

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE BIOTECNOLOGIA
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

EMPREGO DE FERRAMENTAS BIOINFORMÁTICAS NO ESTUDO DE VIAS METABÓLICAS E DE FUNÇÕES ECOLÓGICAS DA *ACETILACTOBACILLUS JINSHANENSIS* E DA *BACILLUS CEREUS*, ENCONTRADAS NO ALIMENTO LARVAL DE INTERAÇÕES PROTEÍNA-PROTEÍNA.

NATANAEL BORGES DE AVILA

Monografia apresentada à
Coordenação do Curso de
Biotecnologia, da
Universidade Federal de
Uberlândia, para obtenção do
grau de Bacharel em
Biotecnologia.

UBERLÂNDIA – MG
ABRIL – 2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE BIOTECNOLOGIA
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

EMPREGO DE FERRAMENTAS BIOINFORMÁTICAS E DA ANÁLISE DO
INTERACTOMA NO ESTUDO DE VIAS METABÓLICAS E DE FUNÇÕES
ECOLÓGICAS DA *ACETILACTOBACILLUS JINSHANENSIS* E DA *BACILLUS CEREUS*,
ENCONTRADAS NO ALIMENTO LARVAL DE ABELHA SEM FERRÃO.

NATANAEL BORGES DE AVILA

Autor

ANDERSON RODRIGUES DOS SANTOS

Orientador

Monografia apresentada à
Coordenação do Curso de
Biotecnologia, da
Universidade Federal de
Uberlândia, para obtenção do
grau de Bacharel em
Biotecnologia.

Uberlândia – MG

ABRIL – 2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE BIOTECNOLOGIA
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

EMPREGO DE FERRAMENTAS BIOINFORMÁTICAS E DA ANÁLISE DO INTERACTOMA NO ESTUDO DE VIAS METABÓLICAS E DE FUNÇÕES ECOLÓGICAS DA *ACETILACTOBACILLUS JINSHANENSIS* E DA *BACILLUS CEREUS*, ENCONTRADAS NO ALIMENTO LARVAL DE ABELHA SEM FERRÃO.

NATANAEL BORGES DE AVILA

Anderson Rodrigues dos Santos

FACOM

Homologado pela coordenação do Curso de
Biotecnologia em __/__/__

Coordenador (a) do Curso

Uberlândia – MG

ABRIL – 2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE BIOTECNOLOGIA
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

EMPREGO DE FERRAMENTAS BIOINFORMÁTICAS E DA ANÁLISE DO INTERACTOMA NO ESTUDO DE VIAS METABÓLICAS E DE FUNÇÕES ECOLÓGICAS DA *ACETILACTOBACILLUS JINSHANENSIS* E DA *BACILLUS CEREUS*, ENCONTRADAS NO ALIMENTO LARVAL DE ABELHA SEM FERRÃO.

NATANAEL BORGES DE AVILA

Aprovado pela Banca Examinadora em: / / Nota: ____

Nome e assinatura do Presidente da Banca Examinadora

Uberlândia, de de

DEDICATÓRIA

A Deus, Criador do Universo e suas leis, cuja existência se prova na complexidade da sua criação, que me permitiu estudar, neste trabalho, pequena parte dessa complexidade.

AGRADECIMENTOS

Agradeço a Deus por me permitir concluir essa etapa acadêmica e por ter me salvado no meu pior momento.

Agradeço à minha mãe, Teresa Batista Correa de Avila, a meu pai, Francisco de Avila e a meu irmão, Daniel Borges de Avila por todo o apoio durante essa jornada. Não teria chegado aqui sem vocês.

Agradeço aos amigos, familiares e professores por todo suporte e aprendizado.

Agradeço ao orientador deste projeto, Anderson Rodrigues dos Santos, pela paciência em ensinar e confiança depositada.

Agradeço à Universidade Federal de Uberlândia, um bem público, que me permitiu estudar e completar essa importante fase da minha vida.

Agradeço a todos que de alguma maneira contribuiu para esse importante momento da minha vida. Deixo de mencionar nominalmente cada pessoa pelo risco de deixar escapar algum nome.

RESUMO

Ferramentas de bioinformática são capazes de investigar vias metabólicas e funções ecológicas de espécies a fim de estabelecer seus papéis no meio em que são encontradas. O estudo do interactoma compreende obter e estudar as redes de interação proteína-proteína (PPI) de uma espécie. O alimento larval de abelhas sem ferrão (AL) é um importante meio a ser estudado com rica microbiota, possuindo significativo potencial biotecnológico. As bactérias *Acetilactobacillus jinshanensis* e *Bacillus cereus* foram encontradas no AL (a primeira por metagenômica, a segunda por colônia isolada) e são também conhecidas por sua presença na fermentação do vinagre aromático de Zhenjiang. Compreender os processos biológicos dessas espécies com o uso de ferramentas de bioinformática é um relevante passo para o avanço dos estudos do AL e suas aplicações. Este trabalho empregou o GenPPI, um preditor de PPI in silico, para construir redes a partir dos genomas completos das duas espécies, montados bioinformaticamente. Buscou-se compreender os processos bioenergéticos e vias metabólicas relevantes, especialmente em *A. jinshanensis*, utilizando BLAST para identificar genes chave e a rede PPI para inferir a ocorrência e dinâmica dos processos. Os resultados mostraram que a *A. jinshanensis* não possui enzimas para fermentação acética e alcoólica, mas possui enzimas da fermentação heterolática e parte das enzimas da via Embden-Meyerhof (EMP). A *B. cereus*, por outro lado, possui genes das enzimas de fermentação alcoólica e da via EMP, mas não possui das fermentações acética e heterolática, nem da via Entner-Doudoroff (ED). O interactoma estudado comprovou a dupla via de oxidação da glicose na *A. jinshanensis*. Além disso, o interactoma montado das espécies e o rastreo de proteínas pontes são promissores para trabalhos futuros. Estes resultados corroboram estudos anteriores, sugerindo papéis distintos no vinagre chinês: *A. jinshanensis* na fermentação heterolática e *B. cereus* na alcoólica, o que pode contribuir para a otimização desse processo. O estudo das vias metabólicas da *A. jinshanensis* abre caminhos para a compreensão de seu papel no AL e aprofunda os estudos para aplicações industriais. Validou-se a metodologia baseada no GenPPI como ferramenta eficaz para estudos

de genômica funcional. Estes achados estabelecem uma base sólida para futuros estudos focados em vias metabólicas com relevância biotecnológica nas duas espécies estudadas.

Palavras-chave: Rede de interação proteína-proteína; GenPPI; alimento larval; vinagre de Zhenjiang; interactoma.

SUMÁRIO

1. INTRODUÇÃO	1
1.1. Rede de interação proteína-proteína	1
1.2. Estudo do interactoma	4
1.3. Métodos de determinação do interactoma	5
1.4. Métodos computacionais de determinação do interactoma	6
1.4.1. Modelos baseados em redes	6
1.4.2. Modelos baseados em dados biológicos: sequência e estrutura	7
1.4.3. Modelos baseados no genoma	8
1.4.4. Modelos que utilizam machine learning	9
1.5. GenPPI: preditor PPI baseado em dados do genoma	11
1.6. <i>Acetilactobacillus jinshanensis</i>	11
1.7. <i>Bacillus cereus</i>	12
1.8. <i>Acetilactobacillus jinshanensis</i> e a <i>Bacillus cereus</i> no vinagre de Zhenijang	12
1.9. Produtos biotecnológicos e prospecção	13
2. OBJETIVOS	14
2.1. Objetivo Geral	14
3. METODOLOGIA	15
3.1. Montagem do genoma	16
3.2. Montagem da rede de interação proteína-proteína	18
3.3. Estudo da topologia da rede	21
3.4. Análise individual dos nós	22
3.5. Estudo das interações	23
3.5.1. Escolha das sementes	24
3.5.2. Validação da presença das sementes	28
3.6. Análise local subcelular	29
3.7. Identificação das principais vias metabólicas do genoma	30
3.8. Busca dos genes na <i>Bacillus Cereus</i> CP174498	31
4. RESULTADOS E DISCUSSÃO	31

4.1. Proteínas das vias estudadas	33
4.1.1. BLAST como ferramenta de filtragem e direcionamento.....	39
4.1.2. Estudo direcionado pelas funções biológicas conhecidas.....	40
4.2. Filtragem por BriCe.....	42
4.2.1. Uma segunda filtragem: coeficientes de centralidade	46
4.3. Estudo das interações individuais.....	48
4.3.1. Interactoma local confere respaldo a análises biológicas	49
4.4. Estudo das vias metabólicas principais	50
4.4.1. Análise de vias e as funções bioecológicas no alimento larval de abelha	52
4.5. Genes de proteínas relevantes presentes na no genoma da <i>Bacillus cereus</i> CP174498	52
. 4.5. Proposta de metodologia	59
6.CONCLUSÕES.....	60
7.REFERÊNCIAS	61

LISTA DE FIGURAS

Figura 1-Interactoma de consórcio de bactérias redutoras de metano, contendo 1887 nós e 232217 arestas. Cortesia do grupo de pesquisa da Comp2Bio (FACOM/IBTEC).....	3
Figura 2-Interactoma de proteína de interesse (azul), nós e algumas arestas. Cortesia do grupo de pesquisa da Comp2Bio (FACOM/IBTEC).....	3
Figura 3 -funções do RagTag. O algoritmo pode corrigir trechos faltantes com base na referência; criar gaps com base na comparação com a referência (empilhar); remendar trechos para que se enquadre em um genoma buscado (alvo/pesquisa); ou combinar trechos. Adaptado de (Alonge et al., 2019).	17
Figura 4- via da fosfocetolase. (1) hexoquinase, (2) glicose-6-fosfato desidrogenase, (3) 6-fosfogluconolactonase, (4) 6-fosfogluconato desidrogenase, (5) ribose-5-fosfato epimerase, (6) pentose fosfocetolase, (7) lactato desidrogenase, (8) fosfotransacetilase, (9) aldeído desidrogenase, (10) álcool desidrogenase, (11) acetil fosfato quinase. Adaptado de Spector (2009).	26
Figura 5-via Embden-Meyehof-Parnas (EMP). Enzimas da via são (1) hexoquinase, (2) fofoglucoisomerase, (3) fosfofrutoquinase, (4) frutose bifosfato aldolase, (5) triose fosfato isomerase,	

(6) gliceraldeído-3-fosfato desidrogenase, (7) fosfogliceroquinase, (8) fosfoglicerato mutase, (9) enolase e (10) piruvato quinase. Adaptado de (Spector, 2009).	27
Figura 6- via Entener-Doudoroff (ED). Enzimas da via são (1) hexoquinase, (2) glicose-6-fosfato desidrogenase, (3) 6-fosfogluconoolactonase, (4) 6-fosfogluconato desidratase e (5) 2-ceto-3-deoxi-6-fosfogluconato aldolase. Adaptado de (Spector, 2009)	27
Figura 7-metodologia utilizada nesse trabalho	29
Figura 8- observação da rede da <i>Acetilactobacillus jinshanensis</i> CP171343 no Gephi com nós indicando proteínas e arestas indicando interações. Nós destacados com tamanhos maiores indicando nós com maiores valores de Bridge Centrality (BriCe) e cores indicando locais subcelular: citoplasma, membrana, potencialmente exposta na superfície e secretada. Fonte: elaboração própria.	32
Figura 9-observação da rede de interação da <i>Bacillus cereus</i> CP174498, com nós indicando proteínas e arestas indicando interações. Nós destacados com tamanhos maiores indicando nós com maiores valores de Bridge Centrality (BriCe) e cores indicando locais subcelular: citoplasma, membrana, potencialmente exposta na superfície e secretada. Fonte: elaboração própria.....	33
Figura 10- via da fosfocetolase. Enzima pentose fosfocetolase (6) forma gliceraldeído-3-fosfato que entra na via EMP. Adaptado de (Spector, 2009)	36
Figura 11 -reação feita pela lactato desidrogenase (7), convertendo piruvato -um produto da via EMP- em lactato.	46
Figura 12- rede de interações observadas no Gephi com proteínas estudadas. Os nós coloridos indicam as proteínas e suas vias, enquanto as arestas indicam as interações. Nós com tarjas repetidas indicam que o BLAST encontrou mais de um trecho da sequência como resultado. Fonte: elaboração própria.	48
Figura 13-Gráfico demonstrando as ocorrências de funções das 30 principais vias do genoma. Fonte: elaboração própria.....	51

Tabela 7- Enzimas de fermentação alcoólica e acética e sua disponibilidade no genoma da B.cereus CP174498.....	53
Figura 14-etapas da fermentação do vinagre chinês e seus principais pontos. Adaptado de (Liu et al., 2023; Xu et al., 2024b, 2011b).	58
Figura 15-proposta de metodologia para trabalhos futuros	59

LISTA DE TABELAS

Tabela 1- espécies utilizadas para montagem de rede de interação	19
Tabela 2-Enzimas das vias de fermentação alcoólica e acética.....	34
Tabela 3- Enzimas da via da fermentação heterolática (via da fosfocetolase).....	34
Tabela 4-enzimas da via EMP.	36
Tabela 5-Enzimas da via ED	37
Tabela 6-proteínas com maior Bridging Centrality	42
Tabela 7- Enzimas de fermentação alcoólica e acética e sua disponibilidade no genoma da B.cereus CP174498	53
Tabela 8-via da fosfocetolase e sua presença na B. cereus CP174498.....	53
Tabela 9- enzimas da via EMP e sua disponibilidade no genoma da B. cereus CP174498.	54
Tabela 10- Enzimas da via ED e sua disponibilidade no genoma da B. cereus CP174498	55

1. INTRODUÇÃO

1.1. Rede de interação proteína-proteína

Um sistema biológico se baseia num emaranhado de redes metabólicas as quais possuem um conjunto também vasto de proteínas com funções diversas. As conexões do sistema vivo demanda um conjunto de interações entre vias, entre suas proteínas constituintes e entre o conjunto de sistemas de um mesmo ser vivo. A elucidação dessas interações é o objeto de estudo de diversos grupos de pesquisas ao longo da história e que empregaram majoritariamente métodos experimentais para obter conclusões razoáveis. O propósito inclui compreender os processos metabólicos do organismo, obter informações relevantes sobre doenças, assim como pensar em mecanismos de supressão de possíveis patógenos ou reversão de anomalias biológicas.

No nível molecular, uma interação proteína-proteína (PPI-protein protein interaction) diz respeito a relação que duas proteínas podem estabelecer, sendo essas interações de aspectos físico-químicos ou biológicos (Wang *et al.*, 2022). Estima-se que 80% das proteínas realizam suas funções em complexos (Berggård, Linse e James, 2007). Essas interações podem ser classificadas em diversos tipos (Nooren e Thornton, 2003): protômeros de um complexo oligomérico, em que cada subunidade realiza interações físico-químicas se mantendo unidos (repressor ACR e proteassoma). Esses complexos podem ser obrigatórios ou não obrigatórios, isto é, seus protômeros podem existir independente do complexo (não obrigatórios) ou inexistir *in vivo* como estruturas singulares (obrigatórios). Complexos sinalizadores intracelulares e ligação anticorpo-antígeno são exemplos de interações não obrigatórias, enquanto o repressor ACR, exemplos de interações obrigatórias. Além disso, as interações podem ser permanentes ou transientes, o que significa que permanecem unidas durante toda vida em conexões estáveis ou se associam e dissociam a depender do contexto. Comumente, complexos obrigatórios funcionais e estruturais são permanentes, enquanto interações não obrigatórias podem ser

ambas. A proteína heterotrimérica G é um exemplo de interação permanente, enquanto a lisina do espermatozoide é um exemplo de interação transitória. Essas interações podem alternar a depender do contexto fisiológico. Sendo assim, interações transitórias podem se tornar permanentes conforme marcadores do ambiente ou sinalizadores biológicos.

Compreender o interactoma, nomenclatura dada ao conjunto de interações proteínas-proteínas de um organismo, é objeto de estudo de diversos grupos de pesquisa com fim a obtenção de informações sobre processos biológicos, vias metabólicas e dados evolutivos. O interactoma é representado graficamente por grafos, que são observados em softwares específicos, tais como Cytoscape (Shannon *et al.*, 2003) e Gephi (Bastian, Heymann e Jacomy, 2009), o que permite observar a rede de interações formadas. Na prática, o estudo das interações se dá pelo estudo da topologia dos grafos. Um grafo possui nós, os quais representam as proteínas daquele organismo, e arestas, que representam as interações formadas entre as proteínas. A topologia pode ser estudada pelo todo, em uma observação ge-ral das interações, ou localmente, em que o interactoma de uma proteína é explorado. As proteínas locais escolhidas para terem seu interactoma estudado são nomeadas de sementes (seeds).

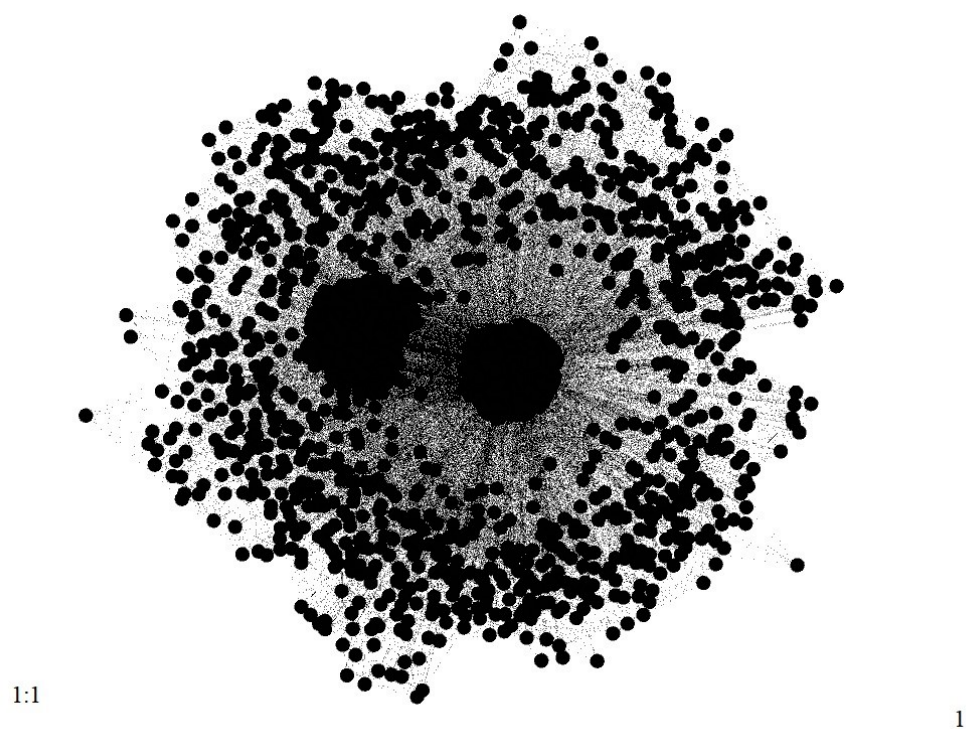


Figura 1-Interactoma de consórcio de bactérias redutoras de metano, contendo 1887 nós e 232217 arestas. Cortesia do grupo de pesquisa da Comp2Bio (FACOM/IBTEC).

