UNIVERSIDADE FEDERAL DE MINAS GERAIS

Programa Interunidades de Pós-Graduação em Bioinformática

Miquéias Fernandes

ANÁLISE DE SPLICING ALTERNATIVO DURANTE O PROCESSO DE AMADURECIMENTO DE FRUTOS: aplicação em café e tomate

Belo Horizonte

2023

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Programa Interunidades de Pós-Graduação em Bioinformática

Miquéias Fernandes

ANÁLISE DE SPLICING ALTERNATIVO DURANTE O PROCESSO DE AMADURECIMENTO DE FRUTOS: aplicação em café e tomate

Tese apresentada ao programa de pós-graduação da UFMG como requisito para obtenção do título de Doutor em Bioinformática.

Orientador: Prof. Dr. Tiago Antônio de Oliveira Mendes

Belo Horizonte

2023

043 Fernandes, Miquéias.
Análise de splicing alternativo durante o processo de amadurecimento de frutos: aplicação em café e tomate [manuscrito] / Miquéias Fernandes. – 2023.
78 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Tiago Antônio de Oliveira Mendes.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Café. 3. Lycopersicon esculentum. 4. Processamento Alternativo I. Mendes, Tiago Antônio de Oliveira. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

ATA DE DEFESA DE TESE

MIQUEIAS FERNANDES

Às nove horas do dia **25 de agosto de 2023**, reuniu-se, através de videoconferência, a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Análise de Splicing Alternativo Durante o Processo de Amadurecimento de Frutos: Aplicação em Café e Tomate"**, requisito para obtenção do grau de Doutor em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, **Dr. Tiago Antonio de Oliveira Mendes**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Professor(a)/Professora(a)	**Instituição**	**Indicação**
Dr. Tiago Antonio de Oliveira Mendes - Orientador	Universidade Federal de Viçosa	**Aprovado**
Dra. Eveline Teixeira Caixeta Moura - Coorientadora	Empresa Brasileira de Pesquisa Agropecuária	**Aprovado**
Dra. Glória Regina Franco	Universidade Federal de Minas Gerais	**Aprovado**
Dr. José Miguel Ortega	Universidade Federal de Minas Gerais	**Aprovado**
Dra. Mariana Lima Boroni Martins	Instituto Nacional de Câncer	**Aprovado**
Dr. Sérgio Vale Aguiar Campos	Universidade Federal de Minas Gerais	**Aprovado**
Dr. Jose Cleydson Ferreira Silva	University of Florida	**Aprovado**

Pelas indicações, o candidato foi considerado: **Aprovado**

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 25 de agosto de 2023.

Documento assinado eletronicamente por **Jose Cleydson Ferreira da Silva**, **Usuário Externo**, em 25/08/2023, às 15:04, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Tiago Antônio de Oliveira Mendes**, **Usuário Externo**, em 25/08/2023, às 15:22, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Mariana Lima Boroni Martins**, **Usuária Externa**, em 25/08/2023, às 18:23, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Eveline Teixeira Caixeta Moura**, **Usuário Externo**, em 28/08/2023, às 09:07, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Sergio Vale Aguiar Campos**, **Professor do Magistério Superior**, em 28/08/2023, às 11:39, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Gloria Regina Franco**, **Professora do Magistério Superior**, em 28/08/2023, às 19:21, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jose Miguel Ortega**, **Servidor(a)**, em 01/09/2023, às 12:31, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2570832** e o código CRC **33A81FED**.

Referência: Processo nº 23072.251550/2023-41 SEI nº 2570832

À minha bióloga Alda,

Isaac e Ruth.

AGRADECIMENTOS

Agradeço a Deus, o supremo regente do universo pela dádiva de alcançar uma gota da sua sabedoria iluminando minha mente por sua benevolência. Apesar de eu ter entrado no doutorado em plena pandemia não me desamparou. Enviou recursos para viabilizar minha pesquisa. Sobretudo, me deu toda saúde para chegar vivo e apetitoso ao sucesso desse trabalho.

Agradeço minha família, em especial à minha esposa bióloga geneticista Dra. Alda que me inspirou ainda mais a correr ao encontro de meus objetivos. Nossos filhos Isaac e Ruth sempre compreensivos com os compromissos do papai. São meu tudo❤️, sempre tratando-me com todo carinho, preenchendo-me da verdadeira boa companhia que precisei durante os desafios.

Agradeço ao PPG Genética e Melhoramento da UFES, por ter me fornecido todo conhecimento biológico que precisei para desenvolver pesquisa com gene e transcrito. Do trajeto que começa com a Lidiane me indicando o laboratório de Biometria até o Mestrado, todos compartilharam algo que usei aqui, somado ao que aprendi na graduação em sistemas de informação na mesma instituição.

Agradeço ao Google, ao Galaxy/EU, ao Periódicos Capes, e a tantas outras ferramentas tecnológicas que foram desenvolvidas e são disponibilizadas gratuitamente na internet. Nesse trabalho fiz muito uso de várias dessas ferramentas. O mundo é aberto, o mundo é livre.

Agradeço ao Banestes, na pessoa do coordenador Cássio que durante a realização das disciplinas me autorizou alteração na escala (para 7 às 13) de modo que eu consegui cumprir os créditos necessários. Estendo, claro, aos demais colegas que compartilharam comigo a boa ansiedade de finalizar com sucesso esse trabalho.

Agradeço ao professor Tiago por me orientar, e me aguentar, desde 2018 quando fui desnorteado à UFMG realizar uma disciplina condensada de Empreendedorismo em Bioinformática. Assim também agradeço ao excelente programa de pós em Bioinfo, secretários e todos envolvidos pela pessoa do Professor Ari com quem também pude aprender metagenômica. Em geral todo corpo docente que me transmitiu um colossal dilúvio de conhecimento, sou mais que feliz por mesmo nas minhas limitações ter absorvido uma boa parte.

Agradeço aos colegas da UFMG, colegas dos laboratórios de biologia sintética e Bioagro da UFV e as fundações estadual e federal de apoio a pesquisa.

Aos que estão ao meu lado e torcem por mim, como meus pais e irmãs, gratidão.

Do nada obtém conhecimento,
Com sabedoria gera matéria,
Toma molécula e dá vida,
Só de perfeição faz tudo.
É a própria energia sendo o próprio criador.

RESUMO

O Splicing Alternativo (AS) é um mecanismo cotranscricional que habilita o eucarioto estender seu proteoma mesmo com uma quantidade limitada de genes. A maquinaria celular realiza o AS pela combinação de regiões alternativas das isoformas, tanto em transcritos produtivos quanto em não produtivos. Os eventos de AS podem ser preditos com dados de RNASeq. Em plantas o evento mais frequente é o intron retention (IR) que pode ter efeito regulatório, por exemplo, ao inserir um códon de parada prematuro (PTC) que leva a degradação do transcrito pela via nonsense-mediated decay (NMD). Para estudar potenciais eventos de AS no processo biológico de amadurecimento de grãos de café e tomate conduzimos um estudo de diferencial AS (DAS) utilizando dados de RNASeq obtidos para experimento de expressão diferencial e o genoma de referência para identificar DAS. Para isso, nós desenvolvemos pipelines em Python afim de realizar uma curadoria automatizada dos 202 genes alvo que foram identificados pelo rMATS em 241 eventos de DAS nos contrastes de frutos verde com estágio intermediário de amadurecimento (yellow), e, maduro com yellow. Realizamos em seguida uma curadoria manual dos 241 eventos enriquecidos com a anotação do Interproscan5 que levou ao interesse de validar experimentalmente os genes Potassium channel AKT1 e Apyrase 7 com auxílio de PCR e qPCR. Os desafios identificados durante as análises de AS em dados de RNASeq nos induziram a desenvolver um aplicativo de interface amigável para explorar dados de DAS por RNASeq. O aplicativo foi desenvolvido composto de três módulos GeneAPPScript, GeneAPPServer, GeneAPPExplorer. O modulo GeneAPPScript é um wrapper poderoso que permite executar uma análise completa de DAS desde a obtenção dos dados em rede até a anotação funcional dos genes sob DAS. Esse modulo pode ser executado em distros Debian, como no ambiente Google Colaboratory (Colab) onde foi desenvolvido. Já o modulo GeneAPPServer é um backend Flask que permite integrar saídas de diferentes softwares de análise de DAS que geram dados em outputs tabulares. Usando o GeneAPPExplorer o usuário pode gerar dezenas de gráficos para visualizar graficamente importantes resultados implícitos nas tabelas técnicas exportadas pelos softwares de análises de DAS. Além disso pelo webapp o pesquisador tem acesso a tabelas enriquecidas de dados de anotação funcional, estrutural e atributos dos eventos. O GeneAPP tem potencial de contribuir na análise de AS em vários outros trabalhos depositados em bancos de dados públicos onde só foi analisado expressão diferencial no nível de gene, permitindo expandir o entendimento dos mecanismos moleculares e explorar o transcriptoma no nível de isoformas.

Palavras-chave: Amadurecimento, Café, Splicing Alternativo, Tomate, webapp.

ABSTRACT

Alternative Splicing (AS) is a co-transcriptional mechanism that enables the eukaryote to extend its proteome even with a limited number of genes. The cellular machinery performs AS by combining alternative regions of the isoforms in both productive and non-productive transcripts. AS events can be predicted using RNASeq data. In plants, the most frequent event is the retention of introns (IR) that can have a regulatory effect, for example, inserting a premature stop codon (PTC) that leads to degradation of the transcript through the nonsense-mediated decay (NMD) pathway. To study potential AS events in the biological process of coffee bean and tomato ripening, we conducted a DAS study using RNASeq data obtained for differential expression experiment and the referential coffee genome to identify differential AS (DAS). For this, we developed pipelines in Python to perform an automated curation of the 202 target genes that were identified with 241 DAS events by rMATS in the comparisons of early green fruits, intermediate yellow and final red ripening stage. We then carried out a manual curation of the 241 events enriched with the Interproscan5 annotation with further experimentally validating of Potassium channel AKT1 and Apyrase 7 genes under differential alternative expression during coffee grain ripening using conventional PCR and qPCR. Due to challenges identified during this analysis associated with the relationship of AS events and RNASeq data processing, we built an application of user-friendly APP to predict AS in RNASeq data. The application development is composed of three modules named GeneAPPScript, GeneAPPServer, GeneAPPExplorer. The GeneAPPScript module is a powerful wrapper that enables to perform a complete DAS analysis from obtaining network data to functional annotation of genes under DAS. This module can run on Debian distros, such as the Google Collaboratory (Colab) environment where it was developed. The GeneAPPServer module is a Flask backend that allows you to integrate outputs from different DAS analysis software that generate data in tabular outputs. Using GeneAPPExplorer, the user can generate dozens of graphs to graphically visualize important results implicit in technical tables exported by DAS analysis software. In addition, through the webapp, the researcher has access to tables enriched with functional and structural annotation data and event attributes. GeneAPP will contribute to the analysis of AS in several other works deposited in public databases where only differential expression at the gene level was analyzed, allowing further explanation when exploring the transcriptome at the isoform level.

Keywords: Alternative splicing, coffee, ripening, tomato, webapp.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

A3SS ..... Alternative 3’ Alternative splicing site
A5SS ..... Alternative 5’ Alternative splicing site
AI ..... Artificial Intelligence
AS..... Alternative Splicing
BP..... base pairs
CLI..... Command line interface
CV ..... cultivar
DAS ..... Differential AS
DEG..... Differentially expressed gene
ES..... Exon skipping
GUI ..... Graphical user interface
HTS ..... High throughput sequencing
IR..... Intron Retention
ML ..... Machine Learning
MXE..... Mutually exclusive event
NMD ..... Nonsense-mediated mRNA decay
PSI..... Percent spliced in
PTC ..... Premature terminator codon

LISTA DE SÍMBOLOS

Δ ..... diferença entre duas quantidades
Ψ..... PSI – percent spliced in

SUMÁRIO

Capítulo 1.

