

**UNIVERSIDADE FEDERAL DE VIÇOSA**

**WESLEY BAILON EVANGELISTA**

**CONSTRUÇÃO DE UM MODELO METABÓLICO COMPUTACIONAL  
CONTEXTO-ESPECÍFICO PARA *Solanum lycopersicum***

**VIÇOSA - MINAS GERAIS**

**2025**

**WESLEY BAILON EVANGELISTA**

**CONSTRUÇÃO DE UM MODELO METABÓLICO COMPUTACIONAL  
CONTEXTO-ESPECÍFICO PARA *Solanum lycopersicum***

Monografia, apresentada ao Curso de Ciências  
Biológicas da Universidade Federal de Viçosa  
como requisito para obtenção do título de  
bacharel em Biologia.

Orientador: Tiago Antônio de Oliveira Mendes

Coorientador: Filipe Augusto Teixeira

**VIÇOSA - MINAS GERAIS**

**2025**

**WESLEY BAILON EVANGELISTA**

**CONSTRUÇÃO DE UM MODELO METABÓLICO COMPUTACIONAL  
CONTEXTO-ESPECÍFICO PARA *Solanum lycopersicum***

Trabalho de Conclusão de Curso apresentado  
ao Curso de Ciências Biológicas para obtenção  
do título acadêmico de Bacharel em Biologia,  
sob orientação do Prof. Tiago Antônio de  
Oliveira Mendes.

APROVADO: 15 Janeiro de 2025.

Assentimento:

Documento assinado digitalmente  
 **WESLEY BAILON EVANGELISTA**  
Data: 31/01/2025 19:07:39-0300  
Verifique em <https://validar.iti.gov.br>

---

Wesley Bailon Evangelista  
Autor

Documento assinado digitalmente  
 **TIAGO ANTONIO DE OLIVEIRA MENDES**  
Data: 03/02/2025 12:53:27-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Tiago Antônio de Oliveira Mendes  
Orientador

## Agradecimentos

Primeiramente, agradeço a Deus por todas as oportunidades concedidas ao longo desta caminhada e por estar sempre abençoando meu percurso.

Agradecer ao meus pais, que foram, sem dúvidas, a base para suportar todos os obstáculos e desafios enfrentados nesta fase. A minha mãe, Nilza, por toda confiança, suporte e apoio depositados em mim. Ao meu falecido pai, Daniel, por todo suporte e ajuda em vida; como eu gostaria que o senhor estivesse presente nesta conquista. Agradeço também a minha irmã, Jessica, e meu sobrinho, Gean, por toda torcida e apoio.

A meu tio Amarildo e meu amigo Viuber, por todo suporte, minha gratidão a vocês é eterna. Agradecer também aos demais familiares que sempre torceram pelo meu sucesso.

Aos amigos do 2132, tanto os que passaram - Jonas, Israel, Alor, Rodrigo, Guilherme, Matheus, Carlos – quanto os presentes - Higor, Paulo, Gustavo, Douglas, Edvan, Pedro, Lucas - obrigado por todos os momentos compartilhados. Suas amizades foram essenciais durante esses anos.

Agradecer em especial a minha amiga Michele, por toda ajuda ao longo desses anos, sua amizade tornou esta caminhada um pouco menos difícil. A minha amiga Gisele, pela amizade e companheirismo durante todos esses anos. Agradecimento às minhas amigas da biologia, Carol e Maria Eduarda, pelo companheirismo durante os anos que estivemos juntos no curso, vocês fizeram as aulas mais divertidas. A minha amiga Marisa, por todo companheirismo e torcida. Aos demais amigos da biologia e de outros cursos, que não foram citados, cada um de vocês foram especiais.

Agradeço a professora Karla, pela oportunidade que me deu de realizar meu primeiro estágio, meu carinho e gratidão por você serão eternos.

Agradeço em especial, ao professor Tiago, pela oportunidade de ser orientado por uma pessoa que tanto admiro. Sou profundamente grato pelas oportunidades concedidas durante o tempo em que fui seu estagiário. Serei eternamente grato ao senhor. Agradecer ao Filipe por ter aceitado ser meu coorientador. Aprendi muito com você durante esses anos e sou imensamente grato por tudo.

Ao pessoal do laboratório do LBM, agradeço por sempre me ajudarem e me receberam tão bem durante as minhas idas ao laboratório. Estarei na torcida pelo sucesso de cada um de vocês.

Enfim, quero agradecer a todos, que de alguma forma, contribuíram para que eu conseguisse alcançar meu objetivo.

*"A ciência é mais que um corpo de conhecimento, é uma forma de pensar, uma forma cética de interrogar o universo, com pleno conhecimento da falibilidade humana."*

*Carl Sagan*

## Resumo

O tomate (*Solanum lycopersicum*) é um dos vegetais mais consumidos e populares no mundo, apresentando uma grande diversidade de fruto, que variam desde cor, textura, forma e tamanho. Por isso ele tem sido amplamente estudado devido ao seu valor econômico, importância cultural e compostos benéficos à saúde humana. Além disso, o fruto é o principal modelo para estudo de metabolismo e no desenvolvimento de frutas frescas pois apresenta fácil cultivo e curtos ciclos de vida. A compreensão das vias metabólicas relacionadas a características comerciais importantes ainda não são tão bem esclarecidas, como é no caso do amadurecimento do fruto. Este é um processo altamente complexo, envolvendo drásticas mudanças durante todo o processo. Um melhor entendimento dos processos bioquímicos associados ao amadurecimento pode viabilizar novas abordagens a serem utilizadas na conservação pós-colheita do fruto. A estratégia “Genome-scale metabolic model” (GEM) tem possibilitado o desenvolvimento de modelos *in silico* por meio da construção e simulação de um modelo metabólico, resultando em um maior entendimento do funcionamento das redes bioquímicas. O GEM *i* HY3410 é um modelo reconstruído para o *Solanum lycopersicum* e foi utilizado como ponto de partida para a integração de dados de transcriptoma, com o intuito de gerar um modelo contexto-específico para compreender as principais vias e enzimas envolvidas no amadurecimento do fruto ao longo de todo o estágio de maturação. A integração de dados de transcriptoma a um GEM é etapa crucial dessa estratégia. Para isso, foi selecionado um conjunto de dados de transcriptoma do tecido pericarpo de tomate no tempo medium green (MG - verde maduro). O conjunto de dados forma submetidos a pré-processamento e análise de qualidade antes de serem implementadas ao modelo. Por meio do pacote Gene Inactivity Moderated by Metabolism and Expression (GIMME), os conjuntos de dados foram integrados, gerando um modelo contexto-específico. Esse modelo permitiu a identificação das vias e enzimas relacionadas ao processo de amadurecimento. Foi realizada algumas edições no modelo para a análise de fluxo e produção de fluxo metabólico. O modelo gerado foi validado, demonstrando que a produção de fluxo metabólico é apresentada em diferentes cenários testados. A compreensão do modelo gerado pode contribuir para um melhor entendimento dos genes e das redes bioquímicas associados ao amadurecimento do *S. lycopersicum*, possibilitando a elaboração de estratégias a fim de aumentar a vida pós-colheita do fruto

**Palavras chaves:** modelo metabólico; transcriptoma; tomate.

## ABSTRACT

The tomato (*Solanum lycopersicum*) is one of the most consumed and popular vegetables in the world, with a great diversity of fruit, varying in color, texture, shape and size. For this reason, it has been widely studied due to its economic value, cultural importance and beneficial compounds for human health. In addition, the fruit is the main model for studying metabolism and the development of fresh fruit because it is easy to grow and has short life cycles. The understanding of the metabolic pathways related to important commercial characteristics is still not as well understood as in the case of fruit ripening. This is a highly complex process, involving drastic changes throughout. A better understanding of the biochemical processes associated with ripening could enable new approaches to be used in post-harvest fruit conservation. The "Genome-scale metabolic model" (GEM) strategy has enabled the development of in silico models through the manipulation of a metabolic model, resulting in a greater understanding of the functioning of biochemical networks. GEM i HY3410 is a reconstructed model for *Solanum lycopersicum* and was used as a starting point for the integration of transcriptome data, with the aim of generating a context-specific model to understand the main pathways and enzymes involved in fruit ripening throughout the entire ripening stage. Integrating transcriptome data into a GEM is a crucial step in this strategy. To this end, a set of transcriptome data was selected from tomato pericarp tissue at medium green time (MG - mature green). The dataset underwent pre-processing and quality analysis before being implemented in the model. Using the Gene Inactivity Moderated by Metabolism and Expression (GIMME) package, the datasets were integrated, generating a context-specific model. This model enabled the identification of pathways and enzymes related to the ripening process. Some edits were made to the model to analyze metabolic flow and production. The model generated was validated, showing that the production of metabolic flux is presented in the different scenarios tested. Understanding the model generated can contribute to a better understanding of the genes and biochemical networks associated with the ripening of *S. lycopersicum*, enabling the development of strategies to increase the post-harvest life of the fruit.