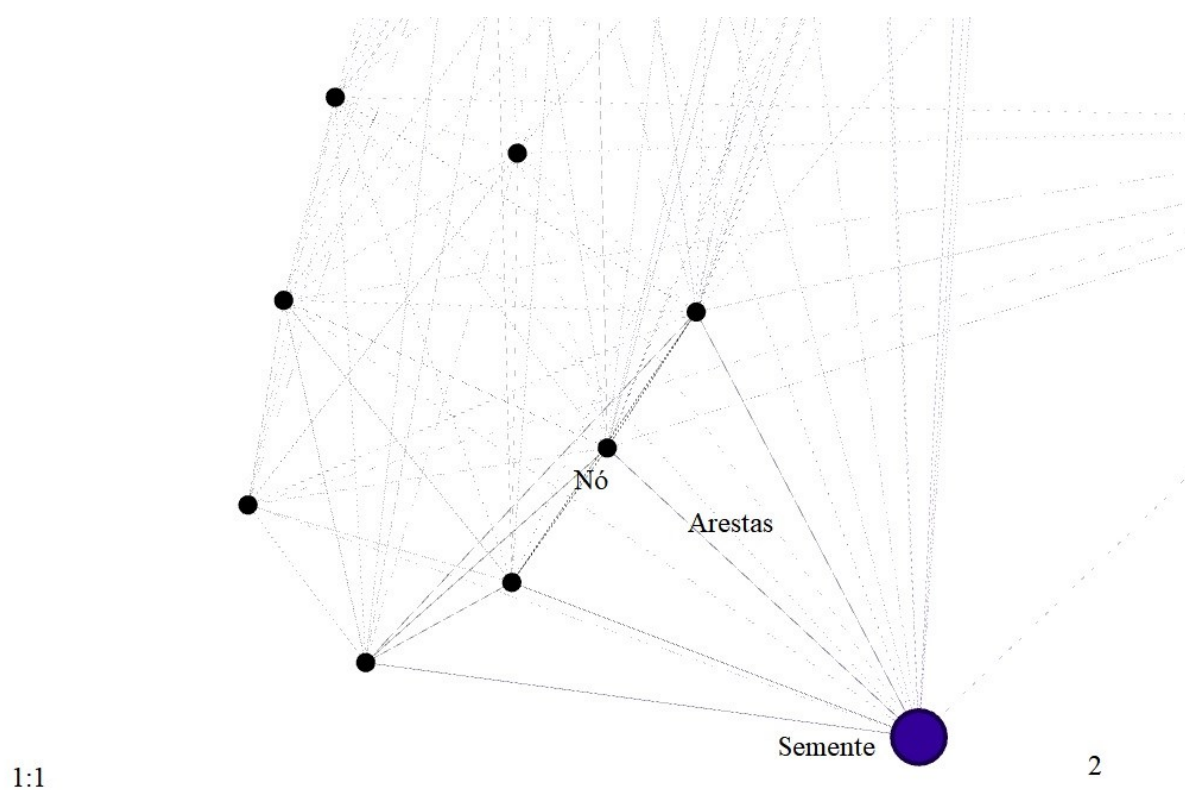


Figura 2-Interactoma de proteína de interesse (azul), nós e algumas arestas. Cortesia do grupo de pesquisa da Comp2Bio (FACOM/IBTEC).

1.2. Estudo do interactoma

O estudo do interactoma compreende estudar a topologia formada pelas interações de proteínas, obter padrões específicos dessas interações e elaborar hipóteses que se adequem ao contexto biológico. Esse conjunto de procedimentos sumarizam a teoria dos grafos, a qual é vastamente empregada no estudo do interactoma (Pavlopoulos *et al.*, 2011). Aspectos dessa teoria fundamentam ferramentas úteis que conferem maior número de abordagens possíveis para o pesquisador quando se propor a elaborar hipóteses. Já que o estudo do interactoma é extenso, repleto de redundâncias e de interações que demandam extensa curadoria, o estudo é sempre direcionado com essas ferramentas. Aqui, menciona-se medidas de centralidade, densidade do grafo e análise de clusters como sendo úteis ferramentas apropriadas na análise (Pavlopoulos *et al.*, 2011).

Transportar os dados de grafos para uma análise biológica é um desafio que coloca as pesquisas de PPI na vanguarda de uma importante ciência de base. Um complexo proteico formado por oligômeros de proteínas é um exemplo claro e notório de PPI estabelecido (Jones e Thornton, 1996). Análises que se focam em apontar os complexos são mais triviais e metodologicamente simples. No entanto, diversos outros processos biológicos podem ser mencionados como sendo operacionalizados por PPI, tais como interação antígeno-anticorpo, enzima-substrato, inibição de vias, adesão celular e transporte celular (Jones e Thornton, 1996; Nooren e Thornton, 2003). Sendo assim, elucidar essas complexas vias é um processo com múltiplas possibilidades de interações proteicas, que vão desde as proteínas participantes diretas do processo analisado até mediadores e reguladores improváveis.

Avanços importantes já foram obtidos com o estudo de PPI. Tomkins e Manzoni (2021) destacaram que o estudo do interactoma vem sendo aplicado para desvendar a doença do Parkinson com uso em revelar novas funções celulares, estudos genômicos, estudos focados em

genes específicos e estudos comparados com outras doenças neurológicas. Ilgova et al. (2022) estudaram sítios de ligação entre HIV-1 integrase e proteína celular Ku70, ligação necessária para replicação, e prospectaram ligantes para servirem como drogas contra a infecção. Kotelnikova et al. (2021) empregaram o estudo de redes de PPI para decifrar fibrose cardíaca, tendo encontrado relações mais importantes entre 54 proteínas da cascata de fibrose cardíaca. Tendiuk et al. (2022) elucidaram o papel da proteína Svx na virulência de bactérias *Pectobacterium* utilizando análises de interações. Interações entre o Sars-Cov-2 e as proteínas humanas foram estudadas em modelos *in silico* e validadas em 75% das interações encontradas em experimentos laboratoriais (Bell et al., 2022). Shan et al. (2022) utilizou modelo de interação de proteínas e identificou que o anticorpo humano P36-5D2 se liga eficazmente a várias variantes de Sars-Cov-2, demonstrando a aplicação relevante do estudo da interação de proteínas para o desenvolvimento de drogas específicas. Crisman et al. (2023) descreveu como o estudo de interação de proteína vem sendo empregado para encontrar composto que se direcionem a inibição da interação KEAP1-NRF2, a qual possui papel biológico relevante relacionados com resposta antioxidante e citoproteção.

1.3. Métodos de determinação do interactoma

A construção de uma rede de interação pode ser feita por métodos experimentais ou computacionais. Os métodos experimentais são os mais utilizados no estudo do interactoma. São eles, mas não somente, ensaio de levedura híbrida 2 (Fields e Sternglanz, 1994), purificação por afinidade em tandem (Bürckstümmer et al., 2006), chip de proteína (Tong et al., 2002), análise de letalidade sintética (Tong et al., 2001), correção perfil de expressão de mRNA (Ge et al., 2001). Algumas metodologias como coimunoprecipitação e transferência de energia por ressonância de bioluminescência (no inglês BRET) são apontadas como possíveis de serem também empregadas na formação de PPI (Durham et al., 2023). Uma vez que as interações resultam em complexos proteicos oligoméricos, métodos como microscopia crioelétrica,

cristalografia em raio X e ressonância magnética nuclear são utilizadas para observar interações resultadas em complexos proteicos (Durham *et al.*, 2023). Esses métodos experimentais sofrem com o alto custo e a quantidade extensa de força de trabalho demandada (Hu *et al.*, 2021), o que pesa negativamente na sua escolha.

1.4. Métodos computacionais de determinação do interactoma

Realizar a predição de PPI utilizando modelos computacionais como preditores é uma tarefa impositiva e que angaria boa parte do esforço recente da pesquisa na área. As interações, conforme já mencionado possuem padrões, uma vez que são frutos de processos bioquímicos cujas leis atuantes são interações de ordem físico-químicas. Nesse sentido, modelos computacionais de predição de redes podem buscar por padrões de interações já descritos e extrapolar para novos proteomas. Esses padrões podem ser biológicos ou de rede.

1.4.1. Modelos baseados em redes

Os modelos baseados em redes buscam os padrões de conectividade de outras redes depositadas em bancos de dados. Há diversos bancos de dados que recebem, classificam e depositam PPI. BIND (Bader, Betel e Hogue, 2003), DIP (Xenarios *et al.*, 2002), MINT (Licata *et al.*, 2012), BioGRID (Chatr-Aryamontri *et al.*, 2017), HPRD (Keshava Prasad *et al.*, 2009), IntAct (Orchard *et al.*, 2014) e STRING (Szklarczyk *et al.*, 2023) são exemplos relevantes de redes PPI armazenadas. Alguns bancos de dados, como o STRING, IntAct e MINT dispõem de scores para descrever o índice de confiabilidade daquela interação ou rede descrita. Uma vez existindo esses bancos de dados, os modelos computacionais baseados em redes preveem um comportamento de interações com base no que já foi descrito, isto é, no padrão de conectividade. Nesse sentido, Zeng (2016) criou um modelo em que um índice estima a probabilidade de ligações entre dois nós com base na informação de ligação dos nós vizinhos imediatos, (Kovács *et al.*, 2019) elaborou modelo em que se utiliza da observação evolutiva de que proteínas não

interagem com base na similaridade entre os nós, mas sim conforme a similaridade que uma proteína possui com o parceiro de outra proteína e Symeonidis et al. (2013) propôs utilização de um método de clustering na rede de interação para capturar elementos topológicos não observáveis em uma análise global.

1.4.2. Modelos baseados em dados biológicos: sequência e estrutura

Os modelos baseados em dados biológicos se utilizam também da noção de comportamento observável para extrapolar a modelos futuros. Aqui são utilizadas informações como a sequência proteica, conformação da proteína, informações sobre o genoma e ontologia gênica. A estrutura primária das proteínas são aminoácidos, contendo cada qual uma carga elétrica, hidrofobicidade, polaridade, composição e posição na cadeia. Essas informações são úteis para traçar interações entre distintos aminoácidos e, por consequência, entre proteínas. Da mesma forma, as estruturas secundárias, terciárias e quaternárias podem servir de parâmetro condutor para estabelecer possíveis interações. Informações acuradas sobre sequências e estruturas podem ser encontradas em bancos de dados de proteínas como UniProt (Bateman et al., 2023), SWISS-PROT (Boeckmann et al., 2003) e PDB (Velankar et al., 2021).

Os modelos baseados em sequência, utilizam de classificadores como support vector machine (Cortes, Vapnik e Saitta, 1995) e random forest (Breiman, 2001), os quais estabelecem valores que descrevem a similaridade entre as proteínas com base nas características (features) utilizadas por cada modelo (polaridade, hidrofobicidade, tensão superficial). Uma operação de concatenação é necessária para que os resultados dos classificadores sejam convertidos em valores comparáveis, isto, em valores que sejam passíveis de quantificar a similaridade. Nesse sentido, (Bock e Gough, 2001) empregaram o SVM para reconhecer interações de proteínas baseando na sequência primária e nas propriedades físico-químicas.

A estrutura de proteínas contém úteis informações para a compreensão do organismo, assim como para definir interações. Desse modo, modelos baseados em estruturas são desenvolvidos e utilizam os bancos de dados existentes, assim como os modeladores de proteínas *in silico*. De modo similar ao que ocorre com os modelos de sequência, os modelos de estrutura criam valores que descrevem a similaridade de proteínas com base em informações da estrutura tridimensional. Utilizando estatística Bayseiana e escores empíricos da estrutura e sequência de proteínas, Zhang et al. (2012) desenvolveram o PrePPI, modelo capaz de prever interações. De modo similar, o InterPred, modelo que utiliza as características de estrutura combinadas com massiva comparação de estrutura, docagem molecular e random forest como classificador, foi eficiente em criar redes de interação (Mirabello e Wallner, 2017). Dados sobre domínios e loops locais também se provaram como eficazes em determinar PPIs (Planas-Iglesias et al., 2013).

1.4.3. Modelos baseados no genoma

Informações do genoma são utilizadas para construção de modelos de predição. Nessa categoria, pode-se destacar três principais enfoques: fusão gênica, ordem gênica e perfil filogenético. A fusão gênica prevê que duas proteínas as quais possuem homologia com organismo em que proteínas são fundidas (complexo proteico de interação) também potencialmente realizará interações. Uma vez que há um complexo no organismo de gene homólogo, o modelo considera que potencialmente esse complexo se concretizará nos homólogos. A ordem gênica se utiliza do conceito de que a ordem gênica se mantém em baixo nível entre pares de genes conservados, embora pares de gene conservados possuam uma melhor chance de interagirem entre si (Dandekar et al., 1998). Aqui também se aplica noções sobre a localização no gene e como essa influencia as interações, com genes próximos tendendo a realizarem interações. Por fim, o perfil filogenético é uma ferramenta bastante útil nos modelos baseados no genoma. Ela se baseia na história evolutiva de uma proteína e é apresentada em

uma árvore filogenética. Informações desse tipo se demonstram como relevantes já que funções e interações costumam se manter em espécies filogeneticamente próximas (Pellegrini *et al.*, 1999). A coevolução também é empregada nesse interim, uma vez que alterações ocorridas por pressões evolutivas em uma proteína tende a causar alterações nas proteínas que entregam aquela rede de interações de modo que sejam mantidas as interações.

A ontologia gênica descreve componentes celulares, funções moleculares e processos biológicos de um gene. Essas informações são armazenadas em banco de dados como o banco de dados GO (Carbon *et al.*, 2017) e QuickGO (Binns *et al.*, 2009). Os modelos que se baseiam em ontologia gênica utilizam a análise das funções biológicas das proteínas para determinar suas potenciais interações, já que proteínas que interagem tendem a se localizarem em um mesmo local celular e realizarem funções biológicas similares. Para isso, os modelos elaboram vetores de features que descrevam as similaridades entre um par de proteínas (um valor de similaridade), de modo parecido com o que é feito para os outros modelos, quantificando, assim, a possibilidade da interação. Esse modelo foi utilizado por (Bandyopadhyay e Mallick, 2017), que tratou os pares de proteínas como documentos (conjunto de palavras) e os termos da anotação das proteínas como palavras. O cálculo para determinar o valor da feature se baseia em um valor atribuído a palavra multiplicado pelo peso daquele termo dentro do documento (interação).

1.4.4. Modelos que utilizam machine learning

Intensificado nos últimos anos, a aplicação de aprendizagem de máquina tem sido de grande valia para a predição de redes de interação. A ideia central é utilizar algoritmos dotados de capacidade de aprender a identificar padrões e presumir novas interações. Utiliza-se dados de treinamento experimentais obtidos de bancos de dados que “ensinam” o modelo para que o sistema seja capaz de identificar padrões gerais e extrapolar esses padrões para prever novas

informações (Tang *et al.*, 2023). Na prática, os modelos são ensinados com grande quantidade de dados sobre a sequência, estrutura, características funcionais e genômicas das proteínas. Da mesma maneira que ocorre nos modelos acima mencionados, os modelos que utilizam aprendizagem de máquina necessitam de receber, como entrada, dados numéricos como representantes das features. Os dados que alimentam os algoritmos preditores de PPI são obtidos dos crescentes bancos de dados biológicos. Um impeditivo observável está no fato de que os bancos de dados são alimentados apenas com informações positivas, isto é, interações de proteínas que ocorrem (no caso de banco de dados de PPI), estruturas tridimensionais observáveis, sequências de aminoácidos. Informações negativas, como casos em que interações não ocorrem ou que a estrutura seja diferente da observável não estão depositados nesses bancos. Os sistemas de aprendizagem de máquinas dependem de dados positivos e negativos para seu modelo de treinamento. Desse modo, obter dados confiáveis para treinar algoritmos em casos negativos é um destaque importante nessas pesquisas. As soluções encontradas normalmente são dos próprios grupos de pesquisa e empregam interações sem sentido para serem os dados negativos (Tang *et al.*, 2023).

Para modelos baseados em sequência que empreguem aprendizado de máquina, os dados podem ser convertidos em representação numérica por uma gama de metodologias, tais como frequência k-mer, auto-covariância, autocorrelação de Moran e matriz de escore baseada em posição específica (Tang *et al.*, 2023). Essas metodologias conferem valores de features para as proteínas ou aminoácidos com base em dados físico-químicos, como polaridade, hidrofobicidade, volume de cadeia lateral; em dados evolutivos da proteína, como análises de conservação; em dados de posição na sequência. Dos diversos métodos existentes, pode-se mencionar o DeepPPI (Du *et al.*, 2017) cria features com base em propriedades dos aminoácidos e constrói redes neurais profundas, além do DeepFE-PPI (Du *et al.*, 2017) que

utiliza também redes neurais profundas para capturar features sobre a distribuição dos resíduos na cadeia proteica.

A estrutura tridimensional também carrega informações padronizadas que são capazes de ensinar algoritmos. Features numéricas são extraídas em diversas metodologias e se utilizam de dobramentos dos resíduos de aminoácidos, alinhamentos de estruturas, distância entre estruturas, ligações dissulfeto entre cadeias e dados de solvatação e protusão (Tang *et al.*, 2023). Nesse sentido, Baranwal et al. (2022) desenvolveu uma ferramenta baseada na estrutura tridimensional e na teoria dos grafos que prevê interações. (Northey, Barešić e Martin, 2018) desenvolveram o InPred, que utiliza dados do PDB para criar valores de features representativas das dobragens de proteínas e emprega o Random Forest como classificador, obtendo resultados com boa performance quando comparados com métodos tradicionais.

1.5. GenPPI: preditor PPI baseado em dados do genoma

O GenPPI se insere no universo de ferramentas preditoras de PPI. Trata-se de um software capaz de prever redes de proteínas utilizando informações sobre vizinhança gênica, perfil filogenético e fusão de genes (Anjos *et al.*, 2021a). O algoritmo passou por um aprimoramento para integrar um modelo de aprendizado de máquina baseado em propriedades biofísicas utilizando o random forest (Silva *et al.*, 2024). O GenPPI permite a inserção de genomas de procariotos e retorna redes de interação a partir desses. Demonstra-se como uma promissora ferramenta para estudar o interactoma de procariotos pouco descritos, obtendo relevantes informações biológicas, e mesmo anotar genomas recém depositados e, portanto, vagamente conhecidos.

1.6. *Acetilactobacillus jinshanensis*

Nesse interim, a *Acetilactobacillus jinshanensis* é uma recente espécie de bactéria gram-positiva, anaeróbica facultativa, não esporulante e não móvel isolada de cultura de

fermentação aberta do vinagre aromático de Zhenjiang (Yu *et al.*, 2020a). Isolada e descrita inicialmente em 2019, foi batizada como *Lactobacillus jinshani* e posteriormente renomeada e reclassificada como *Acetilactobacillus jinshanensis* na toada da grande reclassificação taxonômica que a família *Lactobacillaceae* sofreu em 2020 (Qiao *et al.*, 2022; Zheng *et al.*, 2020). Estudos metabólicos e anotações genômicas do organismo seguem em fase embrionária. Na data da escrita deste texto, o NCBI contava apenas com duas anotações de genoma da espécie (“Genome - NCBI - NLM”, 2024), sendo uma feita pelo orientador deste trabalho como parte do estudo deste projeto. Além disso, o número de artigos sobre a espécie ainda permanece baixo. Demonstra-se, portanto, que novas pesquisas são necessárias para desvendar mais informações sobre a espécie, sobre suas interações ecológicas e seus processos metabólicos que potencialmente possuam relevância biotecnológica.

