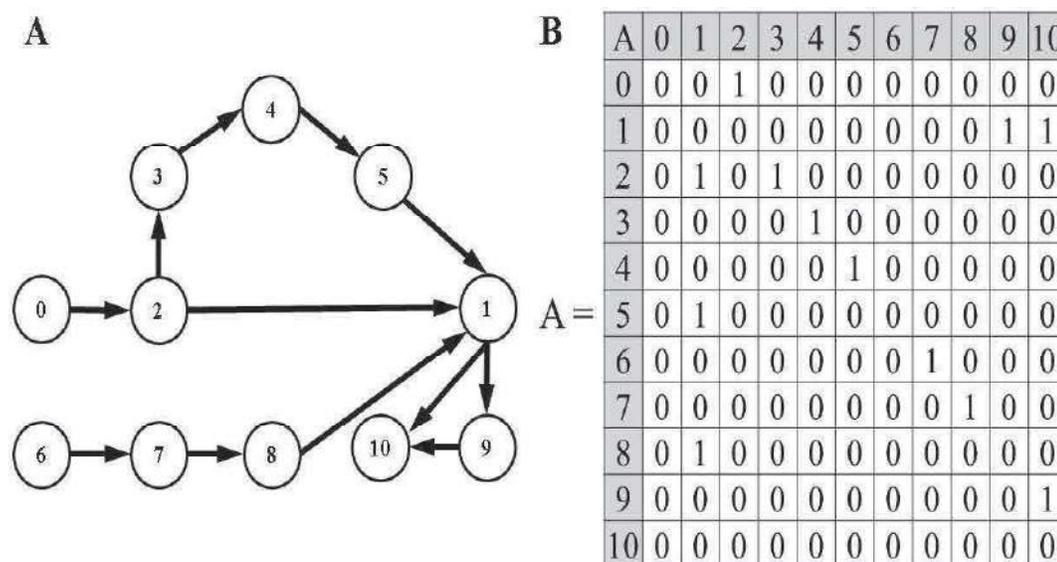


Figura 19 – Grafo de pesquisa e sua matriz de adjacência

Fonte: Autoria própria

A partir da matriz de adjacência foi proposta a utilização das potências desta matriz e a busca dos caminhos dentro do grafo. As potências foram até o número total de vértices no grafo, neste caso 11 vértices. A potência da matriz n representa o comprimento do caminho presente (Figuras 20 e 21), onde são apresentados os vértices originais do grafo e a quantidade de caminhos de n comprimento entre vértices.

