

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICA DE RIBEIRÃO PRETO

HENRIQUE MARCEL YUDI DE OLIVEIRA TSUJI

Modelos metabólicos em escala genômica: compreensão, geração, validação e aplicação em um contexto biotecnológico.

Ribeirão Preto

2024

HENRIQUE MARCEL YUDI DE OLIVEIRA TSUJI

Modelos metabólicos em escala genômica: compreensão, geração, validação e aplicação em um contexto biotecnológico.

Trabalho de Conclusão de Curso
apresentado à Faculdade de Ciências
Farmacêuticas de Ribeirão Preto

Nome do orientador: Ricardo Roberto da
Silva

Ribeirão Preto

2024

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Tsuji, Henrique

Modelos metabólicos em escala genômica: compreensão, geração, validação e aplicação em um contexto biotecnológico, Ribeirão Preto, 2024

Número de páginas: 35 p.

Trabalho de Conclusão de Curso apresentado à Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo.

Nome do orientador: Ricardo Roberto da Silva

1-Bioinformática; 2-Biofármacos; 3-Genômica; 4-Modelagem metabólica em escala genômica

Nome do autor: Tsuji, Henrique

Título do trabalho: Modelos metabólicos em escala genômica: compreensão, criação, validação e aplicação em um contexto biotecnológico.

Trabalho de conclusão de curso apresentado à Faculdade de Ciências Farmacêuticas de Ribeirão Preto.

Aprovado em:

Banca examinadora

Título e Nome: _____

Instituição: _____

Julgamento: _____

Título e Nome: _____

Instituição: _____

Julgamento: _____

Título e Nome: _____

Instituição: _____

Julgamento: _____

AGRADECIMENTOS

Agradeço aos meus pais, Eloi e Sueli, por todo o apoio e incentivo incondicional ao longo da minha vida. Por sempre acreditarem no meu potencial e me oferecerem muito mais do que lhes foi oferecido para que eu pudesse seguir o meu caminho.

Agradeço a minha namorada, Giovanna, pelo amor, paciência e apoio incondicional ao longo desta jornada. Seu companheirismo foi essencial nos momentos mais desafiadores, e sua presença tornou cada conquista ainda mais especial.

Agradeço ao meu orientador, Ricardo, cuja orientação e confiança em mim foram fundamentais para o desenvolvimento deste e outros trabalhos que realizamos. E por providenciar o café tão necessário para as atividades bioinformáticas do laboratório.

Agradeço aos membros do CCBL, Ana, Gabriel, João, Luiz, Pedro e Tiago por todo o apoio, seja com ideias, ajudas e parcerias nos meus projetos ou com conselhos de vida valorosos nos cafés da tarde.

Agradeço aos meus amigos Erasmo, Felipe, Jonathas e Pedro pelo apoio e camaradagem ao longo da faculdade.

Agradeço aos meus amigos não humanos Aria, Britto, Fiona, Frajola, Giselle, Haelga, Horácio, Kimmy, Sansa e Yoko.

RESUMO

TSUJI, Henrique. Modelos metabólicos em escala genômica: compreensão, criação, validação e aplicação em um contexto biotecnológico.

Os modelos metabólicos em escala genômica (GEMs) são ferramentas poderosas que permitem simular e otimizar o metabolismo de microrganismos para aplicações industriais e farmacêuticas. Os GEMs são tipicamente construídos a partir de sequências completas de genomas, e requerem uma série de etapas complexas que devem ser executadas antes da construção do modelo. Para genomas originais obtidos pelas técnicas modernas de sequenciamento são necessárias uma série de etapas prévias à modelagem, como a montagem do genoma, inspeções de qualidade e anotação funcional. Essas etapas requerem conhecimento de técnicas sofisticadas de bioinformática, dificultando o acesso de profissionais das ciências farmacêuticas aos modelos metabólicos que poderiam ser utilizados para otimização em diferentes processos produtivos com aplicações na área da saúde. Este trabalho teve como objetivo construir modelos metabólicos em escala genômica (GEMs) preliminares de bactérias isoladas de ambientes marinhos, por meio de ferramentas integradas a um fluxo de trabalho automatizado. O fluxo de trabalho Pardal foi desenvolvido para automatizar o processamento de dados de sequenciamento genômico gerados por meio da plataforma MinION. O fluxo contém etapas de atribuição de bases, montagem de genomas, anotação, predição de agrupamentos gênicos biossintéticos (BGCs), e modelagem metabólica em escala genômica, gerando GEMs preliminares que podem ser refinados manualmente para análises mais precisas. Este estudo reforça o potencial das GEMs como plataformas de desenvolvimento de biofármacos, ao mesmo tempo que evidencia a importância da curadoria manual e de validações experimentais para aumentar a precisão das previsões.

Palavras-chave: Bioinformática; Biofármacos; Genômica; Modelagem metabólica em escala genômica

SUMÁRIO

1. INTRODUÇÃO.....	6
1.1 O GENOMA E O METABOLISMO CODIFICADO.....	6
1.2 MODELOS METABÓLICOS EM ESCALA GENÔMICA.....	9
1.3 ANÁLISE DE BALANÇO DE FLUXO E DE VARIABILIDADE DE FLUXO.....	10
1.5 BIOTECNOLOGIA E BIOFÁRMACOS.....	14
2. OBJETIVOS.....	16
3. MÉTODOS.....	17
4. RESULTADOS E DISCUSSÃO.....	22
4.1. COMPARAÇÃO DO CARVEME COM MODELOS CURADOS MANUALMENTE.....	22
4.2. ANÁLISE DOS GEMs CRIADOS A PARTIR DE DADOS ORIGINAIS.....	25
4.3. PRODUÇÃO DE BIOFÁRMACOS USANDO OS GEMs PRELIMINARES.....	27
5. CONCLUSÃO.....	32
BIBLIOGRAFIA.....	33

1. INTRODUÇÃO

1.1. O GENOMA E O METABOLISMO CODIFICADO

Atualmente o estudo do metabolismo ainda apresenta uma série de desafios tecnológicos e científicos. Um dos maiores problemas relacionados ao tema é a grande complexidade das redes metabólicas, pois cada organismo possui milhares de reações bioquímicas interligadas, formando um sistema dinâmico e interdependente. Além disso, integrar possíveis variações, como as diferentes condições fisiológicas e ambientais, é uma tarefa homérica, exigindo uma quantidade massiva de dados experimentais (ZELEZNIAK et al., 2018). Fatores como a variabilidade metabólica entre indivíduos de uma mesma espécie também dificultam o entendimento, considerando que variações genéticas em uma mesma espécie podem levar a diferentes fenotipos metabólicos (LARDIES; CATALÁN; BOZINOVIC, 2004).

A construção de modelos metabólicos necessita da sequência completa do genoma do organismo de interesse. Existem diferentes tecnologias de sequenciadores, cujo desenvolvimento é associado às primeira, segunda (ou nova geração) e terceira gerações. Atualmente a tecnologia se encontra em sua terceira geração, representada por plataformas como *PacBio* e *MinION*. Há uma evolução significativa entre as gerações, tanto em qualidade dos dados, como em portabilidade e tempo de processamento (HEATHER; CHAIN, 2016). Enquanto as gerações anteriores desempenharam papéis cruciais ao lançar as bases da genômica, as tecnologias de terceira geração emergiram com capacidades notáveis, como obtenção de sequências mais longas e custos reduzidos, permitindo avanços substanciais na análise genômica (LU; GIORDANO; NING, 2016).

Para obter um modelo metabólico em escala genômica, além do genoma, é necessário obter a anotação do organismo em questão, e para isso há diversas etapas a serem realizadas. A montagem é o processo de reconstruir a sequência de um genoma a partir de fragmentos de DNA sequenciados, utilizando abordagens como a montagem *de novo* e a montagem híbrida, que combina leituras curtas e longas. O objetivo é gerar uma representação precisa do genoma, facilitando análises genéticas (WICK et al., 2017). O resultado final é um arquivo .FASTA, que contém a sequência completa do genoma.

A anotação genômica é o processo pelo qual os dados de uma sequência de DNA são caracterizados para identificar genes, elementos regulatórios e suas respectivas funções. Essa etapa inclui similaridade de sequência a bases de dados, bem como a detecção de quadros abertos de leitura (do inglês *Open Reading Frame*) e outros padrões associados aos genes (STEIN, 2001). Além de fornecer uma compreensão detalhada da biologia do organismo, a anotação pode revelar informações cruciais para o desenvolvimento de novos tratamentos ou para a exploração de vias metabólicas específicas (SEEMANN, 2014). No entanto, com esses avanços e pelo fato de existirem inúmeras ferramentas *open-source* que são capazes de realizar essas funções, vêm novos desafios, especialmente no que diz respeito à análise e interpretação dos dados gerados por essas tecnologias de última geração. Técnicas avançadas de bioinformática, como a utilização de gerenciadores de fluxos de trabalho, como o NextFlow (DI TOMMASO et al., 2017), permitem que múltiplas ferramentas possam ser utilizadas em conjunto, garantindo consistência e eficiência às etapas da montagem e anotação genômica.

Partindo de um modelo preliminar, como será demonstrado na seção dos resultados, boa parte do que é apresentado poderá ser utilizado, porém existirão dados que deverão ser adicionados, outros a serem retirados e dados que deverão ser corrigidos. Para isso, a melhor forma de detectar erros é utilizando uma ferramenta que avalia GEMs como no caso do *Memote* (LIEVEN et al., 2020). Essa ferramenta foi projetada para avaliar a qualidade de GEMs, realizando uma série de testes automáticos que verificam a consistência, completude e conformidade do modelo com padrões estabelecidos. Com ela é possível verificar, por exemplo, se as reações metabólicas estão balanceadas, se há vias metabólicas incompletas, ou se faltam anotações importantes.

Após a análise, a ferramenta gera um relatório detalhado, apontando os problemas encontrados e sugerindo melhorias. Esse *feedback* é essencial para garantir que o modelo esteja corretamente ajustado antes de ser utilizado em simulações ou análises biológicas mais complexas. Além disso, é necessário ter amplo conhecimento acerca do organismo em estudo, para conseguir compreender com auxílio da ferramenta o que provavelmente está errado ou falta no modelo.