Revisão bibliográfica do processo de amadurecimento de frutos e sua relação com o mecanismo de splicing alternativo definindo elementos, escopo, justificativas e objetivos deste trabalho.

1. INTRODUÇÃO

Frutos são ovários amadurecidos, uma importante evolução das plantas que florescem por viabilizar sua dispersão em ambientes terrestres (FORLANI; MASIERO; MIZZOTTI, 2019). Charles Darwin já trouxe indícios da função deles a um triplo proposito: (i) proteger as sementes e (ii) atrair frugívoros para (iii) dispersá-las (PESARESI *et al.*, 2014). Considerando frutos como órgãos das plantas especializados em dispersar sementes, conservado nas Angiospermas (KARLOVA *et al.*, 2014), no contexto evolutivo são umas das últimas evoluções nas plantas (CERRI; REALE, 2020) constituído de diferentes soluções para se tornar um mecanismo eficiente de transporte das sementes em vetor biótico (ZWOLAK; SIH, 2020) e abiótico (SEALE; NAKAYAMA, 2020).

Frutos são importantes tanto para os frugívoros quanto para as próprias plantas. Quanto às plantas, eles são úteis para dispersar as sementes, evitar competição entre parentes e espécies e reduzir a depressão por endogamia (SPENGLER, 2020). É importante observar que mesmo uma semente que não germina, pode contribuir como fonte de matéria orgânica. Já para frugívoros como o peixe tambaqui (*Colossoma macropomum* em sua fase adulta), a ave Fim-fim (*Euphonia chlorotica*), macacos-aranha (primatas da *Família Atelidae*) e morcegos da família Phyllistomidae, as frutas são a base da alimentação diária. A fruticultura é importante e movimenta bilhões de dólares no mundo todo. Além disso, com contexto ecológico, os frutos são fontes de alimentos de muitas comunidades Ameríndias e animais da fauna de diferentes Biomas do planeta.

Para humanos os frutos são deliciosos, nutritivos e podem ser ter uso terapêutico no tratamento de várias doenças (KARASAWA; MOHAN, 2018). Tendo em vista este múltiplo benefício deles, várias espécies frutíferas foram alvo de domesticação, tornando seus frutos comuns na dieta humana. Eles são o alimento de muitos outros organismos frugívoros além dos humanos como vertebrados, invertebrados e micróbios que consomem todo tipo de fruto (WHITEHEAD *et al.*, 2022). Para os humanos estão entre os prediletos: tomate, banana, melancia, maçã e laranja conforme o apêndice 7.1 Produção mundial de frutos. Os frutos são utilizados como alimento tanto *in natura* quanto processados, e para atender a essa demanda, são reservados cerca de 13 bilhões de hectares para plantio (FAO, 2021). Ainda

segundo a FAO a produção mundial em 2019 foi de mais de 796 mil toneladas, o que é um aumento de 52% com relação a mesma quantidade produzida em 2000.

Considerando o crescimento populacional nas duas últimas décadas de 6 bilhões para 8 bilhões e o mesmo crescimento em produção de frutos, segundo os dados da FAO, houve um aumento diário de 240 gramas para 270 gramas de consumo de fruto por pessoa, o que corresponde a apenas uma maçã e uma laranja. No sentido de atender e aprimorar o atendimento a essa demanda mundial programas de melhoramento vegetal tem buscado obter ganho em facilidade de manejo, resistência a stress biótico e abiótico, plasticidade, variedade, importância medicinal, e fatores de qualidade como sabor, textura e tempo de prateleira (BYRNE, 2012). Ao atingir seus objetivos, esses programas beneficiam a humanidade em vários aspectos como: menor custo do alimento, melhor qualidade e menos desperdício o que viabiliza uma dieta mais saudável. Uma vez que a boa alimentação favorece o bem-estar (MUJCIC; OSWALD, 2016), o melhoramento vegetal de frutíferas impacta positivamente na quantidade e na qualidade dos frutos.

Parte da quantidade de bilhões de toneladas de frutos produzidos pode não chegar ao consumidor devido a perdas pós-colheita de 20% ou mais, ocasionadas em parte por danos fisiológicos causados pela senescência do fruto (BOSE *et al.*, 2021; PAUL; PANDEY, 2014). Considerando seu consumo *in natura,* esse período de validade é chamado de tempo de prateleira. Uma vez que o fruto entra no processo de amadurecimento, ele passa por alterações irreversíveis (BAPAT *et al.*, 2010; GAPPER; MCQUINN; GIOVANNONI, 2013), o que exige que os processos de produção, armazenamento e entrega sejam sincronizados com o processo biológico de amadurecimento do fruto. Para conseguir isso, os produtores usam vários métodos artificiais para controlar ou retardar o amadurecimento com recursos físicos como: atmosfera controlada ou modificada, radiação UV, baixa temperatura; recursos químicos como: 1-MCP e hormônios (BOSE *et al.*, 2021; PAUL; PANDEY, 2014; PAYASI; SANWAL, 2010); ou ainda melhoramento ou edição genica.

Pelo exposto, os frutos são importantes fonte de alimento para os animais, para a própria planta e para a nutrição e saúde humana. Contudo seu valor está restrito ao estágio maduro que é atingido por um processo biológico complexo de ser manipulado. Nesse sentido o melhorista vegetal tenta obter frutos com a qualidade de maior tempo de prateleira, enquanto o produtor tenta entregar os frutos ao seu

consumidor em tempo de mercado que não se limita ao processo biológico de amadurecimento. Em outras palavras é desejável, mas não é trivial postergar o amadurecimento de frutos. Para entender melhor a complexidade desse processo, as próximas seções visam trazer uma revisão da literatura com o que existe de conhecimento para explicar a fisiologia dos frutos, alterações que suas partes são submetidas ao longo dos estágios de amadurecimento, e mecanismos trascritptômicos envolvidos nesse processo biológico.

1.1 A BIOLOGIA DOS FRUTOS

A transformação do ovário em fruto para guardar os óvulos como sementes e atrair animais é característico das angiospermas (PESARESI *et al.*, 2014). Esse processo complexo tem início com disparo de sinais que indicam que fertilização ocorreu, sendo seguido pelo desenvolvimento do fruto (FORLANI; MASIERO; MIZZOTTI, 2019). As sementes dependem da fertilização do óvulo na maioria das espécies. A quantidade delas que se desenvolve determina o tamanho do fruto, e quando elas estão maduras o fruto entra no processo de amadurecimento (FORLANI; MASIERO; MIZZOTTI, 2019; PESARESI *et al.*, 2014).

O desenvolvimento dos frutos envolve 3 fases básicas: i) desenvolvimento do ovário e iniciação da divisão celular, ii) divisão celular intensa que ocorre após a fertilização e iii) expansão celular que faz o fruto aumentar de tamanho (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). A Fig. 1 ilustra as fases que o fruto passa até se desprender da árvore. Durante o desenvolvimento a parede do ovário torna-se o pericarpo que que envolve a semente, o pericarpo por sua vez é composto de três camadas de diferentes espessuras e composição química: exocarpo, mesocarpo e endocarpo (CERRI; REALE, 2020; KARLOVA *et al.*, 2014). Como muitos processos moleculares que ocorrem no amadurecimento também ocorrem na senescência é difícil distingui-los, entretanto o amadurecimento é caracterizado como um processo de desenvolvimento (GAPPER; MCQUINN; GIOVANNONI, 2013).

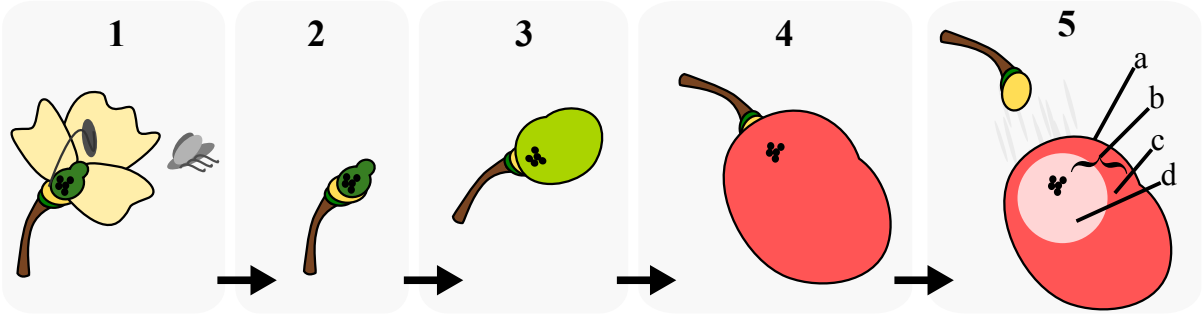


Fig. 1 As fases de desenvolvimento dos frutos, amadurecimento e senescência. A região 1 na imagem apresenta a primeira fase que consiste no desenvolvimento do ovário e a iniciação da divisão celular. Na segunda fase representada na região 2 na imagem, ocorre uma divisão celular mais intensa após a polinização. Na terceira fase apresentado na região 3 ocorre a expansão celular que aumenta o tamanho do fruto até atingir o estágio "Verde" totalmente desenvolvido. Quando as sementes estão maduras o fruto entra no processo de amadurecimento apresentado na região 4 da figura. Assim que está madura ele se desprende automaticamente da planta e entra em senescência conforme apresentado na região 5. As partes do fruto estão indicadas pela representação do fruto visto em corte transversal a) exocarpo, b) pericarpo, c) mesocarpo, d) endocarpo. Fonte: O Autor.

Os frutos podem ser classificados de várias maneiras considerando sua morfologia, deiscência, números de carpelos, anatomia, origem e constituição do pericarpo, número de sementes etc. (CERRI; REALE, 2020; KARLOVA *et al.*, 2014). Anatomicamente, o fruto de uma planta pode ser formado de uma única flor ou uma inflorescência (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). Este último é denominado múltiplo e um exemplo é o abacaxi. Fruto agregado como o morango é um conjunto de carpelos (KARLOVA *et al.*, 2014) em uma única flor. Frutos simples são classificados pela textura do pericarpo onde existem dois grupos: secos e carnosos. Há evidências fosseis e moleculares que os frutos carnosos evoluíram dos frutos secos (SEYMOUR *et al.*, 2008). Os frutos secos podem ainda ser classificadas como deiscentes, que são mais evoluídos pois abrem-se e dispersam suas sementes como as ervilhas por exemplo, ou indeiscente que não se abrem (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011; KARLOVA *et al.*, 2014; PESARESI *et al.*, 2014).

1.2 O PROCESSO BIOLÓGICO DE AMADURECIMENTO

O processo de amadurecimento envolve alterações profundas no metabolismo do pericarpo, o que causa modificações claramente perceptíveis na cor, textura, sabor e aroma do fruto (KARLOVA *et al.*, 2014; PAUL; PANDEY, 2014). Esse complexo processo é regulado no espaço-tempo por fatores genéticos, epigenéticos, hormonais, transcricionais e outros moleculares (CHEN; QIN; TIAN, 2020; HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). Esse processo tem sido foco de muitas pesquisas já que o amadurecimento tem relação com a dispersão de sementes e este é um objetivo fim no desenvolvimento da planta (FORLANI; MASIERO; MIZZOTTI, 2019). Nessas pesquisas, a observação de trocas gasosas em alguns frutos que foram classificados como climatéricos, permitiu identificar o gás etileno como hormônio que induz e acelera o processo de amadurecimento (ALEXANDER; GRIERSON, 2002; PAUL; PANDEY, 2014).

Frutos classificados como climatéricos possuem um pico de respiração com etileno conforme a Fig. 2. O tomate é o organismo modelo para estudo de amadurecimento de frutos climatéricos há mais de 40 anos (LI, SHAN; CHEN; GRIERSON, 2019), isso é justificado pela existência de vários mutantes com alterações em traços relacionados com o amadurecimento, o ciclo de vida curto que pode ocorrer durante o ano todo, além de possuir vários estudos de transcriptômica, proteômica e metabolômica já realizados com variedades da espécie (CHEN; QIN; TIAN, 2020; PESARESI *et al.*, 2014). O morango por sua vez é mais usado para estudo de amadurecimento de frutos não-climatéricos, nesses este processo tem pouca ou nenhuma relação com o etileno (GIOVANNONI, 2004).