**KEYWORDS:** metabolic model; transcriptome; tomato.

## LISTA DE TABELAS

|                                                                                     |    |
|-------------------------------------------------------------------------------------|----|
| <b>Figura 1</b> – Gráfico com a produção mundial de tomates ao longo dos anos ..... | 15 |
| <b>Figura 2</b> – Tabela dos SRR do <i>Solanum lycopersicum</i> .....               | 24 |
| <b>Figura 3</b> - FASTQC do SRR5724292.....                                         | 25 |
| <b>Figura 4</b> – tabela transcrito por milhão .....                                | 26 |
| <b>Figura 5</b> - ilustração da equação de biomassa editada.....                    | 27 |
| <b>Figura 6</b> – Ilustração da geração do modelo .....                             | 28 |
| <b>Figura 7</b> – Modelo final do <i>Solanum lycopersicum</i> .....                 | 29 |

## **LISTA DE ABREVIATURAS E SIGLAS**

GEM - Genome-scale Metabolic Model

MG - Medium Green

FBA - Flux Balance Analysis

TPM - Transcrito Por Milhão

|                                             |           |
|---------------------------------------------|-----------|
| <b>1 Introdução .....</b>                   | <b>12</b> |
| <b>2 Revisão da literatura .....</b>        | <b>14</b> |
| <b>2.1 Tomate .....</b>                     | <b>14</b> |
| <b>2.2 Modelo Metabólico .....</b>          | <b>16</b> |
| <b>2.3 Transcriptomas .....</b>             | <b>18</b> |
| <b>3 Objetivos.....</b>                     | <b>18</b> |
| <b>4 Objetivos específicos.....</b>         | <b>19</b> |
| <b>5 Métodos: .....</b>                     | <b>19</b> |
| <b>5.1 Obtenção do modelo Genoma .....</b>  | <b>19</b> |
| <b>5.2 Obtenção dos transcriptomas.....</b> | <b>20</b> |
| <b>5.3 Análise dos TPM .....</b>            | <b>21</b> |
| <b>5.4 Identificação das enzimas.....</b>   | <b>21</b> |
| <b>5.5 Construção do modelo .....</b>       | <b>21</b> |
| <b>5.6 Edição do modelo .....</b>           | <b>22</b> |
| <b>5.7 Validação do modelo .....</b>        | <b>23</b> |
| <b>6 Resultados: .....</b>                  | <b>23</b> |
| <b>7 Discussão: .....</b>                   | <b>29</b> |
| <b>8 Conclusões:.....</b>                   | <b>31</b> |
| <b>Referências: .....</b>                   | <b>32</b> |

## 1 Introdução

O *Solanum lycopersicum*, conhecido popularmente no Brasil como tomate, é um dos vegetais mais populares e consumidos ao redor de todo o mundo, sendo uma das hortaliças mais cultivadas mundialmente. De acordo com a Organização das Nações Unidas para Alimentação e Agricultura (FAO), no ano de 2020, houve uma produção de aproximadamente 186 milhões de toneladas, o que demonstra seu impacto econômico em todo mundo.

Pertencentes à família das Solanáceas, que possui cerca de 3000 espécies, entre as quais podemos destacar espécies como o pimentão (*Capsicum annum*), a batata (*Solanum tuberosum*) e a berinjela (*Solanum melongena*), exemplo de grande importância mundial (BRANDÃO FILHO et al., 2018). O gênero *solanum* é possivelmente o maior e mais importante economicamente do mundo (TAMBURINO et al., 2020). O *Solanum lycopersicum* é originário da América do Sul nas regiões do Peru, Chile e Bolívia.

O tomate é apreciado em diversas culturas ao redor de todo mundo, sendo consumido de diversas maneiras. Devido ao seu sabor adocicado e levemente ácido, ele pode ser utilizado de diferentes formas, seja consumindo a fruta fresca, produtos processados, cozido ou como parte de diferentes tipos de preparações. Além de ser bem nutritivo, o *Solanum lycopersicum* é rico em vitaminas A, B e principalmente, vitamina C, licopeno, carotenoides e minerais. Esses compostos têm uma grande gama de qualidades fisiológicas, envolvendo efeitos antialérgico, antimicrobianos, vasodilatadores entre outros (QUINET et al., 2019).

Atualmente existe uma grande diversidade do fruto, que varia desde a cor, tamanho, sabor e até mesmo na sua textura. Essa grande diversidade tem como objetivo promover o melhoramento genético, visando características importantes para atender a diferentes tipos de demandas do mercado. Por exemplo, aumentar a tolerância a algum tipo de estresse (KRISHNA et al., 2022), resistentes a pragas (MACIEL et al., 2024), produtividade (AL-AMRI, 2013), entre outras. Outras características que acabam contribuindo para a grande popularidade deste fruto são sua capacidade de adaptação a variedades climáticas, um baixo custo de produtividade e principalmente a facilidade de se cultivar o fruto, podendo se cultivar a hortaliça em pequena escala. O tomate pode ser plantado ao ar livre, em estufas ou até mesmo em vasos.

O *Solanum lycopersicum* possui um ciclo de vida curto, que pode ser dividido em 4 etapas. Na primeira etapa, ocorre a germinação das sementes, que dura cerca de 10 dias após o plantio da semente. Na segunda etapa, ocorre o crescimento vegetativo, marcado pelo rápido desenvolvimento das hastes e das folhas da planta, com duração entre 10 a 30 dias após o plantio. A terceira etapa é a floração, que marca o início da fase reprodutiva da planta, no momento da aparição das primeiras flores, ocorrendo entre 30 e 45 dias após o plantio da semente. Por fim, tem-se a etapa da frutificação, marcada pelo desenvolvimento dos frutos, que ocorre aproximadamente entre 45 e 60 dias após o plantio.

Além de toda importância descrita acima, o tomate é apontado como o principal modelo, sendo o ideal para estudo de frutos carnosos e climatérico (KLEE; GIOVANNONI, 2011) (YUAN et al., 2016), devido ao fácil cultivo, genoma relativamente pequeno e pelo ciclo de vida curto, como já mencionado. Devido a essas razões, vários estudos já foram e permanecem sendo conduzidos sobre diversas características comerciais do fruto, como, por exemplo, pesquisas relacionadas à genética (VARGAS; AGUIRRE; CORONADO, 2020), bioquímica (VOLOSHINA et al., 2023) e fisiologia (MIRZAYEVA et al., 2023), a fim de se entender melhor sobre as características do fruto.

O processo de amadurecimento do *Solanum lycopersicum* é bastante complexo e envolve mudanças drásticas durante todo o estágio, incluindo a regulação estruturada de inúmeras mudanças fisiológicas e bioquímicas, tendo modificações na expressão de milhares de genes. Durante a fase de amadurecimento do fruto, decorre a ativação de vias metabólicas que influenciam no nível de açúcar, pigmentos, ácidos e voláteis (QUINET et al., 2019). Estudos sobre algumas dessas vias já estão sendo realizados, como no caso de pesquisas referente à via do etileno, que busca estudar a relação do etileno e fatores de transcrição que controlam o processo de amadurecimento (JIA et al., 2024); estudos relacionados à formação do sabor, que objetivam compreender os mecanismos subjacentes à sua formação durante o amadurecimento (PAN et al., 2023) e alguns estudos sobre alguns genes que buscam entender os padrões de expressões ao longo do amadurecimento (EDRIS et al., 2023).

Apesar dos avanços nos estudos dos genes relacionados ao amadurecimento, o entendimento das complexas redes de expressão genética envolvidas durante todo o processo ainda é um grande desafio, em grande parte por não serem completamente compreendidas (PAN et al., 2023). Compreender essas vias pode contribuir bastante a obter um maior entendimento e consequentemente um maior controle sobre o processo de amadurecimento

do fruto, além de possibilitar a criação de novas abordagens de conservação pós-colheita do *Solanum lycopersicum*.

Os modelos metabólicos em escala genômica (GEMs) podem ser utilizados para contribuir no entendimento de como essas vias funcionam durante toda a fase do amadurecimento, uma vez que esses modelos são uma representação *in silico* dos fluxos metabólicos complexos de um conjunto de reações que ocorrem em uma célula. Esses modelos são uma excelente ferramenta, pois possibilitam compreender e prever como determinados organismos acabam reagindo na variação tanto dos parâmetros genéticos como nos parâmetros ambientais(WANG et al., 2018). Os GEMs vêm demonstrando grande importância na descoberta de novas estratégias de engenharia metabólica(CARTER; CONSTANTINIDOU; ALAM, 2024; DEL CARRATORE et al., 2022; SIMEONIDIS; PRICE, 2015). Sendo indispensáveis para o estudo do metabolismo(THIELE; PALSSON, 2010).

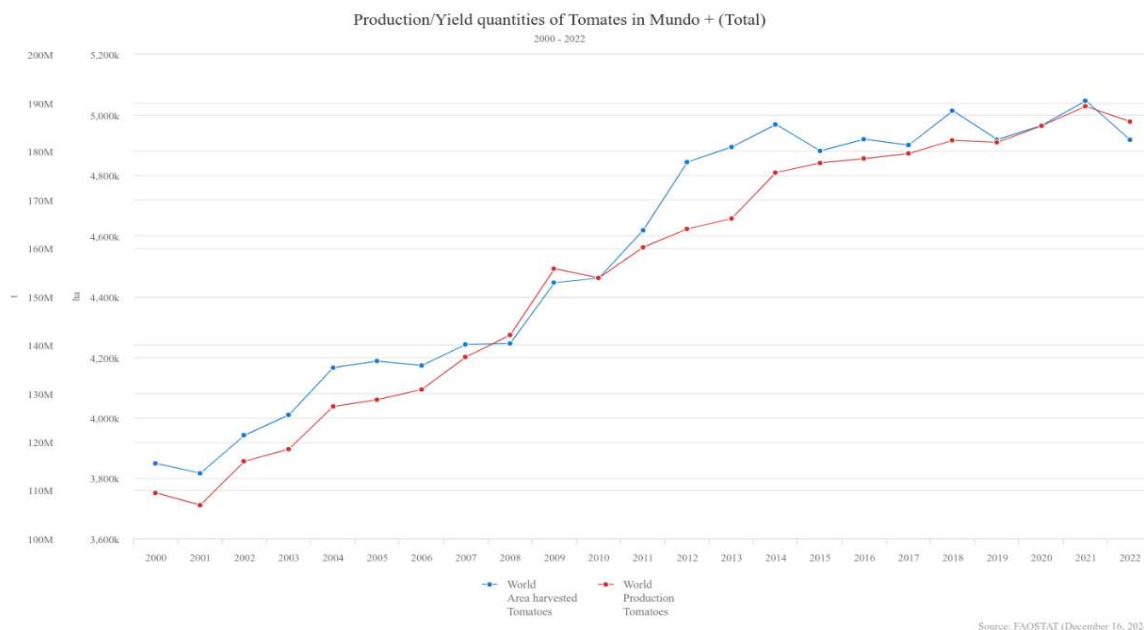
Portanto, propomos, através deste trabalho, a manipulação de um modelo *in silico* para o *Solanum lycopersicum*, realizando a integração com dados de transcriptoma com objetivo alvo de se entender as vias metabólicas envolvidas durante o processo de amadurecimento. Em futuros estudos, após validação experimental do modelo, o entendimento destas vias poderá ser aplicado na engenharia genética para a conservação pós-colheita do fruto.