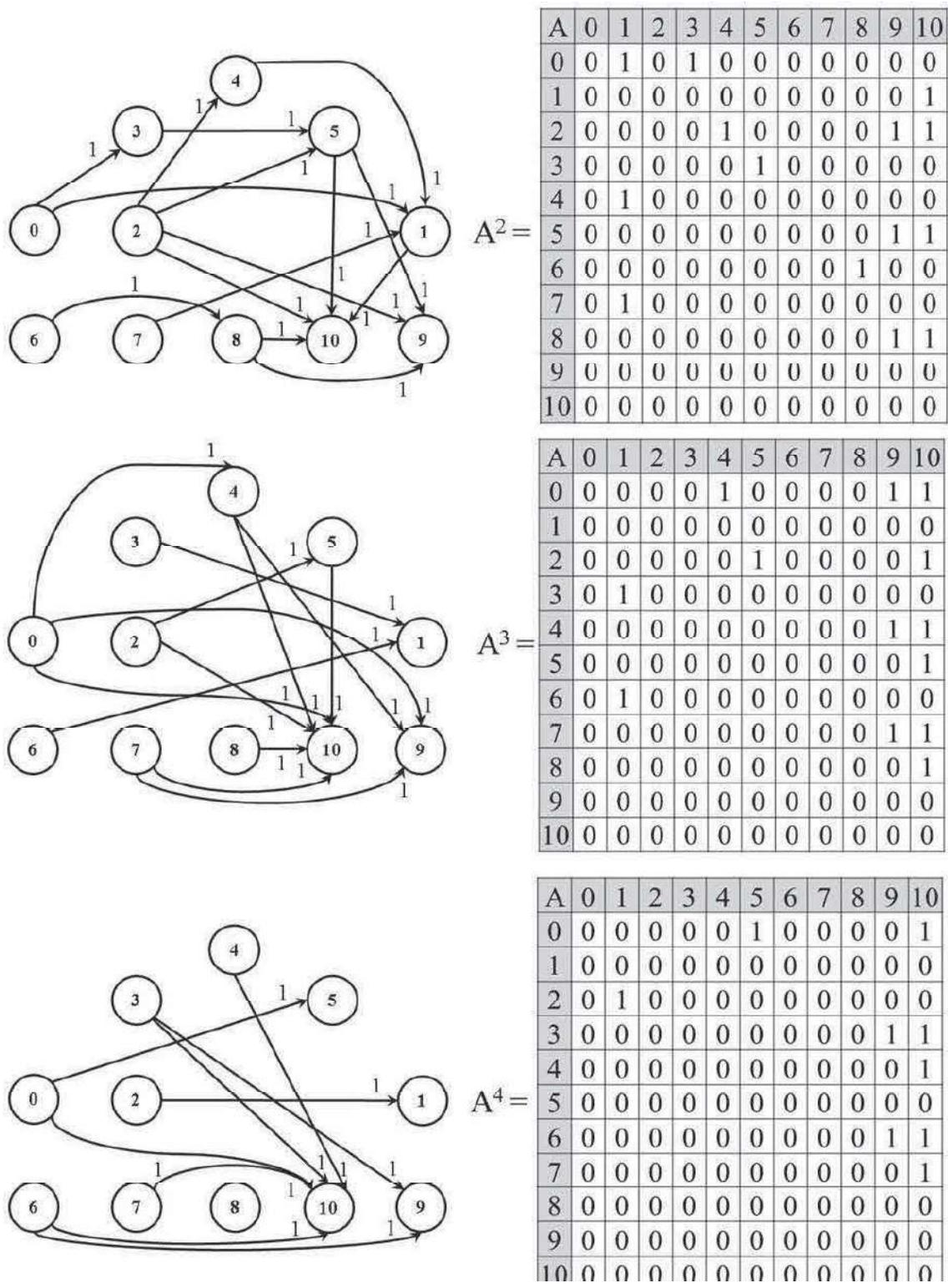


Figura 20 – Potências 2, 3 e 4 da matriz de adjacência

Fonte: Autoria própria

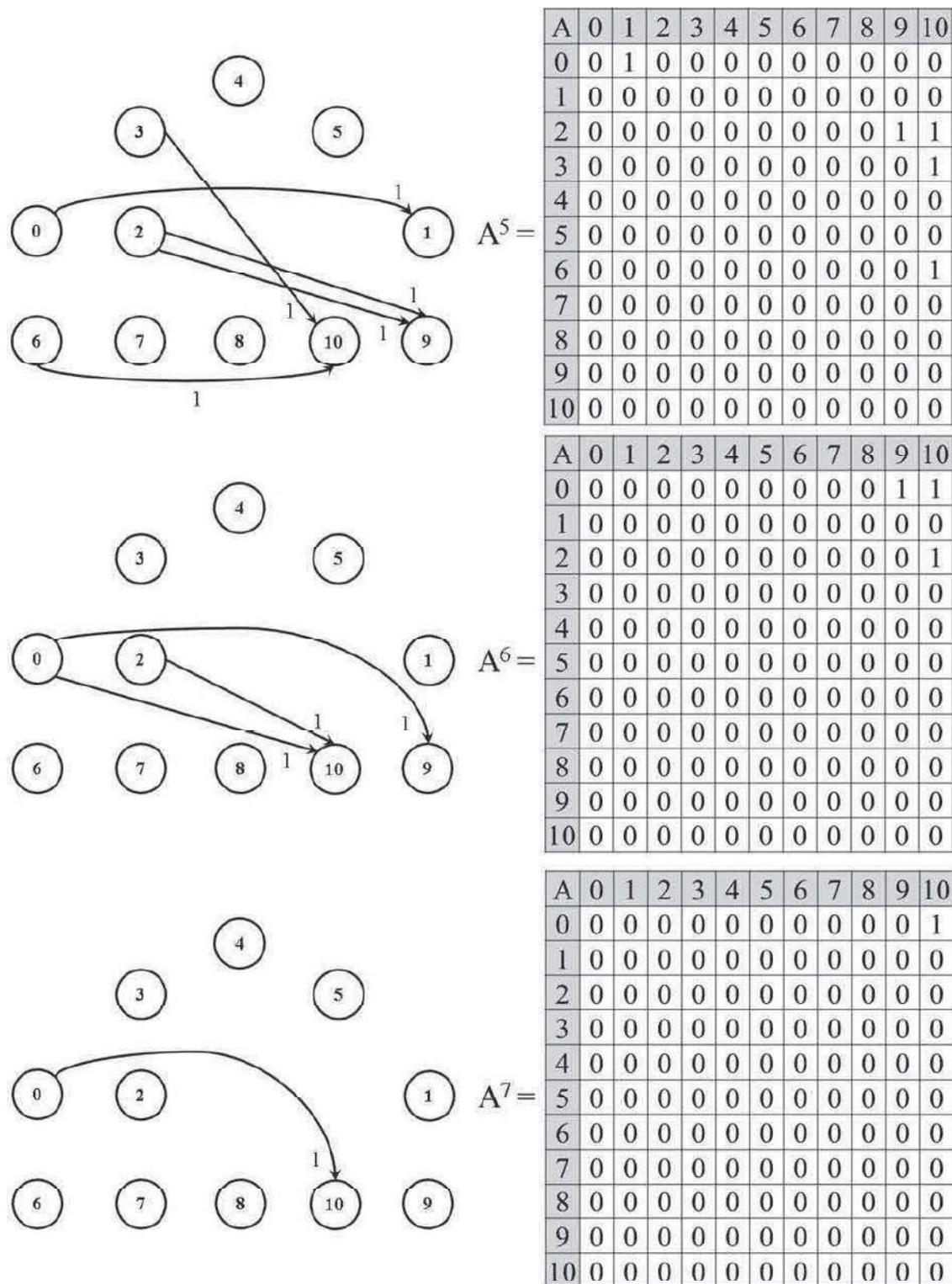


Figura 21 – Potências 5, 6 e 7 da matriz de adjacência

Fonte: Autoria própria

APÊNDICE C – Aplicação de DNA computing

Neste apêndice é apresentada a utilização da técnica de DNA computing baseada em experimentos de biologia molecular para solucionar problemas de algoritmos em grafos (SETUBAL; MEIDANIS, 1997; ADLEMAN, 1994). A estratégia é gerar os possíveis caminhos de comprimento $(n - 1)$, seja $(1 \leq i \leq n)$ onde n é o número total de vértices no grafo de interesse. Além disso, identificar qual dos possíveis caminhos são a solução do problema abordado. O procedimento principal nesta técnica é codificar caminhos e vértices do grafo em forma de sequências de DNA. Foi utilizada essa técnica para exemplificar outra forma de obter os fluxos dentro de um grafo (Fig. 22). O grafo problema contém 4 vértices e 3 arestas. O procedimento foi realizado *in silico* devido aos custos dos reagentes e a falta de equipamento necessário para executar na bancada.