1.7. *Bacillus cereus*

A *Bacillus cereus* é uma espécie mais bem descrita e estudada. Trata-se de uma gram-positiva, anaeróbica facultativa, esporulante e produtora de toxinas (Parihar, 2014). Sua presença está normalmente relacionada com infecções alimentares (Kotiranta, Lounatmaa e Haapasalo, 2000; Mendes, Coelho e Azeredo, de, 2011). Curiosamente, essa espécie também foi encontrada no vinagre de Zhenjiang como partícipe do processo fermentativo (Liu *et al.*, 2023). A mesma espécie foi encontrada no alimento larval de abelhas sem ferrão, o que levanta diversas questões sobre papéis bioecológicos.

1.8. *Acetilactobacillus jinshanensis* e a *Bacillus cereus* no vinagre de Zhenjiang

A *Acetilactobacillus jinshanensis* e a *Bacillus cereus* foram isoladas de um processo de fermentação de estado sólido aberta para produzir vinagre aromático de Zhenjiang. O vinagre de Zhenjiang, produzido na província de Jiangsu na China, representa um produto alimentício de grande importância econômica e cultural, com um mercado anual estimado em centenas de milhões de dólares e exportações para mais de 60 países. O método de fermentação é

subdividido em sacarificação do amido, fermentação alcoólica e oxidação do etanol a ácido acético. Os substratos são trigo e arroz e o processo é conduzido com o uso de um vasto consórcio de bactérias (Xu *et al.*, 2024a). A inoculação do meio é feita com o mosto que já fora fermentado por 7 dias, o que realimenta o processo e o mantém estável (Xu *et al.*, 2011a). Todos os passos da fermentação usualmente levam 20 dias. O conjunto de produtos dos processos fermentativos presentes na fermentação aberta do vinagre promove uma mistura diversa de sabores e produtos com riqueza nutricional e potencial biotecnológico.

1.9. Produtos biotecnológicos e prospecção

A prospecção de espécies novas e de biomoléculas com fim a exploração para bioeconomia é tarefa recorrente e relevante na pesquisa de base (Oliveira, De, 2023; Pessoa, 2024; Verma e Shah, 2022). Produtos naturais, fitoterápicos e enzimas são passíveis de serem extraídos de ambientes diversos. O potencial que se pode descobrir a respeito desses compostos permanece incalculado, sendo presumido que apenas uma pequena parte dos microrganismos e bioativos estão claramente descritos (Scherlach e Hertweck, 2021). Ferramentas recentes como tecnologias ômicas vêm sendo empregadas no sentido de incrementar a capacidade exploratória dessas substâncias.

1.10. O alimento larval de abelhas sem ferrão como alvo de prospecção e a presença das bactérias *Acetilactobacillus jinshanensis* e *Bacillus cereus*

O alimento larval de abelhas sem ferrão possui um grande potencial biotecnológico (Santos, 2022). Sua composição é de uma mistura de pólen fermentado, mel e secreções glandulares de um grupo diverso de abelhas cuja principal característica é a ausência de um ferrão funcional. Diversos estudos indicam uma microbiota diversa (Vásquez *et al.*, 2012) e compostos bioativos com propriedades antimicrobianas, anti-inflamatórias, antioxidante, além de outras potenciais aplicações farmacológicas dos compostos constituintes do alimento larval (Al-

Hatamleh *et al.*, 2020; Mohammad, Mahmud-Ab-Rashid e Zawawi, 2020; Santos *et al.*, 2022). Tanto a *Acetilactobacillus jinshanensis* quanto a *Bacillus cereus* foram encontrados no alimento larval de abelhas sem ferrão. Esse achado abre caminhos para um estudo que envolva a compreensão dos papéis bioecológicos das duas espécies dentro do sistema de interações bioecológicas dentro das colônias dessas abelhas, assim como para um estudo que enfoque a prospecção de produtos biotecnológicos utilizando essas espécies como referenciais.

1.11. Uso da bioinformática para prospecção de produtos biotecnológicos

O emprego de ferramentas bioinformáticas modernas para colaborar com a prospecção de microrganismos e de bioativos é uma aplicação que se demonstra promissora e inovadora. Além disso, essas ferramentas possibilitam uma elaboração de hipóteses holísticas que englobem a compreensão das possíveis relações ecológicas, das vias metabólicas relevantes para aplicações futuras e da melhor compreensão de organismos pouco descritos. As bactérias *Acetilactobacillus jinshanensis* e a *Bacillus cereus* foram encontradas e isoladas do alimento larval de abelhas sem ferrão. Dado o baixo conhecimento que ainda se possui sobre a *Acetilactobacillus jinshanensis*, as pouco intuitivas relações entre a *Bacillus cereus* com o alimento larval e com o vinagre chinês, o potencial biotecnológico desse alimento larval e as possíveis interações biológicas existentes, observa-se que há um vasto campo para o emprego das ferramentas bioinformáticas, incluindo-se o estudo do interactoma.

2.OBJETIVOS

2.1. Objetivo Geral

Empregar ferramentas bioinformáticas para montar e estudar o genoma da bactéria *Acetilactobacillus jinshanensis* CP171343 e da *Bacillus cereus* CP174498, assim como compreender suas interações ecológicas, seu potencial biotecnológico e seus processos metabólicos relevantes.

2.2. Objetivos específicos

- Construir rede de interações de proteínas da *Acetilactobacillus jinshanensis* CP171343 *in silico* a partir do genoma coletado do alimento larval de abelhas e montado em ferramentas biotecnológicas;
- Construir rede PPI da *Bacillus cereus* CP174498 *in silico* a partir do genoma isolado do alimento larval de abelhas e montado em ferramentas biotecnológicas;
- Estudar a topologia da rede da da *Acetilactobacillus jinshanensis* CP171343 formada, rastrear interações relevantes e verificar implicações biológicas dessas interações;
- Proceder análises das principais vias metabólicas presentes no conjunto de proteínas estudadas;
- Testar metodologia desenvolvida para estudo de redes PPI;
- Elaborar hipóteses que englobem o estudo da topologia da rede de PPI, das vias e do conhecimento experimental presente na literatura.

3. METODOLOGIA

O presente trabalho utilizou material genético sequenciado a partir de amostras do alimento larval de abelhas sem ferrão. O genoma da *Acetilactobacillus jinshanensis* CP171343 foi encontrada por metagenômica nesse substrato, enquanto o genoma da *Bacillus cereus* CP174498 adveio de uma colônia isolada, a qual também foi retirada do alimento larval. As amostras do alimento larval foram retiradas de colmeias de 4 espécies de abelha sem ferrão: *Frieseomelitta varia*, *Melipona quadrifasciata*, *Melipona scutellaris*, e *Tetragonisca angustula*. O sequenciamento do material foi realizado, em que se encontrou a *Acetilactobacillus jinshanensis* CP171343. Para a *Bacillus cereus*, o alimento larval foi preparado e incubado para o crescimento das bactérias. O cultivo das colônias selecionadas por caracterização de gram e por morfologia foi em LB (Luria Bertani) suplementado com 2% de ácido acético. Posteriormente, foi plaqueado em ágar LB com ácido acético 2% para isolar as colônias características da *Bacillus cereus*. Procedeu-se com a extração do material genético e

sequenciamento. Ferramentas bioinformáticas para montagem de genoma foram empregadas. Posteriormente submetidos ao GenPPI, os genomas devidamente montados resultaram em uma rede de interação por metodologia que será mais bem descrita a seguir. O cultivo e sequenciamento não foi parte desse trabalho, sendo realizado pelo grupo de pesquisa do Laboratório de Genética (IBTEC/UFU) e por empresa contratada para realizar o sequenciamento. A metodologia da montagem do genoma será descrita, mas foi realizada pelo orientador deste projeto, o Dr. Anderson Rodrigues dos Santos, conjuntamente com o grupo de pesquisa do Laboratório de Genética. O enfoque desse trabalho foi a montagem da rede e sua interpretação biológica

3.1. Montagem do genoma

A montagem de genoma consiste no processo de obter uma sequência com sentido biológico a partir dos dados do sequenciamento. As leituras (reads) do sequenciamento são manipuladas para que um genoma de organismo vivo seja descrito. Para isso, foi utilizando as ferramentas SPAdes (Prjibelski *et al.*, 2020) e RagTag (Alonge *et al.*, 2022).

O SPAdes (St. Petersburg genome Assembler) é um algoritmo de montagem de genomas projetado para montagem de bactérias unicelulares, mas que pode ser adaptado para montagem de organismos multicelulares. A ferramenta aceita dados de sequenciamento de IonTorrent, assim como montagem híbrida de genoma de leituras curtas e longas (PacBio e Oxford Nanopore). O algoritmo opera por meio da montagem de Bruijn e a montagem de sobreposição. De início, o algoritmo corrige erros de leitura de sequência, segue com a montagem *de novo* da sequência com uso da construção de contigs (sequências contíguas de DNA) usando um grafo de Bruijn, que é uma representação compacta das sobreposições entre as sequências. Após isso, o SPades expande os contigs com base em informações de sobreposição entre as leituras, reduz erros nas montagens, refinando as sobreposições e corrigindo inconsistências e gera o genoma

final montado, que consiste em contigs e scaffolds (contigs conectados por regiões não resolvidas) que representam o genoma original.

O RagTag opera aprimorando a sequência consenso que foi obtida pelo SPAdes. Nessa etapa, o algoritmo recorta, remenda e abre gaps na sequência para que ela reflita o mais próximo possível de um genoma de referência, que aqui é a *Acetilactobacillus jinshanensis* e o *Bacillus cereus*. O RagTag também modifica as posições dos trechos quando essa for necessária para a adequação no genoma a ser montado.



Figura 3 -funções do RagTag. O algoritmo pode corrigir trechos faltantes com base na referência; criar gaps com base na comparação com a referência (empilhar); remendar trechos para que se enquadre em um genoma buscado (alvo/pesquisa); ou combinar trechos. Adaptado de (Alonge et al., 2019).

O conjunto de ferramentas empregadas nessa etapa busca identificar qual genoma é predominante nas leituras advindas do sequenciamento e montar esse genoma utilizando a referência depositada no NCBI. A montagem do genoma é, portanto, uma montagem que utiliza o conhecimento de uma referência. Contra essa referência, o genoma que está sendo montado é comparado para confirmar se há predominância dele no conjunto das leituras. Não raras as vezes que são encontrados outros genomas nas leituras advindos de um processo de metagenômica ou contaminação. Como mencionado, dois genomas foram montados: o da *Acetilactobacillus jinshanensis* CP171343 e da *Bacillus cereus* CP174498.

Em etapa subsequente, foi feita a anotação desses genomas com o uso do PANNOTATOR, que consiste numa ferramenta baseada em web que prediz os genes, funções e produtos de um genoma tendo recebido como entrada uma sequência de nucleotídeos (Santos,

A. R. *et al.*, 2013). O output do programa é um arquivo com a predição das proteínas que estão presentes nesse genoma.

3.2. Montagem da rede de interação proteína-proteína

Para a montagem da rede de interação, foi utilizado o GenPPI (Anjos *et al.*, 2021b; Silva *et al.*, 2024). O algoritmo executa a criação de uma rede de interações de proteínas utilizando como entrada de genomas. O algoritmo emprega informações sobre perfil filogenético conservado, vizinhança gênica conservada e fusão de genes, além do aprimoramento recente ter empregado aprendizado de máquina utilizando o random forest como classificador, no qual se faz um perfil de similaridade entre as proteínas com nas características biofísicas dos aminoácidos constituintes. Por padrão, a entrada no software é com conjuntos de genomas taxonomicamente próximos, assim como do próprio genoma da *Acetilactobacillus jinshanensis* CP171343 e do *Bacillus cereus* CP174498 montados nas etapas anteriores.

Para a *Acetilactobacillus jinshanensis* CP171343, foram escolhidas bactérias amplamente estudadas e do gênero *Lacticaseibacillus*. Seus genomas depositados no NCBI foram baixados e uma filtragem foi feita para que os tamanhos de genomas escolhidos fossem similares ao da *A. jinshanensis* CP171343. Essa filtragem por tamanho é necessária para que o GenPPI realize a predição de redes de forma correta, uma vez que genomas com tamanhos dissimilares geram uma explosão combinatória de interações, o que torna a análise de redes inviável. Além disso, genomas de tamanhos parecidos tendem a possuírem proteomas mais próximos ao do organismo estudado. As espécies possuem, normalmente, mais de um genoma depositado no NCBI. Desse modo, foram escolhidas 9 espécies mais a *A. jinshanensis*. Para a *Bacillus cereus* CP174498, foram utilizados 85 genomas de 85 espécies próximas taxonomicamente e a *Bacillus cereus* CP174498, montada nesse projeto.

Tabela 1- espécies utilizadas para montagem de rede de interação

Rede	Número de espécies	Espécies
<i>Acetilactobacillus jinshanensis</i> CP171343	10 (da <i>A. jinshanensis</i> CP171343 inclusa)	<i>Lacticaseibacillus casei</i> , <i>Lactobacillus acidophilus</i> , <i>Lactobacillus apis</i> , <i>Lactobacillus bombicola</i> , <i>Lactobacillus crispatus</i> , <i>Lactobacillus delbrueckii</i> , <i>Lactobacillus intestinalis</i> , <i>Lactobacillus johnsonii</i> , <i>Lactobacillus paragasseri</i> . <i>A. jinshanensis</i>
<i>Bacillus cereus</i> CP174498	85 (<i>Bacillus cereus</i> CP174498 inclusa)	<i>Bacillus aerius</i> , <i>Bacillus albus</i> , <i>Bacillus clarus</i> , <i>Bacillus cytotoxicus</i> , <i>Bacillus manliponensis</i> , <i>Bacillus mesophilus</i> , <i>Bacillus pumilus</i> , <i>Bacillus rugosus</i> , <i>Bacillus subtilis</i> , <i>Bacillus tequilensis</i> , <i>Bacillus aerolatus</i> , <i>Bacillus altitudinis</i> , <i>Bacillus coahuilensis</i> , <i>Bacillus daqingensis</i> , <i>Bacillus mesophilum</i> , <i>Bacillus methanolicus</i> , <i>Bacillus rhizoplanae</i> , <i>Bacillus safensis</i> , <i>Bacillus taeanensis</i> , <i>Bacillus thermotolerans</i> , <i>Bacillus amyloliquefaciens</i> , <i>Bacillus anthracis</i> , <i>Bacillus dicomae</i> , <i>Bacillus mojavenensis</i> , <i>Bacillus mobilis</i> , <i>Bacillus salipaludis</i> , <i>Bacillus salacetis</i> , <i>Bacillus tianshenii</i> , <i>Bacillus thuringiensis</i> ,

Bacillus aquiflavi, *Bacillus*
ectoiniformans, *Bacillus mycoides*,
Bacillus salitolerans, *Bacillus toyonensis*,
Bacillus arachidis, *Bacillus fungorum*,
Bacillus nakamurai, *Bacillus*
seohaeensis, *Bacillus tropicus*,
Bacillus atrophaeus, *Bacillus*
gaemokensis, *Bacillus norwichensis*,
Bacillus shivajii, *Bacillus velezensis*,
Bacillus australimaris, *Bacillus*
glycinifermentans, *Bacillus oleivorans*,
Bacillus smithii, *Bacillus weihaiensis*,
Bacillus badius, *Bacillus gobiensis*,
Bacillus pacificus, *Bacillus solimangrovi*,
Bacillus wudalianchiensis, *Bacillus*
benzoevorans, *Bacillus haikouensis*,
Bacillus pakistanensis, *Bacillus*
songklensis, *Bacillus xiamenensis*,
Bacillus cabrialesii, *Bacillus*
halotolerans, *Bacillus paramobilis*,
Bacillus sonorensis, *Bacillus*
zhangzhouensis, *Bacillus canaveralius*,
Bacillus hominis, *Bacillus paramycoides*,
Bacillus spizizenii, *Bacillus cereus*,
Bacillus cereus CP174498, *Bacillus*

capparidis, Bacillus horti, Bacillus
paranthracis, Bacillus spongiae, Bacillus
carboniphilus, Bacillus inaquosorum,
Bacillus pretiosus, Bacillus stercoris,
Bacillus changyiensis, Bacillus
infantis, Bacillus proteolyticus, Bacillus
suaedae, Bacillus chungangensis,
Bacillus licheniformis, Bacillus
pseudomycooides, Bacillus suaedaesalsae

O algoritmo prevê uma rede de interações com base nesses genomas. Após a rodagem, o algoritmo retorna arquivos com as redes, mas não possui dentro dele ferramenta de observação da rede. Pode-se empregar o software R para visualizar detalhes das redes formadas; o Cytoscape (Shannon *et al.*, 2003) e o Gephi (Bastian, Heymann e Jacomy, 2009) para observar a topologia da rede com seus nós e arestas.

3.3. Estudo da topologia da rede

Buscou-se estabelecer interpretações biológicas para topologia das redes obtidas. A topologia da rede compreende a estrutura visual da rede PPI, conforme indicada na figura 1. Ela foi observada no Gephi e o R foi empregado para obter informações complementares sobre interações. A topologia de rede compreende um conjunto de nós e arestas, sendo as proteínas representadas pelos nós e as interações representadas pelas arestas. Por se tratar um genoma inteiro, há um número significativo de nós e arestas. A análise individual é inviável. Portanto, a escolha por uma filtragem é uma abordagem adotada neste projeto.

Para essa filtragem ser efetiva, foi empregado alguns aspectos da teoria dos grafos. Nesse sentido, os nós são rastreados utilizando algumas ferramentas de medidas de centralidade, densidade do grafo e análise de clusters (Pavlopoulos *et al.*, 2011). O Gephi possui plug-ins

que permitem essas análises. Aqui foi empregado o Bridging Centrality (BriCe). Trata-se de uma medida que combina outras duas: a Betweenness Centrality (BetCe) e Bridging Coefficient (BriCo). A Betweenness Centrality se refere a uma medida que mede a importância de um nó com base em quantas vezes ele aparece nos caminhos mais curtos entre outros pares de nós. Isso significa que mensura como aquele nó serve de intermediário na rede, ao mensurar quantos caminhos curtos lhe cruzam. O Bridging Coefficient quantifica o quão bem-posicionado um nó está entre nós com altos graus, isto é, com grandes números de interações. Um maior número de interações representa a formação de clusters, que é um conjunto local de nós com mais interações dentro de uma rede de PPI maior. O Bridging Centrality refina os dois conceitos e quantifica o potencial se um nó servir como ponte entre dois clusters, isto é, entre nós com alto grau. O esperado na análise de BriCe para redes biológicas é que haja proteínas que integrem processos biológicos maiores dentro do organismo. Cabe detalhar que uma proteína com alto BriCe potencialmente estaria ligando conjuntos de clusters de uma rede. Na interpretação biológica, seria uma proteína que possibilita que haja outros processos metabólicos em funcionamento, sendo esses processos parcialmente independentes. A descrição matemática e a implementação foi descrita por Pereira, Ghosh e Santos, (2021). Os autores também utilizaram a medida para traçar hipótese entre uma possível ligação entre proteínas do grupo glicosiltransferase e o sistema de secreção tipo VII do *Mycobacterium tuberculosis* H37Rv. O BriCe foi utilizado para ranquear os nós com mais altos valores. Esse consistiu em um dos processos de filtragem utilizados para escolher quais proteínas seriam estudadas por metodologia descrita no tópico 3.4.