## **2 Revisão da literatura**

### **2.1 Tomate**

O tomate, cientificamente conhecido como *Solanum lycopersicum*, é uma das principais hortaliças comerciais cultivada em todo mundo (Figura 1). O fruto é amplamente consumido e possui implicações econômicas, sociais e culturais de grande importância mundial. A planta pode ser cultivada em hortas ou até mesmo em um simples vaso, o que facilita a sua produção por qualquer pessoa. A planta selvagem é originária da América do Sul, tendo como centro de origem a região conhecida como Peru(ANWAR; FATIMA; MATTOO, 2019). Pertencentes ao gênero *Solanum*, este é considerado o gênero mais importante economicamente do mundo.

**Figura 1** – Gráfico com a produção mundial de tomates ao longo dos anos



Em vermelho indica a produção mundial do tomate de 2000 a 2022. Em azul está demonstrando a área total da produção do tomate no mesmo período.

FONTE: FAOSTAT, 2024.

O tomate ou compostos a base da fruta são de grande relevância para a saúde humana, por serem fontes de vitaminas, potássio, carotenoides e polifenóis (SILASTE et al., 2007). Outros fatores que fazem da fruta um bom alimento para a saúde humana são seu baixo teor de calorias e gorduras, além de ser uma boa fonte de fibra (SOTO-ZAMORA et al., 2005). Estudos sobre esses compostos demonstraram a eficácia e importância de ingestão deste alimento. Por exemplo, os efeitos dos carotenoides, podem exercer grande influência na prevenção de doenças como o câncer, conforme demonstrado no trabalho de Wang, que avaliou a associação entre absorção dos carotenoides e o baixo risco de desenvolvimento de câncer de mama ao longo da vida (WANG et al., 2015).

As frutas podem ser divididas em dois grupos relacionadas ao amadurecimento: climatéricos, nos quais envolve um pico de respiração e uma explosão na produção do hormônio etileno, desencadeando mudanças na expressão de milhares de genes; e os não climatéricos, nos quais não há nenhuma mudança drástica na produção desse hormônio (ALEXANDER; GRIERSON, 2002). O *Solanum lycopersicum* é considerado o principal modelo de estudo para compreensão das nuances intrincadas dos desenvolvimentos das frutas climatéricas, incluindo todos os estágios de amadurecimento, pós-colheita e senescência, tornando-o um excelente

modelo de estudo(ANWAR; FATIMA; MATTOO, 2019). O fruto é sensível a temperaturas baixas e temperaturas altas, o que ocasiona serias perdas durante todo o ano e, consequentemente, um grande prejuízo econômico mundial. Nos últimos anos, diversos estudos vêm sendo conduzidos em relação a genética, fisiologia e bioquímica para tentar compreender as vias envolvidas durante todo o processo de amadurecimento e, assim, propor estratégias para aumentar a vida pós-colheita do fruto. Essas estratégias incluem tratamento físico, químicos ou até mesmo manipulação do genoma do *Solanum lycopersicum*.

Estudos como o trabalho de Yoon buscou elucidar, utilizando-se da transcriptômica, quais os mecanismos estão sendo afetados durante a irradiação de raios gama, o que, consequentemente, influencia positivamente o tempo de amadurecimento do fruto, além de inibir o crescimento microbiano durante o tempo de armazenamento conferindo uma maior vida pós-colheita do fruto(YOON et al., 2023). Outro estudo, como o de Jia, investigou os principais efeitos regulatórios que determinados fatores de transcrição exercem no metabolismo do *Solanum lycopersicum*, além de realizar análise de metabolômica e sequenciamento de transcriptomas. Os resultados desse estudo indicaram que as modificações nos fatores de transcrição afetaram o tempo de amadurecimento do fruto.

Apesar de vários estudos estarem sendo conduzidos no decorrer desses últimos anos, a compreensão das diversas vias e genes em todos os sistemas relacionados durante todo o processo de amadurecimento do fruto não são tão bem compreendidas. Portanto, novas abordagens vêm sendo utilizadas para ajudar a compreender acerca dos estágios de desenvolvimento do fruto. Uma dessas abordagens consiste na manipulação de um modelo *in silico* conhecida como “Genome-scale metabolismo”(YUAN et al., 2016).

## 2.2 Modelo Metabólico

Modelos metabólicos em escalas genômicas (GEMs) são ferramenta baseada em redes que integram todas as informações metabólicas conhecidas de um sistema biológico, como as enzimas, reações metabólicas, genes do organismo, entre outras. Essas redes genômicas reconstruídas abrangem informações sistematizadas e curadas sobre as reações e metabólitos conhecidos de uma determinada célula, baseando-se no genoma anotado e na literatura(BORDBAR et al., 2014). Os GEMs possibilitam prever como os organismos estudados acabam reagindo com modificações tanto no seu parâmetro genético quanto ambientais, possibilitando a compreensão das interações das diversas vias metabólicas, e definindo quantitativamente a relação entre genótipo e fenótipo.

Nos últimos anos, houve um grande aumento na produção de GEMs, como demonstra o trabalho de revisão de Antonakoudis (ANTONAKOUDIS et al., 2020). Esse aumento deve-se, principalmente, ao sequenciamento e a publicação do genoma de milhares de organismos, tornando os modelos metabólicos fundamentais para análise das interações metabólicas dos organismos, ou seja, do sistema biológico completo. Os GEMs podem ser reconstruídos através de ferramentas automáticas ou ferramentas semiautomáticas.

As redes metabólicas podem ser convertidas em um formato matemático, conhecido como matriz estequiométrica. Essa matriz é o ponto central de um modelo baseado em restrições de fluxo, no qual todos os estados fenotípicos viáveis existem, de acordo as restrições impostas, possibilitando a verifique simultaneamente de todos os processos que atuam sobre a célula (BORDBAR et al., 2014).

Utiliza-se dados de transcriptomas do organismo alvo, o que resulta em uma melhor compreensão da fisiologia do organismo e num GEM de alta qualidade. A obtenção de um GEM de alta qualidade é refinado pode gerar previsões da taxa de crescimento e de produção de metabolitos, permitindo a compreensão sobre o balanço de fluxo, que só podem ser confirmadas posteriormente após validação experimental do modelo. Essa capacidade foi demonstrada pelo estudo de Rau, que avaliou os principais aspectos da fisiologia e metabolismo para *S. thermophilus*, por meio da combinação de abordagens *in silico* e experimentais (RAU et al., 2022).

A primeira reconstrução de um modelo *in silico* em escala genômica foi realizada para *Escherichia coli* MG1655, após seu genoma ser completamente sequenciado, e se utilizando de informações específicas. Edwards e Bernhard desenvolveram a reconstrução de um modelo para definir a capacidade do metabolismo e as características do sistema. O FBA foi utilizado para avaliar a capacidade e o desempenho da *Escherichia coli* (EDWARDS; PALSSON, 2000). Desde então, o número de modelos metabólicos publicados vem aumentando consideravelmente a cada ano.

Em relação ao *Solanum lycopersicum*, foram publicados alguns modelos gnômicos após o sequenciamento completo do seu genoma, o qual foi publicado por Sato em 2012 (SATO et al., 2012). Em 2015, foi publicado um modelo estequiométrico em média escala para o tomate, como principal objetivo de descrever todo o estágio de desenvolvimento do tomate, modelando os fluxos de metabolismo central (COLOMBIÉ et al., 2015). Em 2016, foi publicado o primeiro artigo de um modelo metabólico em escala genômica em larga escala para o tomate, que buscou

simular e avaliar as mudanças no metabolismo da folha do *Solanum lycopersicum* em resposta a estresse abiótico, integrando dados da literatura e banco de dados. Esse foi o modelo utilizado no presente trabalho para integrá-lo com dados de transcriptomas, com o alvo de compreender as principais vias e genes relacionados ao processo de amadurecimento do fruto.

### 2.3 Transcriptomas

A integração de um transcriptoma a um rascunho de modelo é uma etapa de grande importância, pois a incorporação desses dados ômicos auxiliam nas predições. Os transcriptômicos fornecem informações sobre as expressões genicas que ocorrem em todo o genoma do organismo na atual condição de estudo. Com estas informações, torna-se possível identificar genes ativos que estão associados aos processos metabólicos estudados sob determinadas condições, no nosso caso, sobre todo o processo de amadurecimento.(KAMSEN et al., 2021).

O trabalho desenvolvido por Shinozaki proporcionou um atlas abrangente de transcriptomas do fruto, que foi gerado por meio do sequenciamento de RNA (RNAseq), envolvendo o crescimento e todo o processo de amadurecimento. Esse trabalho caracterizou a distribuição e o tempo das redes estruturais e regulatórias dos genes ao longo de todo o tempo de desenvolvimento do *Solanum lycopersicum*. Esses conjuntos de dados podem ajudar a elucidar diversos aspectos da biologia(SHINOZAKI et al., 2018). Os transcriptomas utilizados no presente trabalho foram resultados desse trabalho.

Os transcriptomas têm sido utilizados para entender outras características do fruto. Um é o trabalho de Liu, que utilizou do metaboloma e transcriptomas durante a etapa de amadurecimento para explorar os mecanismos de regulação responsáveis pelo sabor volátil da fruta(LIU et al., 2024). Os transcriptomas vem contribuindo para compreender os diversos aspectos da biologia do fruto.

## 3 Objetivos

Este projeto tem como objetivo a reconstrução de um modelo contexto-específico para *Solanum lycopersicum*, por meio da integração de dados de transcriptomas a um modelo já existente. Buscando validar o modelo reconstruído e se realizar análise de balanço de fluxo.

## 4 Objetivos específicos

- Recuperar de bancos de dados públicos dados de transcriptomas do *Solanum lycopersicum* e realizar análise e quantificações dos conjuntos de dados coletados;
- Obter o modelo existente de *Solanum lycopersicum* para integrá-los aos dados de transcriptomas coletados;
- Realizar a integração dos transcriptomas ao modelo;
- Editar o modelo por meio da escrita e inserção de equações, viabilizando a realização da análise de fluxo;
- Realizar análise de fluxo por meio da abordagem de análise de balanço de fluxo, identificando as vias e enzimas ativas no amadurecimento;
- Validar o modelo.