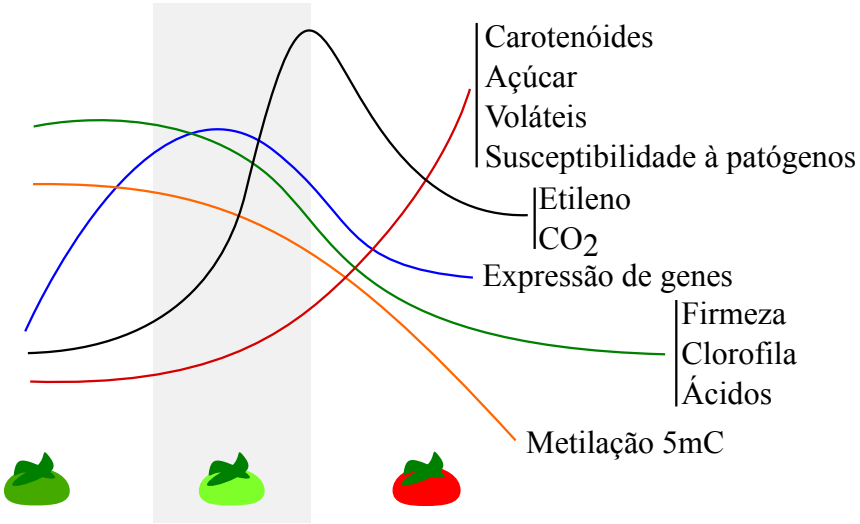


Fig. 2 Alterações que ocorrem no fruto durante o amadurecimento. Adaptado de CHEN; QIN; TIAN, (2020); FORLANI; MASIERO; MIZZOTTI, (2019); GIOVANNONI, (2004); LIU et al., (2015); PAUL; PANDEY, (2014); WANG et al., (2020). O perfil de metilação do DNA em vários loci com 5-methylcytosine (5mC) ocorre de maneira inversa na laranja. Em geral os ácidos tornam-se menos intensos à medida que a fruta fica mais doce, a clorofila se degrada e a fruta muda de cor, um pico de troca gasosa (inclusive com etileno intenso nas frutas climatéricas) dá início ao amadurecimento e à medida que ela fica com o pericarpo menos rígido torna-se mais suscetível a patógenos. Genes que transcrevem sinais de sinalização e biossíntese de compostos necessários para o amadurecimento (estes são listados no apêndice 7.2 Tabela de genes envolvidos no processo de amadurecimento) tem intensidade maior quando a fruta está entrando no processo de amadurecimento.

Ao se desprender da planta, o fruto não recebe água para compensar a transpiração, o que induz ao estresse hídrico levando à senescência (PAUL; PANDEY, 2014). Em senescência, alterações irreversíveis que o fruto sofre o torna mais suscetível a patógenos, o que facilita a dispersão das sementes; entretanto como as possibilidades de controle são limitadas – passar além do estágio de amadurecimento – torna-se um efeito negativo quando causa perdas na colheita (FORLANI; MASIERO; MIZZOTTI, 2019). Consumidores querem frutas maduras, mas de qualidade, com aparência atraente e isentas de infecção de patógenos. Para obter frutos resistentes a patógenos, a comunidade científica tenta entender como eles atacam as frutas, como crescem, qual mecanismo de patogenicidade usam e que condições oferecidas pelas frutas a eles são favoráveis ou restritivas (DÍAZ RICCI *et al.*, 2020).

1.2.1 ALTERAÇÕES NO FENÓTIPO CAUSADAS PELO AMADURECIMENTO

As alterações mais expressivas em frutos maduros alteram sua cor, aroma, textura e sabor. A cor de vários frutos é alterada quando eles estão maduros para sinalizar o estado do fruto aos frugívoros (LI, SHAN; CHEN; GRIERSON, 2021), com isso o fruto maduro outrora camuflado entre as folhas pode ser distinguido de longe. O aroma produzido pelo fruto quando maduro também é um importante indicador do estado de maturação para os predadores. Alterações na parede das células que compõem o pericarpo causadas por um processo complexo durante o amadurecimento causa mudanças na textura do fruto. Já o sabor é um traço de dupla importância no fruto tornando-o impróprio para o consumo enquanto imaturo, protegendo as sementes (proteger) de serem dispersadas antes que elas terminem seu desenvolvimento, mas torna-se atrativo aos frugívoros (atrair) para consumir o fruto e dispersar as sementes quando o fruto está maduro.

No tomate, o gene PHYTOTENE SYNTASE (PSY) tem sido bastante estudado pois o phytotene é usado como precursor do lycopeno que é o pigmento vermelho, já a anthocyanin apresenta-se roxo e azul, o rosa é gerado pela alteração na expressão do fator de transcrição MYB que regula flavonoides (KARLOVA *et al.*, 2014). Alguns mutantes de tomate que exibem cor diferente do tipo selvagem são usados para entender como essas cores são formadas, entre eles o mutante YELLOW-FLESH que possui coloração amarela, o mutante TANGERINE que apresenta coloração laranja, o mutante GREEN-FLESH de cor marrom, o mutante CRIMSON, HI-PIGMENT e INTENSE PIGMENT de cor vermelho intenso (GIOVANNONI, 2004; HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). O mutante COLORLESS NON RIPENING (Cnr) que apresenta fenótipo vermelho claro foi uma das primeiras evidências de influência epigenética no amadurecimento (CHEN; QIN; TIAN, 2020; TANG; GALLUSCI; LANG, 2020). Essa simples alteração na cor causada basicamente pelo acúmulo de carotenoides coloridos licopeno e β-caroteno e degradação da clorofila tem uma implicação ecológica grande, já que atrai predadores para dispersão das sementes (LI, SHAN *et al.*, 2020; PESARESI *et al.*, 2014).

O fruto exala aroma quando maduro e utiliza-o como recurso para que predadores sem visão ou em condição noturna possam percebê-lo quando verde, maduro ou ruim. Cerca de 400 compostos voláteis constituem o aroma do tomate maduro. Em geral os frutos possuem os seguintes voláteis: etileno, etanol,

acetaldeído, metanol, acetona, butanol, etano, hexanol, hexenol, 3-methyl butanal, ethyl acetate, propyl acetate, butyl acetate, propanol, acetate esters, ethyl butyrate, geraniol, octenal, octenol, citral, terpenos, carboxylic acids, sulphur compounds, amonia, jasmonatos, benzaldehyde, methyl salicylate, iso-, sec- ou tert- alcohols, hydrocarbons, ketones, esters, aldehydes e álcoois de alto carbono (PAUL; PANDEY, 2014). Os frutos produzem alguns desses voláteis para resistir a patógenos. A combinação e concentração desses voláteis produzem diferentes aromas, e afeta na preferência de consumidor (GAPPER; MCQUINN; GIOVANNONI, 2013).

A parede celular é uma matriz celulose-hemicelulose estruturada em pectinas, alteração nessas pectinas e outras transformações que variam de fruto para fruto tornam a parede celular macia (GAPPER; MCQUINN; GIOVANNONI, 2013). Nesse sentido enzimas modificadoras de pectina como polygalacturonase (PG), β-galactosidase (β-Gal), pectate lyase (PEL), pectin esterase (PE) e modificadoras de celulose/hemicelulose como 1,4-β-glucanase, xyloglucan trnansglycosylase / hydrolase (XTH) e expansin (EXP) estão associadas a alterações que tornam o fruto mais macio (KARLOVA *et al.*, 2014; LI, SHAN; CHEN; GRIERSON, 2021; WHITE, 2002). Em frutos carnosos como o tomate e a uva a pressão interna causada pelo desenvolvimento e acúmulo dos compostos também aumenta a firmeza do fruto (GAPPER; MCQUINN; GIOVANNONI, 2013). O pericarpo perde a firmeza quando a fruta perde água causando redução da pressão interna. Somado a isso, a perda da integridade da parede celular torna o fruto mais suscetível a patógenos (FORLANI; MASIERO; MIZZOTTI, 2019; GIOVANNONI, 2004). Estudos dos mecanismos que regulam o processo de alteração da textura têm atraído a comunidade cientifica, uma vez que o período de pericarpo integro determina o tempo de prateleira, que é alvo de interesse comercial (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). Assim estudos recentes com aplicação exógena de fitormônios como ABA e etileno são realizados para entender como induzem genes nas vias de remodelagem da parede celular (FORLANI; MASIERO; MIZZOTTI, 2019).

A alteração no sabor ocorre, em geral, pelo fato de durante desenvolvimento, o amido acumulado nas folhas durante a fotossíntese ser transportado para o fruto e convertido em açúcares como sucrose, glucose e frutose por enzimas de degradação. As enzimas de degradação são traduzidas por cerca de 11 genes regulados por fatores de transcrição como *zinc finger* durante o amadurecimento (LI, SHAN; CHEN;

GRIERSON, 2021). A biossíntese de açúcares e ácidos afeta a qualidade nutricional além do sabor (GIOVANNONI, 2004). O espectro de sabor dos frutos é obtido com a combinação de açucares, ácidos e aproximadamente 30 compostos derivados de aminoácidos, ácidos graxos, carotenoides, e outros terpenos (GAPPER; MCQUINN; GIOVANNONI, 2013). Segundo PAUL; PANDEY (2014) os compostos voláteis também contribuem com o sabor doce e a diferença de sabores (WHITE, 2002). É importante notar que mesmo sem valor sensorial, compostos alterados durante o amadurecimento tem importância econômica, já que determinam a qualidade do fruto com relação ao sabor do produto derivado dele como acontece no caso do café (CHENG *et al.*, 2016) e do cacau (KONGOR *et al.*, 2016). Nesses dois, por exemplo, o sabor do produto depende de compostos que são formados ou alterados durante o processo de amadurecimento do fruto.

1.2.2 FATORES QUE INFLUENCIAM NO PROCESSO DE AMADURECIMENTO

Estudos no padrão de respiração dos frutos identificaram o volátil etileno em alta intensidade no início do amadurecimento de vários frutos que foram classificadas então como climatéricos (ALEXANDER; GRIERSON, 2002; HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). Os primeiros registros da percepção do etileno como fitormônio intermediando o amadurecimento no contexto fisiológico, bioquímico e molecular datam 1962 (LIU, MINGCHUN *et al.*, 2015). Desde então, várias evidências sugerem-no como o principal hormônio regulador do amadurecimento em frutos climatéricos (FORLANI; MASIERO; MIZZOTTI, 2019; PAUL; PANDEY, 2014). Verificou-se ainda que a aplicação externa de etileno em frutos induz o amadurecimento. Entretanto há um limite, imposto por um mecanismo desconhecido, que impede o amadurecimento precoce do fruto (LI, SHAN; CHEN; GRIERSON, 2021; PESARESI *et al.*, 2014).

1.2.2.1 Fatores hormonais

A biossíntese de etileno inicia com a conversão de metionina para S-adenosilmetionina (SAM), em seguida SAM é convertido para ácido 1-aminociclopropano-1-carboxílico (ACC) que será transformado em etileno pela enzima aminociclopropanocarboxilato oxidase (ACO) (ALEXANDER; GRIERSON, 2002; PAUL; PANDEY, 2014). Apesar do etileno ser produzido em baixos níveis em vários processos biológicos da planta e do fruto em um mecanismo denominado

sistema-1, quando o fruto entra no processo de amadurecimento uma quantidade massiva do hormônio é produzida no fruto pelo sistema-2, onde o próprio etileno induz sua biossíntese (ALEXANDER; GRIERSON, 2002; LI, SHAN; CHEN; GRIERSON, 2021). Nas frutas climatéricas o etileno é o sinal para dar início e continuidade no processo de amadurecimento (LI, SHAN; CHEN; GRIERSON, 2021).