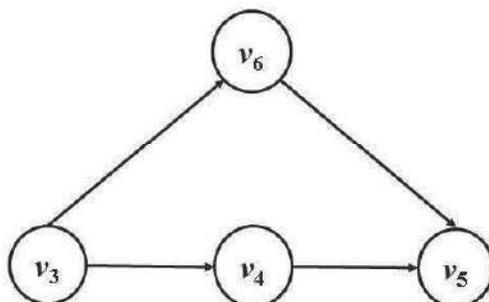


Figura 22 – Subgrafo exemplo do problema utilizado na aplicação de DNA computing

Fonte: Autoria própria

A implementação de DNA computing tem três passos importantes: 1) Codificação dos vértices, 2) Codificação das arestas e 3) Construção dos caminhos. Na codificação dos vértices as sequências de DNA foram escolhidas arbitrariamente, de comprimento 8 pb, para cada vértice v_i do grafo, onde i é o número de vértice. Para cada aresta $v_i v_j$ no grafo a codificação foi realizada de acordo a sequência formada pelo sufixo de 4 bp da sequência de v_i concatenada com o prefixo de 4 bp da sequência de v_j . Além disso, sequências complementares \bar{v}_i aos vértices foram utilizadas, de acordo ao conceito de complementariedade do dogma central da biologia molecular. A função destas sequências é gerar uma ligação entre duas arestas que contenham o trecho de sequência complementar. Por último a codificação dos caminhos é gerada pela união química da concatenação das arestas com a sequência complementar \bar{v}_i .