Com a aplicação *web Fluxer* (HARI; LOBO, 2020) pode-se visualizar o modelo como um mapa a fim de permitir o entendimento do modelo metabólico em escala genômica de forma mais ilustrativa. Com a aplicação também é possível realizar

simulações e interagir com os fluxos metabólicos por meio de diferentes tipos de gráficos.

O software *OptFlux* (ROCHA et al., 2010) também pode ser utilizado para realizar simulações no modelo ao longo do processo de refinamento, tanto em organismos normais quanto modificados. Com as informações indicadas pela ferramenta *Memote* e após averiguar possíveis alterações que podem melhorar a qualidade dos GEMs a serem desenvolvidos, um meio de modificar o modelo é utilizando o pacote *Python COBRApy* (EBRAHIM et al., 2013). Com ele é possível retirar e adicionar reações, genes e metabólitos ao modelo, e também realizar várias operações comuns à análise de GEMs como o FBA e o FVA que serão explicados no decorrer do trabalho.

Outro desafio importante é a integração das diferentes ciências ômicas, como genômica, transcriptômica, proteômica e metabolômica. Muitas vezes, os dados dessas áreas são gerados em condições diferentes, por pesquisadores diferentes, e apresentam escalas distintas, dificultando sua harmonização e análise conjunta (WALLEY et al., 2016). Essa limitação impacta a capacidade de compreender de forma holística como o metabolismo é regulado e modulado em diferentes condições ambientais ou patológicas (GU et al., 2019).

Diante dos desafios mencionados, a bioinformática emergiu como uma ferramenta essencial, e muitas vezes obrigatória, para a análise do metabolismo. Uma de suas principais contribuições é a modelagem do metabolismo em escala genômica, que permite simular e prever o comportamento metabólico de organismos em diferentes condições. Operações computacionais, como a análise de balanço de fluxo (FBA - *Flux Balance Analysis*), são amplamente utilizadas (ORTH; THIELE; PALSSON, 2010).

Além disso, plataformas de integração e armazenamento de dados bioquímicos e ômicos, como o KEGG (KANEHISA, 2002), MetaCyc (KARP et al., 2002) e BioCyc (KARP et al., 2019), facilitam a organização e exploração de informações relacionadas. Com essas plataformas, é possível automatizar a anotação de modelos metabólicos preliminares, utilizando ferramentas como o *CarveMe* (MACHADO et al., 2018), e cruzar dados genômicos com reações bioquímicas conhecidas, no caso de uma abordagem manual. A bioinformática também auxilia com métodos como o aprendizado de máquina, que tem sido

aplicado, por exemplo, para prever modificações possíveis em genes e no metabolismo de bactérias ao longo de gerações (KONNO; IWASAKI, 2023).

1.2. MODELOS METABÓLICOS EM ESCALA GENÔMICA

A modelagem metabólica em escala genômica (do inglês *genome-scale metabolic model* - GEM) é uma abordagem de se estudar um organismo, que por sua natureza multidisciplinar, vem sendo adotada lentamente pela comunidade científica. (GU et al., 2019). Esses modelos são construídos com base na anotação completa do genoma, incluindo todas as reações metabólicas conhecidas e potenciais. Isso permite uma análise precisa da contribuição de cada gene para a rede metabólica global (Lewis et al., 2012). Modelos metabólicos em escala genômica representam, de forma computacional, as interações entre genes, metabólitos e reações que envolvem todo o metabolismo de um organismo. Esses modelos permitem simular fluxos metabólicos e prever o comportamento de sistemas em variados cenários (GU et al., 2019).

O grande diferencial dos GEMs é a sua capacidade de permitir realizar a previsão de respostas fenotípicas a partir de alterações genéticas. Com esses modelos é possível modificar virtualmente genes específicos, dessa forma, pesquisadores podem prever quais vias metabólicas serão afetadas devido a essas mudanças. Diversas áreas se beneficiam dessa abordagem, por exemplo, na biotecnologia GEMs são usados para otimizar o metabolismo microbiano, tornando possível projetar microrganismos capazes de produzir compostos de interesse industrial de forma mais eficiente (ZHANG; HUA, 2016).

Outro benefício dos GEMs é sua aplicação na ecologia microbiana, ajudando a entender como diferentes espécies interagem em comunidades complexas. Esses modelos podem ser combinados com dados ambientais para prever a dinâmica populacional e como diferentes fatores impactam o fluxo de carbono e nutrientes em ecossistemas (ZUÑIGA; ZARAMELA; ZENGLER, 2017).

A presente monografia se dedica a criar um material acessível para introdução à modelagem metabólica em escala genômica. Em geral, GEMs já foram aplicados para os seguintes propósitos: a exploração da essencialidade gênica, definição de contexto metabólico para bioinformática integrativa, estudo da evolução de vias metabólicas, predição de estratégias de engenharia metabólica, predição de

resultados de evolução adaptativa e interpretação de dados de fermentação (SANTOS; BOELE; TEUSINK, 2011). Nas próximas seções os conteúdos fundamentais para o entendimento e utilização dos GEMs serão explorados.

1.3. ANÁLISE DE BALANÇO DE FLUXO E DE VARIABILIDADE DE FLUXO

ANÁLISE DE BALANÇO DE FLUXO

A Análise de Balanço de Fluxo (do inglês *Flux Balance Analysis* - FBA) é uma técnica que utiliza programação linear para estudar redes metabólicas, permitindo a predição de consumo ou produção de metabólitos em um sistema biológico (ORTH; THIELE; PALSSON, 2010). O fluxo metabólico representa o nível de consumo ou produção de um metabólito em uma via, ou seja, a taxa *in vivo* de uma reação enzimática, e é definido como o número de moléculas convertidas por unidade de tempo por célula ($\text{mmol gDW}^{-1} \text{h}^{-1}$, DW é do inglês *dry weight*, peso seco)(STEPHANOPOULOS; ARISTIDOU; NIELSEN, 1998).

Para que os fluxos metabólicos possam ser utilizados computacionalmente é necessário que os mesmos sejam representados como um dado estruturado, que servirá de entrada para os modelos computacionais preditivos. Essa representação é ilustrada a seguir:

1) Matriz (S)

A base do FBA é a representação matemática das reações metabólicas em uma matriz conhecida como matriz estequiométrica (S). Esta matriz é construída da seguinte forma: A matriz possui tamanho m por n , ou seja, cada linha da matriz representa um dos m metabólitos (compostos químicos), e cada coluna representa uma das n reações químicas. As entradas em cada coluna são os coeficientes estequiométricos dos metabólitos que participam de uma reação. Há um coeficiente negativo para cada metabólito consumido e um coeficiente positivo para cada metabólito produzido. Um coeficiente estequiométrico de zero é utilizado para cada metabólito que não participa de uma determinada reação (RAMAN; CHANDRA, 2009).

A matriz S abaixo será utilizada para exemplificar esse conceito:

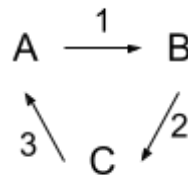
$$S = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

Neste exemplo:

A primeira linha representa o metabólito A, que é consumido na primeira reação, indicado pelo valor negativo na primeira coluna, não participa da segunda reação, indicado pelo valor zero na segunda coluna, e é produzido na terceira reação, indicado pelo valor positivo na terceira coluna.

Desse pequeno conjunto de metabólitos e reações é possível determinar: O consumo de A produz B na reação 1, o consumo de B produz C na reação 2 e o consumo de C produz A na reação 3. Formando desta forma um ciclo de reações, como o indicado abaixo:

Figura 1: Esquema ilustrativo de uma reação cíclica.



Fonte: De autoria própria.

S é uma matriz esparsa, possui muitos zeros, pois a maioria das reações envolvem apenas alguns metabólitos (ORTH; THIELE; PALSSON, 2010).

Uma matriz S utilizada em um modelo metabólico em escala genômica envolve todas as reações e metabólitos do metabolismo de um organismo.

2) Equação de Balanceamento de Fluxo

O fluxo através de todas as reações em uma rede é representado pelo vetor v , que tem comprimento n . A equação fundamental do FBA é dada pela equação de balanceamento de fluxo dada por (RAMAN; CHANDRA, 2009):

$$Sv=0$$

na qual:

S é a matriz estequiométrica.

v é um vetor que representa as taxas de fluxo (fluxos) de cada reação.

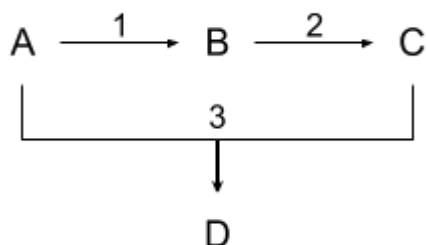
Para essa equação, há inúmeros valores de v possíveis, e para cada v que satisfaz a equação é dito que ele faz parte do espaço nulo de S (ORTH; THIELE; PALSSON, 2010).

No caso da matriz S citada anteriormente, qualquer v resolve a equação, por se tratar de um ciclo isolado no qual o aumento ou diminuição do fluxo de qualquer metabólito influencia nos outros de forma diretamente proporcional. Porém, a matriz S abaixo funciona de forma diferente:

$$S = \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

O conjunto de reações que representa essa matriz pode ser apresentada dessa forma:

Figura 2: Esquema ilustrativo de uma reação com quatro compostos e três reações.



Fonte: De autoria própria.

Uma situação com 4 compostos e 3 reações. Caso o objetivo seja aumentar o fluxo da reação 1, então, logicamente, o fluxo da reação 3, em que há o consumo do metabólito responsável pela reação 1, teria de ser interrompido, formando um vetor como esses:

$$v = \begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix} \quad v = \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix} \quad v = \begin{bmatrix} 20 \\ 22 \\ 0 \end{bmatrix}$$

Todos esses valores de v são possíveis para resolver o problema, que é aumentar o fluxo de formação de B.