## 5 Métodos:

### 5.1 Obtenção do modelo Genoma

O modelo gnômico base utilizado como ponto de partida inicial deste trabalho foi reconstruído por Yuan(YUAN et al., 2016), que utilizou o genoma publicado pelo *Tomato Genome Consortium*(SATO et al., 2012). A pesquisa, tinha como alvo principal o desenvolvimento de uma rede metabólica em escala genômica do tomate, pretendendo utilizar este modelo para-se avaliar como o estresse hídrico afetaria o metabolismo do *Solanum lycopersicum*, especificamente o metabolismo da folha do tomate, simulando o modelo através de vários cenários metabólicos. Apesar de esse modelo ter sido desenvolvido com a finalidade de estudar e analisar o comportamento da folha, ele também é um excelente ponto de partida para se explorar o comportamento metabólico do fruto do tomate, como sugere o próprio autor.

O modelo resultante, *i* HY3410, é compreendido de 3410 genes, 1998 metabólitos e 2143 reações. Dentre as 2143 a grande maioria representa conversões bioquímica, enquanto a minoria representa processo de transporte metabólitos. A síntese de biomassa foi representada por um conjunto de transporte independente. Um ponto importante se deu pela inclusão de associação gene-proteína-reação (GPR), possibilitando que o *i* HY3410 possa ser utilizado para a integração e contextualização de dados proteômico e transcriptômicos.

## 5.2 Obtenção dos transcriptomas

Os dados de RNAseq utilizados foram publicados no trabalho de Shinozaki (SHINOZAKI et al., 2018), que através de sua pesquisa buscou apresentar uma extensa variedade de dados de transcriptomas, abrangendo todo o processo de amadurecimento do fruto, através dos tecidos, tipos de célula, desenvolvimento e topografia do *Solanum lycopersicum*. Esses conjuntos de dados foram utilizados para caracterizar toda a organização e o estágio das redes estruturais e regulatórias de genes ao longo de todo desenvolvimento, possibilitando uma base para diversos outros estudos, o qual se investigue características comerciais importantes, como textura, cor, aroma entre outras.

Para o presente trabalho, o tecido escolhido foi o tecido *total pericarp* (pericarpo), devido a ser parte significativa de todo o fruto, possibilitando uma análise de todo o estágio de desenvolvimento do fruto. O tempo escolhido para a coleta foi o *medium green* (MG-verde maduro). Para aquisição do conjunto de transcriptomas (RNAseq), utilizou-se o banco de dados do NCBI (National Center for Biotechnology Information). A coleta dos transcriptomas (RNAseq) foi realizada na forma de triplicatas de três distintas regiões da placenta do fruto: Equatorial of Placenta; Stem of Placenta e Stylar of Placenta.

Posteriormente as coletas dos transcriptomas, se deu a realização do controle de qualidade e o pré-processamento de cada um dos conjuntos de dados brutos, com a finalidade de preparar os dados para a integração junto ao modelo existente. Utilizando a plataforma Galaxy/Europe (<https://usegalaxy.eu/>), foi realizado controle de qualidade e o pré-processamento dos RNAseq, utilizando o parâmetro de phred superior a 30. Primeiramente, implementou a trimagem dos dados (Trim Galore, The Babraham Institute - <https://github.com/FelixKrueger/TrimGalore>), sendo possível realizar a remoções dos adaptadores ou qualquer tipo de sequencias indesejadas ligadas às sequências de leitura oriundas do sequenciador em cada um dos conjuntos de dados. Logo após, foi verificado o controle de qualidade de cada uma das sequências (FastQC, The Babraham Institute - <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

### 5.3 Análise dos TPM

Por fim, foram estabelecidos o parâmetro de transcrito por milhão (TPM) e o mapeamento com o genoma de referência devidamente anotado, utilizando o algoritmo do Salmon quant. Efetuamos uma análise abundância gênica. Um dos pontos principais dessa análise é a indexação dos transcritos de referência, identificando a relação dss read (leitura de uma sequência realizada pelo sequenciador) com o transcrito alvo. Por fim, para se obter uma melhor estimativa dos TPM, calcula-se a expressão genica normalizada, proporcionando uma quantificação precisa entre a amostra e os genes, além de gerar para visualização relatórios e gráficos detalhados sobre a qualidades dos dados analisados.

### 5.4 Identificação das enzimas

A identificação das possíveis enzimas ativas em cada estágio do amadurecimento, que só pode ser confirmada após validação experimental, foi baseado no trabalho de Becker(BECKER; PALSSON, 2008). Em seu trabalho, ele descreve o uso de dados de transcriptômicos em escala genoma para restringir reações bactérias e células humanas, permitindo a reconstrução de redes metabólicas específicas.

Posteriormente a todo o processamento realizado nos RNAseq, foi implementado a associação gene-proteína-reação (GPR) com o objetivo de o programa identificar as possíveis enzimas ativas, e as reações que estas enzimas catalisam no estado de maturação em estudo do fruto. Através do mapeamento da expressão genica das reações metabólicas, foram atribuídas as enzimas que estes genes codificam com base no GPR. O algoritmo GIMME, através de uma pontuação, que considera os dados de expressão dessas enzimas e as reações que elas catalisam, permite que o modelo selecione as reações com maior grau de atividade, durante a expressão genica.

### 5.5 Construção do modelo

A reconstrução de um modelo em escala genômica para tomate foi realizada através do pacote Gene Inactivity Moderated by Metabolism and Expression (GIMME)(BECKER; PALSSON, 2008) através da ferramenta COBRA Toolbox (CONstraint-Based Reconstruction and Analysis)(SCHELLENBERGER et al., 2011) no software MATLAB. A geração desse modelo genômico abrange todas as reações conhecida para o organismo, que, no caso do presente trabalho, é o tomate.

Primeiramente, os dados de transcriptomas devidamente tratados foram incorporados ao modelo *i* HY3410, gerando um rascunho de modelo inicial das redes metabólicas para cada um dos conjuntos de dados. Essa integração desses dados é importante pois, possibilitou observar quais as vias e enzimas estariam ativas durante os estágios de expressão estudados neste trabalho.

O refinamento do modelo foi realizado através da aplicação de alguns algoritmos de restrição, como FBA, que permite calcular, através da rede metabólica, o fluxo de metabólitos(ORTH; THIELE; PALSSON, 2010), prevendo o fluxo de crescimento e a taxa de metabólitos importante. Esse algoritmo utiliza dados ômicos para restringe o modelo a certas vias e genes que são considerados ativos, de acordo com limites pré-estabelecidos. Além do mais, foi utilizado um algoritmo de remoção de reações indesejáveis, responsável pela exclusão das reações que não são expressas ou não possui qualquer tipo de relevância para o estudo, com base nos dados ômicos. Por fim, o modelo passou por uma etapa de validação para confirmar sua compatibilidade com os dados experimentais observados.

## 5.6 Edição do modelo

A edição do modelo se deu por duas significativas modificações relacionadas a inserção de dados. A primeira grande modificação do modelo ocorreu pela escrita de uma equação geral da biomassa, baseado no material suplementar do trabalho de Yuan (YUAN et al., 2016). Essa equação de biomassa é constituída pela interação de 46 reações da massa seca da folha do *Solanum lycopersicum*, conforme definido por experimentos El-Sayed e Schauer(SAYED et al., 2015; SCHAUER; ZAMIR; FERNIE, 2005).

A realização de uma análise de fluxo no modelo resultou em um valor negativo, o que indicou os denominados metabólitos dead-end (DEM)(MACKIE et al., 2013), apontando que uma ou mais reações estariam suspendendo o fluxo geral. Em função disso, foi necessário se

investigar o fluxo destas reações, sendo possível detectar os pontos desconexo pela identificação das reações que não geravam nenhum tipo de fluxo. A segunda e última grande modificação se deu pela adição de oito reações de *exchange*, responsáveis pela eliminação dos metabólitos dead-end que acabavam impedindo o fluxo metabólico geral do modelo.

## 5.7 Validação do modelo

A funcionalidade do modelo foi avaliada usando o FBA para analisar o balanço de fluxo. Para isso, foi desenvolvido uma equação de biomassa, pois o modelo não possuía uma equação geral de biomassa que nos permitisse observar o fluxo em uma simulação. Essa equação gerada é unidirecional e, configurada pelas reações bioquímicas responsáveis pela geração de biomassa, ou seja, demonstrando que o modelo é válido. Essa reação é composta pela inserção de 46 componentes da massa seca da folha do *Solanum lycopersicum*, conforme definidos pelos experimentos de Sayed e Schaur (SAYED et al., 2015; SCHAUER; ZAMIR; FERNIE, 2005). Compreende aminoácidos, metabólitos solúveis, sais, bases nitrogenadas, carboidratos, pigmentos que resultam em ADP, fosfato e difosfato.

Após a realização da análise de fluxo, obteve-se um resultado negativo, indicando que o fluxo geral estaria sendo interrompido. Foram efetivadas modificações no modelo, como já demonstrado acima, a fim de se solucionar esse problema. Posteriormente as essas modificações, o modelo passou por uma nova análise que resultou em um resultado positivo, demonstrando sua validade.

## 6 Resultados:

A **Figura 2** apresenta uma tabela contendo todos os nove conjuntos de RNAseq (SSR) de *Solanum lycopersicum*, os quais foram coletados através do NCBI (National Center for Biotechnology Information), com a finalidade de realizar a análise e quantificação desses dados, visando integrá-lo ao modelo existente i HY3410 do tomate. Esses nove conjuntos de transcriptomas representam três diferentes regiões do pericarpo, e os RNAseq coletados de cada uma dessas regiões foram obtidos em forma de triplicatas, ambos no tempo MG.