Tomando o tomate como organismo modelo para estudo do amadurecimento em frutas climatéricas, verificou-se que o etileno sozinho é suficiente para despertar o processo de amadurecimento (LI, SHAN *et al.*, 2020), enquanto em frutos do tipo não-climatéricos sua síntese é indiferente ao amadurecimento. É intrigante que o uso do etileno pelo fruto para sinalizar o início do amadurecimento não tem relação filogenética, uma vez que há espécie dos dois tipos entre monocotiledôneas e dicotiledônias, há espécies dos dois tipos próximas filogeneticamente ou ainda os dois tipos podem estar em variedades da mesma espécie (GIOVANNONI, 2004; HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). É importante notar que a expressão diferencial dos genes durante o amadurecimento pode ser regulada em ambos os tipos climatérico e não-climatérico pelo etileno, o que pode indicar que diferentes aspectos do amadurecimento podem ser classificados dos dois tipos no mesmo fruto (LI, SHAN; CHEN; GRIERSON, 2019). Em geral é classificado como climatérico: abacate, banana, café, caqui, cherimólia, damasco, goiaba, kiwi, mamão, manga, maracujá, mirtilo, pêssego, tomate, maçã; mas os frutos: abacaxi, cairela, amora, caju, cereja, framboesa, laranja, lichia, limão, melancia, morango, nêspera, pepino, romã, tangerina, tomate-de-arvore, toranja, uva são não-climatéricos (BAPAT *et al.*, 2010; FORLANI; MASIERO; MIZZOTTI, 2019; LI, SHAN *et al.*, 2020; PAUL; PANDEY, 2014; SÁGIO *et al.*, 2013). Mas melão, goiaba, ameixa, pera e pimenta são um modelo alternativo para estudo do amadurecimento devido a coexistência de variedade climatérica e não-climatérica na mesma espécie (FORLANI; MASIERO; MIZZOTTI, 2019; PEREIRA *et al.*, 2020). Independentemente do tipo, de acordo com GAO et al. (2020) fatores de transcrição ativados por etileno (ERF) apresentam funções reguladoras na biossíntese de etileno, coloração, formação do aroma, entre outros durante o processo de amadurecimento do fruto.

Há outros fitormônios envolvidos no processo de amadurecimento além do etileno, e ainda evidências de interações entre eles na regulação (CHEN; QIN; TIAN, 2020; FORLANI; MASIERO; MIZZOTTI, 2019). O fitormônio auxina (AUX) tem efeito

inverso ao exibido pelo etileno, inibindo o amadurecimento do fruto (LI, SHAN *et al.*, 2020; LI, SHAN; CHEN; GRIERSON, 2021). A AUX pode ser usada como substituta para sinais de polinização e fertilização (PESARESI *et al.*, 2014). Já o fitormônio ácido abscísico (ABA) tem a função de modular a transcrição dos genes de biossíntese de ACC e ACO, para isso ele é acumulado antes da manifestação do etileno, há indícios que ele tem mais importância em frutas não-climatéricas (CHEN; QIN; TIAN, 2020; FORLANI; MASIERO; MIZZOTTI, 2019; LI, SHAN *et al.*, 2020; LI, SHAN; CHEN; GRIERSON, 2021; PAYASI; SANWAL, 2010). Jasmonato é um fitormônio com funções em vários processos biológicos das plantas, estudos de sua função na maçã identificaram indução da produção de etileno causado por ele induzindo a de síntese de ACO e ACC (LI, SHAN; CHEN; GRIERSON, 2021). Os brassinosteróides (BR) são fatores de crescimento que têm mostrado efeito em induzir o amadurecimento pelo brassinolide e reprimir pelo brassinazole (LI, SHAN; CHEN; GRIERSON, 2021). Além dos citados, outros fitormônios como citocinina e giberelina, e ainda, compostos químicos como ácido salicílico (SA) e o 1-Metilciclopropeno (1-MCP) possuem atividade de atraso ou inibição de amadurecimento de frutos (FORLANI; MASIERO; MIZZOTTI, 2019; LI, SHAN; CHEN; GRIERSON, 2021; PAYASI; SANWAL, 2010; PESARESI *et al.*, 2014).

1.2.2.2 Fatores ambientais

O ambiente influencia no processo de amadurecimento. Câmaras de ambiente controlado são usadas para acelerar ou atrasar o amadurecimento. O controle da respiração da planta submetendo a em atmosfera controlada, atmosfera modificada limitando trocas gasosas ou preenchendo com baixo oxigênio, alto dióxido de carbono, ozônio, umidade entre outros tem sido usado para preservação do fruto em maior período (BOSE *et al.*, 2021; PAUL; PANDEY, 2014). A temperatura do ambiente também influencia no processo, uma vez que a redução dela dificulta o metabolismo, este é inclusive o método mais usado para preservação de frutas estocadas (BOSE et al., 2021). O amadurecimento devagar causado pelo microambiente no café resulta em bebida de melhor qualidade, quando causa supressão da degradação da clorofila durante o início do amadurecimento do grão (CHENG *et al.*, 2020).

A luz regula parte do processo de amadurecimento, ela tem uma contribuição regulatória significante na síntese de carotenoides (LI, SHAN; CHEN; GRIERSON, 2021). Os mutantes de tomate high-pigment têm mutações em fatores de transcrição

sinalizados pela luz, isso contribuiu para o entendimento do papel da luz (LI, SHAN; CHEN; GRIERSON, 2021). Com a comparação do tipo selvagem com o mutante foram encontrados fatores de transcrição relacionados com a luz e pigmentação (ROHRMANN *et al.*, 2012). A luz regula o acúmulo de flavonoides e carotenoides, que além de serem pigmentos de cor, possuem significância nutricional (PAYASI; SANWAL, 2010). Além de carotenoides outros oxidantes são regulados pela luz (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). Uma vez que a luz está relacionada com acúmulo de carotenoides, manipular a sinalização celular da luz pode ser interessante para obter frutos mais nutritivos e de cores mais vivas (GIOVANNONI, 2004).

1.2.2.3 Fatores genéticos e moleculares

A genética clássica é usada para entender e alterar traços e interações genicas com efeito no processo de amadurecimento. Cruzamentos de mutantes do tomate já são usadas comercialmente para estender o tempo de prateleira, mas o uso de engenharia genética para manipulação da qualidade do fruto pode ser útil para obter variedades transgênicas visando outras características ainda mais interessantes, tendo em vista a saúde humana e não só o objetivo financeiro (WHITE, 2002). Há linhas transgênicas de maçã, melão, tomate, mamão e abacaxi com alteração em genes relacionados a biossíntese de etileno e consequentemente a cor, textura e aroma (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011). A caracterização de um loco dominante no mutante de tomate *never-ripe* (*NR*) permitiu identificar o primeiro receptor de etileno (GIOVANNONI, 2004). Interações epistáticas no mutante *ctrl* do tomate permitiram identificar a existência de uma família genica capaz de complementar a função de genes dessa família que perdem de função (GAPPER; MCQUINN; GIOVANNONI, 2013).

No tomate muitos loci foram caracterizados para vários traços como peso e tamanho, por exemplo FRUIT WEIGHT, fasciated e lucule number (PESARESI *et al.*, 2014). Cultivares melhoradas obtidas pelo cruzamento de variedades com essas características aprimoradas no processo de amadurecimento é de ganho comercial grande. Cruzamento entre variedades climatéricas e não-climatéricas de melão permitiram identificar dois loci responsáveis pelo fenótipo climatérico, indicando que o alelo não climatérico é recessivo (KARLOVA *et al.*, 2014). Atualmente os genes MADS-RIN, NAC-NOR e SPL-CNR são considerados principais reguladores do processo de

amadurecimento, outros genes e suas funções estão listados no apêndice 7.2 Tabela de genes envolvidos no processo de amadurecimento. De acordo com WANG et al. (2020) há diferentes modelagens propostas para a interação entre esses genes, entre elas: (i) redundância: onde vários genes têm mesma função (ii) aditividade: onde vários genes regulam características complementares (iii) dependência epistática: onde dois genes agem na mesma via e um depende da ação do outro (iv) dependência com compensação genética: um gene substituto compensa a perda de função de um dos reguladores.

O polimorfismo epigenético tem implicações no processo de amadurecimento, e é potencial para novas variações herdáveis dispensando alteração na sequência genética. Essas marcas epigenéticas são conservadas entre frutas secas e carnosas (FORLANI; MASIERO; MIZZOTTI, 2019). A utilização de modificações epigenéticas: metilação do DNA, modificação pós-traducional de histonas, remodelação da cromatina e rna não-codificadores, como epialelos naturais e induzidos, combinados, possibilita obter variedades melhoradas com frutos de qualidade superior (TANG; GALLUSCI; LANG, 2020). No tomate dezenas de milhares de sites em seu epigenoma estão sujeitos a modificação durante o desenvolvimento do fruto (PESARESI *et al.*, 2014). No tipo selvagem do mutante *Cnr*, por exemplo, o promotor do gene SPL-CNR tem poucas bases metiladas, enquanto o mutante *Cnr* tem muitas bases metiladas que reprime sua expressão (SEYMOUR *et al.*, 2008). Essa hipermetilação que altera 286pb em cerca de 2300 bp antes do domínio SPB denota que a metilação e demetilação do DNA é um componente crucial na regulação de amadurecimento (WANG, RUFANG *et al.*, 2020). Estudos do epigenoma do tomate sugerem que a metilação do DNA influencia na transição da biossíntese de etileno do sistema-1 para o sistema-2 (GAPPER; MCQUINN; GIOVANNONI, 2013). Em geral a metilação na extremidade 5’ dos genes diminuiu ao longo do desenvolvimento da fruta (LIU, MINGCHUN *et al.*, 2015), assim é possível induzir o amadurecimento precoce com demetilação artificial (KARLOVA *et al.*, 2014). A acetilação/desacetilação de histonas e metilação de mRNA podem também contribuir na regulação do processo de amadurecimento (CHEN; QIN; TIAN, 2020). A modificação pós-traducional de histonas pode regular a expressão dos genes podendo facilitar ou dificultar a expressão deles com remodelagem da cromatina, ambos têm estudos no tomate (LI, SHAN; CHEN; GRIERSON, 2021). Um estudo recente revelou o link entre o gene de demetilação de

histona SlJMJ6 atuando em 11 genes relacionados ao amadurecimento, entre eles RIN, ACO, ACS, PL, β-GAL (LI, ZHIWEI *et al.*, 2020). Isso ocorre porque a metilação de DNA e modificação de histonas tem impacto na rede de controle de amadurecimento (LI, SHAN; CHEN; GRIERSON, 2019). A metilação varia para fruta não climatérica no morango, por exemplo, a metilação diminui enquanto na laranja aumenta à medida que o fruto avança no amadurecimento (CHEN; QIN; TIAN, 2020). Apesar de já se saber sobre epigenética e amadurecimento, o impacto do epitranscriptoma na regulação do processo de amadurecimento é uma área ainda pouco explorada (WANG, RUFANG *et al.*, 2020).

Modificações pós-transcricionais e pós-traducionais também regulam o processo de amadurecimento, esses mecanismos regulatórios interagem em uma rede apresentada na Fig. 3. Como exemplo de modificação pós-traducional o óxido nítrico (NO) promove uma modificação pós-traducional (PTM) com nitração e S-nitrosação de proteínas durante o amadurecimento (PALMA *et al.*, 2019). Estudos para entender evolução de receptores de etileno durante o amadurecimento identificaram que a quantidade de transcritos não tem relação com a de proteínas, sugerindo que a sinalização de etileno é modulada pela fosforilação de resíduos modulando o sinal transdução etileno (LIU, MINGCHUN *et al.*, 2015).