Em casos pequenos é simples imaginar como se comportaria o v que gera uma solução, mas ao analisar todo o metabolismo de um organismo, com centenas a milhares de reações ocorrendo com inúmeros metabólitos, obter soluções é algo impraticável sem utilizar meios computacionais. Por isso, a análise de FBA é executada computacionalmente, demandando poucos minutos para encontrar soluções que aumentem ou diminuam o fluxo do objetivo, geralmente encontrando a solução ideal, sendo essa, a que mais aumenta o fluxo de uma reação, por exemplo, em um caso no qual esse seria o objetivo.

3) Restrições

As restrições (em inglês *constraints*) em GEMs são limitações matemáticas que garantem previsões realistas ao regular a atividade das reações metabólicas, evitando fluxos irrealistas e reproduzindo a fisiologia celular. Elas reduzem o espaço de soluções, facilitando a interpretação dos resultados e permitem simular condições específicas, como restrição de nutrientes ou variações ambientais (STALIDZANS et al., 2018). Há três tipos principais de restrições: (1) De balanço de massa em estado estacionário, que define que a produção e o consumo de cada metabólito sejam iguais, evitando acúmulo ou depleção; (2) Que garante a reversibilidade de reações, e assegura que reações irreversíveis ocorram apenas na direção correta, com base em dados bioquímicos ou estimativas termodinâmicas; e (3) De capacidade enzimática, que impõe limites superiores e inferiores aos fluxos em reações nas quais as capacidades de fluxo são conhecidas, como taxas de crescimento e absorção de nutrientes (REED, 2017).

As restrições são representadas de duas maneiras, como equações que equilibram as entradas e saídas das reações e como desigualdades que impõem limites ao sistema (ORTH; THIELE; PALSSON, 2010).

Por exemplo, caso exista uma reação chamada v_1 que não pode possuir um fluxo menor que 0, uma forma de limitar a busca por soluções seria informar que:

$$0 \leq v_1 \leq \infty$$

E essa seria uma limitação do espaço de possíveis soluções.

4) Função objetivo

A função objetivo é uma combinação linear de fluxos, determinada pela equação abaixo:

$$Z = c^T v$$

O FBA tem como objetivo maximizar ou minimizar a função objetivo. Para isso é necessário utilizar c , um vetor de pesos, que indica o quanto cada reação contribui para a função objetivo. Quando apenas uma reação é o objetivo, c é um vetor de zeros, com um valor de 1 na posição da reação de interesse (ORTH; THIELE; PALSSON, 2010).

Sendo assim, todo v em que o fluxo da reação alvo não é zero será considerada como uma solução no espaço de soluções possíveis.

ANÁLISE DE VARIABILIDADE DE FLUXO

A análise de variabilidade de fluxo (do inglês *Flux Variability Analysis* - FVA) parte de um FBA, mas como já foi explicado, um FBA geralmente revela mais de uma solução. O FVA é um método que expande a análise de FBA, calculando a faixa de variabilidade dos fluxos de reação que atendem aos requisitos ótimos definidos, proporcionando uma visão mais abrangente do comportamento metabólico do sistema. Tipicamente o FVA é realizado em duas etapas, sendo que a primeira é a realização de um FBA, e na segunda são calculadas as extensões dos fluxos da rede metabólica, representando o intervalo de valores possíveis para cada fluxo de reação. Isso é feito resolvendo 2 problemas de programação linear (LPs) adicionais com uma restrição que permite valores quase ótimos e outra que impõe valores ótimos exatos do problema FBA. Essa abordagem permite avaliar a flexibilidade do metabolismo e identificar reações críticas, fornecendo uma visão mais abrangente do comportamento metabólico do sistema (KENEFAKE et al., 2022).

1.4. BIOTECNOLOGIA E BIOFÁRMACOS

Um biofármaco é um tipo de medicamento cujo princípio ativo é obtido a partir de processos biotecnológicos, em contraste com os medicamentos convencionais,

que são usualmente produzidos por meio de síntese química (BRANDÃO; SOUZA, 2015). Geralmente eles são projetados para agir de maneira muito específica, diretamente sobre os mecanismos biológicos envolvidos em doenças, o que aumenta a precisão e a eficácia em comparação com os medicamentos de síntese química. Essas características fazem dos biofármacos uma opção promissora para o tratamento de várias doenças, embora existam também desafios e desvantagens. Por exemplo, biofármacos são tipicamente fabricados por meio de biorreatores (COONEY, 1983), no entanto, testar hipóteses em biorreatores pode ter um custo financeiro elevado (LO; MCADAM; JUDD, 2015), considerando a necessidade de otimização das culturas por múltiplos testes por um longo período de tempo.

As bactérias desempenham um papel crucial na descoberta de novas moléculas bioativas, especialmente no desenvolvimento de biofármacos. A título de exemplo ao longo das últimas décadas, compostos produzidos por microrganismos têm sido explorados para criar antibióticos (MARWICK; WRIGHT; BURGESS, 1999), e agentes antitumorais (NIRAULA et al., 2010).

Diante do exposto, o uso de GEMs na produção de biofármacos se apresenta como estratégia promissora, uma vez que é possível simular o metabolismo de um organismo e modificar os fluxos de produção de diferentes compostos, possibilitando a otimização da produção de biofármacos (RICHELLE et al., 2020). Além disso, com o uso dos GEMs é possível encontrar vias que levam a produção de produtos tóxicos no processo de obtenção de um composto biotecnológico. O projeto Pardal, que será descrito a seguir, desenvolveu um fluxo de trabalho computacional para processamento de dados de sequenciamento genômico, produção e inspeção de GEMs, possibilitando a utilização dos modelos metabólicos na prospecção de biofármacos.

2. OBJETIVOS

Objetivo geral:

A presente monografia teve por objetivo construir os modelos metabólicos em escala genômica preliminares de um conjunto de bactérias isoladas de ambientes marinhos pertencentes ao escopo de um projeto maior.

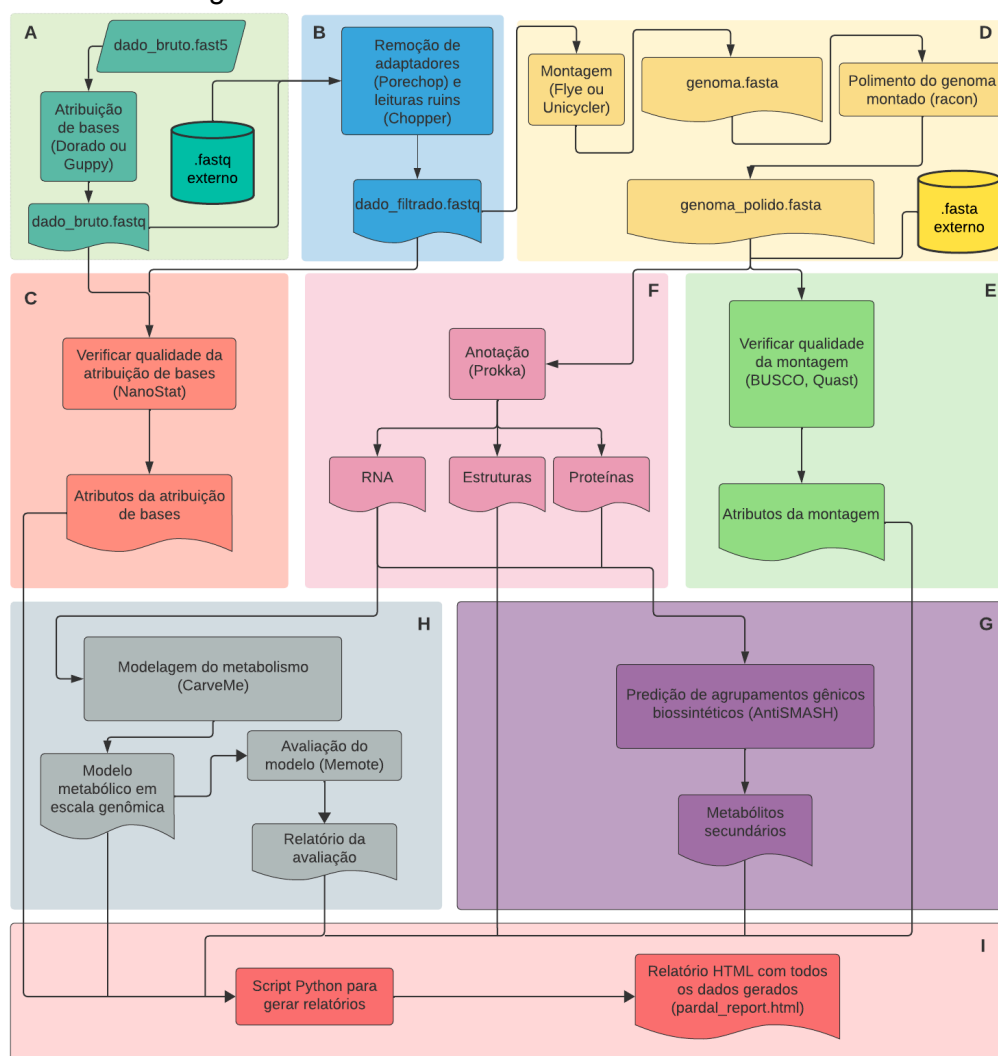
Objetivos específicos:

- 1) Pré-processar, aferir a qualidade, montar e anotar sequências de genomas obtidas pela plataforma MinION;
- 2) Elaborar uma revisão sobre modelos metabólicos em escala genômica e sobre sua utilização com foco na produção e elaboração de biofármacos;
- 3) Criar modelos metabólicos preliminares a partir dos genomas montados;
- 4) Elaborar uma revisão sobre métodos para refinar modelos preliminares, incluindo a identificação e correção de erros.