A implementação de DNA computing representa outra maneira de obter os caminhos no grafo. Nesta seção foi apresentado o processo de codificação dos vértices, arestas e a construção dos caminhos. A Figura 23 mostra os 4 vértices envolvidos no grafo com a codificação arbitraria de sequencias de DNA assinadas para cada vértice. Na Fig 23B são apresentadas as arestas codificadas pelo sufixo de 4 pb da sequência v_i concatenada com o prefixo de 4 pb de sequência v_j . Além disso, as sequências complementares aos vértices (Figura 23C) tem a função de realizar a ligação entre as arestas que onde tenham sequências complementares. As sequências complementares tem comprimento de 8 pb e nesta exemplificação foram usadas 2 sequências complementares. Todos os elementos anteriormente codificados foram utilizados para codificar dos caminhos presentes desde o vértice 3 até vértice 5 na Figura 23D, passando pelos vértices v_4 e analogamente pelo vértice v_6 .

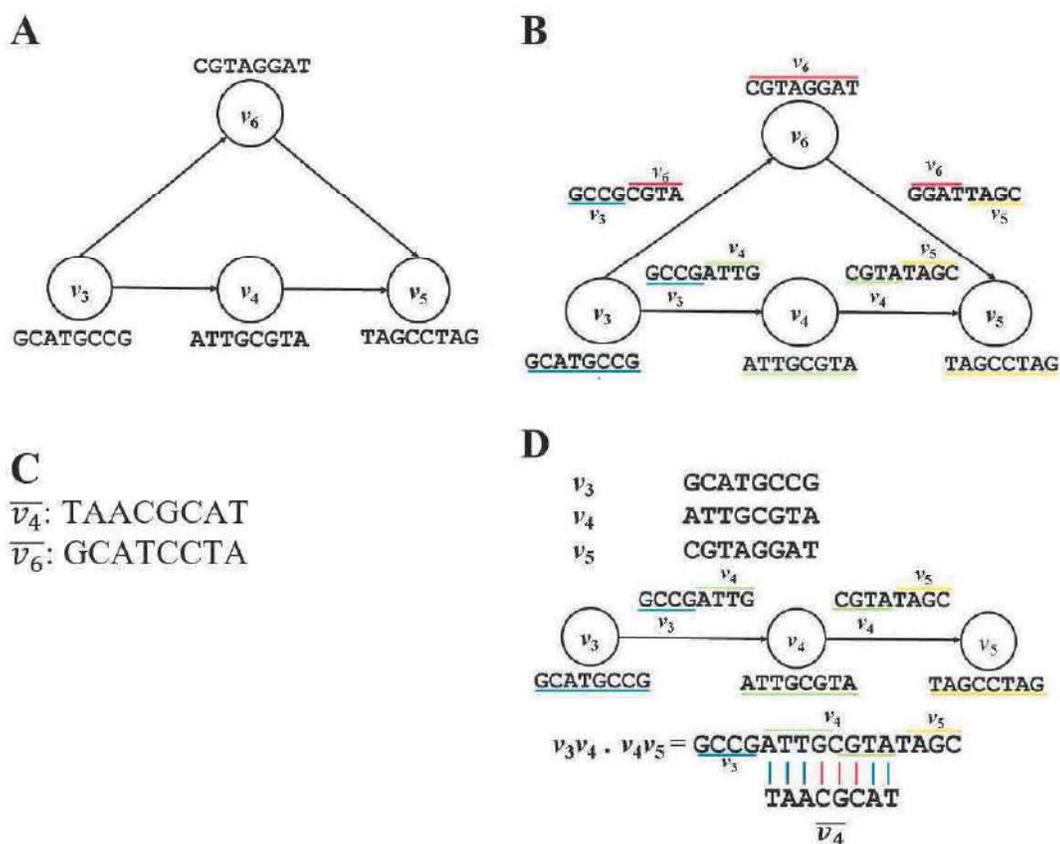


Figura 23 – Codificação do grafo problema de acordo a técnica de DNA computing

C.1 Uso de DNA computing

O primeiro passo para realizar a aplicação de DNA computing é construir os caminhos de interesse presentes no grafo problema. Para isso, foi necessário realizar a reação de ligação (Fig 24), onde tem presença das arestas e sequências complementares em forma de sequências de DNA de comprimento 8 pb. A reação de ligação tem como resultado a formação de sequências de DNA codificando diferentes caminhos dentro do grafo problema.