3.4. Análise individual dos nós

Os nós da rede são as proteínas do genoma estudado. Duas abordagens foram empregadas para estudar essas proteínas. A primeira foi buscar sua anotação nos próprios depósitos do NCBI. Lá é possível encontrar a anotação da função do gene, o produto da

expressão e a sequência proteica e de ácido nucleico referente a proteína estudada. Esse dado já permite elaboração de hipótese sobre a função daquela proteína e seu potencial papel biológico. A segunda abordagem foi utilizar o InterProScan. Trata-se de uma ferramenta que integra vários bancos de dados sobre funções de proteínas, famílias e domínios (Quevillon *et al.*, 2005). O InterProScan recebe sequências das proteínas estudadas e retorna domínios conservados, famílias e funções principais. Essas informações são relevantes para o estudo individual dos nós. Seu uso nesse trabalho ficou restrito a um menor grupo de proteínas do genoma, que foram escolhidas após um processo de filtragem. Essa filtragem está descrita nos tópicos 3.5.1 e 3.3. Apenas esse número menor de proteínas foi submetido ao InterProScan para o estudo, já que o estudo de todos os nós individualmente é impossibilitado, dado a quantidade de nós e interações normalmente presentes em um genoma. Na prática, tem-se algumas centenas de proteínas e centenas de milhares de interações, dado que a análise individual é humanamente inviável.

3.5. Estudo das interações

O estudo das interações consistiu em analisar individualmente as interações da rede que foram consideradas relevantes. Como mencionado, o genoma possui um número significativo de proteínas e interações, tornando a análise individual impossível. Já foi apontado que uma abordagem adequada seria partir de proteínas específica, as seeds (sementes), e estudar suas interações (Pavlopoulos *et al.*, 2011; Tomkins e Manzoni, 2021). Estudar as interações, nesse sentido, é elaborar hipóteses sobre qual a razão de duas proteínas estarem realizando interações no genoma, dado que se sabe quais são as funções e produtos dos genes tanto das seeds, quanto das proteínas que interagem com essas seeds, uma vez que esse dado é obtido pela metodologia descrita no tópico 3.4. Escolher nós específicos para o estudo foi uma decisão desse projeto. Esse processo reduz o número de nós avaliados e potencializa novas descobertas.

3.5.1. Escolha das sementes

Nesse tópico, será descrito como se deu a seleção dos nós que serão estudados. Essa escolha compreende a segunda etapa de filtragem de proteínas. A priori, destaca-se que esses nós seriam estudados se a análise da etapa seguinte comprovar a presença dessas proteínas no genoma. Por conveniência, aqui será descrito quais foram os critérios para que fossem selecionadas as proteínas que seriam estudadas. Não há na literatura um critério bem estabelecido para a escolha dos nós que serão ponto de partida para o estudo da rede. Aqui propõe-se um conjunto de práticas que foram norteadoras deste projeto.

A origem das bactérias foi considerada. A metodologia aqui empregada demanda um ponto de partida relativamente conhecido, para o qual seja possível estabelecer um conjunto de proteínas conhecidas para o estudo das interações. As bactérias *A. jinshanensis* e *B. cereus* foram isoladas de uma fermentação de estado sólido aberta que produz vinagre aromático de Zhenjiang (Yu *et al.*, 2020a). Esse produto alimentício contém uma diversidade de microrganismos envolvidos e uma composição química diversa que está relacionada com odores e sabores do produto (Wang *et al.*, 2024; Xu *et al.*, 2011b). O conhecimento de que processos fermentativos ocorrem no local em que se isolou a bactéria levou a teorizar sobre a possibilidade desses organismos participarem ativamente da fermentação acética. Há sempre a possibilidade de se tratar de uma contaminação, já que se está diante de uma fermentação aberta. Sendo assim, as enzimas tradicionalmente presentes em fermentação acética foram um ponto de investigação. Com o emprego do Metacyc, foi buscada as vias e enzimas que participam da produção de ácido acético. O Metacyc é um grande banco de dados sobre vias biológicas (Caspi *et al.*, 2020). Diversas vias são responsáveis por essa produção. No entanto, dado que a *A. jinshanensis* possui proximidade de clado com as bactérias do ácido acético, optou-se pela via mais descrita do grupo (Mamlouk e Gullo, 2013). Nessa via, etanol é convertido a acetaldeído

pela ação da álcool desidrogenase e o acetaldeído é convertido a ácido acético pela ação da aldeído desidrogenase.

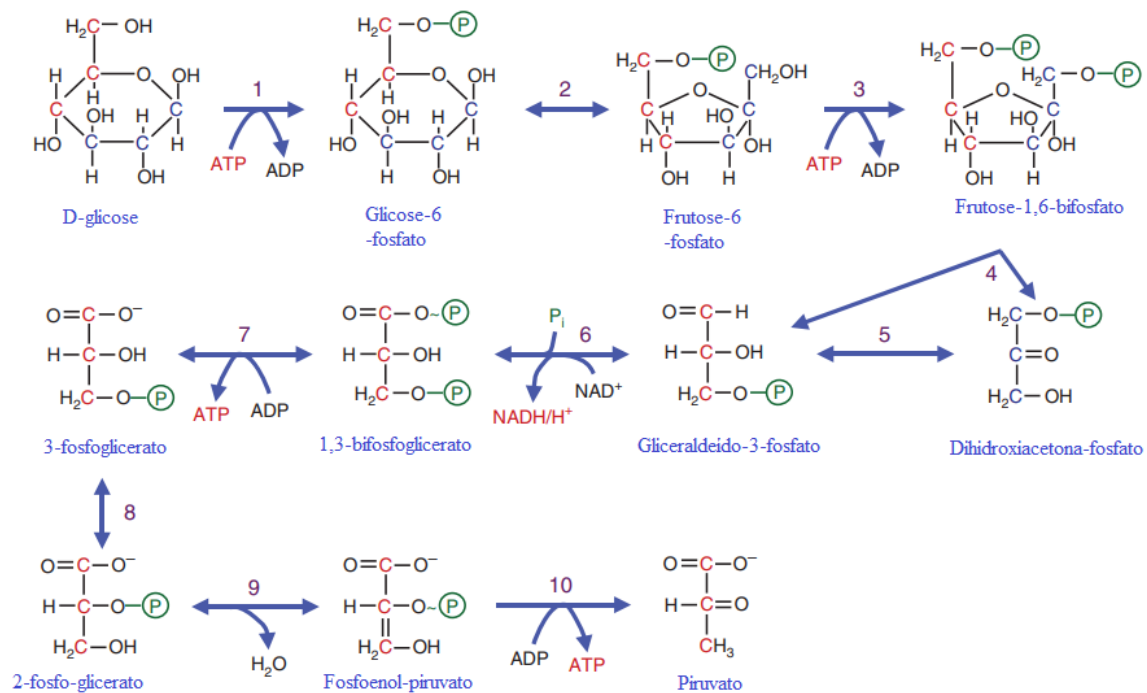
Uma outra questão que se levanta é se a bactéria realizaria a fermentação alcoólica para produzir o etanol, que é demandando para fermentação acética. A investigação mirou enzimas da fermentação alcoólica nessa etapa. Conhecidamente a piruvato descarboxilase, que converte piruvato a etanol. É descrito também que o ácido acético é transportado para o exterior da célula, sendo necessário a presença de um transportador (Mamlouk e Gullo, 2013). Ao final dessa etapa, foram selecionadas para estudo as seguintes enzimas: proteína FPS1 facilitadora de transporte de glicerol (transportador), álcool desidrogenase, acetaldeído desidrogenase e piruvato descarboxilase.

Atrelado a isso, propôs-se conhecer os processos bioenergéticos dos organismos. Uma vez que o ponto de partida foi o conhecimento da fermentação que seria investigada, buscou-se investigar as vias bioenergéticas. Essa escolha também se baseou na existência de prova experimental de que a *A. jinshanensis* realiza um processo de fermentação heterolática, também chamada de via da fosfoacetolase (Yu *et al.*, 2020b). Em todas essas etapas, parte-se de um conhecimento existente, normalmente obtido por prova experimental, para explorar relações não descritas ou pouco conhecidas.



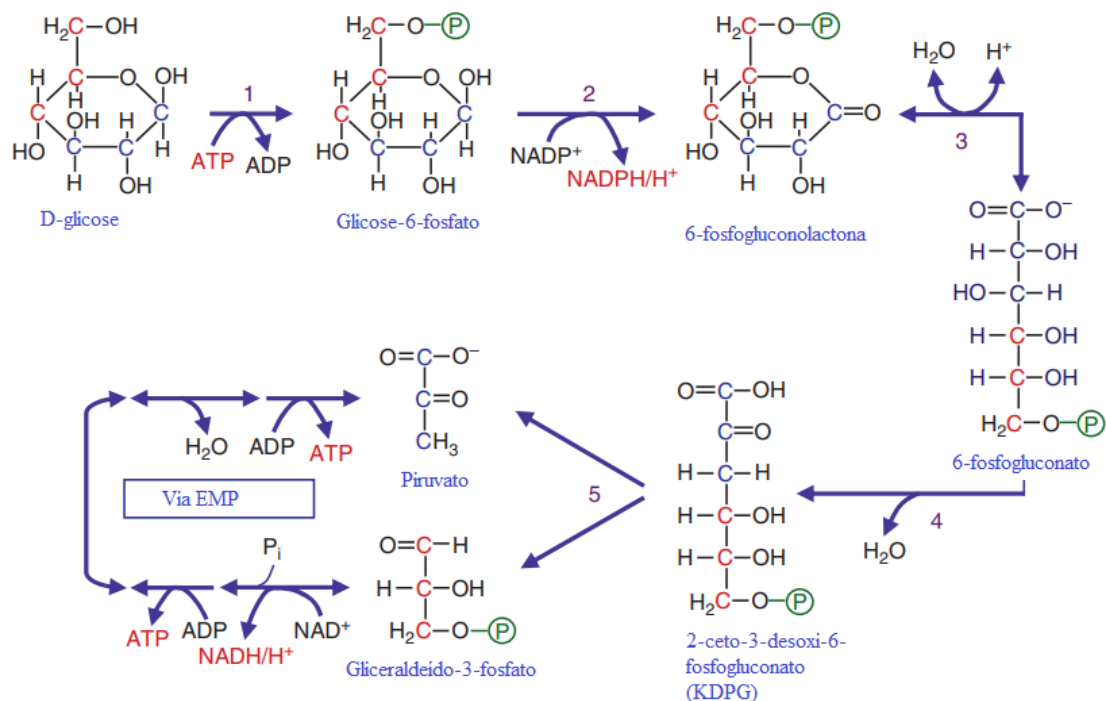
Os processos bioenergéticos são bons pontos de partidas para o estudo dos organismos. Há vasta descrição das principais vias e bancos de dados completos que permitem um endereçamento relevante para as análises. A literatura descreve 3 principais vias de oxidação da glicose: a via Entner-Doudoroff, (ED) a via Embden-Meyerhof (EMP) e a via da fosfoctolase (Shimizu, 2011). Buscou-se investigar essas vias rastreando a presença das enzimas participante dessas vias.

26



5

Figura 5-via Embden-Meyehof-Parnas (EMP). Enzimas da via são (1) hexoquinase, (2) fofoglucisomerase, (3) fosfofrutoquinase, (4) frutose bifosfato aldolase, (5) triose fosfato isomerase, (6) gliceraldeído-3-fosfato desidrogenase, (7) fosfoglicericoquinase, (8) fosfoglicerato mutase, (9) enolase e (10) piruvato quinase. Adaptado de (Spector, 2009).



6

Figura 6- via Entener-Doudoroff (ED). Enzimas da via são (1) hexoquinase, (2) glicose-6-fosfato desidrogenase, (3) 6-fosfogluconoolactonase, (4) 6-fosfogluconato desidratase e (5) 2-ceto-3-deoxi-6-fosfogluconato aldolase. Adaptado de (Spector, 2009)

As enzimas de todas essas vias foram alvos de estudos. Após comprovar a presença de parte dessas enzimas no genoma da *A. jinshanensis* CP171343, procedeu-se com estudo individual das interações que estabelecem. Em síntese, são as enzimas estudadas: hexoquinase, glicose-6-fosfato desidrogenase, 6-fosfoglucono-lactonase, 6-fosfogluconato desidrogenase, ribose-5-fosfato epimerase, petose fosfocetolase, lactato desidrogenase, fosfotransacetilase, aldeído desidrogenase, álcool desidrogenase, acetil fosfato quinase, fofoglucoisomerase, fosfofrutoquinase, frutose bifosfato aldolase, triose fosfato isomerase, gliceraldeído-3-fosfato desidrogenase, fosfogliceroquinase, fosfoglicerato mutase, enolase, piruvato quinase, glicose-6-fosfato desidrogenase, 6-fosfogluconato desidratase e 2-ceto-3-deoxi-6-fosfogluconato aldolase.

3.5.2. Validação da presença das sementes

Essa etapa é parte da anterior e ocorreu simultaneamente. Como a pesquisa ficou focada em desvendar os processos bioenergéticos e as fermentações que as bactérias poderiam participar, buscou-se na literatura os principais processos metabólicos de oxidação da glicose e as enzimas participantes. Uma vez encontradas todas as enzimas, foi empregado a ferramenta BLAST, utilizando como busca cada proteína mencionada na etapa anterior. As sequências de cada proteína foram retiradas do banco de dados do NCBI. Buscou-se as proteínas de espécies de clados próximos, preferencialmente do gênero *Lacticaseibacillus*. Uma vez em posse dessas sequências, o BLAST foi executado. Utilizou-se o genoma montado da *A. jinshanensis* CP171343 para proceder com a busca pelas enzimas dos bioprocessos mencionados. O intuito era identificar quais enzimas estavam presentes e, uma vez presentes, estabelecê-las como nós cujas interações serão estudadas por metodologia descrita no tópico 3.4.



Figura 7-metodologia utilizada nesse trabalho

3.6. Análise local subcelular

A localização subcelular de uma proteína permite inferir relevantes informações sobre suas funções. Aqui, empregou-se o Mature Epitope Density-Medpipe (Santos, Anderson R. *et al.*, 2013) para inferir sobre a localização subcelular das proteínas. O Medpipe é um preditor de epítomos para o desenvolvimento de vacinas. Nesse trabalho, a informação que é relevante é

sua predição sobre o local subcelular das proteínas: citoplasma, membrana, secretada e potencialmente exposta na superfície.

3.7. Identificação das principais vias metabólicas do genoma

Uma segunda etapa bastante relevante desse projeto foi o estudo das principais vias existentes no genoma. A Ontologia Gênica se refere a anotação de um gene, em que se reúne informações sobre a sequência, os produtos desse gene, as funções desempenhadas e as vias metabólicas que participam. Essas informações estão reunidas no InterproScan, que possui ferramentas para rastrear a Ontologia Gênica. As proteínas do genoma que foi montado nesse projeto da *A. jinshanensis* CP171343 foram submetidas ao InterproScan para predição das principais vias. Aqui também foi empregado a clusterização. Está descrito na literatura que o estudo da rede por clusters, isto é, por subconjuntos de redes menores de interação os quais estão inseridos na rede completa de interação, é uma ferramenta capaz de prover pistas sobre processos biológicos e análises locais (Pavlopoulos *et al.*, 2011; Tomkins e Manzoni, 2021). A clusterização aqui foi efetuada criando quatro clusters da rede de interação formada. O parâmetro para criação dos clusters foi a densidade de interações. Há plug-in do Gephi capaz de identificar, ordenar e apresentar os clusters dentro da rede conforme a densidade de interação entre os nós variam. Uma vez em posse dos clusters, as proteínas envolvidas nas interações da cada cluster foram submetidas ao InterproScan para predição das principais vias metabólicas de cada cluster.

3.7. Estudo das vias metabólicas

O estudo das vias metabólicas se refere a análise das vias principais que foram encontradas no genoma da *A. jinshanensis* CP171343. Aqui, buscou-se realizar interpretações holísticas à luz do conhecimento das interações ecológicas que a *A. jinshanensis* CP171343 realiza, dada sua presença no alimento larval de abelhas sem ferrão e de seu potencial papel na fermentação do vinagre chinês. Essas considerações levam em conta que foram isoladas da

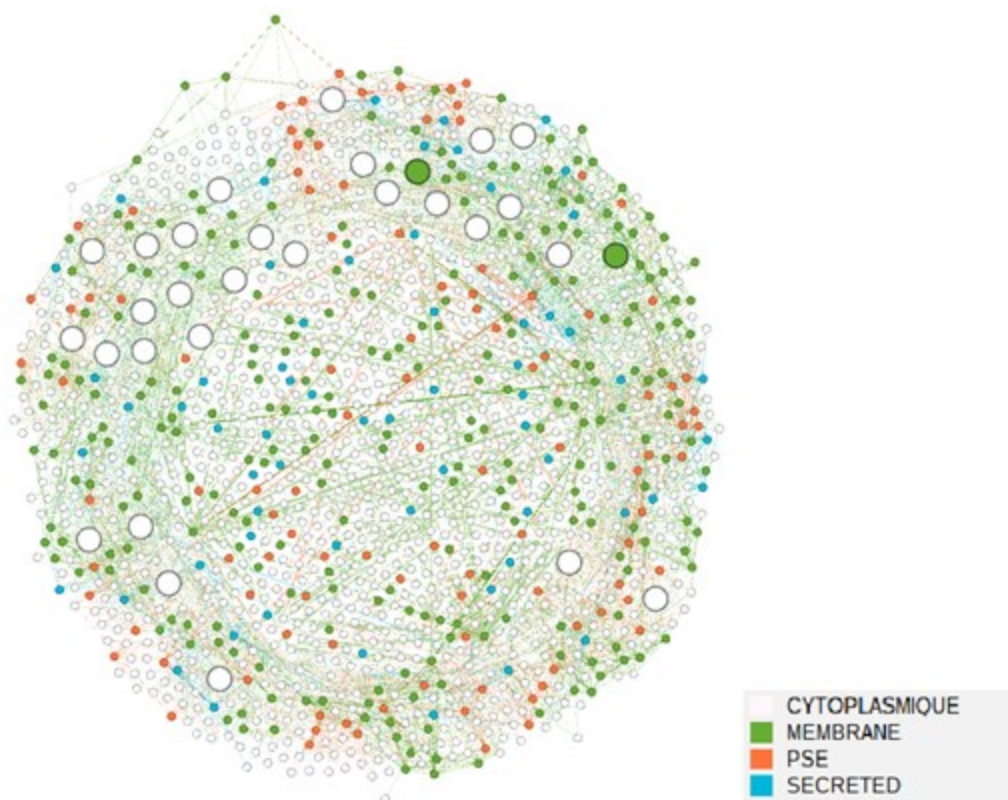
fermentação em estado sólido do vinagre aromático de Zhenjiang (Yu *et al.*, 2020a) e no alimento larval de abelhas, cuja potencial biotecnológico e ecológico na manutenção da homeostase ambiental foram apontados (Santos, 2022). A análise de vias busca compreender os papéis bioecológicos da espécie dentro do sistema da colmeia de abelhas sem ferrão. A análise das vias em clusters busca encontrar possíveis nuances não encontradas na análise do genoma completo. O conjunto de dados levantados permitem hipóteses sobre os papéis biológicos similares ocorridos no alimento larval e no vinagre chinês.

3.8. Busca dos genes na *Bacillus Cereus* CP174498

As enzimas de fermentações e dos processos bioenergéticos estudados também foram buscadas no genoma montado da *B. cereus* CP174498. As sequências das enzimas foram obtidas do NCBI e escolhidas aquelas cujos genomas de origem fossem a proximidade taxonômica com a *B. cereus* CP174498. São as mesmas enzimas dos processos fermentativos investigados na *A. jinshanensis* CP171343 com a distinção de que nesse processo não se utilizou essa etapa para seleção de seeds dado que esse estudo da rede não foi escrutinado nesse projeto.