3. MÉTODOS

A ferramenta Pardal, desenvolvida no projeto de iniciação científica, elaborado pelo autor da presente monografia, tem como objetivo gerar, por meio de um fluxo de trabalho automatizado, a montagem e anotação do genoma sequenciado pela plataforma Oxford Nanopore MinION (LU; GIORDANO; NING, 2016; MAGI et al., 2018) e, ao final, modelar o metabolismo em escala genômica e informar sobre agrupamentos gênicos biossintéticos. Em cada um dos passos do fluxo de trabalho, descritos abaixo, a eficiência foi testada analisando os genomas de bactérias isoladas de ambientes marinhos no escopo do projeto Jovem Pesquisador (JP) FAPESP 17/18922-2 (ANDRADE, [s.d.]; SILVA et al., 2017). As etapas do fluxo de trabalho da ferramenta Pardal são ilustradas na Figura 3.

Figura 3 - Fluxo de trabalho da ferramenta Pardal.



Fonte: De autoria própria.

A) Atribuição de bases (*basecalling*)

A análise começa pela atribuição de bases do arquivo de sequenciamento inicial usando o conjunto de ferramentas *Dorado* (“Data analysis (DATD_5000_v1_revU_22Aug2016)”, 2017), ou, alternativamente, o *Guppy* (“Guppy protocol (GPB_2003_v1_revAX_14Dec2018)”, 2018). Isso converterá os dados de corrente elétrica dos arquivos .fast5 em arquivos .fastq, que incluem as leituras geradas pelo sequenciador, juntamente com detalhes sobre a qualidade de cada base individualmente.

Alguns passos diferenciam a atribuição realizada pelo *Dorado* da realizada pelo *Guppy*. Quando o *Dorado* é utilizado é necessário a conversão dos arquivos .fast5 em um arquivo .pod5, sendo necessário realizar essa operação com o pacote *python pod5* (“pod5”, [s.d.]) criado pela *Oxford Nanopore*.

B) Remoção de adaptadores e leituras ruins

Para eliminar sequências muito curtas ou de baixa qualidade, bem como adaptadores usados durante a preparação da biblioteca para sequenciamento, empregamos as ferramentas *chopper* (DE COSTER; RADEMAKERS, 2023) e *Porechop* (WICK, 2024) nos dados do *MinION*.

A etapa de remoção de adaptadores é automaticamente realizada pelo *Dorado*, sendo assim, por padrão, o uso da ferramenta *Porechop* é dispensado para atribuições geradas por essa ferramenta.

C) Verificar qualidade da atribuição de bases

O controle de qualidade pré-montagem dos dados de leituras longas Nanopore é executado através da ferramenta *NanoStat* (DE COSTER et al., 2018). Cada etapa anterior foi avaliada separadamente, gerando 3 tabelas no relatório final com as informações geradas.

D) Montagem

A montagem foi realizada por meio da ferramenta *Flye* (KOLMOGOROV et al., 2019) ou *Unicycler* (WICK et al., 2017), gerando um arquivo .fasta com o genoma montado. É possível, por meio de parâmetros personalizados, realizar um processamento dos dados pela ferramenta *Racon* (VASER et al., 2017), aumentando a qualidade da montagem com uma etapa adicional de polimento.

E) Verificar qualidade da montagem

De forma semelhante a etapa de verificação de qualidade anterior, nesse momento foram gerados relatórios de qualidade da montagem pelas ferramentas

BUSCO (SIMÃO et al., 2015) e *Quast* (“QUAST: quality assessment tool for genome assemblies - PubMed”, [s.d.]).

F) Anotação

A anotação no fluxo de trabalho foi gerada pela ferramenta *Prokka* (SEEMANN, 2014), encontrando regiões com atribuição funcional presentes no genoma, como regiões codificadoras de DNA (do inglês *coding dna sequence* - CDS) que levam a anotação de proteínas produzidas pela bactéria.

G) Predição de agrupamentos gênicos biossintéticos

Nessa etapa, o fluxo de trabalhos utiliza da ferramenta *antiSMASH* (BLIN et al., 2021) para realizar a predição de agrupamentos gênicos biossintéticos (do inglês *Biosynthetic Gene Clusters* - BGCs), que pode ser descrita como uma busca por regiões codificadoras de vias para a produção de metabólitos secundários, que são de grande interesse para aplicações biotecnológicas (MEDEMA et al., 2014).

H) Modelagem do metabolismo

O resultado que sai da anotação é usado como entrada da ferramenta *CarveMe* (MACHADO et al., 2018), responsável por gerar um modelo metabólico em escala genômica preliminar, permitindo a obtenção de informações em relação às vias metabólicas que são geradas pelas enzimas codificadas no genoma. Uma curadoria manual é usualmente necessária, uma vez que o modelo gerado pela ferramenta *CarveMe* é apenas uma base para o usuário com conhecimentos de modelagem consiga editar. Análises como a análise de balanço de fluxo, que consiste na otimização de uma via de interesse para a maior produção de um composto alvo na via, são possíveis com o modelo curado. Na documentação da ferramenta Pardal há instruções de como começar a inspeção manual do GEM, com ferramentas que permitem a edição e a visualização do modelo e artigos que explicam de forma detalhada a respeito da modelagem.

Além disso, Pardal faz uso da ferramenta *Memote* (LIEVEN et al., 2020), que gera um relatório acerca da qualidade do modelo metabólico, disponibilizando uma fonte alternativa de informações sobre como iniciar a curadoria do modelo.

I) Geração de relatório

Após a execução de todas as etapas definidas, se têm a geração de um relatório no formato .HTML, utilizando um módulo *Python* desenvolvido no projeto. Os relatórios das análises geradas pelo fluxo de trabalho Pardal ao longo do projeto podem ser convenientemente disponibilizados em uma interface *web*.

Muitas etapas são necessárias para se obter um modelo metabólico em escala genômica, sendo a ferramenta Pandal capaz de criar os arquivos necessários para que um modelo seja gerado. O fluxo de trabalho Pandal é uma das formas possíveis de se obter um GEM, seja a utilizando ou seguindo a sequência de etapas descritas em seu fluxo de trabalho usando cada ferramenta apresentada. É possível obter um GEM de inúmeras formas diferentes, adicionando ou retirando etapas, trocando as ferramentas ou partindo de um ponto diferente, mas iniciando de um genoma recém sequenciado, as etapas em sua essência serão semelhantes às da ferramenta Pandal. Ademais, existem diversas ferramentas além do *CarveMe* para produzir um modelo metabólico em escala genômica preliminar, algumas delas são: *GEM system*, *merlin*, *Model SEED*, *Pathway tools*, *Raven toolbox* e *SuBLiMinal toolbox* (FARIA et al., 2018).

No entanto, mesmo que uma ferramenta seja capaz de desenvolver um GEM automaticamente partindo de seu genoma ou anotação, como no caso do *CarveMe*, no estado atual dos algoritmos de modelagem ainda não é possível confiar totalmente em um GEM produzido automaticamente, sendo necessário etapas de curadoria manual do modelo. Um modelo refinado manualmente e com uma boa confiabilidade é pauta de tal relevância como no caso da construção do GEM para a bactéria *Helicobacter pylori* (SCHILLING et al., 2002).

Tabela 1 - Meio padrão M9 utilizado pela ferramenta CarveMe.

medium	description	compound	name
M9	M9 minimal medium	ca2	Ca2+
M9	M9 minimal medium	cl	Cl-
M9	M9 minimal medium	cobalt2	Co2+
M9	M9 minimal medium	cu2	Cu2+
M9	M9 minimal medium	fe2	Fe2+
M9	M9 minimal medium	fe3	Fe3+
M9	M9 minimal medium	glc__D	D-Glucose
M9	M9 minimal medium	h2o	H2O
M9	M9 minimal medium	h	H+
M9	M9 minimal medium	k	K+

M9	M9 minimal medium	mg2	Mg
M9	M9 minimal medium	mn2	Mn2+
M9	M9 minimal medium	mobd	Molybdate
M9	M9 minimal medium	na1	Na+
M9	M9 minimal medium	nh4	Ammonium
M9	M9 minimal medium	ni2	Ni2+
M9	M9 minimal medium	o2	O2
M9	M9 minimal medium	pi	Phosphate
M9	M9 minimal medium	so4	Sulfate
M9	M9 minimal medium	zn2	Zn2+

Fonte: Adaptado da tabela com os meios, disponível no GitHub do CarveMe.

Para criar os modelos preliminares, a ferramenta CarveMe faz uso de um meio padrão conhecido como M9. Esse meio fornece os nutrientes básicos necessários para o crescimento microbiano, permitindo que a ferramenta construa modelos metabólicos iniciais ao identificar as vias essenciais para a sobrevivência do organismo no ambiente definido.

Os modelos preliminares fomentados pelo *Pardal* foram verificados com a ferramenta *Memote*, verificando a consistência, completude e conformidade dos modelos. A ferramenta *web Fluxer* (HARI; LOBO, 2020) foi utilizada para obter visualizações dos modelos e o pacote python *CobraPy* (EBRAHIM et al., 2013) para as simulações dos modelos preliminares. O aplicativo *OptFlux* (ROCHA et al., 2010) também foi utilizado para verificar quais reações são críticas para o funcionamento do metabolismo da bactéria que foi utilizada para as simulações.