Uma vez que muitos mRNA de fatores de transcrição são alvo de miRNA análises com degradoma de tomate identificaram a importância deles na regulação durante o desenvolvimento do fruto, em especial os pares miR156-CNR e miR172-AP2a podem ser reguladores no amadurecimento do tomate (KARLOVA *et al.*, 2014). Os miRNA atuam como mecanismo de regulação delicado e afetam a floração e alteração morfológica, sendo um importante controle a ser estudado (SEYMOUR *et al.*, 2008). Além desses RNA citados diversos outros tipos como pequenos RNA interferentes (siRNA), trans-atuadores siRNA (tasiRNA), RNA não codificadores longos (lncRNA), RNA circular (circRNA) têm sido reportados como relacionados a desenvolver papel na rede de regulação genica de amadurecimento (LI, SHAN; CHEN; GRIERSON, 2021). Os fatores de transcrição e a expressão diferencial estão entre os mais visados entre os fatores transcricionais, entretanto os pós transcricionais têm sido pouco estudados. Vários fatores de transcrição foram publicados (LI, SHAN; CHEN; GRIERSON, 2019; ROHRMANN *et al.*, 2012), pesquisas recentes têm

mostrado seu importante papel na regulação de vias no processo de amadurecimento (TEROL *et al.*, 2019; WANG, WEIHAO *et al.*, 2021).

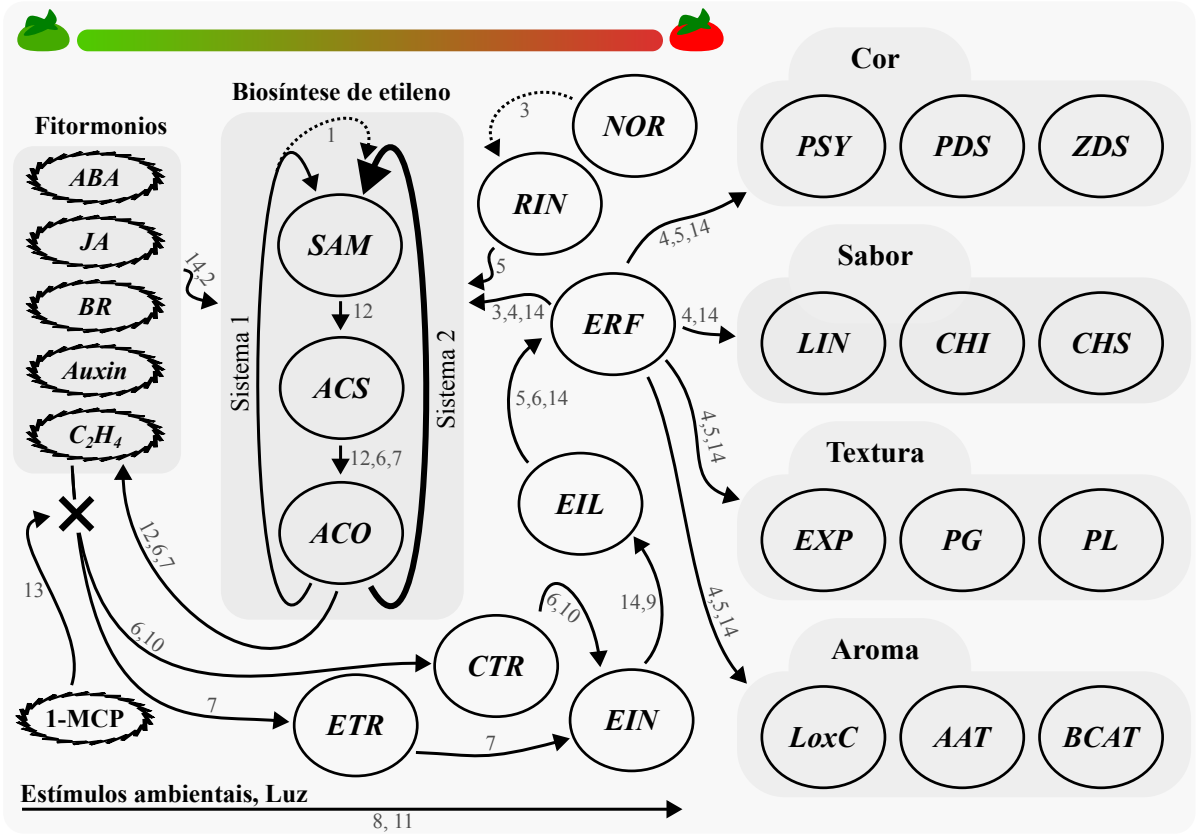


Fig. 3 Rede de regulação de genes envolvidos no processo de amadurecimento. Os números em cinza são as referências: [1] (LI, ZHIWEI *et al.*, 2020), [2] (LI, SHAN; CHEN; GRIERSON, 2021), [3] (LI, SHAN; CHEN; GRIERSON, 2019), [4] (GAO *et al.*, 2020), [5] (KARLOVA *et al.*, 2014), [6] (GAPPER; MCQUINN; GIOVANNONI, 2013), [7] (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011), [8] (GIOVANNONI, 2004) , [9] (TANG; GALLUSCI; LANG, 2020), [10] (BAPAT *et al.*, 2010), [11] (PALMA *et al.*, 2019) , [12] (PAUL; PANDEY, 2014) , [13] (CHEN; QIN; TIAN, 2020), [14] (LIU, MINGCHUN *et al.*, 2015) . Apenas 3 dos principais genes listados no apêndice 7.2 Tabela de genes envolvidos no processo de amadurecimento estão apresentados nos grupos Cor, Sabor, Textura e Aroma. As setas tracejadas representam efeitos epigenéticos.

Frequentemente a literatura tem visto as alterações moleculares que ocorrem durante o amadurecimento considerando a expressão diferencial dos genes (GAPPER; MCQUINN; GIOVANNONI, 2013; LI, SHAN; CHEN; GRIERSON, 2019; PESARESI *et al.*, 2014; ROHRMANN *et al.*, 2012). A análise global para explicar aspectos do amaduramento do tomate levou em conta a alteração na expressão dos genes para entender o amadurecimento (SHINOZAKI *et al.*, 2018). Análises recentes de RNA-seq durante o amadurecimento do café apresentaram vários genes com expressão diferencial (CHENG; FURTADO; HENRY, 2018).

Outras frutas também foram reportadas com análise de expressão diferencial para investigar o amadurecimento por experimento de transcriptômica, como laranja (FENG; WU; YI, 2019), melancia (ZHU *et al.*, 2017), morango (YI *et al.*, 2021), maçã (NAWAZ *et al.*, 2021) e uva (GUO, DA LONG *et al.*, 2020). Nessas análises de transcriptômica são reportados genes diferencialmente expressos, entretanto mesmo o gene sem expressão diferencial pode ocorrer de o gene usar de splicing alternativo para que, com outra isoforma o gene possa atuar regulando o processo de amadurecimento (SLUGINA *et al.*, 2021), ou até mesmo ocorrer ambos expressão diferencial e splicing alternativo (GUO, WENBIN *et al.*, 2021). Uma publicação recente mostra o splicing alternativo diferencial regulando a degradação de amido durante amadurecimento da banana (JIANG *et al.*, 2021) e análises in silico realizadas por YAN et al. (2021) e outras apresentadas nos capítulos 2. RNASeq-based alternative splicing prediction reveals isoform alteration of potassium channel and apyrase associated with coffee fruit ripening e 3. Screening of Differential alternative splicing events in tomato fruit ripening desse trabalho evidenciam a influência do splicing alternativo na regulação do amadurecimento de frutos.

Tendo em vista que o processo biológico pode fazer uso de splicing alternativo (AS) para realizar o amadurecimento (WANG, WEIHAO *et al.*, 2023), no próximo tópico desse capítulo será exposto como é esse mecanismo de controle co-transcricional e como ele é usado. Nesta tese os dados de RNA-seq de café e tomate citados anteriormente que foram usados para explicar o amadurecimento com expressão diferencial gênica (DEG) serão usados para identificar splicing alternativo diferencial (DAS) durante o processo de amadurecimento.

1.3 O SPLICING ALTERNATIVO (AS)

O processamento alternativo do RNA mensageiro transcrito do gene nuclear, denominado *alternative splicing* (AS), é a transformação do transcrito primário em diferentes isoformas pelo spliceossomo ao alternar sítios do pre-mRNA entre essas isoformas (SIBLEY; BLAZQUEZ; ULE, 2016; WILKINSON; CHARENTON; NAGAI, 2020). Esse mecanismo habilita os eucariotos, em seus genes multi-exons, estender a diversidade de seu transcriptoma e proteoma por meio de diferentes isoformas mesmo com um número limitado de genes (MARASCO; KORNBLIHTT, 2022; NILSEN; GRAVELEY, 2010). Não é exigido que todo gene multi-exon esteja sujeito a AS (TRESS; ABASCAL; VALENCIA, 2017), contudo ter essa possibilidade é importante para viabilizar o transcriptoma plástico.

Além de produzir isoformas diferentes, o AS pode ser um mecanismo para regular a expressão genica criando um dinamismo proteômico qualitativo e quantitativo também (LING *et al.*, 2019). Esse dinamismo permite o organismo se adaptar em diferentes ambiente e condições, o que faz do AS uma possível explicação para a evolução das plantas para habitar ambientes terrestres (MELO; KALYNA; DUQUE, 2020). Uma vez que o AS tem papel importante na plasticidade das plantas, cerca de 60% de seus genes que contêm íntrons estão sujeitos a AS (FILICHKIN *et al.*, 2015; SYED *et al.*, 2012). Em humanos AS anormais podem causar doenças (PARK *et al.*, 2018; VUONG; BLACK; ZHENG, 2016), logo, compreender o AS em sua complexidade é fundamental para entender organismo com um todo. Apesar de não ser possível determinar éxons alternativos apenas pelo pré-mRNA, vários trabalhos foram publicados reportando tentando determinar o responsável pelo processamento não canônico do pré-mRNA (BARASH *et al.*, 2010).

No processamento canônico o pré-mRNA é intervalado por íntrons que são removidos dando origem a uma sequência só de éxons que é o mRNA maduro. A maioria dos íntrons e exons são constitutivos, ou seja, seu tamanho e processamento são invariáveis para (i) aumentar a expressão, (ii) alocar sequências de regulação, (iii) persistir o AS tornando-o constitutivo como evolução, e (iv) modular a expressão dos genes (DING; ELOWITZ, 2019). A remoção dos íntrons do pré-mRNA é feita pelo complexo de ribonucleoproteínas chamado de spliceosomo em duas etapas: primeiro a sequência a ser removia é transformada em laço pelo reconhecimento e aproximação de sítios de processamento, na segunda etapa esse íntron laçado é

desconectado do pré-mRNA quando as extremidades 5' e 3' dos exons próximos se conectam (HERZEL *et al.*, 2017; SHARP; ROBERTS; SHI, 2017). O grau de afinidade do spliceosomo com o sítio de processamento torna essa região mais forte, a disposição de sítios fortes e fracos dá origem a regiões alternativas que causa os tipos básicos de AS: exon removido (ES), íntron retido (RI), sítio de processamento 3' (A3SS) e 5' (A5SS) alternativo (KORNBLIHTT *et al.*, 2013).