4. RESULTADOS E DISCUSSÃO

Os genomas da *A. jinshanensis* CP171343 e da *B. cereus* CP174498 foram montados com êxito com o uso do SPades e RagTag, conforme descrito no tópico 3.1. Esses dois genomas foram anotados com o PANNOTATOR. Para formação das redes, os 131 genomas descritos no tópico 3.2 foram enviados para o GEPPi para predição de rede da *A. jinshanensis* CP171343; em processo paralelo, os 85 genomas descritos no tópico 3.2 foram também enviados para o GEPPi para a montagem da rede da *B. cereus* CP174498. Como resultado, obteve-se as redes das duas espécies, *Acetilactobacillus jinshanensis* CP171343 e a *Bacillus cereus* CP174498, utilizando os genomas isolados e montados na metodologia desse trabalho. A rede da *A. jinshanensis* CP171343 resultou em 1454 nós e 246670 arestas. A rede da *Bacillus cereus* CP174498 resultou em 5313 nós e 163409 arestas.



1:1

8

Figura 8- observação da rede da *Acetilactobacillus jinshanensis* CP171343 no Gephi com nós indicando proteínas e arestas indicando interações. Nós destacados com tamanhos maiores indicando nós com maiores valores de Bridge Centrality (BriCe) e cores indicando locais subcelular: citoplasma, membrana, potencialmente exposta na superfície e secretada. Fonte: elaboração própria.

A rede da *A. jinshanensis* CP171343 foi destrinchada para obter elementos como local subcelular, valores de BriCe, estudo dos nós e das interações individuais. A rede da *Bacillus cereus* CP174498 não foi objeto de intenso escrutínio, sendo seu genoma montado analisado apenas para obter informações complementares. Mas sua rede está aqui representada.

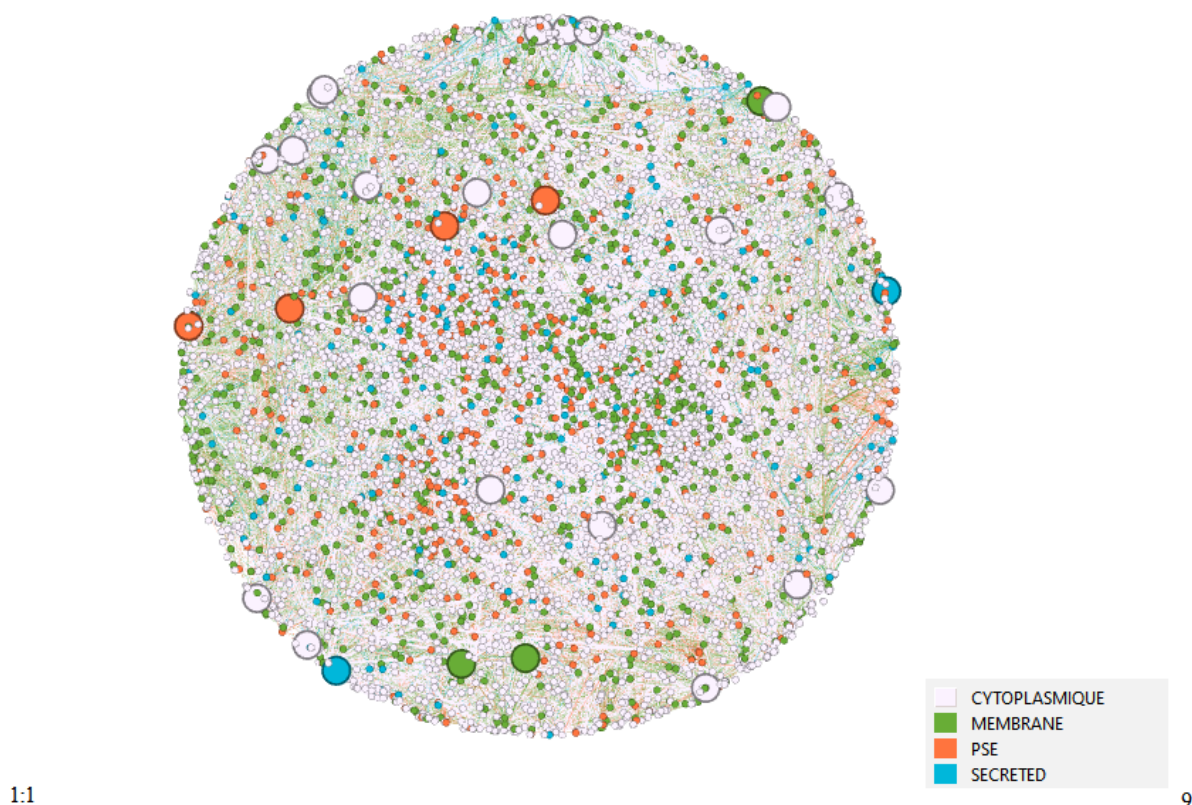


Figura 9-observação da rede de interação da *Bacillus cereus* CP174498, com nós indicando proteínas e arestas indicando interações. Nós destacados com tamanhos maiores indicando nós com maiores valores de Bridge Centrality (BriCe) e cores indicando locais subcelular: citoplasma, membrana, potencialmente exposta na superfície e secretada. Fonte: elaboração própria.

4.1. Proteínas das vias estudadas

Enzimas da fermentação alcoólica e acética não estão totalmente presentes na via. A fermentação alcoólica demanda a presença da álcool desidrogenase e acetaldeído desidrogenase. Não há no genoma anotado da *A. jinshanensis* CP171343 a acetaldeído desidrogenase e o transportador do ácido acético. A bactéria seria incapaz de converter acetaldeído a ácido acético sem essa enzima. Já para a fermentação alcoólica, a bactéria não possui a piruvato descarboxilase no seu genoma, o que impossibilita a conversão de piruvato a etanol. A *A. jinshanensis* CP171343 não faz fermentação acética, nem alcoólica. Pelos resultados dessas análises, acredita-se que a contribuição da bactéria na fermentação do vinagre chinês não se dá pela produção de ácido acético.

O BLAST com as enzimas das principais vias bioenergéticas demonstrou que a *A. jinshanensis* CP171343 possui a maior parte das enzimas da fermentação heterolática, da via Embden-Meyerhof (EMP) e não possui enzimas importantes da via ED (Entner-Doudoroff), tais como 6-fosfogluconato desidratase e 2-deshidro-3-deoxi-fosfogluconato aldolase.

Tabela 2-Enzimas das vias de fermentação alcoólica e acética

VIAS	ENZIMAS	PROCESSO	STATUS
FERMENTAÇÃO ALCOÓLICA	Piruvato descarboxilase	Piruvato→Etanol	Ausente
	Álcool desidrogenase	Etanol→Acetaldeído	Presente
FERMENTAÇÃO ACÉTICA	Acetaldeído desidrogenase	Acetaldeído→Ácido acético	Ausente

Fonte: elaboração própria

A ausência de enzimas chaves possibilita uma hipótese de que não há a realização de fermentação acética ou alcoólica como metabolismo principal. Elimina-se uma questão inicial desse trabalho que era compreender o papel da bactéria na fermentação do vinagre chinês. Embora ausente do metabolismo principal, a base conjugada do ácido acético (o acetato) é um dos produtos na fermentação heterolática. O BLAST com essas enzimas demonstrou que estão presentes no genoma, conforme mostrado na tabela abaixo.

Tabela 3- Enzimas da via da fermentação heterolática (via da fosfocetolase)

VIA	ENZIMA	PROCESSO	STATUS
	Hexoquinase	Glicose→Glicose-6-fosfato	Presente

FERMENTAÇÃO HETEROLÁTICA	Glicose-6-fosfato desidrogenase	Glicose-6-fosfato→6-fosfogluconolactona	Presente
	6-fosfogluconolactonase	6-fosfogluconolactona→6-fosfogluconato	Presente
	6-fosfogluconato desidrogenase	6-fosfogluconato→Ribulose-5-fosfato	Presente
	ribulose-5-fosfato epimerase	Ribulose-5-fosfato→Xilulose-5-fosfato	Presente
	Pentose fosfocetolase	Xilulose-5-fosfato→Gliceraldeído-3-fosfato e acetil-fosfato	Presente
	Lactato desidrogenase	Piruvato→Lactato	Presente
	Fosfotransacetilase	(I)Acetil-fosfato→acetaldeído	Presente
	Aldeído desidrogenase	(II)Acetil-fosfato→acetaldeído	Presente
	Álcool desidrogenase	Acetaldeído→Etanol	Presente
	Acetil fosfato quinase	Acetil-fosfato→Acetato	Ausente

Fonte: elaboração própria

Esses resultados indicam que a *A. jinshanensis* CP171343 realiza a fermentação heterolática (via da fosfocetolase), convertendo glicose a etanol, lactato e acetato. A via EMP é parcialmente parte da via da fosfocetolase, conforme demonstrado pela representação da via.

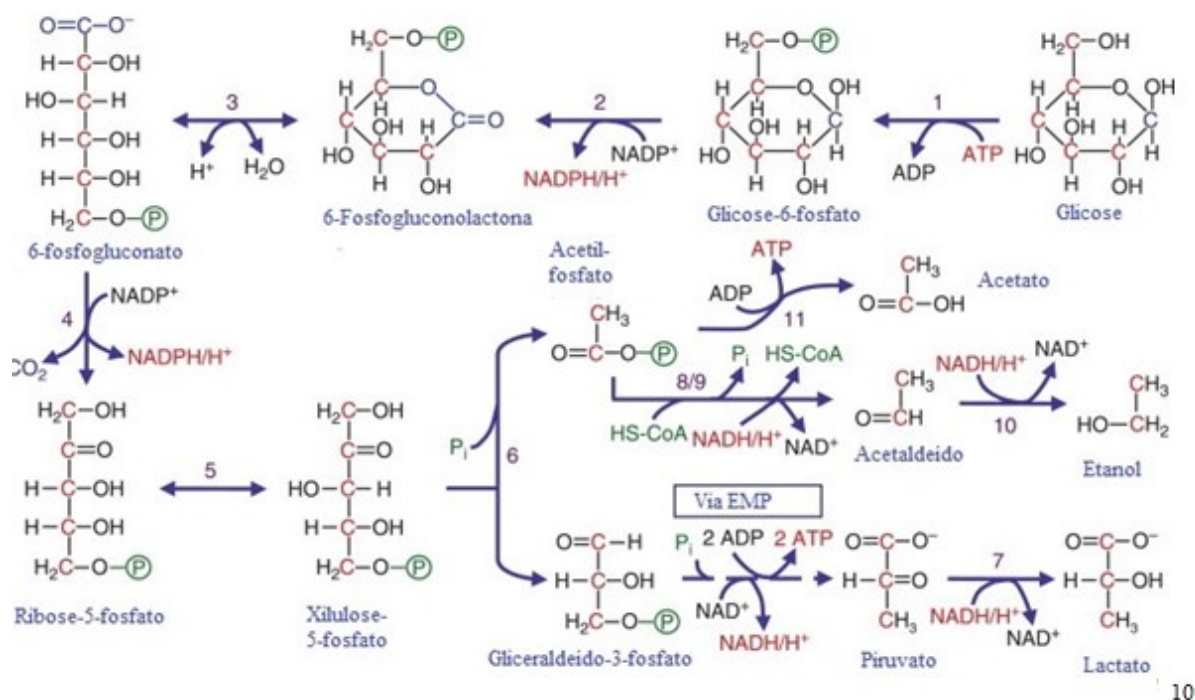


Figura 10- via da fosfoctolase. Enzima pentose fosfoctolase (6) forma gliceraldeído-3-fosfato que entra na via EMP. Adaptado de (Spector, 2009)

Tabela 4-enzimas da via EMP.

VIA	ENZIMA	PROCESSO	STATUS
VIA EMBDEN- MEYERHOF- PARNAS	Hexoquinase	Glicose→Glicose-6-fosfato	Presente
	Fosfoglicoisomerase	Glicose-6-fosfato→Frutose-6-fosfato	Presente
	Fosfofrutoquinase*	Frutose-6-fosfato→Frutose-1,6-bifosfato	Ausente
	Frutose bifosfato aldolase*	Frutose-1,6-bifosfato→Dihidroxiacetona-fosfato e Gliceraldeído-3-fosfato	Ausente
	Triose fosfato isomerase	Dihidroxiacetona-fosfato→Gliceraldeído-3-fosfato	Presente
	Gliceraldeído-3-fosfato desidrogenase**	Gliceraldeído-3-fosfato→1,3-bisfosfoglicerato**	Presente

	Fosfogliceroquinase	1,3-bifosfoglicerato→3-fosfoglicerato	Presente
	Fosfoglicerato mutase	3-fosfoglicerato→2-fosfoglicerato	Presente
	Enolase	2-fosfoglicerato→fosfoenolpiruvato	Presente
	Piruvato quinase	Fosfoenolpiruvato→Piruvato	Presente

*Enzimas ausentes fortalecem a hipótese de inexistência da via. **Etapa em que ocorre a entrada do gliceraldeído-3-fosfato, que é produto da reação indicada pela enzima 6 na via da fosfoctolase. Fonte: elaboração própria

Os achados apontam para duas ausências importantes: não haveria frutose bifosfato aldolase e fosfofrutoquinase. Essas duas enzimas possuem papéis cruciais na via EMP, sem as quais não ocorre a continuidade das reações. Diante dessas ausências, é conclusivo dizer que a via EMP não ocorre integralmente na bactéria como rota bioenergética. O BLAST também resultou na presença de uma parte das enzimas da via EMP. Como demonstrado na tabela 3, todas enzimas que conduzem o gliceraldeído-3-fosfato até piruvato estão presentes no genoma. Esse caminho até piruvato ocorre na via da fosfoctolase, conforme indicado na figura 8. Esses dados permitem afirmar que a via EMP ocorre parcialmente, convertendo gliceraldeído-3-fosfato produzido na via da fosfoctolase a piruvato.

A última via escrutinada foi a Entner-Doudoroff (ED).

Tabela 5-Enzimas da via ED

VIA	ENZIMA	PROCESSO	STATUS
-----	--------	----------	--------

VIA ENTNER-DOUDOROF	Hexoquinase	Glicose→Glicose-6-fosfato	Presente
	Glicose-6-fosfato desidrogenase	Glicose-6-fosfato→6-fosfogluconolactona	Presente
VIA ENTNER-DOUDOROF	6-fosfogluconolactonase	6-fosfogluconolactona→6-fosfogluconato	Presente
	6-fosfogluconato deshidratase*	6-fosfogluconato→2-ceto-3-deoxi-6-fosfogluconato	Ausente
	2-ceto-3-deoxi-6-fosfogluconato	2-ceto-3-deoxi-6-fosfogluconato→piruvato e gliceraldeído-3-fosfato	Ausente
	aldolase*		

*Ausência impede a via de ocorrer. Fonte: elaboração própria.

O BLAST dessa via apontou duas ausências importantes, sem as quais a via não ocorre. A conversão de 6-fosfogluconato a 2-ceto-3-deoxi-6-fosfogluconato (KDPG) e 2-ceto-3-deoxi-6-fosfogluconato a piruvato e gliceraldeído-3-fosfato impediriam a ocorrência dessa via. Se trata de porções cruciais e incontornáveis da via.

Diante desses achados, conclui-se pela hipótese de que a *A. jinshanensis* CP171343 não fermenta a etanol, nem a ácido acético, mas produz etanol, lactato e acetato por via da fosfocetolase. Além disso, possui um metabolismo duplo da glicose, pela via da fosfocetolase e da via EMP.

Conforme mencionado, o estudo de um interactoma demanda a seleção de proteínas específicas que servem como sementes (seeds). As enzimas das vias bioenergéticas que foram encontradas no BLAST serão as seeds dos estudos de interações posterior. Trata-se de uma

filtragem que foi exitosamente empregada neste trabalho. Para fins de apresentação, aqui será empregado o nome da proteína e não o seu nome anotado no genoma, já que ocorreram hits com porções das proteínas, o que resultou, eventualmente, em uma proteína obter mais de um hit no genoma. Quando for apresentado o resultado das interações, a Álcool desidrogenase, por exemplo, será o HIT que se obteve no BLAST. Nesse caso, os três trechos que foram resultados aceitáveis do BLAST da enzima Álcool desidrogenase receberá a alcunha de Álcool desidrogenase para fins de melhor apresentação. As seeds selecionadas nessa etapa são: álcool desidrogenase, hexoquinase, glicose-6-fosfato desidrogenase, 6-fosfogluconolactonase, 6-fosfogluconato desidrogenase, ribulose-5-fosfato epimerase, pentose fosfocetolase, lactato desidrogenase, fosfotransacetilase, aldeído desidrogenase, álcool desidrogenase, fosfoglicoisomerase, triose fosfato isomerase, gliceraldeído-3-fosfato desidrogenase, fosfogliceroquinase, fosfoglicerato mutase, enolase, piruvato quinase.

4.1.1. BLAST como ferramenta de filtragem e direcionamento

A escolha por explorar os processos fermentativos potencialmente realizados pela *A. jinshanensis* CP171343 no vinagre chinês iniciou com a hipótese de que o organismo potencialmente realizaria fermentação acética. Adotar o BLAST como etapa inicial foi relevante já que endereçou as análises futuras, filtrando o que poderia ser analisado. No caso da *A. jinshanensis* CP171343, descobriu-se que não realizava fermentação acética, contrariando as hipóteses traçadas inicialmente. No caso da *B. cereus* CP174498, o BLAST foi capaz de indicar a presença de genes que potencialmente indicam a realização de fermentação alcoólica. Esse trabalho constatou que um ponto de partida crucial e viável para início das análises é empregar o BLAST para verificar a presença de genes estudados. Aqui foi o caso da fermentação acética, porém acredita-se que pode ser expandido para outros processos biológicos estudados. Servirá sempre ao propósito de explorar unicamente as vias que se sabe estar presentes e expande os conhecimentos a respeito dessas vias.

4.1.2. Estudo direcionado pelas funções biológicas conhecidas

A escolha por estudar as interações a partir de uma ou mais vias metabólicas se demonstrou uma etapa incontornável do processo. A rede de interação sempre será humanamente impossível de ser analisada em sua totalidade. Ao propor que seja escolhida uma via, notou-se que novos conhecimentos podem ser obtidos, expandindo o que se sabe sobre aquele processo metabólico e biológico. Baranwal et al. (2020) apontou essa complexidade de interações das vias metabólicas e como o estudo individual dos componentes das vias pode resultar em valiosas informações sobre o processo como um todo. Dessa maneira, propor que se estude o conjunto das interações utilizando uma via metabólica como ponto inicial é coerente.

Nesse trabalho, surpreendeu inicialmente que uma bactéria isolada no vinagre chinês e descrita como partícipe do processo não tenha um papel direto na fermentação acética em si. O inesperado abriu campos para novas buscas que, conforme ficou demonstrado, mostraram significantes novas descobertas do microrganismo e abre portas para estudos de bancada complementares os quais poderão ser promissores. Escolher um ponto de partida para estudo pode soar um processo arbitrário. No entanto, é plausível que esse direcionamento para estudo seja estratégico para evitar um estudo extenso e humanamente impossível do genoma. Também é razoável que a escolha de uma via metabólica para estudo possa partir de um ponto em que se possua mais conhecimentos experimentais. No caso da *A. jinshanensis* CP171343, estudar os processos fermentativos era partir de um experimento de bancada que demonstrou haver a fermentação heterolática. Sendo assim, a análise fica direcionada, com menor risco de arbitrariedades.