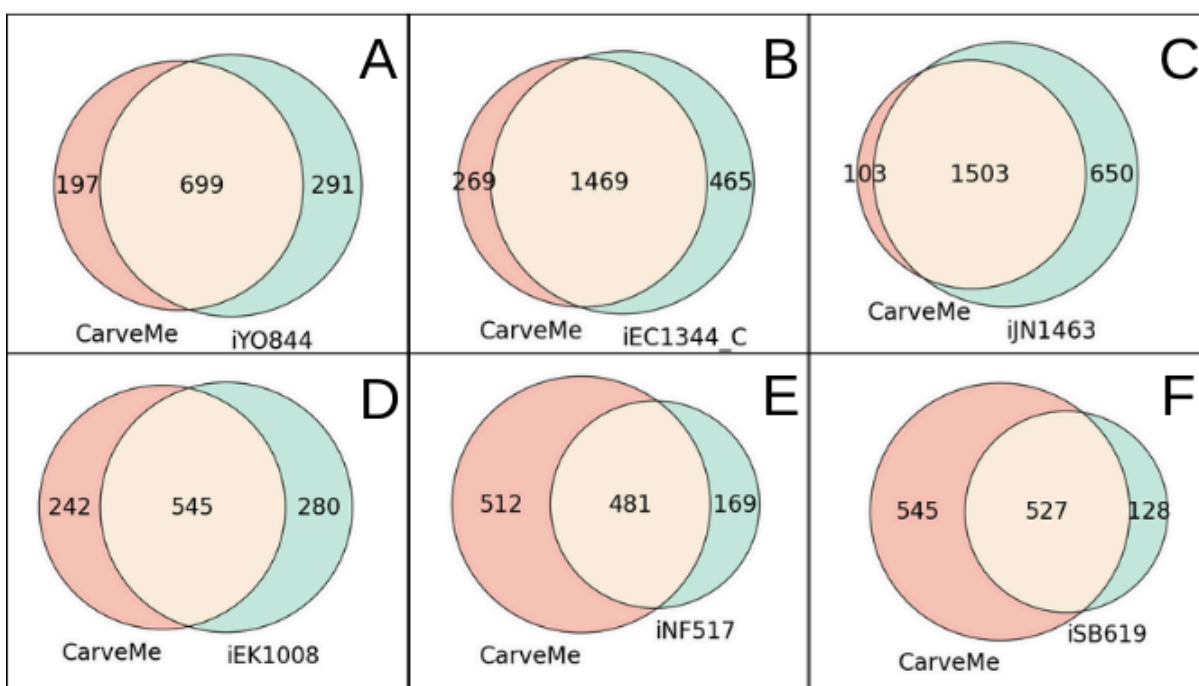
4. RESULTADOS E DISCUSSÃO

4.1. COMPARAÇÃO DO CARVEME COM MODELOS CURADOS MANUALMENTE

Para comparar os modelos criados pelo *CarveMe* foram utilizados diferentes modelos curados, para diferentes bactérias, disponíveis na literatura. Os seguintes modelos foram utilizados para a comparação: A) Modelo iYO844 (YK et al., 2007) B) Modelo iEC1344_C (MONK et al., 2016) C) Modelo iJN1463 (NOGALES et al., 2020) D) Modelo iEK1008 (KAVVAS et al., 2018) E) Modelo iNF517 (FLAHAUT et al., 2013) F) Modelo iSB619 (BECKER; PALSSON, 2005). A partir do genoma das bactérias que correspondem a esses modelos curados, o fluxo de trabalho Pardal foi executado a partir da etapa de anotação (Figura 3, F). Os modelos gerados foram refinados por meio da ferramenta *CarveMe* e comparados com os modelos curados. Os resultados são apresentados nas Figuras 4 e 5.

A Figura 4 ilustra o número de metabólitos comum entre os modelos. A média dos metabólitos comuns entre os modelos originados por Pardal/*CarveMe* e os modelos curados foi de 54% dos metabólitos obtidos pelas duas abordagens. Os modelos B e C possuem as maiores sobreposições, ambos com aproximadamente 67% de sobreposição. É interessante notar que para os genomas E e F Pardal/*CarveMe* atribuíram um número significativamente maior de metabólitos ao modelo. Uma hipótese, para explicar essa observação, é que atualizações nos bancos de dados utilizados para anotação do genoma tenham capturado novas funções gênicas que estavam ausentes quando os modelos foram inicialmente construídos. Essa observação necessita de uma inspeção manual para confirmação de novos metabólitos.

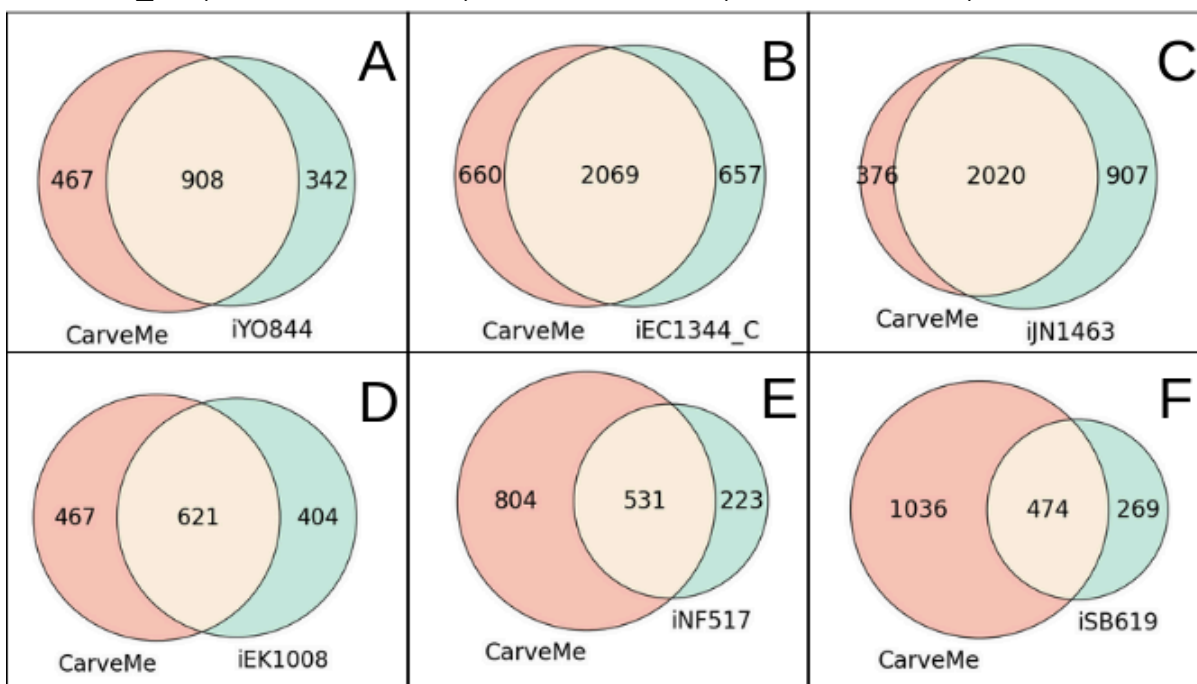
Figura 4 - Diagramas de venn comparando o número de metabólitos dos modelos gerados pela ferramenta *CarveMe* com diferentes modelos curados A) Modelo iYO844 B) Modelo iEC1344_C C) Modelo iJN1463 D) Modelo iEK1008 E) Modelo iNF517 F) Modelo iSB619.



Fonte: De autoria própria.

De forma semelhante, a Figura 5 mostra as reações compartilhadas entre os modelos proporcionados por Pardal/*CarveMe* e os modelos curados. Seguindo a mesma proporção, em média 46% das reações são compartilhadas entre as duas abordagens, sendo que a maior proporção de reações compartilhadas foi observada para os modelos B e C, ambos com aproximadamente 61% de sobreposição. Os modelos E e F originados por Pardal/*CarveMe* também apresentaram um número significativamente maior de reações.

Figura 5 - Diagramas de venn comparando o número de reações dos modelos gerados pela ferramenta *CarveMe* com os diferentes modelos curados. A) Modelo iYO844 B) Modelo iEC1344_C C) Modelo iJN1463 D) Modelo iEK1008 E) Modelo iNF517 F) Modelo iSB619.



Fonte: De autoria própria.

Considerando que os modelos manualmente inspecionados contêm o número esperado de metabólitos e reações, as proporções recuperadas pelo *Pardal/CarveMe* são apresentadas na Tabela 1.

Tabela 2 - Informações gerais sobre a comparação entre os modelos curados e os modelos gerados pela ferramenta *CarveMe*.

Bactéria	Modelo	Metabólitos totais	Metabólitos iguais	Cobertura metabólitos (%)	Reações totais	Reações iguais	Cobertura reações (%)
Bacillus subtilis	iYO844	990	699	70,61	1250	908	72,64
Escherichia coli	iEC1344_C	1934	1469	75,96	2726	2069	75,90
Pseudomonas putida	iJN1463	2157	1503	69,68	2927	2020	69,01
Mycobacterium tuberculosis	iNJ661	825	545	66,06	1025	621	60,59
Lactococcus lactis	iNF517	650	481	74,00	754	531	70,42

Fonte: De autoria própria.

A média de cobertura de metabólitos foi de 72,79% e a média de cobertura de reações foi de 68,72%, isso demonstra que há uma cobertura considerável realizada pelo *CarveMe*. Porém, ainda são necessárias melhorias para que haja uma maior

cobertura dos modelos curados, para afirmar que um modelo gerado por uma ferramenta automática não precisa de uma etapa de curagem manual. Além disso, diversos outros aspectos de um modelo criado automaticamente indicam a necessidade de uma curadoria manual para os modelos gerados pelo *CarveMe*. Um dos principais é o fato da função objetivo padrão, gerado para a produção de biomassa pela bactéria, ser extremamente alta, indicando que há discrepâncias no modelo, sendo necessário verificar cada reação que está sendo desenvolvida, separar as que fazem sentido para o organismo, utilizando dados da anotação e revisão de literatura, e adicionar outras reações que não foram encontradas e que se espera que ocorram no organismo.

4.2. ANÁLISE DOS GEMs CRIADOS A PARTIR DE DADOS ORIGINAIS

Os genomas de bactérias isoladas de ambientes marinhos no escopo do projeto Jovem Pesquisador (JP) FAPESP 17/18922-2 (ANDRADE, [s.d.]; SILVA et al., 2017)), foram processados pelo fluxo de trabalho *Pardal* e foram utilizados para criação de GEMs preliminares. Oito genomas passaram por todas as etapas da ferramenta *Pardal*, obtendo *outputs* relacionados a sua montagem, anotação, predição de agrupamentos gênicos biossintéticos e modelagem metabólica em escala genômica. Os *scores* dos GEMs preliminares criados são apresentados na tabela a seguir:

Tabela 3 - *Scores* gerado pela ferramenta *Memote* dos GEMs das bactérias do projeto.