O que torna um sítio mais forte ou fraco são os cis atuadores, que são regiões do pré-mRNA como elementos de regulação de processamento (SRE), e os fatores trans atuadores que são proteínas (RBPs) que interagem com os SREs de modo e intensidade a depender da condição, resultando em AS (FU, XIANG DONG; ARES, 2014). Apesar de nem todas as proteínas (RBPs) serem conhecidas o avanço nas tecnologias de ômicas tem viabilizado a caracterização de muitas delas e contextualizado sua interação na rede de regulação de AS (ULE; BLENCOWE, 2019). Em especial os íntrons - peculiares de eucariotos e longos em animais - são regiões que podem possuir múltiplas funções acumuladas, entre elas: (i) registrar de elementos de regulação cotranscricional (ii) causar mais variação quando extensos sob recursive splicing (iii) codificar partes inesperadas de proteína gerando quimeras (iv) incluir stop códon prematuro fazendo regulação pós-transcricional ao truncar proteínas ou submeter o mRNA maduro à via NMD (GEORGOMANOLIS; SOFIADIS; PAPANTONIS, 2016; SIBLEY; BLAZQUEZ; ULE, 2016). Nesse último item, a via Nonsense-mediated mRNA decay (NMD) degrada o mRNA maduro que contêm stop códon prematuro, assim o controle de qualidade do RNA descarta isoformas com anomalias que faz de um AS do tipo RI uma forma de regulação pós-transcricional (LYKKE-ANDERSEN; JENSEN, 2015). Enquanto o processamento de íntrons frequentemente é cotranscricional, o ES acontece após a transcrição (BRAUNSCHWEIG *et al.*, 2013). Nesse sentindo o AS é regulado, mas também é um mecanismo de regulação qualitativo (NOBLE *et al.*, 2020; SCHAEFKE *et al.*, 2018) e quantitativo, conforme é apresentado na Fig. 4. As alterações causadas pelo AS podem ter impacto em nível de mRNA causando a instabilidade dele por exemplo, em nível de proteína alterando sua propriedade enzimática ou localização intracelular por exemplo, em nível de organismo causando diferenças entre indivíduos de mesma espécie ou doença em algum (KELEMEN *et al.*, 2013). O AS é crucial para vários

processos biológicos de humanos e plantas (BARALLE; GIUDICE, 2017; DANTAS *et al.*, 2019).

O processamento do mRNA envolve uma série de componentes, incluindo o spliceossomo, o sítio doador, o sítio aceptor e regiões ricas em pirimidinas. O sítio doador é uma sequência de DNA que é reconhecida pelo spliceossomo e que é cortada para iniciar o processo de splicing alternativo (AS). O sítio aceptor é uma sequência de DNA que é reconhecida pelo spliceossomo e que é usada para unir as extremidades dos exons. As regiões ricas em pirimidinas são sequências de DNA que são frequentemente encontradas em introns. Essas regiões podem ajudar a regular o AS, pois podem influenciar a ligação do spliceossomo. O spliceossomo é um complexo proteico que realiza o AS. Ele é composto por cinco pequenas subunidades (U1, U2, U4, U5 e U6) e uma grande subunidade (U12) conforme pode-se visualizar na figura https://www.genome.jp/kegg-bin/show_pathway?ko03040. O complexo U1 é o complexo inicial que se liga ao sítio doador e inicia o processo de AS. O complexo U2 é o complexo que reconhece o sítio aceptor e ajuda a unir as extremidades dos exons. O complexo U4/U5/U6 é o complexo que ajuda a estabilizar o spliceossomo e a auxiliar na remoção dos introns.

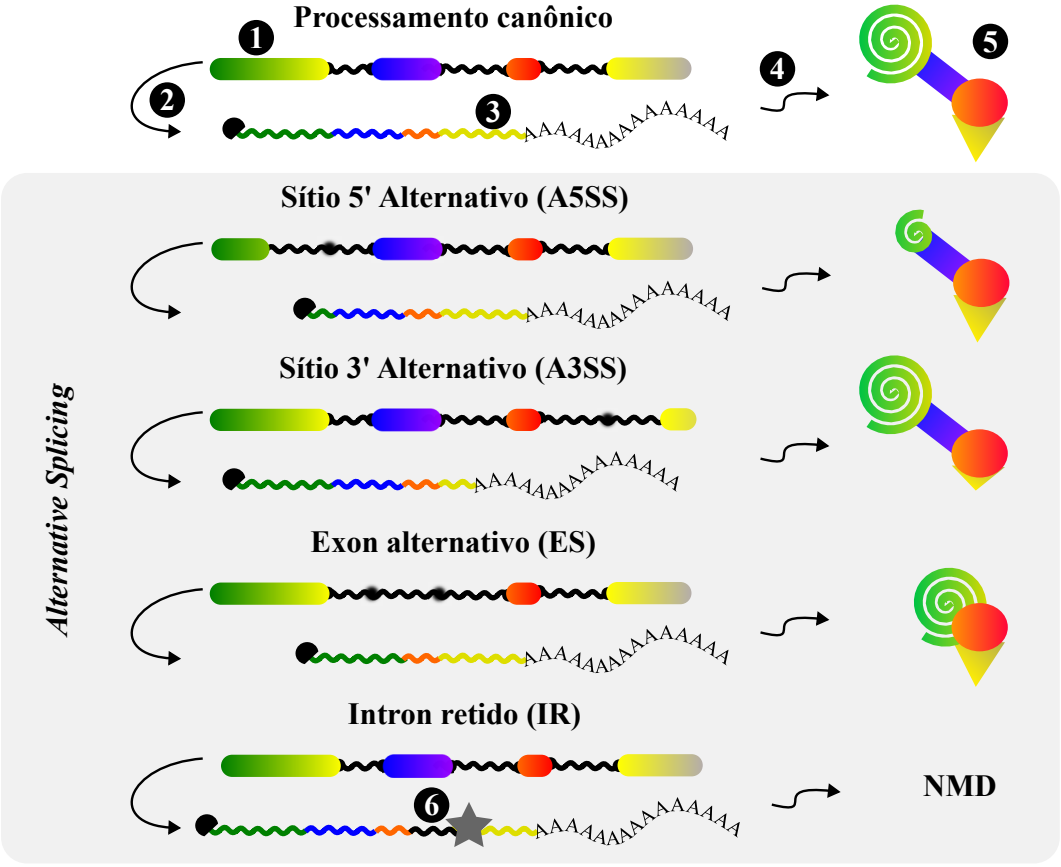


Fig. 4 Tipos básicos de eventos de AS. O transcrito precursor do mRNA maduro, indicado em 1 tem processamento canônico quando seus íntrons são removidos e éxons mantidos em suas posições normais. Após o splicing indicado em 2, o transcrito maduro recebe um "capacete" e uma cauda poli-A conforme indicado em 3. Com a tradução indicada em 4 a informação contida na sequência do mRNA maduro é transformada em proteína indicada em 5. A depender de regulação de fatores intrínsecos e extrínsecos ao mRNA um éxon pode ser removido ou um íntron pode ser mantido, se um éxon ou íntron não tem seu processamento alterado em nenhuma isoforma ele é dito como sendo constitutivo, aqui representado pelo éxon laranja. Se o sítio de splicing é fraco pode acontecer de o éxon ser removido (evento denominado ES) ou seu início (A3SS) ou fim (A5SS) ser alterado durante o splicing. Pode ainda acontecer de um íntron deixar de ser removido, e permanecer no transcrito maduro (IR). Como os íntrons frequentemente são longos eles podem ser processados em partes (recursive splicing) isso pode levar ao aparecimento de uma parte que não foi removida (exintron), ou ainda pode inserir um códon de parada prematuro (PTC) no transcrito maduro, representado pela estrela indicada em 6, esse PTC é percebido na via NMD e então é degradado antes da tradução. Qualitativamente o produto do gene pode ser diferente conforme as proteínas. Quantitativamente o IR, por exemplo, pode levar a quantidade de produto do gene a valores muitos baixos pela degradação de todas isoformas do gene por eventos de IR, ou diminuir o tamanho da proteína do gene por truncar a proteína traduzida pelo códon de parada prematuro.

De acordo com experimentos registrados na literatura, as plantas podem realizar AS em vários processos biológicos. A exposição da planta a diferentes condições térmicas induz AS que aprimora a performance dela, resultando em adaptação (JOHN; OLAS; MUELLER-ROEBER, 2021). O AS é usado em resposta a estresse abiótico (LALOUM; MARTÍN; DUQUE, 2018). Em coqueiros, a variedade anã leva 3-5 anos para florescer enquanto a alto leva 8-10 anos para floração, essa diferença entre variedades da mesma espécie está sendo estudada considerando AS, que por A3SS produz isoformas diferentes para o gene FT (XIA *et al.*, 2020). Também há estudos de AS em resposta a patógenos em estresse biótico (RIGO *et al.*, 2019). Mesmo que os AS não tenha relação com o processo biológico é importante considerá-lo na análise do transcriptoma, uma vez que se tenha as sequencias dos transcritos. Nesse sentido há uma área crescente de estudo em AS para tentar explicar processos biológicos, resistência, adaptação, entre outros de modo que acompanham inclusive os avanços nas tecnologias de sequenciamento (WANG, MAOJUN *et al.*, 2018) desenvolvendo novos pipelines de análises de bioinformática.

1.3.1 TIPOS DE EVENTOS DE AS

Para encontrar eventos de AS é necessário quantificar as isoformas presentes na amostra para determinar o índice *I* de processamento de cada isoforma. Considerando um gene sujeito a AS, e a princípio um evento SE, esse *I* pode ser determinado pelo índice de inclusão do éxon (*percent spliced in* ou *PSI* ou Ψ) ou de abundância da isoforma (percent spliced ou PS). No caso da determinação do Ψ, o *I* pode ser calculado partir do mapeamento de reads com software STAR (DOBIN *et al.*, 2013), por exemplo, que é possível obter pela contagem reads oriundas do transcrito que inclui em relação as do transcrito que exclui o éxon alternativo no evento de SE. Para eventos diferentes de SE basta modelar o sítio alternativo como éxon e proceder a mesma lógica. No caso de PS, a partir da quantificação das isoformas que pode ser realizada com software Salmon (PATRO *et al.*, 2017), por exemplo, o *I* está em função da relação entre as abundancias de transcritos da isoforma e do gene. Tomando Ψ das isoformas do mesmo gene é possível calcular o AS diferencial (ΔPSI ou ΔPS) pela diferença entre eles, uma vez que Ψ está no intervalo [0, 1] $\Delta\Psi$ está entre [-1, 1]. Vale observar que a qualidade do genoma e sua anotação, o controle de qualidade do transcriptoma, a remoção de artefatos da técnica, a aplicação e ajuste de testes

estatísticos, entre outros são determinantes para a acurácia do $\Delta\Psi$ e sua interpretação.

Com eventos de AS a que um gene está sujeito identificados pelo $\Delta\Psi$, eles podem ser classificados como básicos ou complexos, os quatro tipos básicos são apresentados na Fig. 4. Quando um éxon qualquer do gene está ausente em algum transcrito maduro é um evento de éxon removido (SE ou skiped exon ou casset exon). Quando um íntron qualquer está presente em algum transcrito maduro é um evento de íntron retido (RI ou retained íntron). Quando alguma das isoformas tem o início de algum éxon alterado é um evento de sítio 3' alternativo (A3SS - alternative aceptor). Quando alguma das isoformas tem o fim de algum éxon alterado é um evento de sítio 5' alternativo (A5SS - alternative donor).

O aumento do sequenciamento de transcriptomas tornou possível estimar que 70% dos genes multi-exon estão sujeitos a algum evento de AS, sendo RI mais comum em plantas e SE o mais comum em animais (CHAUDHARY; KHOKHAR; *et al.*, 2019). Esses tipos básicos de eventos podem ser combinados ou relacionados formando eventos mais complexos que podem ser modelados em grafo, matriz, entre outros (SAMMETH; FOISSAC; GUIGÓ, 2008; VAQUERO-GARCIA *et al.*, 2016). Em plantas já foi reportado múltiplos eventos do mesmo tipo e de tipos diferentes combinados e relacionado (LI, YEYUN *et al.*, 2020). O éxon mutualmente exclusivo (MXE) ocorre quando dois éxons do gene só aparecem em isoformas diferentes, mas nunca na mesma isoforma (POHL *et al.*, 2013).

1.3.2 SOFTWARES DESENVOLVIDOS PARA ANÁLISES DE AS

Análises computacionais tem um papel crucial na caracterização de splicing alternativo, por isso vários softwares foram e vem sendo desenvolvidos para dar suporte a análise de AS (HAAS, 2008; HOOPER, 2014; KATZ *et al.*, 2015; REDDY *et al.*, 2012; SONG *et al.*, 2019; WEN; MEAD; THONGJUEA, 2020a). Frequentemente os softwares são desenvolvidos em pacotes e disponibilizados em CLI, mas há uma recente preocupação em disponibilizar interface amigável para permitir usuários que não são de a área de computação realizar essas análises, como por exemplo o SpliceDetector desenvolvido com o framework .NET (BAHARLOU HOUREH *et al.*, 2018). Os softwares de análises de AS podem ser categorizados de acordo com MEHMOOD et al. (2020) em orientado a éxon: que calcula o ΔI pela contagem de

reads em sítios do gene (como DEXSeq, edgeR, JunctionSeq, limma), orientado a isoforma: que reconstrói a isoforma e determina sua abundância nas amostras para calcular o ΔI (como cuffdiff2, DiffSplice), orientado a evento: que identifica o ΔI nas regiões onde ocorrem eventos de AS no gene (como dSpliceType, MAJIQ, rMATS, SUPPA).