Compreender as vias bioenergéticas que os organismos realizam é benéfico a uma série de propósitos. Para um processo fermentativo que empregue esse organismo, conhecer as vias metabólicas fermentativas é essencial (Liu, 2020). Da mesma maneira, o cultivo laboratorial demanda o conhecimento das vias bioenergéticas para melhor aproveitamento (Adamberg,

Valgepea e Vilu, 2015). O resultado de BLAST com os genomas da *A. jinshanensis* CP171343 fornecem uma base minimamente sólida para afirmar que a bactéria participa da fermentação heterolática. É corroborado pelas descobertas laboratoriais, que demonstraram a presença da via no microrganismo (Yu et al., 2020b).

Xu et al. (2024) quantificou a abundância de espécies na fermentação do vinagre de Zhenjiang e apontou que a *Acetilactobacillus jinshanensis* começa a se destacar no 11º dia da fermentação até ser predominante no final do processo. (Xu et al., 2011a) demonstraram que a produção de ácido acético se estabiliza entre os dias 10º e 15º. O autor também indica que quantidade do ácido é até consumida no intervalo do 15º dia até o 20º dia. Esse achado reforça o resultado encontrado via BLAST de que a bactéria não realiza fermentação acética. Xu et al.(2011) também indicaram que o ácido láctico atinge seu pico no 5º de fermentação e começa a ser consumido progressivamente, atingindo seu valor mínimo no 20º dia. Esse dado contrasta com a descoberta de que a *A. jinhanensis* potencialmente realize fermentação heterolática, produzindo produz quantidades equimolares de etanol, lactato e dióxido de carbono com ocasionais traços de acetato (“MetaCyc heterolactic fermentation”, 2025). Dado a prevalência da bactéria na fase final da fermentação, não é coerente afirmar que a contribuição que a bactéria possui no processo fermentativo seja produzindo ácido láctico.

A via EMP não ocorre integralmente na *A. jinshanensis* CP171343, conforme indica o BLAST feito no genoma montado da bactéria e no genoma da referência depositado no NCBI. A ausência de duas enzimas importantes levou a essa conclusão definitiva. Logo, a fermentação heterolática é o mecanismo bioenergético predominante. A literatura demonstra que o intermediário chave da via EMP é o gliceraldeído-3-fosfato (Spector, 2009). O produto é produzido em uma etapa da via da fosfocetolase e continua na via EMP até ser produzido piruvato. É relevante que se tenha encontrado no genoma as enzimas que são responsáveis por conduzir o gliceraldeído-3-fosfato na via. Não se trata de uma informação contraditória, já que

apenas esse intermediário entra na via. Esse achado reforça que as duas vias ocorrem simultaneamente, mas não integralmente. Årsköld et al. (2008) descreveu que processo similar ocorre com a *Lactobacillus reuteri*, bactéria taxonomicamente próxima. Foi encontrado que a via do fosfoacetolase é dominante no organismo, o qual contém uma via dual para glicólise. Esse contraste comprova a qualidade da montagem genômica e a capacidade da análise de interações, conforme será mais discutido a frente.

4.2. Filtragem por BriCe

Essa etapa consistiu na busca pelas proteínas com maiores valores de BriCe da rede da *A. jinshanensis* CP171343. O BriCe se fundamenta na teoria dos grafos (Pavlopoulos et al., 2011). Pressupõe, dentre outras coisas, que alguns parâmetros são relevantes para análise de uma rede e capazes de fornecer dados com os quais se pode levantar hipóteses. Entre esses parâmetros estão as medidas de centralidade, que indicam o quanto nós de uma rede servem como elementos conectores entre conjuntos de nós. O BriCe mensura essa capacidade dos nós servirem como pontes. O plug-in do próprio Gephi foi utilizado. Obteve-se os 30 nós com maiores valores de BriCe. Essa etapa consistiu na segunda filtragem de nós. Essas proteínas com maiores valores de BriCe foram buscadas no NCBI e submetidas ao InterProScan para que seus domínios conservados fossem obtidos para análises individuais.

Tabela 6-proteínas com maior Bridging Centrality

Código da proteína (NCBI)	Domínio conservado	Interação com enzimas estudadas
QBP19020.1	Aminotransferase com dobradura de classe I/II e piridoxal [Acetilactobacillus jinshanensis]	Álcool desidrogenase, Lactato Desidrogenase
QBP18937.1	Proteína ligadora de ATP de transportador ABC de aminoácidos [Acetilactobacillus jinshanensis]	Lactato desidrogenase, glicose-6-fosfato desidrogenase, 6-fosfogluconato desidrogenase

QBP18702.1	Racemase de aminoácido [Acetilactobacillus jinshanensis]	Lactato desidrogenase
QBP18828.1	Tioesterase de acil-ACP [Acetilactobacillus jinshanensis]	Lactato desidrogenase
QBP18028.1	Aminotransferase com dobradura de classe I/II [Acetilactobacillus jinshanensis]	Lactato desidrogenase
QBP18418.1*	D-2-hidroxiácido [Acetilactobacillus jinshanensis] -blast lactatodesidrogenase	Não há
QBP17725.1	Glicosiltransferase [Acetilactobacillus jinshanensis]	Lactato desidrogenase Fosfoglicerato mutase
QBP18992.1**	Proteína contendo domínio DUF2187 (domínio de função desconhecida 2187) [Acetilactobacillus jinshanensis]	Não há
QBP19031.1***	NrnA proteína bifuncional com atividade oligoribonuclease e fosfatase de PAP (fosfoadenosina 5'-difosfato) [Acetilactobacillus jinshanensis]	Triose fosfato isomerase
QBP18177.1**	Proteína contendo domínio FRG [Acetilactobacillus jinshanensis]	Lactato desidrogenase
QBP17961.1	TruB, sintetase de pseudouridina(55) de tRNA [Acetilactobacillus jinshanensis]	6-fosfogluconolactonase Glicose-6-fosfato desidrogenase
QBP18782.1**	Proteína hipotética ELX58_06650 [Acetilactobacillus jinshanensis]	Não há
QBP17770.1**	Proteína hipotética ELX58_00965 [Acetilactobacillus jinshanensis]	Fosfoglicerato mutase, glicose-6-fosfato isomerase, glicose-6-fosfato desidrogenase, 6-fosfogluconolactonase
QBP18617.1	PlsX, fosfato aciltransferase [Acetilactobacillus jinshanensis]	6-fosfogluconato desidrogenase, ribulose-5-fosfato epimerase
QBP17928.1***	Proteína ribossômica 30S S20 [Acetilactobacillus jinshanensis]	Não há

QBP19011.1***	Endonuclease de restrição [Acetilactobacillus jinshanensis]	Alcool desidrogenase, 6-fosfoglucono lactonase, hexoquinase, lactonase, glicose-6-fosfato isomerase, fosfoglicerato mutase, enolase
QBP18171.1**	Proteína hipotética ELX58_03250 [Acetilactobacillus jinshanensis]	Não há
QBP17773.1**	Proteína hipotética ELX58_00980 [Acetilactobacillus jinshanensis]	6-fosfoglucono lactonase
QBP18768.1**	DUF2969 (domínio de função desconhecida 2969) [Acetilactobacillus jinshanensis]	Não há
QBP19045.1**	Proteína hipotética ELX58_07310 [Acetilactobacillus jinshanensis]	Triose fosfato isomerase
QBP18189.1	SDR family oxidoreductase [Acetilactobacillus jinshanensis]	Lactato desidrogenase
QBP18130.1**	Proteína hipotética ELX58_07310 [Acetilactobacillus jinshanensis]	Não há
QBP18340.1**	Proteína hipotética ELX58_04140 [Acetilactobacillus jinshanensis]	Não há
QBP18481.1**	Proteína hipotética ELX58_04880 [Acetilactobacillus jinshanensis]	Não há
QBP18016.1	Proteína da família YozE [Acetilactobacillus jinshanensis]	Não há
QBP19016.1	Proteína contendo domínio redox [Acetilactobacillus jinshanensis]	Álcool desidrogenase, 6-fosfoglucono-lactonase, hexoquinase, glicose-6-fosfato isomerase, fosfoglicerato mutase, enolase
QBP17915.1**	Proteína da família DUF1447 (domínio de função desconhecida 1447) [Acetilactobacillus jinshanensis]	Não há

QBP18473.1**	Proteína hipotética ELX58_04840 [Acetilactobacillus jinshanensis]	Não há
QBP19021.1	Hipoxantina fosforribosiltransferase [Acetilactobacillus jinshanensis]	Álcool desidrogenase, glicose-6-fosfato desidrogenase, hexoquinase, lactonase, glicose-6-fosfato isomerase, fosfoglicerato mutase, enolase

*Proteína parte de via estudada nesse projeto. **Proteínas sem funções descritas.***Proteínas com funções muito distintas do estudado nesse projeto. Fonte: elaboração própria

Endonucleases e proteínas ribossômicas, por exemplo, dificilmente possuem relação de funções diretas com os processos bioquímicos estudados aqui, dados os diferentes propósitos e produtos. Há outras proteínas com função desconhecida ou vagamente descrita, com um grau de incerteza muito alto para alegar função. No entanto, a proteína com domínio da lactato desidrogenase está nos achados com maiores valores de BriCe (QBP18418.1).

A Lactato Desidrogenase também interage com as proteínas de maior BriCe. As interações com proteínas relacionadas com transporte de aminoácidos, racemases e oxirredutases são indicadas e podem conferir algum significado biológico. Importante destacar que a Lactato Desidrogenase converte piruvato a lactato na reação da fosfoctolase. O piruvato aqui é produto da via EMP, o que corrobora com o papel de ponte que a enzima possui ao ligar processos biológicos distintos.

Uma endonuclease de restrição (QBP19011.1) aparece com alto BriCe e interagem com muitas proteínas. Não parece haver alguma relação biológica conjecturável dessas interações dadas as funções biológicas distintas que desempenham.

A QBP19016.1, é descrita como sendo uma proteína com função de oxirredução. Aparece interagindo com diversas enzimas estudadas, sendo algumas com funções de oxirredução. Um indicativo de potencial papel complementar e que faz sentido biológico.

A Hipoxantina fosforribosiltransferase (QBP19021.1) possui função de recuperação de bases nitrogenadas, sendo funções na manutenção do material genético. Também aparece realizando uma série de interações aparentemente incoerentes com os processos estudados.

Esses resultados servem ao propósito de alimentar as hipóteses para trabalhos futuros.

4.2.1. Uma segunda filtragem: coeficientes de centralidade

A análise de Bridging Centrality demonstrou que apenas uma proteína das 30 com maior BriCe está relacionada com os processos estudados nesse trabalho. Estabelecer hipóteses com as outras proteínas não relacionadas ao processo é arriscado, sendo necessário maiores análises e estudos conjuntos, os quais devem empregar análises de outros processos metabólicos.

Empregou-se o BriCe como um passo na metodologia baseando-se na teoria dos grafos. Pavlopoulos et al. (2011) propôs o uso de coeficientes de centralidade como parâmetros para estudar as redes PPI. Esses coeficientes podem ser traduzidos para termos biológicos com relativa facilidade.

A lactato desidrogenase aparece como uma das proteínas com alto valor de BriCe, embora o seu estudo de interações não tenha resultado em novas informações sobre o organismo. No entanto, a enzima é participante da via da fosfocetolase e usa o produto da via EMP para o converter a lactato, isto é, a enzima integra a via EMP na via da fosfocetolase.

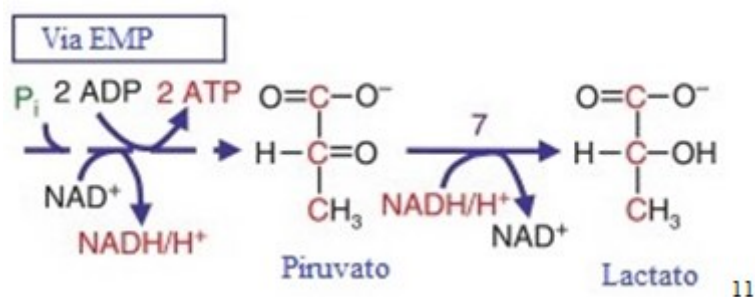


Figura 11 -reação feita pela lactato desidrogenase (7), convertendo piruvato -um produto da via EMP- em lactato.

Esse achado demonstra que a análise de proteínas pontes podem ser de grande relevância para o estudo do interactoma. Aqui, foi possível observar que o software previu corretamente uma proteína que é ponte entre dois processos metabólicos relevantes para a sobrevivência da bactéria. É razoável acreditar que se o estudo desse projeto fosse expandido para que se estudasse outros processos biológicos, seria possível encontrar mais proteínas que sirvam como pontes. Há um grande potencial de aplicações dessa etapa em novos estudos.

Nesse sentido, depositar a informação de que essas proteínas da tabela 5 são pontes no genoma da bactéria é uma informação que poderá ser base de novas descobertas. Os novos experimentos de bancada que porventura demonstrem processos biológicos relacionados com essas enzimas, potencialmente poderão ser complementados com essas informações. Há a possibilidade de novas hipóteses serem testada em bancada ou *in silico* tendo como ponto de partida as proteínas pontes. Nesse sentido, há sempre uma grande contribuição da ciência da base em apontar novos caminhos e buscar teorizar sobre as potencialidades, se limitando ao que os dados concretos podem oferecer no momento. Embora possa soar um conhecimento limitado, é importante apontar que escolher o ponto de partida para análise de interações é uma etapa dispendiosa, que pode ser encurtada com informações sobre as pontes do genoma.

Cabe apontar nessa discussão que os bancos de dados como STRING (Szkarczyk *et al.*, 2023) armazenam diversas interações por vezes pouco esclarecidas. Esse banco de dados é consultado em pesquisas específicas, sendo uma importante ferramenta complementar para estudo. Levantar hipóteses sobre processos biológicos utilizando apenas interações preditas *in silico* pode ser um caminho espinhoso e bastante arriscado. Apontar as possibilidades, no entanto, proporciona a redução de tempo de pesquisa para trabalhos futuros, sobretudo os trabalhos de bancada. Esses dados sobre as interações e as “proteínas pontes” do genoma, embora inconclusivas, possibilita que trabalhos futuros encontrem pistas de quais processos serão mais promissores para elucidar as vias bioquímicas pouco esclarecidas. Os dados obtidos

aqui permitem apontar a lactato desidrogenase como um potencial enzima que integra vias do genoma. Por fim, permite também dizer que essa etapa da metodologia que esse trabalho propõe é coerente com os dados biológicos disponíveis até o momento, sendo bastante razoável seu emprego em novos trabalhos futuros. O fato de a maioria das descobertas serem confirmadas por dados experimentais demonstra a eficiência em predizer interações e processos bioquímicos relevantes.

4.3. Estudo das interações individuais

Foram selecionadas todas as interações realizadas por todas as enzimas dos processos estudados que foram localizadas no genoma. As análises indicam interações entre as enzimas dos processos estudados.

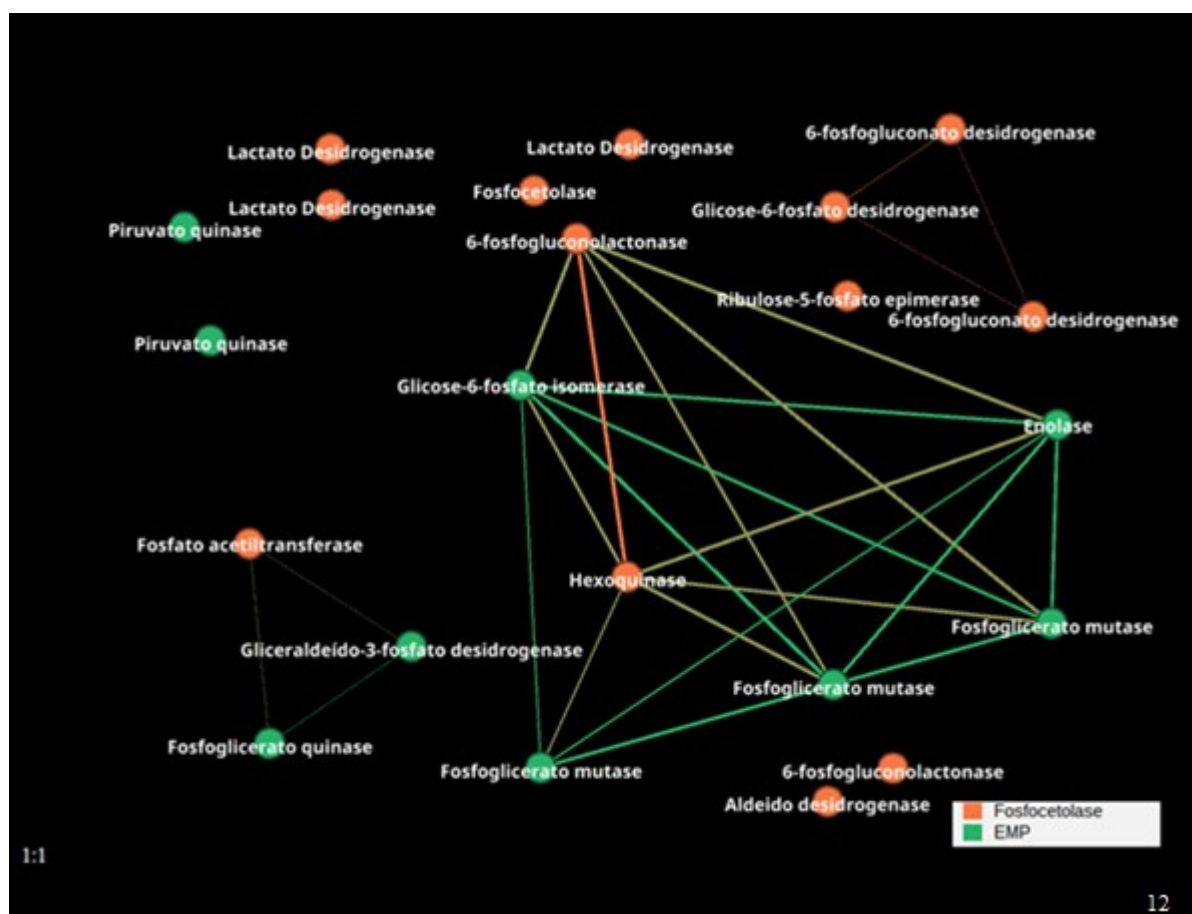


Figura 12- rede de interações observadas no Gephi com proteínas estudadas. Os nós coloridos indicam as proteínas e suas vias, enquanto as arestas indicam as interações. Nós com tarjas repetidas indicam que o BLAST encontrou mais de um trecho da sequência como resultado. Fonte: elaboração própria.

As interações demonstram que as enzimas das vias fosfoglicolase e EMP interagem tanto entre si, quanto de forma cruzada. O clado fosfoglicerato mutase, enolase, hexoquinase, glicose-6-fosfato-isomerase e 6-fosfogluconolactonase apontam uma interconexão de enzimas que são participantes da via da fosfoglicolase e da EMP, ambas potencialmente presentes nessa bactéria. O conjunto de interações que unem esse clado reforça a hipótese dessas vias estarem presentes. Menciona-se também as ligações entre os nós de fosfoglicerato mutase e 6-fosfogluconato desidrogenase. Os nós com a mesma tarja indicam que foram encontrados mais de um hit com o BLAST. A interação entre esses nós de mesma tarja indica a possível formação de oligômeros em um possível complexo proteico que constitui essas enzimas.