Bactéria	<i>Memote Score</i> (%)
BRA_006	69
BRA_177	68
BRB_041	69
BRB_042	69
BRB_234	71
BRB_291	69
BRB_294	69
BRB_341	87

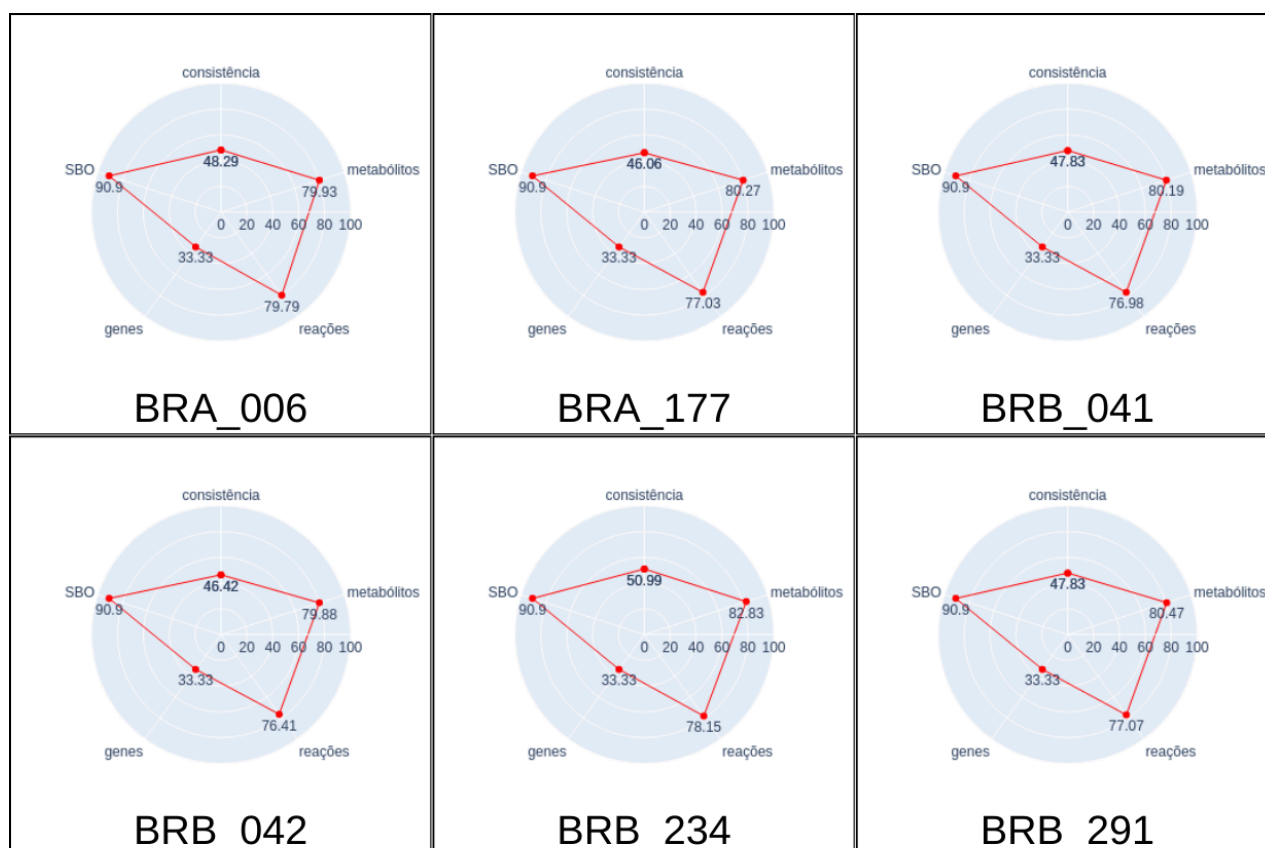
Fonte: De autoria própria.

O *score* médio foi de 71,37%.

Para calcular os *scores* o *Memote* utiliza cinco parâmetros com pesos diferentes para avaliar a qualidade do modelo, sendo esses: Consistência, qualidade da anotação de metabólitos, reações e genes e presença de termos ligados a ontologia de biologia de sistemas (do inglês *Systems Biology Ontology* - SBO)(LIEVEN et al., 2020).

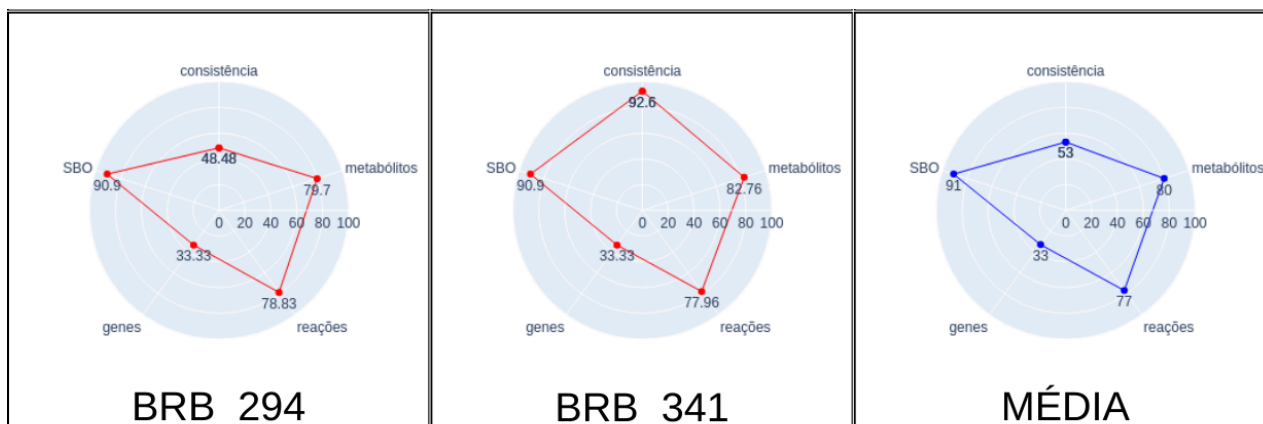
A consistência é medida verificando a integridade dos valores relacionados às massas e cargas e a consistência estequiométrica (GEVORGYAN; POOLMAN; FELL, 2008), enquanto que os *scores* das anotações se referem a presença dos identificadores das anotações em bancos de dados conhecidos. Abaixo, nas figuras 6 e 7, está cada valor dado a esses parâmetros para os GEMs das bactérias do projeto:

Figura 6 - Gráficos de radar com os parâmetros calculados pela ferramenta *Memote*, para as bactérias BRA_006, BRA_177, BRB_041, BRB_042, BRB_234 e BRB_291.



Fonte: De autoria própria.

Figura 7 - Gráficos de radar com os parâmetros calculados pela ferramenta *Memote*, para as bactérias BRB_294 e BRB_341. O último gráfico de radar se refere a média de cada valor entre todas as bactérias.



Fonte: De autoria própria.

Com a análise dos gráficos de radar é possível verificar que as 8 bactérias obtiveram *scores* muito similares em todos os parâmetros, exceto a BRB_341, que também possui um *score* geral maior do que as outras (ver tabela 2).

O valor que apresentou diferença entre as outras bactérias e a BRB_341 é o relacionado a consistência, o que demonstra que a bactéria BRB_341 tem uma consistência calculada maior que as outras pelo *Memote*. Uma hipótese para essa diferença é o fato da BRB_341 ser a mais adaptada ao meio padrão utilizado pelo *CarveMe* para a construção dos GEMs, sendo que a utilização de meios de cultura personalizados para as outras bactérias provavelmente iriam corrigir esse problema.

4.3. INVESTIGAÇÃO DA PRODUÇÃO DE BIOFÁRMACOS USANDO OS GEMs PRELIMINARES

No artigo publicado sobre a BRA_177 (SILVA et al., 2017) compostos anticâncer foram prospectados nessa bactéria, porém, no modelo metabólico gerado pelo *CarveMe*, nenhuma via de formação desses compostos foi obtida, isso se deve, provavelmente, ao fato das vias para a formação não estarem presente na base de dados do *BIGG* que é a utilizada pelo *CarveMe* para gerar os modelos.

Uma forma de começar a compreender como esses compostos poderiam ser adicionados aos modelos, seria utilizar a anotação fornecida da ferramenta *Prokka* (Figura 3, etapa F). O resultado da anotação (Tabela 3) contém as enzimas e os

genes relacionados à formação de compostos de interesse, que não estão mapeados nas vias metabólicas do *BIGG*, como a prodigiosina citada no artigo.

Tabela 4 - Sequências codificadoras de proteínas (CDSs) relacionadas à produção de prodigiosina anotadas utilizando a ferramenta *Prokka*.

locus_tag	ftype	length_bp	product	gene	EC_number	COG
OPDGNOEB_09783	CDS	1512	Prodigiosin synthesizing transferase PigC	pigC_1	6.4.-.-	no value
OPDGNOEB_09784	CDS	1197	Prodigiosin synthesizing transferase PigC	pigC_2	6.4.-.-	no value

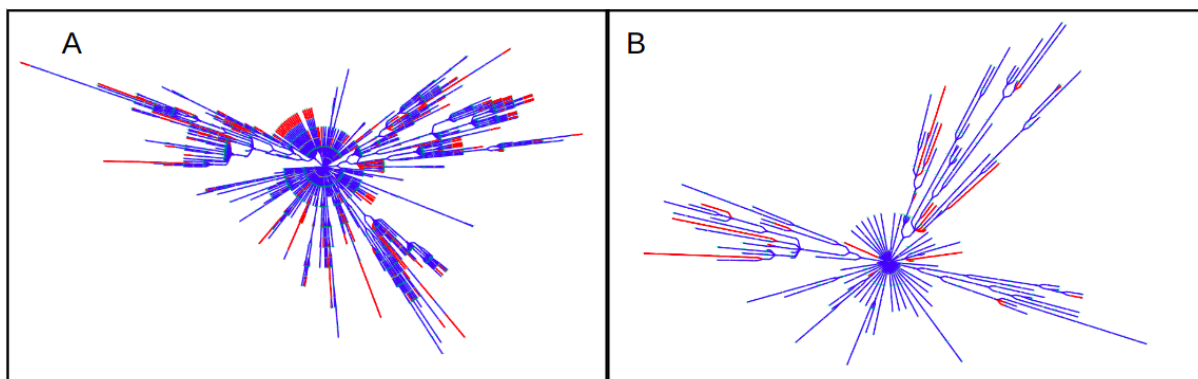
Fonte: De autoria própria.