O rMATS é um software desenvolvido em Python utilizado em pesquisas de AS usando dados de RNA-seq (SHEN *et al.*, 2014). Os pontos positivos rMATS são: ser classificado como estado da arte (DENTI *et al.*, 2018), encontrar eventos novos, permitir plotar o evento em nível de éxon com o pacote rmats2sashimiplot da suíte. Os pontos negativos rMATS são: necessidade de habilidade em CLI uma vez que ele não disponibiliza GUI, ausência de filtros para remover falsos positivos exigindo análises com pacotes adicionais como o MASER/Bioconductor (HUBER *et al.*, 2015), depender de mapeamento que é uma análise de alto custo computacional, e, necessidade de determinar a isoforma do evento AS uma vez que ele aponta a região do evento no gene.

O 3D RNASeq é uma ferramenta nova com interface gráfica, desenvolvida em R fazendo uso do pacote Limma (RITCHIE *et al.*, 2015), disponibilizado como webapp para análises que inclui, mas não se limita a: expressão diferencial, AS, troca de isoforma nas condições, visualização de dados de DEG e AS (GUO, WENBIN *et al.*, 2021). Os pontos positivos do 3D RNASeq são: a facilidade de uso inclusive pela GUI e a documentação, aplicação de métodos do estado da arte, velocidade na análise dos dados, pré-processamento robusto dos dados, ter possibilidade de executar as análises pelo webapp que dispensa instalação de software, geração de um relatório detalhado em formato de artigo científico, e ainda, possuir resultados superiores comparado aos demais (MEHMOOD *et al.*, 2020). Os pontos negativos do 3D RNASeq são: realizar as análises em nível de isoforma o que causa necessidade de classificação do tipo de evento de AS pelo usuário, necessidade de mapeamento para visualizar o evento em nível de éxon, necessidade de análise posterior para determinar e exportar a isoforma e sua sequência de um determinado AS.

Alguns métodos mais novos têm feito uso de IA e ML para análise de splicing alternativo, isso tem acontecido porque análises usando tecnologias HTS dependem da cobertura do sequenciamento para encontrar o AS (ZHANG, ZIJUN *et al.*, 2019). Uma novidade é o uso de IsoSeq (Single-Molecular Long-Read Isoform Sequencing)

nas análises de AS (HU, HONGYIN *et al.*, 2020), o que tem ajudado na análise uma vez que a isoforma lida completa reduz a possibilidade de sua leitura pelo equipamento ser contabilizada em outra isoforma ou gene, uma vez que muitos genomas de plantas são de ploidia duplicada contido de muitos elementos repetitivos isso tem muita importância. O aprimoramento do resultado considerando o nível proteômica é importante nessas análises de AS (LAU *et al.*, 2019). Análises posteriores a identificação do evento tornam possível caracterizá-lo melhor considerando: o impacto quantitativo e qualitativo na proteína, o impacto em regiões regulatórias, o contexto genômico e transcriptômicos, entre outros. Nesse sentido softwares de visualização de AS como IGV (THORVALDSDÓTTIR; ROBINSON; MESIROV, 2013), rmats2sashimiplot (https://github.com/Xinglab/rmats2sashimiplot), AS-Quant (FAHMI *et al.*, 2021), VALERIE (WEN; MEAD; THONGJUEA, 2020b), Vials (STROBELT *et al.*, 2016), Manananggal (BARANN; ZIMMER; BIRZELE, 2017) também são importantes nas análises de AS.

1.3.3 PESQUISAS DE AS REALIZADAS EM FRUTOS

Os avanços nas tecnologias de sequenciamento, em especial as de baixo custo e alto rendimento, permitiram extrapolar análises transcriptômicas de organismo modelo para os demais, o que permitiu incrementar o conhecimento de AS em plantas nas últimas décadas (BEDRE *et al.*, 2019). Em arroz e cevada por exemplo o AS respondeu perguntas sobre a diferença de tolerância a estresse abiótico (sal) entre essas espécies da mesma família (FU, LIANGBO *et al.*, 2019). No sentido de entender o comportamento de AS em *Brassica napus* em estresse abiótico pesquisadores encontraram AS em resposta a estresse térmico (LEE; ADAMS, 2020). Na análise de *Camellia sinensis* o AS também apresentou um papel importante na regulação de genes em resposta ao frio (LI, YEYUN *et al.*, 2020). Na análise em *Artemisia annua* foram encontrados genes envolvidos em vias de biossíntese de metabolitos secundários com isoformas alteradas com RI regulado por luz no espectro azul e vermelho (MA *et al.*, 2021). No *Populus alba* análises de AS foram realizadas com tecnologias de sequenciamento de terceira geração e os resultados reportados estão de acordo com estudos anteriores feitos no sorgo, bambu e algodão (HU, HONGYIN *et al.*, 2020). Para esse último, *Populus alba*, autores também utilizaram tecnologias de sequenciamento de terceira geração, no caso PacBio Iso-Seq, e desenvolveram um pipeline para análise de dados de transcriptômica Iso-Seq para caracterizar AS

em espécie poliploide (WANG, MAOJUN *et al.*, 2018). Nas análises de AS em *Gnetum luofuense* durante desenvolvimento foram reportados 9061 eventos de AS com validação de 2 genes sob RI e SE com qRT-PCR (DENG *et al.*, 2020). Nas análises de transcriptômica de banana realizadas por JIANG e colaboradores (2021), os autores identificaram um fator de transcrição MYB regulado negativamente, e analisando-o descobriram que o AS regula a degradação de amido durante o amadurecimento.

Análises de AS no tomate revelaram que ao menos 59,3% dos genes multi-exon estão sujeitos a AS, o tipo RI é predominante e o SE o menos frequente, e frutos tem mais isoformas alternativas por gene comparado as outras partes da planta (SUN; XIAO, 2015). Na comparação do transcriptoma de variedades de tomate selvagem e cultivado pesquisadores encontraram mais de mil genes com AS entre as variedades, adicionalmente identificaram eventos de AS durante o início do desenvolvimento da fruta (WANG, KETAO *et al.*, 2016). Análises posteriores integrando dados de 27 projetos de RNASeq estimaram que 65% dos genes sujeitos a AS (CLARK, SARAH *et al.*, 2019). Nas análises das isoformas do gene RIN no tomate selvagem e cultivado os autores sugeriram que a isoforma RIN1 tem função de ativar genes relacionados ao amadurecimento, enquanto a isoforma RIN2 com A5SS no último éxon tem o papel de regular esses genes (SLUGINA *et al.*, 2021). Com análises de transcriptômica do tomate, COLANERO e colaboradores (2020) verificaram que o AS no fator de transcrição MYB que produz cor roxa do tomate selvagem leva a uma isoforma alternativa com perda de função, o que pode explicar o tomate vermelho conhecido nas cultivares comercias.

1.4 JUSTIFICATIVA E OBJETIVOS DESTE TRABALHO

As frutas são importantes para as plantas, para animais, para a alimentação humana e para a economia mundial. Para consumo *in natura*, em geral, as frutas possuem mais valor enquanto maduras. Nesse estágio elas são saborosas e nutritivas, entretanto é difícil sincronizar o processo biológico de amadurecimento com o momento do consumo delas, o que leva ao desperdício e prejuízo. O processo de amadurecimento é complexo, constantemente revisado e ainda não totalmente entendido. Uma vez que o AS é um mecanismo regulatório em plantas, uma análise global afim de entender como e quais genes estão sujeitos a AS durante o processo de amadurecimento foi realizada, colaborando assim com o entendimento do processo,

1.4.1 OBJETIVOS GERAIS

Realizar a predição de AS associados ao processo biológico de amadurecimento de frutos de café e tomate e desenvolver uma ferramenta computacional para auxiliar nas análises de AS de modo a atender as necessidades de programação para analise de dados dos resultados encontrados.

1.4.2 OBJETIVOS ESPECÍFICOS

- Predizer a ocorrência de eventos de AS no *Coffea arabica* com o rMATS, utilizando dados de RNASeq disponibilizados no banco de dados SRA.
- Analisar eventos de AS de *Solanum lycopersicum* com o 3D-RNAseq, utilizando e dados de RNASeq disponibilizado no SRA.
- Implementar um webapp para visualizar dados de AS e auxiliar nas pesquisas de DAS, sanando necessidades encontradas nas análises realizadas com café e tomate.

Cada objetivo específico foi explorado em capítulo específico a seguir.

Capítulo 2.

Análise de splicing alternativo no amadurecimento de frutos de café utilizando genoma e dados de RNASeq de banco de dados públicos, através da ferramenta rMATS com eventos de DAS validados por PCR e qPCR.

2. RNASEQ-BASED ALTERNATIVE SPLICING PREDICTION REVEALS ISOFORM ALTERATION OF POTASSIUM CHANNEL AND APYRASE ASSOCIATED WITH COFFEE FRUIT RIPENING

The Alternative splicing (AS) enables eukaryotes to extend the transcriptome and proteome diversity even with a limited gene number. Ripening is a complex biological process accompanied by alteration of gene expression by different mechanisms, including AS-based regulation. However, the impact of AS in coffee fruit ripening and cup quality has not yet been explored. Identification of AS mechanisms specifically associated to coffee fruit will bring a better understanding of the ripening process, including new technologies for process control and optimization. In this work, Coffea arabica RNA-Seq data collected from unripe, half-ripe, and ripe fruit stages were used to predict genes that undergo differential AS. A total of 202 genes under 241 AS events with significant difference during the ripening were found. Among these events, a Potassium channel, and an Apyrase 7 coding genes with variable isoforms between unripe and ripe fruits were validated. In the potassium channel gene, AS reduces the protein productive mRNA level due to increase of isoform level with insertion of premature terminator codon by intron retention. On the other hand, Apyrase gene that often is associated to regulation of the energy metabolism in fruit undergoes AS by exon skipping that change known functional domains of the protein. These results suggest that transcriptional control based on AS is an important mechanism during coffee fruit ripening.

Keywords: Alternative splicing; Apyrase 7; Coffea arabica; fruit ripening; Potassium channel; transcriptome

2.1 INTRODUCTION

The coffee world market quadrupled during the last 30 years (ICO, 2021). Coffee is an important commodity for the world economy with a market estimated on 70 bi US dollars (TALHINHAS *et al.*, 2017). Most of the world's coffee production is based on the species *Coffea arabica* (arabica), with a higher beverage sensorial quality (MARIE *et al.*, 2020) and commercial value. In this context, the arabica genomics and transcriptomics have been studied to increase productivity by breeding (SOUSA *et al.*, 2019) and improve the quality of the aroma and flavor (CHENG; FURTADO; HENRY, 2018). The *Coffea canephora* is the second most valuable species of Coffea genera due to its considerable pathogen resistance (MCCOOK; VANDERMEER, 2015) and low production cost but have a smaller planting area due to cup quality considered lower than arabica (MEHRABI; LASHERMES, 2017). The specie C. arabica is an allopolyploid resulting from the natural hybridization between *Coffea eugenioides* and *Coffea canephora* (CENCI; COMBES; LASHERMES, 2012). The plant is native from Ethiopia, Kenya, and Sudan (ANTHONY *et al.*, 2002).