4.3.1. Interactoma local confere respaldo a análises biológicas

A análise das interações demonstrou a capacidade de confirmar hipóteses sobre os processos biológicos. O BLAST indica se há as enzimas dos processos estudados, enquanto a análise do interactoma local reforça ou refuta as hipóteses iniciais. Os achados desse trabalho, conforme mostrado no tópico 4.3 e na figura 12, demonstram que as enzimas da via EMP estão não somente presentes, como também realizarem interações com as enzimas da via heterolática. Essas interações reforçam um papel conjunto dentro dos processos biológicos, o que é coerente com a via heterolática, conforme a figura 10, e com os trabalhos pretéritos de espécie taxonomicamente próxima. Årsköld et al. (2008) descreveu que processo similar ocorre com a *Lactobacillus reuteri*. Demonstrou-se a eficácia do modelo de predição de redes em resultar dados biológicos coerentes de um genoma recém montado.

Neste trabalho, resta poucas dúvidas sobre a dupla via de oxidação da glicose feita pela bactéria, uma vez que tanto os genes envolvidos estão presentes, quanto as interações de proteínas entre as duas vias foram observadas. Esse método de estudo das interações possui grande relevância para outros organismos, já que é capaz de apontar interações com significados biológicos. Acredita-se que essa etapa possa ser mais proveitosa em processos biológicos para

os quais existam gaps de entendimento sobre o tema. Nesses cenários, a análise individual das proteínas da rede pode conferir novas pistas.

4.4. Estudo das vias metabólicas principais

As principais vias realizadas pelas proteínas do genoma foram obtidas. Aqui cabe mencionar que esse resultado indica quais vias as proteínas do genoma estão potencialmente envolvidas, conforme o que está anotado na ontologia do gene. Isso significa que uma proteína pode estar envolvida em mais de uma via e que haver muitas proteínas envolvidas com a mesma via indica que potencialmente aquela via está presente. Desse modo, os dados foram tratados para ranquear as 30 principais ocorrências, isto é, as 30 vias que possuem maiores números de proteínas do genoma presentes.

Sendo assim, obteve-se que as mais prevalentes vias são de produção de antibiótico, tais como higromicina B, fosfomicina, lincomicina A, precursor de tetraciclina, mupirocina; vias de produção de compostos anticâncer, tais como salinosporamide A, mitramicina, bicaverina; vias de produção de antiprotozoário, tal como apicidina; e vias de produção de antivirais, tais como echinomicina e triostina.

As principais vias foram classificadas conforme sua importância biotecnológica e apresentadas na figura 13.

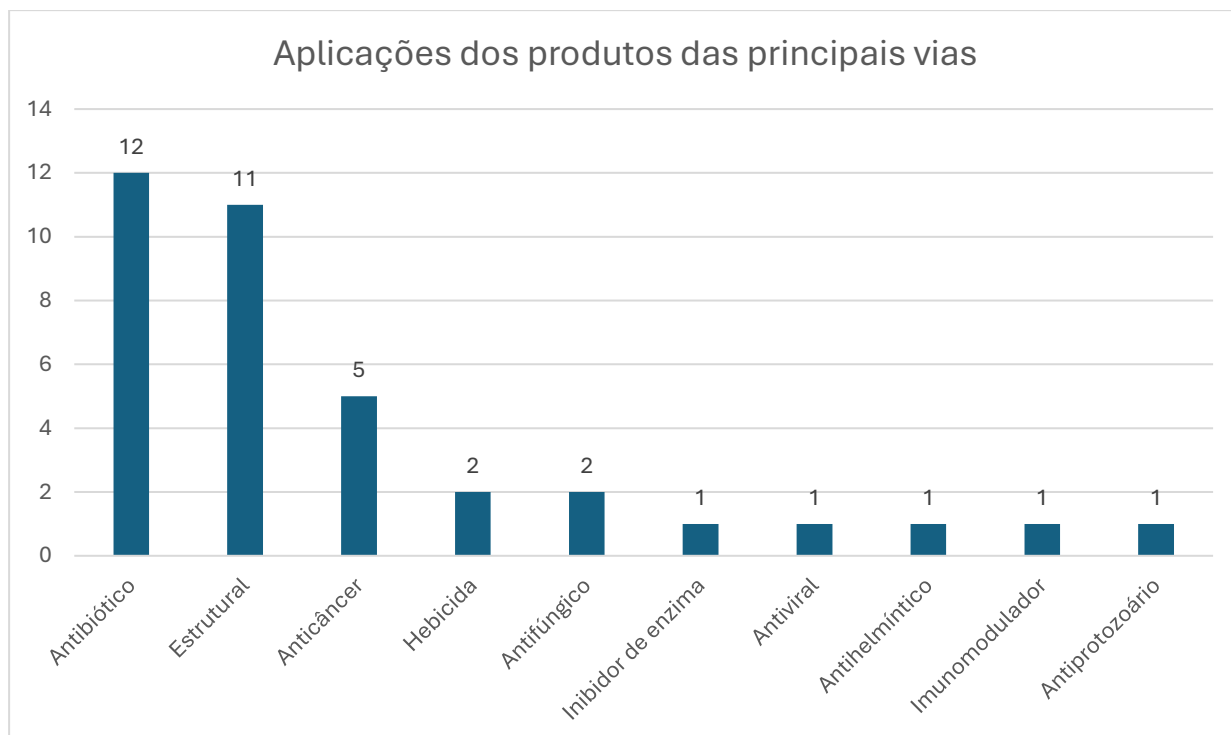


Figura 13-Gráfico demonstrando as ocorrências de funções das 30 principais vias do genoma. Fonte: elaboração própria.

Esses resultados indicam uma prevalência de vias que produzem antibióticos. Vias estruturais se referem a vias cujas importâncias biotecnológicas não estão bem esclarecidas, mas que compõe estruturas como parede celular e processos celulares relevantes.

Esses dados indicam um grande potencial biotecnológico a ser explorado, seja em pesquisas com novas redes, seja em bancada.

Foram também elaboradas as principais vias de clusters do genoma. Inicialmente, obteve-se quatro clusters por meio de plug-in do Gephi. As dez principais vias de cada um dos quatro clusters foram obtidas para análise. Com raras exceções, as vias prevalentes nas análises de clusters não se mostraram distintas da que fora encontrada nas principais vias do genoma. Nem nos clusters, nem no genoma completo foram observadas vias que estivessem relacionadas diretamente com os processos bioenergéticos estudados inicialmente. Isso em parte se justifica pelo número de proteínas necessárias para ocupar as primeiras posições de principais vias. Uma via pode ser prevalente na quantidade de proteínas por ser uma via com

diversas etapas e processos anexos que a inviabilize. Portanto, a análise de vias serve mais ao propósito de observar vias que potencialmente ocorrem no genoma.

4.4.1. Análise de vias e as funções bioecológicas no alimento larval de abelha

A análise das vias levantou diversas hipóteses sobre o papel da bactéria no alimento larval de abelhas, local em que foi colhida a amostra contendo o genoma da *A. jinshanensis* CP171343. Corroborado por outras pesquisas (Mohammad, Mahmud-Ab-Rashid e Zawawi, 2020; Santos, 2022; Santos *et al.*, 2022; Vásquez *et al.*, 2012), o alimento larval de abelhas já foi associado com propriedades antimicrobianas, anti-câncer, antioxidante. O fato da *Acetilactobacillus jinshanensis* CP171343 ter sido isolada no alimento larval de abelhas e possuir, entre suas vias principais, caminhos metabólicos de produção de produtos com essas propriedades potencialmente indica a possibilidade do papel ecológico da bactéria, bem como seu potencial biotecnológico. O estudo de cada via renderia novos trabalhos para se verificar sua real presença no organismo. No entanto, o achado é respaldado com o que está descrito na literatura. A bactéria potencialmente contribui para a produção de um alimento larval com propriedades terapêuticas e com papéis bioecológicos relevantes no sistema colmeia de abelhas sem ferrão

4.5. Genes de proteínas relevantes presentes na no genoma da *Bacillus cereus* CP174498

A análise do genoma montado da *Bacillus cereus* CP174498 via BLAST indicou a presença de genes da fermentação alcoólica (piruvato decarboxilase) e ausências dos genes de fermentação acética (acetaldeído desidrogenase). Essa fermentação está presente no processo de produção do vinagre chinês, constituindo a primeira etapa do processo (Xu *et al.*, 2011b). O vinagre passa por um conjunto de transformações dos seus substratos até ser produzido o produto alimentício. A fermentação alcoólica é uma dessas etapas, que não está presente na *A.*

jinshanensis CP171343, mas aparece na *B. cereus* CP174498. Esses indícios apontam para um papel conjunto dos dois organismos.

Tabela 7- Enzimas de fermentação alcoólica e acética e sua disponibilidade no genoma da *B.cereus* CP174498

VIAS	ENZIMAS	PROCESSO	STATUS
FERMENTAÇÃO ALCOÓLICA	Piruvato descarboxilase	Piruvato→Etanol	Presente
	Álcool desidrogenase	Etanol→Acetaldeído	Presente
FERMENTAÇÃO ACÉTICA	Acetaldeído desidrogenase	Acetaldeído→Ácido acético	Ausente

Fonte: elaboração própria.

A via da fosfocetolase (fermentação heterolática) potencialmente não ocorre na *B. cereus* CP174498. Isso não pode ser levantado como uma hipótese forte dado que a maior parte das enzimas da via está presente. No entanto, falta a fosfocetolase, enzima que dá nome a via. Diante disso, é provável que essa via não ocorra na *B. cereus* CP174498.

Tabela 8-via da fosfocetolase e sua presença na *B. cereus* CP174498.

VIA	ENZIMA	PROCESSO	STATUS
	Hexoquinase	Glicose→Glicose-6-fosfato	Presente
	Glicose-6-fosfato desidrogenase	Glicose-6-fosfato→6-fosfogluconolactona	Presente
	6-fosfogluconolactonase	6-fosfogluconolactona→6-fosfogluconato	Presente
	6-fosfogluconato desidrogenase	6-fosfogluconato→Ribulose-5-fosfato	Presente

FERMENTAÇÃO HETEROLÁTICA	ribulose-5-fosfato	Ribulose-5-	Presente
	epimerase	fosfato→Xilulose-5- fosfato	
	Pentose fosfocetolase	Xilulose-5-	Ausente
	*	fosfato→Gliceraldeído-3- fosfato e acetil-fosfato	
	Lactato desidrogenase	Piruvato→Lactato	Presente
	Fosfotransacetilase	(I)Acetil- fosfato→acetaldeído	Presente
	Aldeído desidrogenase	(II)Acetil- fosfato→acetaldeído	Presente
	Álcool desidrogenase	Acetaldeído→Etanol	Presente
	Acetil fosfato quinase	Acetil-fosfato→Acetato	Presente

*Enzima chave da via. Fonte: elaboração própria.

A via EMP foi investigada na *B. cereus* CP174498 e identificou as enzimas relevantes dessa via como estando presentes no genoma. Diferente do que ocorre com a *A. jinshanensis* CP171343, aqui a via está presente em sua totalidade.

Tabela 9- enzimas da via EMP e sua disponibilidade no genoma da *B. cereus* CP174498.

VIA	ENZIMA	PROCESSO	STATUS
	Hexoquinase	Glicose→Glicose-6-fosfato	Presente
	Fosfoglicoisomerase	Glicose-6-fosfato→Frutose-6- fosfato	Presente
	Fosfofrutoquinase	Frutose-6-fosfato→Frutose-1,6- bifosfato	Presente

VIA EMBDEN– MEYERHOF– PARNAS	Frutose	bifosfato	Frutose-1,6- bifosfato → Dihidroxiacetona- fosfato e Gliceraldeído-3-fosfato	Presente
	Triose	fosfato	Dihidroxiacetona-fosfato → Gliceraldeído-3-fosfato	Presente
	Gliceraldeído-3- fosfato desidrogenase**		Gliceraldeído-3-fosfato → 1,3- bifosfoglicerato**	Presente
	Fosfogliceroquinase		1,3-bifosfoglicerato → 3- fosfoglicerato	Presente
	Fosfoglicerato mutase		3-fosfoglicerato → 2- fosfoglicerato	Presente
	Enolase		2- fosfoglicerato → fosfoenolpiruvato	Presente
	Piruvato quinase		Fosfoenolpiruvato → Piruvato	Presente

Fonte: elaboração própria

Por fim, investigou-se a via ED. Não foram encontradas enzimas relevantes da via, o que indica sua provável inexistência.

Tabela 10- Enzimas da via ED e sua disponibilidade no genoma da *B. cereus* CP174498

VIA	ENZIMA	PROCESSO	STATUS
	Hexoquinase	Glicose → Glicose-6- fosfato	Presente

VIA ENTNER-DOUDOROF	Glicose-6-fosfato desidrogenase	Glicose-6-fosfato→6-fosfogluconolactona	Presente
	6-fosfogluconolactonase	6-fosfogluconolactona→6-fosfogluconato	Presente
	6-fosfogluconato desidratase	6-fosfogluconato→2-ceto-3-deoxi-6-fosfogluconato	Presente
	2-ceto-3-deoxi-6-fosfogluconato aldolase*	2-ceto-3-deoxi-6-fosfogluconato→piruvato e gliceraldeído-3-fosfato	Ausente

*Ausência de enzima chave da via. Fonte: elaboração própria

Portanto, é bem provável que a bactéria participe da fermentação alcoólica, um processo relevante para a fermentação do vinagre de aromático de Zhenjiang. Trata-se da primeira etapa do processo fermentativo para produção do vinagre, o qual possui outras etapas, em que organismos são substituídos naturalmente. Assim, há a etapa de fermentação alcoólica e fermentação acética. Na alcoólica, os organismos dominantes são os capazes de oxidar parcialmente a glicose em etanol. Essa etapa é realizada por diversas espécies que são substituídas com o decorrer do processo por bactérias de fermentação acética (Xu *et al.*, 2024b). A *B. cereus* que sequenciamos, montamos e depositamos no NCBI potencialmente realiza a fermentação alcoólica, de acordo com nossas descobertas das enzimas responsáveis pela fermentação alcoólica no genoma de nossa *B. cereus*. A *A. jinshanensis* CP171343 potencialmente não realiza essa fermentação alcoólica pela ausência do conjunto de enzimas

capazes de executar essa tarefa. No entanto, sua participação no processo fermentativo do vinagre chinês é crucial já que produz compostos como o ácido lático, cuja composição é presente no produto. Já a *B. cereus* CP174498 realiza potencialmente fermentação alcoólica, dada a presença dessas enzimas.

4.4.1. Papel da *Bacillus cereus* no processo fermentativo e sua relação com o alimento larval de abelhas

A bactéria *Bacillus cereus* CP174498 teve sua rede PPI montada e genes dos principais processos bioenergéticos buscados. A *Bacillus cereus* CP174498 está presente no processo fermentativo do vinagre chinês, mas sua abundância é baixa, com a comunidade de *Bacillus* sendo inferior a 1% (Liu *et al.*, 2023). A abundância do gênero *Bacillus* no meio de fermentação altera conforme a estação do ano, o que implica na alteração do sabor e das suas propriedades. É também apontado que seu papel na fermentação é conferir sabor ao vinagre por meio de rotas metabólicas. Um dos compostos que potencialmente é produzido é a acetoína (Liu *et al.*, 2023) e é apontada como composto que confere sabor. Essas rotas de sabores não foram alvos de estudo desse projeto por questão de adequação de tempo. No entanto, trata-se de um promissor campo a ser explorado utilizando a rede montada. Demonstrou-se que a *Bacillus cereus* CP174498, que teve seu genoma montado nesse projeto, possui gene da fermentação alcoólica, o que é corroborado com o fato de que a abundância da bactéria é maior nos primeiros dias de fermentação do mosto alcoólico, quando a fermentação alcoólica ainda está presente, mas cai após alguns dias, com o início de outros processos fermentativos (Liu *et al.*, 2023). A real contribuição da bactéria merece ser estudada posteriormente em trabalho a parte, dado que a abundância é pequena demais para uma contribuição tão relevante na fermentação e essa abundância varia conforme as estações do ano, afetando também o sabor do alimento (Liu *et al.*, 2023).

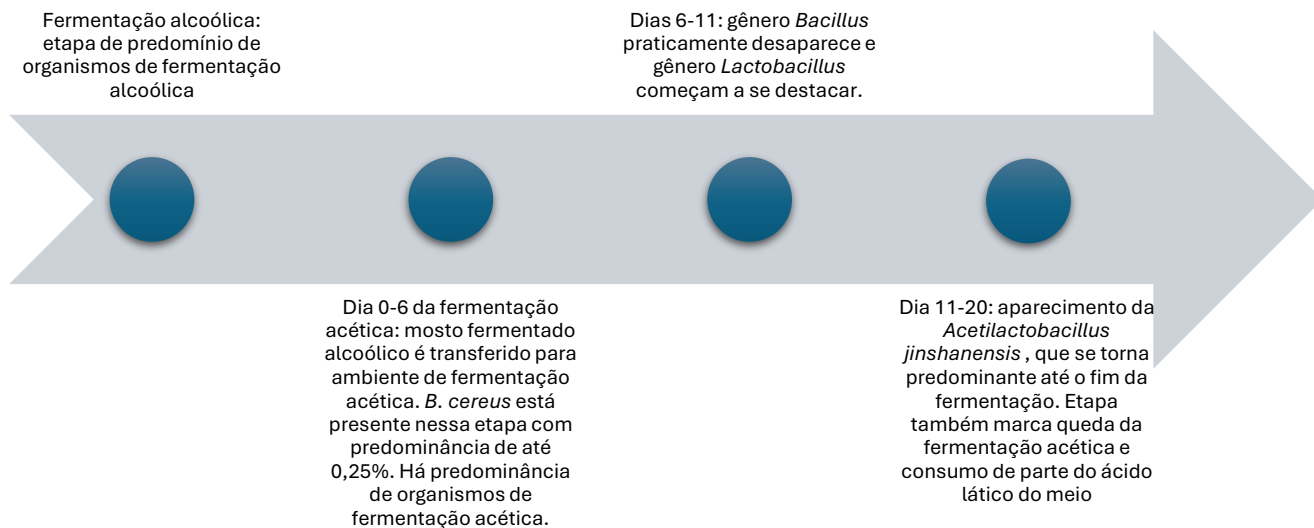


Figura 14-etapas da fermentação do vinagre chinês e seus principais pontos. Adaptado de (Liu et al., 2023; Xu et al., 2024b, 2011b).

Em contraste com a *A. jinshanensis* CP171343, a *B. cereus* CP174498 potencialmente não realiza fermentação heterolática. Sua presença pode ser resíduo da etapa anterior, de fermentação alcoólica, ou mesmo ter atuação nessa fermentação no meio.

Quanto a presença da *B. cereus* no alimento larval, há indícios que papéis bioecológicos relevantes possam ser estudados futuramente. Neste trabalho, o estudo ficou restrito a demonstrar seus processos fermentativos e bioenergéticos, mas é promissor um estudo sobre vias metabólicas e processos bioquímicos que indiquem relações ecológicas relevantes na formação de um alimento larval com bioativos, bem como na manutenção da estabilidade biológica da colmeia.

. 4.5. Proposta de metodologia

O êxito desse trabalho e a falta de metodologias claras para uso de preditores de rede PPI nos fez propor uma metodologia desenvolvida, aplicada e discutida nesse trabalho, conforme a figura 15.