Dessa forma é possível procurar o gene relacionado a produção da prodigiosina, procurar a reação que leva a formação desse composto e propor conexões às vias metabólicas presentes no modelo preliminar da bactéria *BRA_177*. Como mencionado anteriormente, o modelo necessariamente precisa ser refinado manualmente, levando em consideração informações do metabolismo, preferencialmente validadas experimentalmente. Porém, com um modelo curado e com a reação de formação de prodigiosina integrada, seria possível realizar um FBA com a prodigiosina e identificar qual a condição que otimizaria a produção desse composto, por exemplo.

Para a bactéria *BRB_341*, foi possível identificar a produção de ácido succínico, um composto com potencial farmacológico devido a suas propriedades antimicrobianas (NGHIEM; KLEFF; SCHWEGMANN, 2017), tanto pela anotação com *Prokka*, como por sua presença no GEM preliminar criado pelo *CarveMe*. Nesse caso o composto alvo possui sua via presente no banco de dados *BIGG*, sendo possível simular sua otimização com métodos como o FBA utilizando o pacote *CobraPy*.

O primeiro passo foi visualizar o modelo desenvolvido utilizando a ferramenta *Fluxer*, a via disponibilizada preliminarmente é ilustrada na Figura 8, sendo que ao aproximar a imagem no visualizador *web* observamos todas as reações e metabólitos encontrados pelo *CarveMe* (Figura 8A). Também é possível remover todas as reações com fluxo igual a zero, facilitando a exploração do modelo (Figura 8B).

Figura 8 - A) Modelo metabólito da bactéria BRB_341 aberto na plataforma *web Fluxer*. B) Modelo metabólito da bactéria BRB_341 depois de retirar todas as reações em que o fluxo é zero.

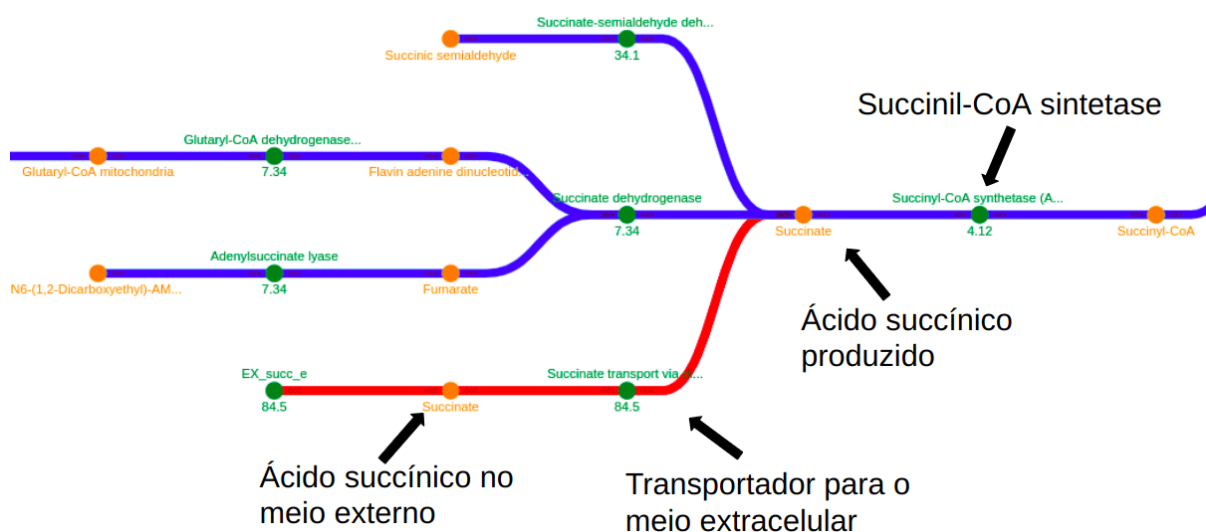


Fonte: De autoria própria.

A primeira informação obtida analisando a Figura 8 é quanto a grande diferença entre a via com os fluxos zerados e sem os fluxos zerados, em que grande parte dos caminhos são cortados. Isso indica a necessidade de uma verificação manual para confirmar se essas reações com fluxo zero realmente fazem parte do metabolismo do organismo.

Aproximando a via é possível separar um trecho com o caminho para a formação e exportação para o meio externo do ácido succínico (Figura 9).

Figura 9 - Visão detalhada da via de produção do ácido succínico obtida pela ferramenta *Fluxer*.



Fonte: De autoria própria.

Na forma da via de produção do ácido succínico obtida no presente modelo é possível verificar, que para realização de um FBA visando maximizar o fluxo de produção de biomassa da bactéria, o fluxo do ácido succínico pelo transportador

para o meio extracelular é de 84,5 mmol gDW⁻¹ h⁻¹, sendo esse o fluxo do ácido succínico que é secretado pela bactéria em condições ótimas. Também é possível verificar, que o ácido succínico produzido está sendo utilizado para a produção de Succinil-CoA pela enzima Succinil-CoA sintetase, levando a uma perda de fluxo de 4,12 gDW⁻¹ h⁻¹ para essa produção.

Visando verificar esse aumento de fluxo, o *knockout* da reação realizada pela Succinyl-CoA sintetase foi realizado utilizando o pacote *python CobraPy*, levando ao aumento do fluxo para 88,62 gDW⁻¹ h⁻¹. Isso possivelmente levaria a uma maior produção de ácido succínico pela bactéria, sendo que o *knockout* do gene que codifica essa enzima poderia aumentar a produção em um contexto industrial.

Figura 10 - Código utilizado para o *knockout* da reação realizada pela enzima succinil-CoA sintetase e verificação da mudança do fluxo.

```
#Verificando o fluxo do metabólito em condições ótimas
modelo.metabolites.succ_e.summary()
```

SUCC_e

C4H4O4

Fluxo de ácido succínico em
84.51 gDW⁻¹ h⁻¹

Producing Reactions

Percent	Flux	Reaction	Definition
100.00%	84.51	SUCct	succ_e <=> succ_c

```
#Realizando o knockout da reação encontrada pela aplicação fluxer realizada
#pela enzima Succinil-CoA Sintetase
modelo.reactions.SUCOAS.knock_out()
modelo.optimize()
#Verificando o fluxo de ácido succínico em condições ótimas após o knockout
modelo.metabolites.succ_e.summary()
```

SUCC_e

C4H4O4

Fluxo de ácido succínico aumentado
para 88.62 gDW⁻¹ h⁻¹ após o knockout

Producing Reactions

Percent	Flux	Reaction	Definition
100.00%	88.62	SUCct	succ_e <=> succ_c

Fonte: De autoria própria.

Além disso, a ferramenta *OptFlux* foi utilizada para verificar se a reação relacionada a enzima Succinil-CoA sintetase é crítica para o fluxo de biomassa.

Essa verificação é baseada em testes que realizam os *knockouts* de todas as reações separadamente e verifica se ainda há a geração de biomassa. Dessa forma, foi verificado que essa reação não está entre as reações críticas para o desenvolvimento da bactéria.

5. CONCLUSÃO

A modelagem metabólica em escala genômica é uma ferramenta poderosa para otimizar o uso de microrganismos na produção de compostos com potencial farmacêutico. No entanto, os modelos iniciais, como os desenvolvidos pelo *CarveMe*, precisam ser refinados manualmente de maneira criteriosa, pois podem conter lacunas devido à dependência de bases de dados incompletas, como o BIGG.

Um desafio comum é a identificação incorreta ou incompleta de vias metabólicas relevantes para compostos bioativos. Ferramentas de anotação como o *Prokka* são fundamentais para complementar o modelo, identificando proteínas, enzimas, e, indiretamente os metabólitos produzidos por elas, que passam despercebidos, ou que foram incorretamente atribuídos pela modelagem automatizada. Além disso, modelos preliminares podem sugerir vias metabólicas com fluxo zero, indicando a necessidade de validação mais aprofundada. Ferramentas como o *Fluxer* ajudam a explorar essas rotas, mas a interpretação e ajustes manuais são essenciais para garantir previsões realistas.

O processo de modelagem exige constante refinamento e validações experimentais, a integração dessas informações permite aumentar a precisão das simulações e tornar os modelos mais precisos para predizer o funcionamento de um organismo em testes *in vivo*. Dessa forma, a modelagem metabólica se torna uma plataforma promissora para acelerar descobertas e direcionar esforços para a produção eficiente de biofármacos.

Neste contexto, o presente trabalho contribuiu significativamente ao construir modelos metabólicos preliminares de bactérias marinhas, oferecendo uma base inicial para estudos futuros que visem o refinamento desses modelos e sua aplicação na bioprospecção de compostos bioativos. Além disso, ao explorar e documentar os desafios e limitações do processo, a pesquisa reforça a importância da modelagem metabólica como uma ferramenta promissora no desenvolvimento de biofármacos, destacando o potencial das bactérias marinhas como fontes ainda pouco exploradas de inovação farmacêutica.

BIBLIOGRAFIA

ANDRADE, Lucas Souza Neves. Identificação e Caracterização de clusters biossintéticos de *Micromonospora* sp. [s.d.].

BECKER, Scott.; PALSSON, Bernhard. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC microbiology*, v. 5, p. 8, 7 mar. 2005.

BLIN, Kai. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research*, v. 49, n. W1, p. W29–W35, 2 jul. 2021.

BRANDÃO, Conceição Zulmira Gomes de Sousa; SOUZA, José Natal. DE. BIOFÁRMACOS: DA PESQUISA AO MERCADO: UMA REVISÃO DA LITERATURA. *SAÚDE & CIÊNCIA EM AÇÃO*, v. 1, n. 1, p. 105–118, 2015.