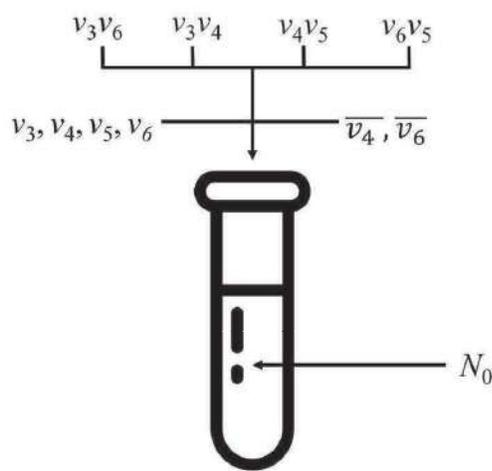


Figura 24 – Reação de ligação

Fonte: Autoria própria

O resultado da mistura dos reagentes gera múltiplas combinações no tubo de ensaio. Os resultados são múltiplas cópias de sequências de arestas e múltiplas cópias de sequências complementares aos vértices. No tubo de ensaio ocorre a ligação, mas algumas delas não representam caminhos no subgrafo. Na Figura 25 são apresentadas múltiplas combinações entre arestas que podem ou não estar presentes no grafo original. As combinações marcadas com vermelho são as combinações de interesse que resolvem o problema estudado.

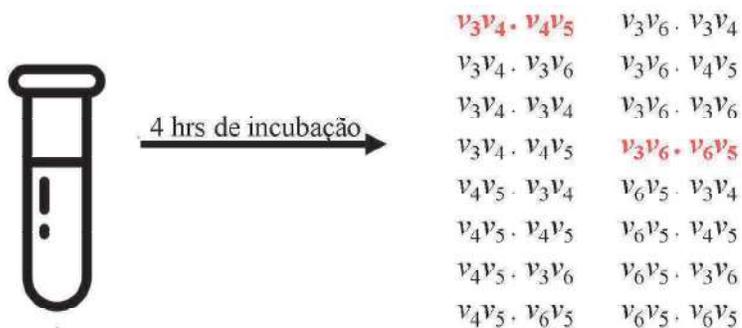


Figura 25 – Produto de reação de ligação

Fonte: Autoria própria

O produto da reação de ligação é o principal reagente para realizar o PCR *in silico*, porque contém as cópias dos possíveis caminhos obtidos. A utilização dos primers v_3 e sequência complementar $\overline{v_5}$ são necessários para procurar só os caminhos que iniciam no v_3 e o destino é a sequência complementar $\overline{v_5}$. O produto desta amplificação identifica os caminhos de comprimento 16 pb correspondentes às duas opções de caminhos dentro do grafo como é apresentado na Figura 26, que mostra a eletroforeses dos produtos de PCR.

- Caminho 1: $v_3 \rightarrow v_4 \rightarrow v_5$
- Caminho 2: $v_3 \rightarrow v_6 \rightarrow v_5$
- Amostra controle (\emptyset): ausência de oligonucleotídeos.

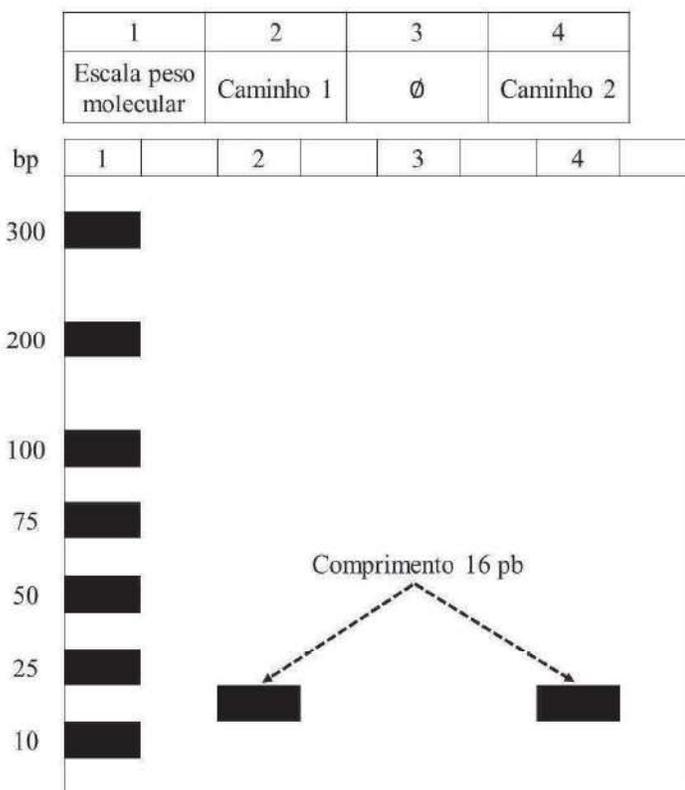


Figura 26 – Electroforeses do PCR *in silico*

Fonte: Autoria própria