**Figura 2** – Tabela dos SRR do *Solanum lycopersicum*

| SRR        | Tecido                 | Tempo |
|------------|------------------------|-------|
| SRR5724280 | Equatorial of Placenta | MG    |
| SRR5724281 | Equatorial of Placenta | MG    |
| SRR5724282 | Equatorial of Placenta | MG    |
| SRR5724399 | Stem of Placenta       | MG    |
| SRR5724400 | Stem of Placenta       | MG    |
| SRR5724284 | Stem of Placenta       | MG    |
| SRR5724278 | Stylar of Placenta     | MG    |
| SRR5724279 | Stylar of Placenta     | MG    |
| SRR5724292 | Stylar of Placenta     | MG    |

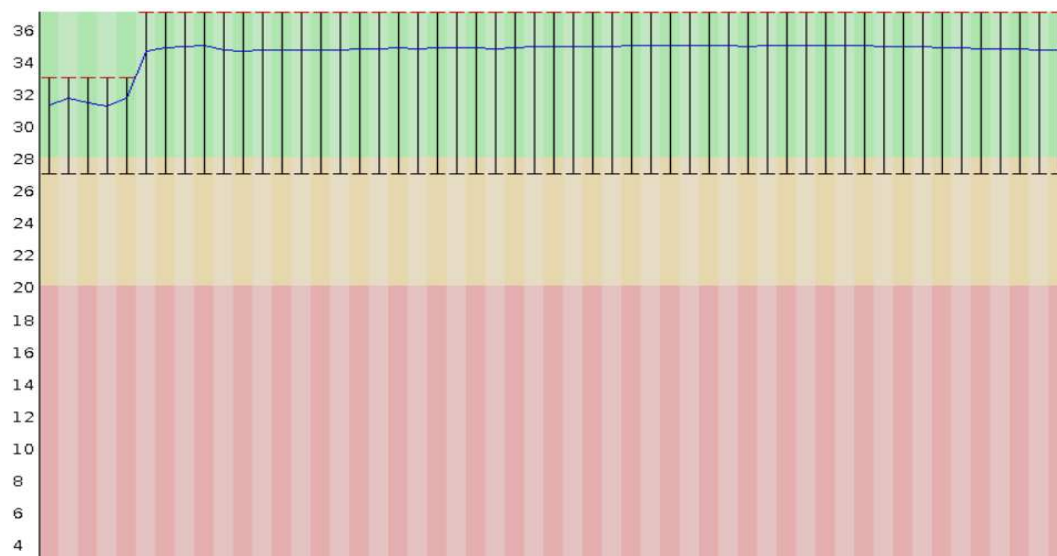
Tabela referente aos nove SRR de *Solanum lycopersicum* coletados no tempo MG

Cada uma das sequências coletadas foi submetida a processos de análise de qualidade e por procedimentos de quantificação, antes delas serem implementadas ao modelo, garantindo a qualidade de cada uma das sequências obtidas. A **Figura 3** apresenta um gráfico gerado para o SRR5724292 por meio do algoritmo do FASTQC. Através do gráfico é possível observar o controle de qualidade da sequência, evidenciado pelo valor de *phred* superior a 30. O valor de *phred* é um parâmetro importante na avaliação da qualidade dos dados de sequenciamentos de RNAseq. Consideramos como uma sequência satisfatória para o nosso objetivo, quando o controle de qualidade apresenta valor de *phred* superior a 30.

Para cada um dos determinados conjuntos de transcriptomas coletados, foram gerados gráficos similares, sendo ambos avaliados pelo valor de *phred* superior a 30, garantindo a qualidade de cada uma dos transcriptomas antes da sua implementação ao modelo.

Essa etapa foi de grande importância, pois uma sequência de boa qualidade garante resultados mais confiáveis e robustos para as etapas subsequentes. Um valor de *phred* superior a 30 assegurou que essas leituras apresentam menor riscos de determinados erros que poderiam de alguma forma, impactar na análise. Sequências abaixo desse valor foram eliminadas através da trimagem desses dados, assegurando maior qualidade ao conjunto final.

**Figura 3 - FASTQC do SRR5724292**



A figura demonstra a qualidade do SSR, através do valor de *phred* superior a 30.

Fonte: GALAXY/EUROPE

A **figura 4** é uma representação parcial dos dados resultante da etapa de quantificação do SRR5724284 utilizando um genoma de referência. Pode-se observar a coluna relacionada à contagem (coluna TPM – Transcrito por milhão), a qual reflete a abundância relativa de cada transcrito. Além disso, é possível observar outra coluna referente aos mapeamentos do *reads* (coluna Name e coluna NumReads) de um transcriptoma, a qual indica o nível de cobertura da RNAseq, mostrando quais genes estão sendo expressos, sendo uma medida brusca de profundidade. Essas duas colunas são de extrema importância, pois evidenciam que os dados sejam sólidos e relevantes, além de auxiliarem na identificação dos genes e vias que estão ativos em determinadas condições. Tabelas semelhantes foram geradas para todas as outras réplicas de transcriptomas referentes aos tecidos.

**Figura 4** – tabela transcrito por milhão

| Name               | Length | EffectiveLength | TPM      | NumReads |
|--------------------|--------|-----------------|----------|----------|
| Solyc00g500018.1.1 | 426    | 176.000         | 0.000000 | 0.000    |
| Solyc00g014290.1.1 | 327    | 77.082          | 0.000000 | 0.000    |
| Solyc00g005280.1.1 | 411    | 161.000         | 0.000000 | 0.000    |
| Solyc00g500019.1.1 | 372    | 122.000         | 0.000000 | 0.000    |
| Solyc00g500020.1.1 | 1305   | 1055.000        | 0.147853 | 4.000    |
| Solyc00g500021.1.1 | 324    | 74.118          | 5.261372 | 10.000   |
| Solyc00g500022.1.1 | 306    | 56.785          | 1.373460 | 2.000    |
| Solyc00g180440.2.1 | 387    | 137.000         | 0.853930 | 3.000    |
| Solyc00g500023.1.1 | 471    | 221.000         | 1.411625 | 8.000    |
| Solyc00g500024.1.1 | 1527   | 1277.000        | 0.305373 | 10.000   |
| Solyc00g500025.1.1 | 195    | 8.348           | 0.000000 | 0.000    |
| Solyc00g500026.1.1 | 207    | 9.766           | 0.000000 | 0.000    |
| Solyc00g500028.1.1 | 450    | 200.000         | 0.000000 | 0.000    |
| Solyc00g500029.1.1 | 273    | 30.760          | 3.803288 | 3.000    |
| Solyc00g500030.1.1 | 417    | 167.000         | 0.467020 | 2.000    |
| Solyc00g500031.1.1 | 771    | 521.000         | 0.149697 | 2.000    |
| Solyc00g500032.1.1 | 357    | 107.001         | 0.364447 | 1.000    |
| Solyc00g500033.1.1 | 507    | 257.000         | 0.000000 | 0.000    |
| Solyc00g500034.1.1 | 1059   | 809.000         | 0.337420 | 7.000    |
| Solyc00g500035.1.1 | 429    | 179.000         | 0.000000 | 0.000    |

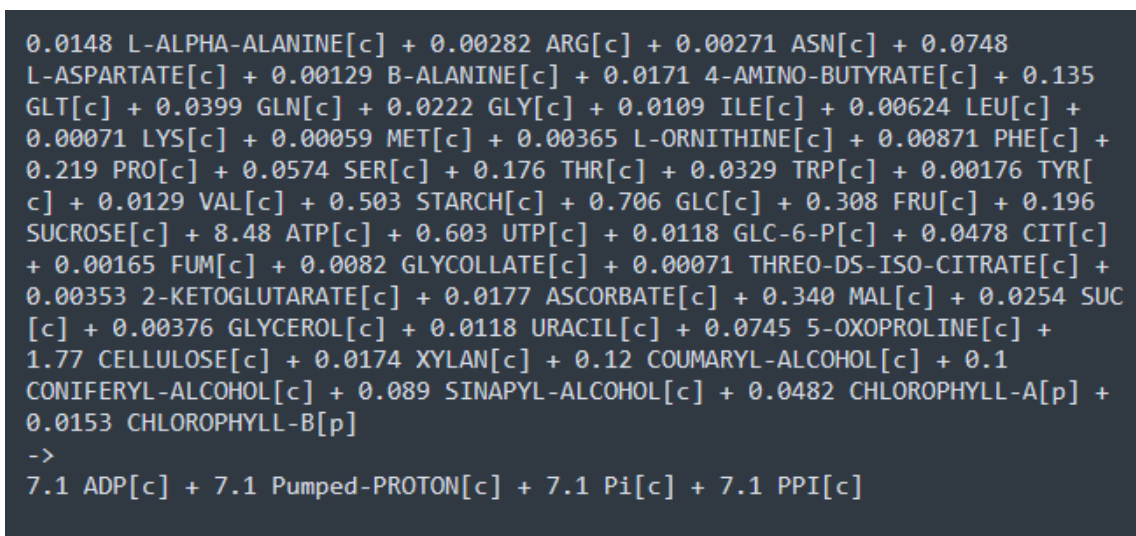
Transcritos por milhão de um transcriptoma mapeados no genoma de referência anotado.

Fonte: GALAXY/EUROPE.

A **figura 5** demonstra uma ilustração da equação geral de biomassa configurada nos padrões de formatação do modelo editado, realizada durante a edição do modelo. A equações escrita e inserida no modelo durante essa fase buscaram eliminar pontos de fluxo que não se conectavam com redes estabelecidas, as quais interferiam no fluxo geral do modelo, além da inserção de reações que não estavam presentes no modelo existente. A equação incluem componentes como carboidratos, aminoácidos, ácidos graxos, metabólitos solúveis, ADP, fosfato e difosfato.

Por meio da edição e inserção dessa reação, foi possível a realização de análise de balanço de fluxo, a qual, depois da inserção dessas reações, apresentou resultados positivos para a produção de biomassa.

**Figura 5** - ilustração da equação de biomassa editada



Equação de biomassa configurada nos padrões de formatações do modelo editado

A **figura 6** apresenta um esquema da etapa da integração dos dados de transcriptoma ao modelo existente, identificando os genes e vias ativas no contexto estudado, de acordo com os níveis de expressão gênica pré-estabelecidos. Esse processo resulta em um modelo reduzido, contendo os genes e vias com os maiores níveis de expressão. O modelo final realiza previsões específicas, demonstrando quais são as principais vias e genes relacionadas ao processo de amadurecimento do *Solanum lycopersic*, que só podem ser confirmados por meio da validação experimental do modelo.