COONEY, Charles. Bioreactors: Design and Operation. *Science*, v. 219, n. 4585, p. 728–733, 11 fev. 1983.

Data analysis (DATD_5000_v1_revU_22Aug2016). Disponível em: <<https://nanoporetech.com/document/data-analysis>>. Acesso em: 10 out. 2024.

DE COSTER, Wouter. et al. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics (Oxford, England)*, v. 34, n. 15, p. 2666–2669, 1 ago. 2018.

DE COSTER, Wouter.; RADEMAKERS, Rosa. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*, v. 39, n. 5, p. btad311, 1 maio 2023.

EBRAHIM, Ali. et al. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, v. 7, n. 1, p. 74, 8 ago. 2013.

FARIA, José. et al. Methods for automated genome-scale metabolic model reconstruction. *Biochemical Society Transactions*, v. 46, n. 4, p. 931–936, 31 jul. 2018.

FLAHAUT, Nicolas. et al. Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. *Applied Microbiology and Biotechnology*, v. 97, n. 19, p. 8729–8739, out. 2013.

GEVORGYAN, Albert.; POOLMAN, Mark.; FELL, David. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, v. 24, n. 19, p. 2245–2251, 1 out. 2008.

GU, Changdai. et al. Current status and applications of genome-scale metabolic models. *Genome Biology*, v. 20, n. 1, p. 121, dez. 2019.

Guppy protocol (GPB_2003_v1_revAX_14Dec2018). Disponível em: <<https://nanoporetech.com/document/Guppy-protocol>>. Acesso em: 10 out. 2024.

HARI, Archana.; LOBO, Daniel. Fluxer: a web application to compute, analyze and visualize genome-scale metabolic flux networks. *Nucleic Acids Research*, v. 48, n. W1, p. W427–W435, 2 jul. 2020.

HEATHER, James.; CHAIN, Benjamin. The sequence of sequencers: The history of sequencing DNA. *Genomics*, v. 107, n. 1, p. 1–8, jan. 2016.

KANEHISA, Minoru. The KEGG Database. Em: 'In Silico' Simulation of Biological Processes. [s.l.] John Wiley & Sons, Ltd, 2002. p. 91–103.

KARP, Peter. et al. The MetaCyc Database. *Nucleic Acids Research*, v. 30, n. 1, p. 59–61, 1 jan. 2002.

KARP, Peter. et al. The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, v. 20, n. 4, p. 1085–1093, 19 jul. 2019.

KAVVAS, Erol. et al. Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC systems biology*, v. 12, n. 1, p. 25, 2 mar. 2018.

KENEFAKE, Dustin. et al. An improved algorithm for flux variability analysis. *BMC Bioinformatics*, v. 23, n. 1, p. 550, 19 dez. 2022.

KOLMOGOROV, Mikhail. et al. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, v. 37, n. 5, p. 540–546, maio 2019.

KONNO, Naoki.; IWASAKI, Wataro. Machine learning enables prediction of metabolic system evolution in bacteria. *Science Advances*, v. 9, n. 2, p. eadc9130, 11 jan. 2023.

LARDIES, Marco.; CATALÁN, Tamara.; BOZINOVIC, Francisco. Metabolism and life-history correlates in a lowland and highland population of a terrestrial isopod. *Canadian Journal of Zoology*, v. 82, n. 5, p. 677–687, maio 2004.

LIEVEN, C. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, v. 38, n. 3, p. 272–276, mar. 2020.

LO, C. H.; MCADAM, Ewan.; JUDD, Simon. The cost of a small membrane bioreactor. *Water Science and Technology*, v. 72, n. 10, p. 1739–1746, 23 jul. 2015.

LU, Hengyun.; GIORDANO, Francesca.; NING, Zemin. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, v. 14, n. 5, p. 265–279, out. 2016.

MACHADO, Daniel. et al. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, v. 46, n. 15, p. 7542–7553, 6 set. 2018.

MAGI, Alberto. et al. Nanopore sequencing data analysis: state of the art, applications and challenges. *Briefings in Bioinformatics*, v. 19, n. 6, p. 1256–1272, 27 nov. 2018.

MARWICK, Jamie. D.; WRIGHT, Philip. C.; BURGESS, J. G. Bioprocess Intensification for Production of Novel Marine Bacterial Antibiotics Through Bioreactor Operation and Design. *Marine Biotechnology*, v. 1, n. 5, p. 495–507, 1 set. 1999.

MEDEMA, Marnix. et al. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLOS Computational Biology*, v. 10, n. 12, p. e1004016, de dez. de 2014.

MONK, Jonathan. et al. Multi-omics Quantification of Species Variation of *Escherichia coli* Links Molecular Features with Strain Phenotypes. *Cell Systems*, v. 3, n. 3, p. 238- 251.e12, 28 set. 2016.

NGHIEM, Nhuan.; KLEFF, Susanne.; SCHWEGMANN, Stephan. Succinic Acid: Technology Development and Commercialization. *Fermentation*, v. 3, n. 2, p. 26, jun. 2017.

NIRAULA, Narayan. et al. Biotechnological doxorubicin production: pathway and regulation engineering of strains for enhanced production. *Applied Microbiology and Biotechnology*, v. 87, n. 4, p. 1187–1194, 1 jul. 2010.

NOGALES, Juan. et al. High-quality genome-scale metabolic modelling of *Pseudomonas putida* highlights its broad metabolic capabilities. *Environmental Microbiology*, v. 22, n. 1, p. 255–269, jan. 2020.

ORTH, Jeffrey. D.; THIELE, Ines.; PALSSON, Bernhard. What is flux balance analysis? *Nature Biotechnology*, v. 28, n. 3, p. 245–248, mar. 2010.

pod5: Oxford Nanopore Technologies Pod5 File Format Python API and Tools. , [s.d.]. Disponível em: <<https://github.com/nanoporetech/pod5-file-format>>. Acesso em: 10 out. 2024
QUAST: quality assessment tool for genome assemblies - PubMed. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/23422339/>>. Acesso em: 10 out. 2024.

RAMAN, Karthik.; CHANDRA, Nagasuma. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, v. 10, n. 4, p. 435–449, 1 jul. 2009.

REED, Jennifer. Genome-Scale Metabolic Modeling and Its Application to Microbial Communities. Em: *The Chemistry of Microbiomes: Proceedings of a Seminar Series*. [s.l.] National Academies Press (US), 2017.

RICHELLE, Anne. et al. Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. *npj Systems Biology and Applications*, v. 6, n. 1, p. 1–5, 13 mar. 2020.

ROCHA, Isabel. et al. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, v. 4, n. 1, p. 45, 19 abr. 2010.

SANTOS, F.; BOELE, J.; TEUSINK, B. Chapter twenty-four - A Practical Guide to Genome-Scale Metabolic Models and Their Analysis. Em: JAMESON, D.; VERMA, M.; WESTERHOFF, H. V. (Eds.). *Methods in Enzymology. Methods in Systems Biology*. [s.l.] Academic Press, 2011. v. 500p. 509–532.

SCHILLING, Christophe. et al. Genome-Scale Metabolic Model of *Helicobacter pylori* 26695. *Journal of Bacteriology*, v. 184, n. 16, p. 4582–4593, 15 ago. 2002.

SEEMANN, Torsten. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, v. 30, n. 14, p. 2068–2069, 15 jul. 2014.

SILVA, Amaro. et al. Bioprospecting Anticancer Compounds from the Marine-Derived Actinobacteria *Actinomadura* sp. Collected at the Saint Peter and Saint Paul Archipelago (Brazil). *Journal of the Brazilian Chemical Society*, v. 28, p. 465–474, mar. 2017.

SIMÃO, Felipe. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)*, v. 31, n. 19, p. 3210–3212, 1

out. 2015.

STALIDZANS, Egils. et al. Model-based metabolism design: constraints for kinetic and stoichiometric models. *Biochemical Society Transactions*, v. 46, n. 2, p. 261–267, 22 fev. 2018.

STEIN, Lincoln. Genome annotation: from sequence to biology. *Nature Reviews Genetics*, v. 2, n. 7, p. 493–503, jul. 2001.

STEPHANOPOULOS, George.; ARISTIDOU, Aristos.; NIELSEN, Jens. *Metabolic Engineering: Principles and Methodologies*. [s.l.] Elsevier, 1998.

VASER, Robert. et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, v. 27, n. 5, p. 737–746, maio 2017.

WALLEY, Justin. et al. Integration of omic networks in a developmental atlas of maize. *Science*, v. 353, n. 6301, p. 814–818, 19 ago. 2016.

WICK, Ryan. rrwick/Porechop. , 9 out. 2024. Disponível em: <<https://github.com/rrwick/Porechop>>. Acesso em: 10 out. 2024.

WICK, Ryan. et al. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, v. 13, n. 6, p. e1005595, 8 jun. 2017.

OH, You-Kwan. et al. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *The Journal of biological chemistry*, v. 282, n. 39, 28 set. 2007.

ZELEZNIAK, Aleksej. et al. Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. *Cell Systems*, v. 7, n. 3, p. 269- 283.e6, 26 set. 2018.

ZHANG, Cheng.; HUA, Qiang. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. *Frontiers in Physiology*, v. 6, 7 jan. 2016.

ZUÑIGA, Cristal.; ZARAMELA, Livia.; ZENGLER, Karsten. Elucidation of complexity and prediction of interactions in microbial communities. *Microbial Biotechnology*, v. 10, n. 6, p. 1500–1522, 2017.



INFORMAÇÃO GRAD 0019/2025

Nome do Interessado(a): Henrique Marcel Yudi de Oliveira Tsuji

Assunto: TCC - Autorização para Biblioteca Digital e Aprovação CG

A Comissão de Graduação da FCFRP, órgão que atua como coordenação do curso Farmácia desta, declara para os devidos fins que:

- Este trabalho foi apresentado e aprovado no dia 29/11/2024;
- O(a) estudante autorizou a reprodução e divulgação total deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte pelo formulário entregue a esta comissão após aprovação.

Ribeirão Preto, 23 de julho de 2025

Comissão de Graduação FCFRP/USP