Figura 15-proposta de metodologia para trabalhos futuros

Esses procedimentos podem ser balizadores para trabalhos futuros, de modo que haja um direcionamento mínimo no estudo de redes PPI. Conforme abordado, a escolha pelo “ponto

de partida” do estudo de interações pode soar arbitrário e desafiador. Esse trabalho demonstrou que partir de conhecimentos consolidados (como foi o caso das enzimas do processo fermentativo) pode ser uma alternativa bastante conveniente. Por vezes, resultados pouco compreensíveis podem ser obtidos. Alguns serão meros artefatos do processo de análise, do funcionamento dos algoritmos ou de erros ocorridos em etapas anteriores. No entanto, como se defendeu nesse trabalho, mesmo dados pouco coerentes com o conjunto de informações existentes sobre o organismo atualmente poderão servir de respaldo para novas descobertas experimentais futuramente, ou mesmo serem melhor compreendidos com o avanço dos trabalhos naquele organismo.

6.CONCLUSÕES

Neste trabalho, foi concluído com êxito a montagem das redes PPI da *Acetilactobacillus jinshanensis* CP171343 e da *Bacillus cereus* CP174498. Sua topologia foi observada e registrada. O estudo da rede da *A. jinshanensis* CP171343 demonstrou a presença de proteínas pontes do genoma e que conferem interpretações biológicas relevantes. Há um grande potencial para novos estudos dessas proteínas sejam experimentais, sejam computacionais. O estudo de interações foi capaz de corroborar com a hipótese da presença da dual via de oxidação da glicose. O rastreo de proteínas por meio do uso de uma via como ponto de partida foi eficaz e resultou em dados biológicos concretos. Essa etapa metodológica reduziu o tamanho da rede que seria analisada, já que se trata de um conjunto de dados cuja análise individual de cada interação é humanamente impossível. Uma nova metodologia foi proposta para uso em trabalhos similares a este. A metodologia foi testada e se mostrou capaz de capturar informações biológicas em trabalhos futuros. Por fim, a análise das vias expandiu o conhecimento sobre o potencial biotecnológico da *Acetilactobacillus jinshanensis* CP171343, abrindo novos caminhos para estudos de redes PPI de cada via. Ficou verificado a capacidade do estudo de redes PPI e suas

futuras aplicações em trabalhos similares. Foi feita uma abordagem holística dos dados obtidos, elaborando hipóteses que integrem os conhecimentos experimentais depositados na literatura.

7.REFERÊNCIAS

ADAMBERG, K.; VALGEPEA, K.; VILU, R. Advanced continuous cultivation methods for systems microbiology. **Microbiology (United Kingdom)**, v. 161, n. 9, p. 1707–1719, 1 set. 2015.

AL-HATAMLEH, M. A. I. *et al.* Antioxidant-Based Medicinal Properties of Stingless Bee Products: Recent Progress and Future Directions. **Biomolecules 2020, Vol. 10, Page 923**, v. 10, n. 6, p. 923, 18 jun. 2020.

ALONGE, M. *et al.* RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. **Genome Biology**, v. 20, n. 1, p. 1–17, 28 out. 2019.

____. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. **Genome Biology**, v. 23, n. 1, p. 1–19, 1 dez. 2022.

ANJOS, W. F. *et al.* GENPPI: standalone software for creating protein interaction networks from genomes. **BMC Bioinformatics**, v. 22, n. 1, p. 1–26, 1 dez. 2021a.

____. GENPPI: standalone software for creating protein interaction networks from genomes. **BMC Bioinformatics**, v. 22, n. 1, p. 1–26, 1 dez. 2021b.

ÅRSKÖLD, E. *et al.* Phosphoketolase pathway dominates in *Lactobacillus reuteri* ATCC 55730 containing dual pathways for glycolysis. **Journal of Bacteriology**, v. 190, n. 1, p. 206–212, jan. 2008.

BADER, G. D.; BETEL, D.; HOGUE, C. W. V. BIND: the Biomolecular Interaction Network Database. **Nucleic Acids Research**, v. 31, n. 1, p. 248–250, 1 jan. 2003.

BANDYOPADHYAY, S.; MALLICK, K. A New Feature Vector Based on Gene Ontology Terms for Protein-Protein Interaction Prediction. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 14, n. 4, p. 762–770, 1 jul. 2017.

BARANWAL, M. *et al.* A deep learning architecture for metabolic pathway prediction. **Bioinformatics**, v. 36, n. 8, p. 2547–2553, 15 abr. 2020.

____. Struct2Graph: a graph attention network for structure based predictions of protein–protein interactions. **BMC Bioinformatics**, v. 23, n. 1, p. 1–28, 1 dez. 2022.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. **Proceedings of the International AAAI Conference on Web and Social Media**, v. 3, n. 1, p. 361–362, 19 mar. 2009.

BATEMAN, A. *et al.* UniProt: the Universal Protein Knowledgebase in 2023. **Nucleic Acids Research**, v. 51, n. D1, p. D523–D531, 6 jan. 2023.

BELL, E. W. *et al.* PEPPI: Whole-proteome Protein-protein Interaction Prediction through Structure and Sequence Similarity, Functional Association, and Machine Learning. **Journal of Molecular Biology**, v. 434, n. 11, p. 167530, 15 jun. 2022.

- BERGGÅRD, T.; LINSE, S.; JAMES, P. Methods for the detection and analysis of protein–protein interactions. **PROTEOMICS**, v. 7, n. 16, p. 2833–2842, 1 ago. 2007.
- BINNS, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. **Bioinformatics**, v. 25, n. 22, p. 3045–3046, 15 nov. 2009.
- BOCK, J. R.; GOUGH, D. A. Predicting protein–protein interactions from primary structure. **Bioinformatics**, v. 17, n. 5, p. 455–460, 1 maio 2001.
- BOECKMANN, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Research**, v. 31, n. 1, p. 365–370, 1 jan. 2003.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, out. 2001.
- BÜRCKSTÜMMER, T. *et al.* An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. **Nature Methods** **2006 3:12**, v. 3, n. 12, p. 1013–1019, 22 out. 2006.
- CARBON, S. *et al.* Expansion of the Gene Ontology knowledgebase and resources. **Nucleic Acids Research**, v. 45, n. D1, p. D331–D338, 4 jan. 2017.
- CASPI, R. *et al.* The MetaCyc database of metabolic pathways and enzymes - a 2019 update. **Nucleic Acids Research**, v. 48, n. D1, p. D445–D453, 8 jan. 2020.
- CHATR-ARYAMONTRI, A. *et al.* The BioGRID interaction database: 2017 update. **Nucleic Acids Research**, v. 45, n. D1, p. D369–D379, 4 jan. 2017.
- CORTES, C.; VAPNIK, V.; SAITTA, L. Support-vector networks. **Machine Learning** **1995 20:3**, v. 20, n. 3, p. 273–297, set. 1995.
- CRISMAN, E. *et al.* KEAP1-NRF2 protein–protein interaction inhibitors: Design, pharmacological properties and therapeutic potential. **Medicinal Research Reviews**, v. 43, n. 1, p. 237–287, 1 jan. 2023.
- DANDEKAR, T. *et al.* Conservation of gene order: A fingerprint of proteins that physically interact. **Trends in Biochemical Sciences**, v. 23, n. 9, p. 324–328, 1 set. 1998.
- DU, X. *et al.* DeepPPI: Boosting Prediction of Protein-Protein Interactions with Deep Neural Networks. **Journal of Chemical Information and Modeling**, v. 57, n. 6, p. 1499–1510, 26 jun. 2017.
- DURHAM, J. *et al.* Recent advances in predicting and modeling protein–protein interactions. **Trends in Biochemical Sciences**, v. 48, n. 6, p. 527–538, 1 jun. 2023.
- FIELDS, S.; STERNGLANZ, R. The two-hybrid system: an assay for protein-protein interactions. **Trends in Genetics**, v. 10, n. 8, p. 286–292, 1 ago. 1994.
- GE, H. *et al.* Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. **Nature Genetics** **2001 29:4**, v. 29, n. 4, p. 482–486, 5 nov. 2001.
- Genome - NCBI - NLM.** Disponível em:
<https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=1720083>. Acesso em: 15 jan. 2025.
- HU, L. *et al.* A survey on computational models for predicting protein–protein interactions. **Briefings in Bioinformatics**, v. 22, n. 5, p. 1–18, 2 set. 2021.

ILGOVA, E. *et al.* Complex of HIV-1 Integrase with Cellular Ku Protein: Interaction Interface and Search for Inhibitors. **International journal of molecular sciences**, v. 23, n. 6, 1 mar. 2022.

JONES, S.; THORNTON, J. M. Principles of protein-protein interactions. **Proceedings of the National Academy of Sciences**, v. 93, n. 1, p. 13–20, 9 jan. 1996.

KESHAVA PRASAD, T. S. *et al.* Human Protein Reference Database—2009 update. **Nucleic Acids Research**, v. 37, n. suppl_1, p. D767–D772, 1 jan. 2009.

KOTELNIKOVA, E. *et al.* Fibrosis Protein-Protein Interactions from Google Matrix Analysis of MetaCore Network. **International Journal of Molecular Sciences** 2022, Vol. 23, Page 67, v. 23, n. 1, p. 67, 22 dez. 2021.

KOTIRANTA, A.; LOUNATMAA, K.; HAAPASALO, M. Epidemiology and pathogenesis of *Bacillus cereus* infections. **Microbes and Infection**, v. 2, n. 2, p. 189–198, 1 fev. 2000.

KOVÁCS, I. A. *et al.* Network-based prediction of protein interactions. **Nature Communications** 2019 10:1, v. 10, n. 1, p. 1–8, 18 mar. 2019.

LICATA, L. *et al.* MINT, the molecular interaction database: 2012 update. **Nucleic Acids Research**, v. 40, n. D1, p. D857–D861, 1 jan. 2012.

LIU, J. *et al.* Exploring the ecological interactions of *Bacillus* and their contribution to characteristic aroma components in Zhenjiang aromatic vinegar. **Food Bioscience**, v. 54, p. 102900, 1 ago. 2023.

LIU, S. **Bioprocess Engineering: Kinetics, Sustainability, and Reactor Design - Shijie Liu - Google Livros**. Cambridge: [s.n.].

MAMLOUK, D.; GULLO, M. Acetic Acid Bacteria: Physiology and Carbon Sources Oxidation. **Indian Journal of Microbiology**, v. 53, n. 4, p. 377–384, 5 dez. 2013.

MENDES, R. A.; COELHO, A. Í. M.; AZEREDO, R. M. C. DE. Contaminação por *Bacillus cereus* em superfícies de equipamentos e utensílios em unidade de alimentação e nutrição. **Ciência & Saúde Coletiva**, v. 16, n. 9, p. 3933–3938, set. 2011.

MetaCyc heterolactic fermentation. Disponível em:
<https://biocyc.org/pathway?orgid=META&id=P122-PWY>. Acesso em: 27 jan. 2025.

MIRABELLO, C.; WALLNER, B. InterPred: A pipeline to identify and model protein–protein interactions. **Proteins: Structure, Function, and Bioinformatics**, v. 85, n. 6, p. 1159–1170, 1 jun. 2017.

MOHAMMAD, S. M.; MAHMUD-AB-RASHID, N. K.; ZAWAWI, N. Probiotic properties of bacteria isolated from bee bread of stingless bee *Heterotrigona itama*. **Journal of Apicultural Research**, v. 60, n. 1, p. 172–187, 2020.

NOOREN, I. M. A.; THORNTON, J. M. Diversity of protein–protein interactions. **The EMBO Journal**, v. 22, n. 14, p. 3486–3492, 15 jul. 2003.

NORTHEY, T. C.; BAREŠIÄ ħ, A.; MARTIN, A. C. R. IntPred: a structure-based predictor of protein–protein interaction sites. **Bioinformatics**, v. 34, n. 2, p. 223–229, 1 jan. 2018.

OLIVEIRA, K. DE. Bioprospecção de bactérias visando ativos biológicos para biofertilizantes. **Dissertação de mestrado**, 7 jun. 2023.

- ORCHARD, S. *et al.* The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. **Nucleic Acids Research**, v. 42, n. D1, p. D358–D363, 1 jan. 2014.
- PARIHAR, H. S. Bacillus cereus. **Encyclopedia of Toxicology: Third Edition**, p. 353–354, 1 jan. 2014.
- PAVLOPOULOS, G. A. *et al.* Using graph theory to analyze biological networks. **BioData Mining**, v. 4, n. 1, p. 1–27, 28 abr. 2011.
- PELLEGRINI, M. *et al.* Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 8, p. 4285–4288, 13 abr. 1999.
- PEREIRA, G.; GHOSH, P.; SANTOS, A. A Bridging Centrality Plugin for GEPHI and a Case Study for Mycobacterium Tuberculosis H37Rv. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 18, n. 6, p. 2741–2746, 2021.
- PESSÔA, L. C. Biorrefinaria de microalgas e valoração de efluentes industriais: um modelo potencial de bioeconomia circular para o Brasil. 20 ago. 2024.
- PLANAS-IGLESIAS, J. *et al.* Understanding Protein–Protein Interactions Using Local Structural Features. **Journal of Molecular Biology**, v. 425, n. 7, p. 1210–1224, 12 abr. 2013.
- PRJIBELSKI, A. *et al.* Using SPAdes De Novo Assembler. **Current Protocols in Bioinformatics**, v. 70, n. 1, p. e102, 1 jun. 2020.
- QIAO, N. *et al.* After the storm—Perspectives on the taxonomy of Lactobacillaceae. **JDS Communications**, v. 3, n. 3, p. 222–227, 1 maio 2022.
- QUEVILLON, E. *et al.* InterProScan: protein domains identifier. **Nucleic Acids Research**, v. 33, n. Web Server issue, p. W116, jul. 2005.
- SANTOS, A. C. C. MICROBIOMA DO ALIMENTO LARVAL DE ABELHAS SEM FERRÃO: DIVERSIDADE E POTENCIAL BIOTENOLÓGICO. **Tese (Doutorado)**, 2022.
- SANTOS, A. C. C. *et al.* Antimicrobial activity of supernatants produced by bacteria isolated from Brazilian stingless bee’s larval food. **BMC Microbiology**, v. 22, n. 1, p. 1–9, 1 dez. 2022.
- SANTOS, A. R. *et al.* PANNOTATOR: an automated tool for annotation of pan-genomes. **Genetics and molecular research : GMR**, v. 12, n. 3, p. 2982–2989, 16 ago. 2013.
- SANTOS, ANDERSON R. *et al.* Mature Epitope Density - A strategy for target selection based on immunoinformatics and exported prokaryotic proteins. **BMC Genomics**, v. 14, n. 6, p. 1–11, 25 out. 2013.
- SCHERLACH, K.; HERTWECK, C. Mining and unearthing hidden biosynthetic potential. **Nature Communications** 2021 12:1, v. 12, n. 1, p. 1–12, 23 jun. 2021.
- SHAN, S. *et al.* Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization. **Proceedings of the National Academy of Sciences of the United States of America**, v. 119, n. 11, p. e2122954119, 15 mar. 2022.
- SHANNON, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. **Genome Research**, v. 13, n. 11, p. 2498–2504, 1 nov. 2003.

- SHIMIZU, K. Metabolic Regulation Analysis and Metabolic Engineering. **Comprehensive Biotechnology, Second Edition**, v. 2, p. 407–420, 1 jan. 2011.
- SILVA, ALISSON *et al.* The GenPPI tool enhanced Protein Interaction Network Generation with Machine Learning-Based Protein Similarity Inference. **bioRxiv**, v. X, p. 2024.08.16.608332, 19 ago. 2024.
- SPECTOR, M. P. Metabolism, Central (Intermediary). **Encyclopedia of Microbiology, Third Edition**, p. 242–264, 1 jan. 2009.
- SYMEONIDIS, P. *et al.* From biological to social networks: Link prediction based on multi-way spectral clustering. **Data & Knowledge Engineering**, v. 87, p. 226–242, 1 set. 2013.
- SZKLARCZYK, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. **Nucleic Acids Research**, v. 51, n. D1, p. D638–D646, 6 jan. 2023.
- TANG, T. *et al.* Machine learning on protein–protein interaction prediction: models, challenges and trends. **Briefings in Bioinformatics**, v. 24, n. 2, p. 1–11, 19 mar. 2023.
- TENDIUK, N. *et al.* Structure-Functional Characteristics of the Svx Protein—The Virulence Factor of the Phytopathogenic Bacterium *Pectobacterium atrosepticum*. **International Journal of Molecular Sciences**, v. 23, n. 13, p. 6914, 1 jul. 2022.
- TOMKINS, J. E.; MANZONI, C. Advances in protein-protein interaction network analysis for Parkinson’s disease. **Neurobiology of Disease**, v. 155, p. 105395, 1 jul. 2021.
- TONG, A. H. Y. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. **Science**, v. 294, n. 5550, p. 2364–2368, 14 dez. 2001.
- TONG, A. H. Y. *et al.* A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. **Science**, v. 295, n. 5553, p. 321–324, 11 jan. 2002.
- VÁSQUEZ, A. *et al.* Symbionts as Major Modulators of Insect Health: Lactic Acid Bacteria and Honeybees. **PLOS ONE**, v. 7, n. 3, p. e33188, 12 mar. 2012.
- VELANKAR, S. *et al.* The Protein Data Bank Archive. **Methods in Molecular Biology**, v. 2305, p. 3–21, 2021.
- VERMA, P.; SHAH, M. P. **Bioprospecting of Microbial Diversity Challenges and Applications in Biochemical Industry, Agriculture and Environment Protection**. 1^a ed. Cambridge: [s.n.].
- WANG, K. *et al.* Investigation of Zhenjiang Aromatic Vinegar Production Using a Novel Dry Gelatinization Process. **Foods** **2024, Vol. 13, Page 1071**, v. 13, n. 7, p. 1071, 31 mar. 2024.
- WANG, S. *et al.* Protein-protein interaction networks as miners of biological discovery. **PROTEOMICS**, v. 22, n. 15–16, p. 2100190, 1 ago. 2022.
- XENARIOS, I. *et al.* DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. **Nucleic Acids Research**, v. 30, n. 1, p. 303–305, 1 jan. 2002.
- XU, L. *et al.* Predicting the multispecies solid-state vinegar fermentation process using single-cell Raman spectroscopy combined with machine learning. **LWT**, v. 192, p. 115708, 15 jan. 2024a.

- ____. Predicting the multispecies solid-state vinegar fermentation process using single-cell Raman spectroscopy combined with machine learning. **LWT**, v. 192, p. 115708, 15 jan. 2024b.
- XU, W. *et al.* Monitoring the microbial community during solid-state acetic acid fermentation of Zhenjiang aromatic vinegar. **Food Microbiology**, v. 28, n. 6, p. 1175–1181, 1 set. 2011a.
- ____. Monitoring the microbial community during solid-state acetic acid fermentation of Zhenjiang aromatic vinegar. **Food Microbiology**, v. 28, n. 6, p. 1175–1181, 1 set. 2011b.
- YU, Y. *et al.* *Lactobacillus jinshani* sp. nov., isolated from solid-state vinegar culture of Zhenjiang aromatic vinegar. **Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology**, v. 113, n. 1, p. 43–54, 1 jan. 2020a.
- ____. *Lactobacillus jinshani* sp. nov., isolated from solid-state vinegar culture of Zhenjiang aromatic vinegar. **Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology**, v. 113, n. 1, p. 43–54, 1 jan. 2020b.
- ZENG, S. Link prediction based on local information considering preferential attachment. **Physica A: Statistical Mechanics and its Applications**, v. 443, p. 537–542, 1 fev. 2016.
- ZHANG, Q. C. *et al.* Structure-based prediction of protein-protein interactions on a genome-wide scale. **Nature**, v. 490, n. 7421, p. 556–560, 25 out. 2012.
- ZHENG, J. *et al.* A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. **International Journal of Systematic and Evolutionary Microbiology**, v. 70, n. 4, p. 2782–2858, 15 abr. 2020.