**Figura 6** – Ilustração da geração do modelo

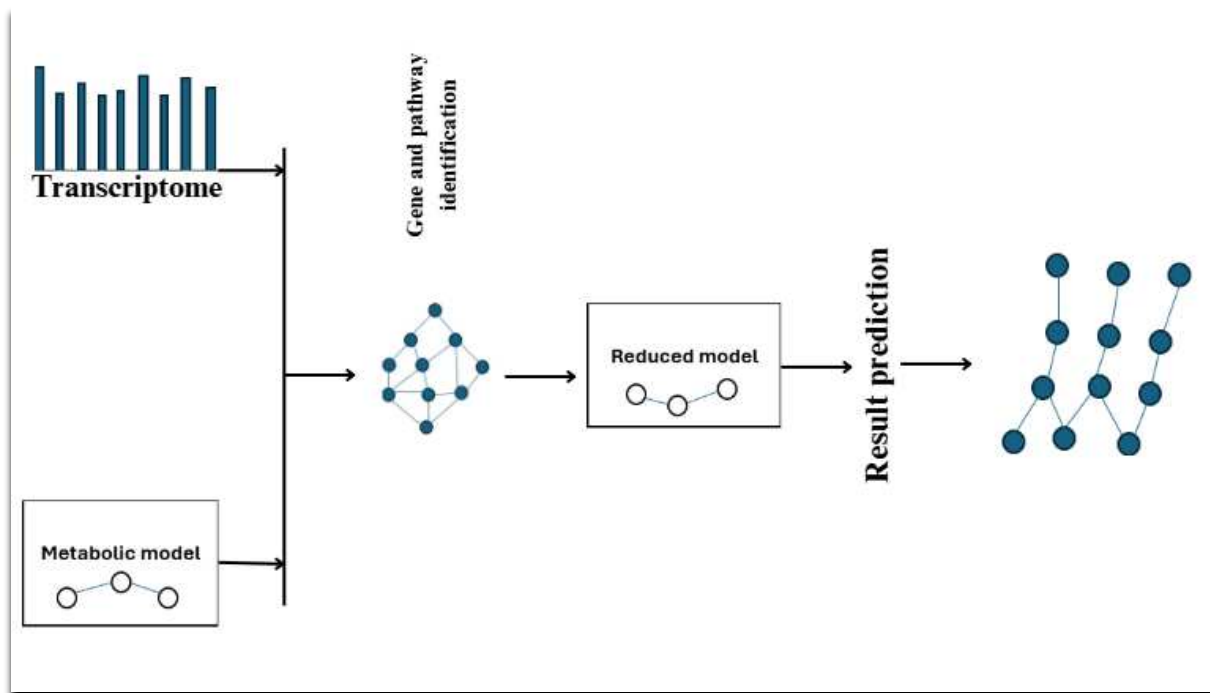


Ilustração esquemática da etapa de integração dos transcriptoma ao modelo

A **figura 7** é uma representação parcial do modelo final de *Solanum lycopersicum* posteriormente a todas as etapas realizadas. O modelo final é um modelo funcional e foi testado através da análise de fluxo, comprovando, que no modelo final do tomate, ocorre a produção de biomassa. Esse modelo pode ser utilizado para passos futuros, para compreender e prever o comportamento de metabólitos, auxiliando na identificação de vias metabólicas, para futuras estratégias de engenharia metabólica. Além disso, permite o estudo de processos biológicos sobre diferentes condições ambientais.

**Figura 7 – Modelo final do *Solanum lycopersicum***

| Abbreviation             | Description                                                                                                                       | Reaction                                                                                                                                        | GPR                                                                                                         | Genes                                                                                                       | Proteins | Subsystem                 |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|----------|---------------------------|
| NEOMENTHOL-DEHYDROGENASE | (+)-neomenthol dehydrogenase                                                                                                      | $\text{NADP}[c] + \text{CPD-1905}[c] \rightarrow \text{NADP}[c] + \text{CP}$                                                                    | $\text{SOLYC10G017570.2}$ or $\text{SOLYC03G059190.2}$                                                      | $\text{SOLYC10G017570.2-MONOMER}$ or $\text{SOLYC03G059190.2-MONOMER}$ or $\text{SOLYC03G059190.2-MONOMER}$ |          |                           |
| MENTHOL-DEHYDROGENASE    | (-)-menthol dehydrogenase                                                                                                         | $\text{NADP}[c] + \text{-MENTHOL}[c] \rightarrow \text{NADP}[c] + \text{C}$                                                                     | $\text{SOLYC03G059190.2}$                                                                                   | $\text{SOLYC03G059190.2-MONOMER}$                                                                           |          |                           |
| 1.1.1.168-RXN            | 2-dehydropantholactone reductase                                                                                                  | $\text{NADP}[c] + \text{PANTOYL-LACTONE}[c] \leftrightarrow \text{NADP}[c] + 2\text{-DEHYDRO-PANTOYL-LACTONE}[c]$                               |                                                                                                             |                                                                                                             |          | PWY-3961                  |
| 1.1.1.178-RXN            | 3-hydroxy-2-methylbutyryl-CoA 2-METHYL-3-HYDROXY-BUTYRYL-COA[m] + NAD[m] $\leftrightarrow$ NADH[m] + 2-METHYL-ACETO-ACETYL-COA[m] |                                                                                                                                                 |                                                                                                             |                                                                                                             |          | ILEUDEG-PWY               |
| 1.1.1.197-RXN            | 15-hydroxyprostaglandin dehydrogenase                                                                                             | $\text{ALPROSTADIL}[c] + \text{NADP}[c] \leftrightarrow 13\text{-E-11-AL} + \text{NADP}[c]$                                                     | $\text{SOLYC12G042370.1}$ or $\text{SOLYC06G0687}$                                                          | $\text{SOLYC12G042370.1-MONOMER}$ or $\text{SOLYC06G068740.2-MONOMER}$                                      |          |                           |
| 1.1.1.255-RXN            | Mannitol dehydrogenase                                                                                                            | $\text{NAD}[c] + \text{MANNITOL}[c] \leftrightarrow \text{MANNOSE}[c] + \text{NAD}[c]$                                                          | $\text{SOLYC11G011330.1}$ or $\text{SOLYC11G011330.1}$                                                      | $\text{SOLYC11G011330.1-MONOMER}$ or $\text{SOLYC11G011330.1-MONOMER}$                                      |          | PWY-3861                  |
| 1.1.1.262-RXN            | 4-hydroxythreonine-4-phosphatase                                                                                                  | $4\text{-PHOSPHONOOXY-THREONINE}[c] + \text{NADP}[c] \rightarrow 2\text{-AMINO-3-OXO-4-PHOSPHONOOXYBUTYRATE}[c] + \text{NADP}[c]$               |                                                                                                             |                                                                                                             |          | PYRIDOXSYN-PWY            |
| 1.1.1.264-RXN(NAD)       | 1.1.1.264-RXN(NAD)[c]                                                                                                             | $\text{NAD}[c] + \text{L-IDONATE}[c] \leftrightarrow \text{NADH}[c] + 5\text{-DEHYDROGLUCONATE}[c]$                                             |                                                                                                             |                                                                                                             |          |                           |
| 1.1.1.264-RXN(NADP)      | 1.1.1.264-RXN(NADP)[c]                                                                                                            | $\text{NADP}[c] + \text{L-IDONATE}[c] \leftrightarrow \text{NADP}[c] + 5\text{-DEHYDROGLUCONATE}[c]$                                            |                                                                                                             |                                                                                                             |          |                           |
| 1.1.1.278-RXN            | 3- $\beta$ -hydroxy-5- $\alpha$ -pregnen-20-one reductase                                                                         | $\text{NADP}[c] + 3\text{-BETA-HYDROXY-5-ALPHA-PREGNANE-20-ONE}[c] \rightarrow \text{NADP}[c] + \text{CPD-293}[c]$                              |                                                                                                             |                                                                                                             |          | PWY-6032                  |
| 1.1.1.34-RXN             | Hydroxymethylglutaryl-CoA reductase                                                                                               | $3\text{-HYDROXY-3-METHYL-GLUTARYL-COA}[c] + \text{NADP}[c] \rightarrow 3\text{-HYDROXY-3-METHYL-GLUTARYL-COA}[c] + \text{NADP}[c]$             | $\text{SOLYC03G032010.2}$ or $\text{SOLYC03G032010.2}$                                                      | $\text{SOLYC03G032010.2-MONOMER}$ or $\text{SOLYC03G032010.2-MONOMER}$                                      |          | PWY-922                   |
| 1.1.3.32-RXN             | (S)-stylophine synthase                                                                                                           | $\text{OXYGEN-MOLECULE}[c] + \text{NADP}[c] + \text{S-CHEILANTHIFOLINE}[c] \rightarrow \text{S-STYLOPHINE}[c] + \text{NADP}[c]$                 |                                                                                                             |                                                                                                             |          | PWY-5287                  |
| 1.14.11.18-RXN           | Phytanoyl-CoA dioxygenase                                                                                                         | $\text{CPD-208}[c] + 2\text{-KETOGLOUTARATE}[c] + \text{OXY} \rightarrow \text{SOLYC05G007970.2}$                                               | $\text{SOLYC05G007970.2-MONOMER}$                                                                           |                                                                                                             |          |                           |
| 1.14.11.20-RXN           | Desacetoxyvindoline 4-hydroxylase                                                                                                 | $\text{DEACETOXYVINDOLINE}[c] + 2\text{-KETOGLOUTARATE}[c] + \text{OXY} \rightarrow \text{SOLYC09G01000.2}$                                     | $\text{SOLYC09G01000.2-MONOMER}$ or $\text{SOLYC09G01000.2-MONOMER}$ or $\text{SOLYC09G01000.2-MONOMER}$    |                                                                                                             |          |                           |
| 1.14.13.36-RXN           | 5-O-(4-coumaroyl)-D-quinic acid 3-O-methyltransferase                                                                             | $\text{OXYGEN-MOLECULE}[c] + \text{NADP}[c] + 4\text{-O-Coumaroyl-D-Quinic acid} \rightarrow \text{SOLYC10G078220.1}$                           | $\text{SOLYC10G078220.1-MONOMER}$ or $\text{SOLYC10G078220.1-MONOMER}$ or $\text{SOLYC10G078220.1-MONOMER}$ |                                                                                                             |          | PWY-6040 PWY-6039 PWY-361 |
| 1.14.13.55-RXN           | Protopine 6-monooxygenase                                                                                                         | $\text{OXYGEN-MOLECULE}[c] + \text{NADP}[c] + \text{PROTOPINE}[c] \rightarrow 6\text{-HYDROXYPROTOPINE}[c] + \text{NADP}[c]$                    |                                                                                                             |                                                                                                             |          | PWY-5287                  |
| 1.14.13.56-RXN           | Dihydroanguinarine 10-monooxygenase                                                                                               | $\text{OXYGEN-MOLECULE}[c] + \text{DIHYDROANGUINARINE}[c] + \text{NADP}[c] \rightarrow 10\text{-HYDROXYDIHYDROANGUINARINE}[c] + \text{NADP}[c]$ |                                                                                                             |                                                                                                             |          | PWY-5287                  |
| 1.14.13.57-RXN           | Dihydrochelirubine 12-monooxygenase                                                                                               | $\text{OXYGEN-MOLECULE}[c] + \text{DIHYDROCHELIRUBINE}[c] + \text{NADP}[c] \rightarrow 12\text{-HYDROXYDIHYDROCHELIRUBINE}[c] + \text{NADP}[c]$ |                                                                                                             |                                                                                                             |          | PWY-5287                  |
| 1.14.13.70-RXN           | Sterol 14-demethylase                                                                                                             | $\text{OBTUSIFOLIOL}[c] + 3.0 \text{ NADP}[c] + 3.0 \text{ OX} \rightarrow \text{SOLYC01G008110.2}$                                             | $\text{SOLYC01G008110.2-MONOMER}$                                                                           |                                                                                                             |          | PWY-2541                  |
| 1.14.13.71-RXN           | N-methylcoclaurine 3-monooxygenase                                                                                                | $\text{OXYGEN-MOLECULE}[c] + \text{NADP}[c] + \text{S-N-Methylcoclaurine} \rightarrow \text{SOLYC08G080380.2}$                                  | $\text{SOLYC08G080380.2-MONOMER}$ or $\text{SOLYC08G080380.2-MONOMER}$ or $\text{SOLYC08G080380.2-MONOMER}$ |                                                                                                             |          |                           |
| 1.14.13.78-RXN           | Ent-kaurene oxidase                                                                                                               | $\text{OXYGEN-MOLECULE}[c] + \text{NADP}[c] + \text{CPD} \rightarrow \text{SOLYC04G083160.1}$                                                   | $\text{SOLYC04G083160.1-MONOMER}$                                                                           |                                                                                                             |          | PWY-5034                  |
| 1.14.13.79-RXN           | Ent-kaurenoic acid oxidase                                                                                                        | $\text{OXYGEN-MOLECULE}[c] + \text{CPD-1F-132}[c] + \text{SOLYC12G006460.1}$                                                                    | $\text{SOLYC12G006460.1-MONOMER}$ or $\text{SOLYC12G006460.1-MONOMER}$ or $\text{SOLYC12G006460.1-MONOMER}$ |                                                                                                             |          | PWY-5034                  |
| 1.14.13.8-RXN            | Dimethylallene monooxygenase                                                                                                      | $\text{OXYGEN-MOLECULE}[c] + \text{NADP}[c] + \text{N-DIMETHYLLAN} \rightarrow \text{SOLYC01G112100.2}$                                         | $\text{SOLYC01G112100.2-MONOMER}$ or $\text{SOLYC01G112100.2-MONOMER}$ or $\text{SOLYC01G112100.2-MONOMER}$ |                                                                                                             |          |                           |
| 1.14.13.93-RXN           | (+)-abscisic acid 8-hydroxylase                                                                                                   | $\text{OXYGEN-MOLECULE}[c] + \text{CPD-693}[c] + \text{NA} \rightarrow \text{SOLYC07G055970.1}$                                                 | $\text{SOLYC07G055970.1-MONOMER}$ or $\text{SOLYC07G055970.1-MONOMER}$ or $\text{SOLYC07G055970.1-MONOMER}$ |                                                                                                             |          | PWY-5271                  |
| 1.14.21.2-RXN            | (S)-cheilanthifoline synthase                                                                                                     | $\text{OXYGEN-MOLECULE}[c] + \text{S-SCOUERINE}[c] + \text{NADP}[c] \rightarrow \text{S-CHEILANTHIFOLINE}[c] + \text{NADP}[c]$                  |                                                                                                             |                                                                                                             |          | PWY-5287                  |
| 1.14.21.6-RXN(NAD)       | 1.14.21.6-RXN(NAD)[c]                                                                                                             | $\text{NADH}[c] + \text{CPD-4186}[c] + \text{OXYGEN-MOLECULE}[c] \rightarrow \text{NAD}[c] + \text{CPD-4187}[c]$                                |                                                                                                             |                                                                                                             |          |                           |
| 1.14.21.6-RXN(NADP)      | 1.14.21.6-RXN(NADP)[c]                                                                                                            | $\text{OXYGEN-MOLECULE}[c] + \text{CPD-4186}[c] + \text{NADP}[c] \rightarrow \text{CPD-4187}[c] + \text{NADP}[c]$                               |                                                                                                             |                                                                                                             |          |                           |

Modelo genômico parcial do *Solanum lycopersicum*

## 7 Discussão:

A aplicação de ferramentas de modelagem metabólica permitiu obter um modelo para predição do comportamento do *Solanum lycopersicum*. A escolha de um modelo de qualidade pode impactar diretamente as funções metabólicas capturadas por ele, impactando diretamente nas predições obtidas.

A reconstrução de um modelo metabólico em escala genômica (GEMs) ocorre em 4 fases principais. Na primeira fase elaborava-se um rascunho de reconstrução, utilizando dados de transcriptomas do organismo-alvo e bancos de dados bioquímicos. Na segunda, realiza-se o refinamento e a curadoria manual, propondo que o rascunho seja reavaliado e refinado. A terceira etapa abrange a conversão para um modelo matemático e modelos específicos. Na quarta etapa, o modelo criado na etapa anterior é testado e validado (THIELE; PALSSON, 2010).

Todas as sequências de dados foram submetidas a processos de controle de qualidade, sendo todas aprovadas nos devidos parâmetros estabelecidos. Essas sequências contribuiriam para a contextualização das redes metabólicas, fornecendo informações sobre as expressões dos

genes no contexto estudado é ajudando nas predições das possíveis vias ativas do modelo. Cada uma das sequências foi validada para as etapas seguintes, através da aplicação de parâmetros de qualidade, como ilustrado na **Figura 3**. Esses resultados são de grande importância, pois certificam que o nosso modelo possa realizar as predições no contexto estudado.

O processo de normalização, realizado pelo método do TPM, que normaliza a expressão genica considerando tanto comprimento do gene quanto a profundidade do sequenciamento, foi uma etapa bem-sucedida para todos os nove conjuntos de dados de transcriptomas. Os resultados obtidos pelo método do TPM permitiram uma comparação direta dos níveis de expressão de diferentes genes com o nosso genoma de referência, garantindo que as sequências estivessem devidamente alinhadas. Isso possibilitou a predição dos genes ativos em cada fase do amadurecimento do fruto em etapas subsequentes. Esses resultados asseguraram que todas as sequências representavam variações biológicas reais antes de serem implementadas ao modelo, demonstram a viabilidade técnica desses conjuntos de dados escolhido para o presente projeto.

A integração dos conjuntos dos dados ao modelo foi realizada por meio do pacote Gene Inactivity Moderated by Metabolism and Expression (GIMME)(BECKER; PALSSON, 2008). Essa integração, utilizando os dados de transcriptômicos, devidamente tratados nas etapas anteriores, foi implementada através da ferramenta COBRA Toolbox (CONstraint-Based Reconstruction and Analysis)(SCHELLENBERGER et al., 2011). A inserção desses dados de expressão gênica em diferentes estados, combinados com a reconstrução de rede metabólica, permitiu entender os conhecimentos acerca dos princípios regulatórios do tomate, abrangendo todas as reações e vias metabólicas. Isso permitiu a aplicação de algoritmos para que se realizasse previsões de análise de balanço de fluxo. A inclusão desses dados é de grande relevância, como demonstrado no estudo de Akesson(ÅKESSON; FÖRSTER; NIELSEN, 2004), que demonstra a importância da integração de dados de transcriptômicos a modelo genômico para previsão do comportamento metabólico, permitindo previsões quantitativas de fluxo.

Aqui demonstramos a formulação de uma equação de biomassa, composta de 46 componentes da massa seca do tomate, uma vez que o modelo base *i* HY3410 não possui uma equação de biomassa. A escrita dessa equação foi um passo crítico para a simulação do crescimento no modelo. Essas equações representam os componentes combinados em quantidades estequiométricas, descrevendo as taxas de produção e consumo de metabólitos que compõem o crescimento da célula. Além disso, exercem papel central durante a análise de

balanço de fluxo. Os componentes dessa equação incluem aminoácidos, carboidratos, bases nitrogenadas, metabólitos solúveis, ácidos graxos e pigmentos que resultam em ADP, fosfato e difosfato.

Não foi obtido um resultado positivo na primeira realização de FBA após a inserção das reações de biomassa. Esse resultado evidenciou a presença de algumas reações no modelo que não estavam produzindo ou não possuíam nenhum tipo de transporte identificado, os chamados metabólitos dead-end (DEM)(MACKIE et al., 2013). Devido a esse cenário, foi realizada uma investigação, a fim de se identificar e eliminar esses metabólitos. Como efeito houve uma significativa modificação no modelo, através da adição de oitos reações de *exchange*. Essas reações foram incluídas com objetivo de eliminar as reações que não estavam produzindo ou não possuíam nenhum tipo de transporte.

Demostramos que foi obtido resultado positivo na segunda realização de FBA. Esse resultado foi de grande relevância, indicando que, na simulação do modelo, havia a produção de fluxo geral sem interrupção, indicando que o nosso modelo podia sintetizar os componentes essenciais para o seu crescimento nas condições simulada.

Devido a reconstrução da rede metabólica do tomate abranger todas as reações e enzimas conhecidas, e por meio da análise de FBA para o cálculo do fluxo metabólico, obtivemos uma visão sobre os genes e vias importantes envolvidas no processo de amadurecimento do *Solanum lycopersicum*. Essas análises foi realizada com base na aplicação de previsão de taxa de crescimento. No futuro, com a validação experimental do modelo, será possível utilizá-lo em estudos direcionados a interferências de engenharia metabólica, com o objetivo de realizar estratégias pós colheita do fruto.

## 8 Conclusões:

As abordagens aplicadas produziram um modelo-contexto específico para o *Solanum lycopersicum*, utilizando como base um modelo já existente *i* HY3410 com a incorporação de dados transcriptômicos. Isso permitiu um entendimento dos principais processos bioquímicos envolvidos durante todo o estágio de amadurecimento do *Solanum lycopersicum*, bem como as predições dos genes e enzimas relacionadas durante todo o processo. Portanto, o modelo resultante mostrou-se viável, como demonstrado pela produção de biomassa, é, em futuros

estudos com a validação experimental, pode ser utilizado para propor estratégias de engenharia metabólica visando aumentar a conservação pós-colheita do fruto.

### **Referências:**

ÅKESSON, M.; FÖRSTER, J.; NIELSEN, J. Integration of gene expression data into genome-scale metabolic models. **Metabolic Engineering**, v. 6, n. 4, p. 285–293, out. 2004.

AL-AMRI, S. M. Improved growth, productivity and quality of tomato (*Solanum lycopersicum* L.) plants through application of shikimic acid. **Saudi Journal of Biological Sciences**, v. 20, n. 4, p. 339–345, out. 2013.

ALEXANDER, L.; GRIERSON, D. Ethylene biosynthesis and action in tomato: A model for climacteric fruit ripening. **Journal of Experimental Botany**, v. 53, n. 377, p. 2039–2055, 1 out. 2002.

ANTONAKOUDIS, A. et al. The era of big data: Genome-scale modelling meets machine learning. **Computational and Structural Biotechnology Journal** Elsevier B.V., , 1 jan. 2020.

ANWAR, R.; FATIMA, T.; MATTOO, A. K. Tomatoes: A Model Crop of Solanaceous Plants. Em: **Oxford Research Encyclopedia of Environmental Science**. [s.l.] Oxford University Press, 2019.

BECKER, S. A.; PALSSON, B. O. Context-specific metabolic networks are consistent with experiments. **PLoS Computational Biology**, v. 4, n. 5, 2008.

BORDBAR, A. et al. **Constraint-based models predict metabolic and associated cellular functions.** *Nature Reviews Genetics*, fev. 2014.

BRANDÃO FILHO, J. U. T. et al. Solanáceas. Em: **Hortalças-fruto.** [s.l.] EDUEM, 2018. p. 37–70.

CARTER, E. L.; CONSTANTINIDOU, C.; ALAM, M. T. **Applications of genome-scale metabolic models to investigate microbial metabolic adaptations in response to genetic or environmental perturbations.** *Briefings in Bioinformatics* Oxford University Press, , 1 jan. 2024.

COLOMBIÉ, S. et al. Modelling central metabolic fluxes by constraint-based optimization reveals metabolic reprogramming of developing *Solanum lycopersicum* (tomato) fruit. **Plant Journal**, v. 81, n. 1, p. 24–39, 1 jan. 2015.

DEL CARRATORE, F. et al. **Biotechnological application of *Streptomyces* for the production of clinical drugs and other bioactive molecules.** *Current Opinion in Biotechnology* Elsevier Ltd, , 1 out. 2022.

EDRIS, S. et al. Early Fruit Development Regulation-Related Genes Concordantly Expressed with TCP Transcription Factors in Tomato (*Solanum lycopersicum*). **Current Issues in Molecular Biology**, v. 45, n. 3, p. 2372–2380, 1 mar. 2023.

EDWARDS, J. S.; PALSSON, B. O. **The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities.** [s.l: s.n.]. Disponível em: <<http://gcrd.ucsd.edu/downloads.html>>.

JIA, H. et al. Key transcription factors regulate fruit ripening and metabolite accumulation in tomato. **Plant Physiology**, v. 195, n. 3, p. 2256–2273, 1 jul. 2024.

KAMSEN, R. et al. Transcriptome integrated metabolic modeling of carbon assimilation underlying storage root development in cassava. **Scientific Reports**, v. 11, n. 1, 1 dez. 2021.

KLEE, H. J.; GIOVANNONI, J. J. Genetics and control of tomato fruit ripening and quality attributes. **Annual Review of Genetics**, v. 45, p. 41–59, 2011.

KRISHNA, R. et al. **Biotechnological Interventions in Tomato (*Solanum lycopersicum*) for Drought Stress Tolerance: Achievements and Future Prospects**. **BioTechMDPI**, , 1 dez. 2022.

LIU, Z. et al. Integration of transcriptome and metabolome reveals regulatory mechanisms of volatile flavor formation during tomato fruit ripening. **Horticultural Plant Journal**, jun. 2024.

MACIEL, G. M. et al. New insights into the use of dwarf tomato plants for pest resistance. **Bragantia**, v. 83, 2024.

MACKIE, A. et al. Dead End Metabolites - Defining the Known Unknowns of the E. coli Metabolic Network. **PLoS ONE**, v. 8, n. 9, 23 set. 2013.

MIRZAYEVA, S. et al. Physiology and Gene Expression Analysis of Tomato (*Solanum lycopersicum* L.) Exposed to Combined-Virus and Drought Stresses. **Plant Pathology Journal**, v. 39, n. 5, p. 466–485, 1 out. 2023.

ORTH, J. D.; THIELE, I.; PALSSON, B. O. **What is flux balance analysis?** *Nature Biotechnology*, mar. 2010.

PAN, F. et al. Transcriptome and Metabolome Provide Insights into Fruit Ripening of Cherry Tomato (*Solanum lycopersicum* var. *cerasiforme*). *Plants*, v. 12, n. 19, 1 out. 2023.

QUINET, M. et al. **Tomato Fruit Development and Metabolism.** *Frontiers in Plant Science* Frontiers Media S.A., , 29 nov. 2019.

RAU, M. H. et al. Genome-Scale Metabolic Modeling Combined with Transcriptome Profiling Provides Mechanistic Understanding of *Streptococcus thermophilus* CH8 Metabolism. *Applied and Environmental Microbiology*, v. 88, n. 16, 1 ago. 2022.

SATO, S. et al. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, v. 485, n. 7400, p. 635–641, 2012.

SAYED, E. et al. **Application of Exogenous Ascorbic Acid on Tomato (*Solanum lycopersicum* L.) Seeds under NaCl Salinity Stress** *Int. J. Curr. Res. Biosci. Plant Biol.* [s.l: s.n.]. Disponível em: <[www.ijcrbp.com](http://www.ijcrbp.com)>.

SCHAUER, N.; ZAMIR, D.; FERNIE, A. R. **Metabolic profiling of leaves and fruit of wild species tomato: A survey of the *Solanum lycopersicum* complex.** *Journal of Experimental Botany. Anais...* jan. 2005.

SCHELLENBERGER, J. et al. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nature Protocols*, v. 6, n. 9, p. 1290–1307, set. 2011.

SHINOZAKI, Y. et al. High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. **Nature Communications**, v. 9, n. 1, 1 dez. 2018.

SILASTE, M. L. et al. Tomato juice decreases LDL cholesterol levels and increases LDL resistance to oxidation. **British Journal of Nutrition**, v. 98, n. 6, p. 1251–1258, dez. 2007.

SIMEONIDIS, E.; PRICE, N. D. **Genome-scale modeling for metabolic engineering. Journal of Industrial Microbiology and Biotechnology** Springer Verlag, , 1 mar. 2015.

SOTO-ZAMORA, G. et al. Effects of postharvest hot air treatments on the quality and antioxidant levels in tomato fruit. **LWT**, v. 38, n. 6, p. 657–663, 2005.

TAMBURINO, R. et al. Cultivated tomato (*Solanum lycopersicum* l.) suffered a severe cytoplasmic bottleneck during domestication: Implications from chloroplast genomes. **Plants**, v. 9, n. 11, p. 1–15, 1 nov. 2020.

THIELE, I.; PALSSON, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. **Nature Protocols**, v. 5, n. 1, p. 93–121, jan. 2010.

VARGAS, J. E. E.; AGUIRRE, N. C.; CORONADO, Y. M. Study of the genetic diversity of tomato (*Solanum* spp.) with ISSR markers. **Revista Ceres**, v. 67, n. 3, p. 199–199, 1 maio 2020.

VOLOSHINA, M. et al. Physiological and Biochemical Responses of *Solanum lycopersicum* L. to Benzo[a]pyrene Contaminated Soils. **International Journal of Molecular Sciences**, v. 24, n. 4, 1 fev. 2023.

WANG, H. et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. **PLoS Computational Biology**, v. 14, n. 10, 1 out. 2018.

WANG, Y. et al. Plasma carotenoids and breast cancer risk in the Cancer Prevention Study II Nutrition Cohort. **Cancer Causes and Control**, v. 26, n. 9, p. 1233–1244, 21 set. 2015.

YOON, K. N. et al. Gamma irradiation delays tomato (*Solanum lycopersicum*) ripening by inducing transcriptional changes. **Journal of the Science of Food and Agriculture**, v. 103, n. 13, p. 6640–6653, 1 out. 2023.

YUAN, H. et al. A genome-scale metabolic network reconstruction of tomato (*Solanum lycopersicum* L.) and its application to photorespiratory metabolism. **Plant Journal**, v. 85, n. 2, p. 289–304, 1 jan. 2016.