About three years after the planting, coffee flowers generally appear resulting in fruits that should be collected after ripening, considering that unripe fruit reduce the quality of the drink. As a climacteric fruit, ethylene plays an important role in coffee fruit ripening (ALEXANDER; GRIERSON, 2002). The fruit ripening time starts with a burst of this plant hormone biosynthesis that triggers chemical interactions taking to physiological transformations. Then ethylene triggers the biochemical production of compounds such as caffeine, trigonelline, chlorogenic acids, and sucrose lipids (CHENG *et al.*, 2016). These compounds are accumulated in coffee fruits during development and processed until the last ripening stage. The chemical composition of coffee fruit, plant genotype, environment, as well as microenvironment play a decisive role in coffee quality (CHENG *et al.*, 2016, 2020; DE SOUZA ROLIM *et al.*, 2020). During ripening there are transcriptional and epigenetic regulation support the expression of ripening-related genes in the complex regulatory network (LI, SHAN *et al.*, 2020). The main up and down-regulated gene lists associated with ripening are usually enriched with ethylene related genes (ALEXANDER; GRIERSON, 2002; CHENG *et al.*, 2016).

Analysis of different plant ripening models has demonstrated that alternative splicing (AS) is an important mechanism to control gene expression and mRNA

structure, resulting in effects in protein cellular location, structure, and activities (BARALLE; GIUDICE, 2017; ULE; BLENCOWE, 2019). In fact, more than 60% of intron-containing genes undergo AS in plants (SYED *et al.*, 2012). AS is a co-transcriptional mechanism that regulates gene expression, which expands proteome diversity by generates different alternative isoforms from a same gene with non-canonical splicing (BEDRE *et al.*, 2019; REDDY *et al.*, 2012; SIBLEY; BLAZQUEZ; ULE, 2016). AS events can be classified into four basic types: i) exon skipping (ES), ii) intron retention (IR), iii) alternative donor (A5SS) and iv) alternative acceptor (A3SS). Combinations or derivations of these principal AS types have been also reported (LI, YEYUN *et al.*, 2020; NILSEN; GRAVELEY, 2010) as mutually exclusive ES (MXE) for example. A vast majority of splice events are IR in plants, but ES is the most common AS mechanism found in animals (CHAUDHARY; KHOKHAR; *et al.*, 2019). Although IR events can be associated to increase of protein function, this mechanism can also be a negative transcriptional regulator by generating protein unproductive mRNAs with premature termination codons (PTC) that are recognized by the nonsense-mediated decay (NMD) machinery and degraded (KORNBLIHTT *et al.*, 2013). Recent works has been conducted to understand the relationship of AS with biological process like stress response (LALOUM; MARTÍN; DUQUE, 2018; LEE; ADAMS, 2020), flowing (XIA *et al.*, 2020), ripening (JIANG *et al.*, 2021), single cell differences in cell populations (WEN; MEAD; THONGJUEA, 2020a), metabolic global overview (CLARK, SARAH *et al.*, 2019; WANG, MAOJUN *et al.*, 2018) and others plant characteristics (KELEMEN *et al.*, 2013).

The prediction of AS may be performed using RNASeq data. Among the most promising software to discover AS events is rMATS in turbo version (SHEN *et al.*, 2014). This software is implemented in Python and allows biologists discover novel events, 3D RNA-seq (GUO, WENBIN *et al.*, 2021) an R implementation with graphical user interface, and DARTS (ZHANG, ZIJUN *et al.*, 2019) that apply deep-learning techniques to infer differential AS in biological samples. This toolset uses high throughput omics data like genomic or transcriptomic sequencing to identify alternative isoforms originated by AS events, enabling understand this mechanism in biologic process as ripening.

In arabica coffee the first fully open access assembled genome was made available by World Coffee Research (SCALABRIN *et al.*, 2020). This work brought the

public arabica genome relatively fragmented with 92,4% of BUSCO completeness and displayed a high level of diversity between the two subgenomes. In addition to this assembly, there are two others genome available so far (NCBI GCA_003713225.1 and Phytozome genome ID 453). About transcriptomic analysis, there are 33 projects deposited on the NCBI database (BioProject search details: "coffea arabica"[orgn] and "transcriptome gene expression"[Filter]). These numbers indicate that the coffee species has little publicly deposited data. In this work we used the published genome by Scalabrin (2020) with its annotation, and transcriptome of fruit pulp (endocarp, perisperm, endosperm and embryo tissues) in triplicate of unripe, half-ripe and ripe development stages deposited on SRA by CHENG et. al (2018) to better understand the role of AS in the coffee fruit ripening process.

Recent studies have shown AS in some fruit developmental processes (WANG, WEIHAO *et al.*, 2021; YAN *et al.*, 2021) but here, we used RNASeq data of coffee fruit to better understand the influence of AS in ripening. We analyzed ripening RNA-seq data SRA PRJEB24137 with rMATS and wrote scripts in python to curate an event set of rMATS, later we annotate them to choose IR and SE events for experimental validation by PCR (polymerase chain reaction) and quantitative real time PCR (qPCR). To our knowledge, this is the first report of AS influence in arabica fruit ripening. The complete understanding of molecular mechanisms can help breeding programs to improve cup quality of drink from arabica coffee.

2.2 MATERIALS AND METHODS

2.2.1 RNA SEQ DATA, READS QUALITY CONTROL AND MAPPING

The FASTA and GFF3 files of the arabica cv. Bourbon Vermelho were downloaded from https://worldcoffeeresearch.org/work/coffea-arabica-genome/ (SCALABRIN *et al.*, 2020). The RNASeq FASTQ files from arabica cv. K7 triplicates of different ripening stages (green: G, yellow: Y and red: R) deposited by project PRJEB24137 (CHENG; FURTADO; HENRY, 2018) in the SRA NCBI database were retrieved. The single-read RNASeq libraries were produced in 2015 on Australia and processed in Illumina HiSeq 4000 instrument (7.3 Supplementary Table S1). The FASTQ files were submitted to GALAXY/EU in https://usegalaxy.eu (COMMUNITY *et al.*, 2022) for quality control. TRIMMOMATIC (BOLGER, ANTHONY M.; LOHSE; USADEL, 2014) v0.36.4 was used with default parameters (AVGQUAL 20) on GALAXY/EU to obtain quality reads, then summary results were inputted in FASTQC (BROWN; PIRRUNG; MCCUE, 2017) after in MULTIQC (EWELS *et al.*, 2016) to visualize Phred score results (7.4 Supplementary Fig. S1). Then STAR v2.7 (DOBIN *et al.*, 2013) was used for mapping the reads with parameters "--outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --outFilterMatchNmin 0", the metrics of mapping is shown in 7.5 Supplementary Fig. S2.

2.2.2 ALTERNATIVE SPLICING DISCOVERY AND ANALYSIS

The GTF file was obtained from the GFF3 annotation file (SCALABRIN *et al.*, 2020) using Agat version v0.8.0 from the Bioconda suite (DALE *et al.*, 2018) and the mapped BAM files were submitted on rMATS v4.1 with parameters “readLength 150 --nthread 4 --tstat 4 -t paired --libType fr-unstranded --cstat 0.0001 --novelSS” in localhost (4 cores x 2,4GHZ and 6GB RAM) for contrasts Green-Yellow (GY) and Red-Yellow (RY). Then rMATS outputs were submitted to the MASER package of Bioconductor (HUBER *et al.*, 2015) to filter significant events. To join and quantify significant events in the ripening stage (GY and RY) and AS types (IR, SE, A3SS, A5SS and MXE) comparison, the gene count versus events count data were inputted in custom python scripts filter_events.py and plot_events.py (7.6 Supplementary Fig. S3). The sashimi plot of significant events was generated with script rmats2sashimiplot.py available on rMATS GitHub repository https://github.com/Xinglab/rmats2sashimiplot.

2.2.3 FUNCTIONAL ANNOTATION AND GO ENRICHMENT ANALYSIS

To obtain protein sequences from coffee genome, the Biopython v1.79 (COCK *et al.*, 2009) was utilized in a script to translate coding regions. Then proteins were annotated with Blastp v2.10.1 (CAMACHO *et al.*, 2009) using non redundant RefSeq (O'LEARY *et al.*, 2016) and SwissProt (BATEMAN, 2019) as a database with param "evalue 1e-5 -outfmt 'std all_sbj_names'" on Galaxy/EU. Next to define gene and protein name, with InterproScan5 (Jones et al. 2014) on Galaxy/EU to find Pfam (EL-GEBALI *et al.*, 2019) motifs and Interpro proteins family and domains, gene ontology (GO) (CONSORTIUM *et al.*, 2023) and pathways associated, and on EggNOG 5.0 (HUERTA-CEPAS *et al.*, 2017) webserver to annotate GO and Reactome (JASSAL *et al.*, 2020) pathways. The protein set was also submitted in Orthoveen2 (XU *et al.*, 2019) webserver to annotate GOs and determine GO enrichment and inparalogs using all genes of the genome as the reference set with evalue param "1e-5". The outputs from four software were collapsed with script annot_genes.py with param "keyword 'coffea arabica'". Finally GO annotation was obtained with GOSlimViewer tool with plant GOSlim Set on Agbase (MCCARTHY *et al.*, 2007) webserver. The GOSlim annotation is detailed in 7.7 Supplementary Table S2. Endly Blastn with evalue param 1e-10 from Galaxy/EU was used to associate genes with your annotated NCBI gene correspondent from Cara_1.0/ GCF_003713225.1 assembly and infer the chromosome located gene.

2.2.4 PRIMER DESIGN STRATEGY

With gene coordinates in the GFF3 file, domains data of InterproScan5 output and rMATS output, we plotted mRNAs with annotation and AS events by custom script plotASdomains.py to score events and determine which occur in coding regions, the figure of SE events is available in 7.8 Supplementary Fig. S4. Genes undergo more with AS because position of AS, size of the alternative region, the coding region has domains, pathways annotated, and difference of percent spliced in (PSI or ψ) among cases of contrast were screened to choose the best target for IR and SE events for experimental validation with use of figures generated by plotASdomains.py and 7.9 Supplementary Table S3. The alternative isoforms were parsed by the custom notebook extract_alternative_sequences.ipynb to primer design in Primer3Plus (UNTERGASSER *et al.*, 2007) for under alternative region, spanning alternative region

and 5' exon/intron junction. The primers for PCR conventional were designed with parameter "Product Size Ranges 200-1500", and "Product Size Ranges 150-300" for quantitative PCR. An in-silico primer verification was performed with web tool Primer-BLAST (YE *et al.*, 2012) to check the presence of regions in *Coffea arabica* cv. Caturra Vermelho genome assembly GCF_003713225.1 available on NCBI database. The primers sequences are available in Table 1. For primer design to validate AS events for each gene at most three primer pairs were design, according to the following strategies:

i. Control: A primer pair inside constitutive region to verify that the gene express any isoform on all compared samples. This pair was length between 100 and 200 bases to uses in PCR, and qPCR also.
ii. Case: A primer pair designed inside alternative region, to verify alternative isoform presence in one of compared samples.
iii. Case confirmation: A primer pair inside the junction: the site downstream the constitutive region and upstream alternative region. If possible was reused forward oligo of step i. or reverse of step ii. This pair only amplify in sample wat express alternative isoform, because this site of the canonical isoform is different.

Table 1 - Details of primers used in PCR/qPCR validation for the two choose targets.

Label	Forward	Reverse
PP1[a]	AGGGTCAGTAAAATGTTTGCAAGG	GTACGTCCACTGCTGTCCAA
PP2[a]	TGATGAGCAGACTTTGTGGG	GTGAGACCAAGGTTGAATAGCATAT
PRT1[b]	TGTTTCTTCTCCGTGGGTGT	GGTTTCTGTGTGCCTGCTTC
PRT2[b]	TGAACTTGAGAAATTTGCCTGC	GCTTTGCACATCGAACCCAA
AP1[a]	CATCCCTCCGTCTGTTGTTT	AGAAGTGCTACGCCACAGGT
AP2[a]	ACATCTGCCCGAGATTATGG	ACAGGTGGTCTCCCTCAATG
AP3[a]	TTGGGAACTTTTCGAGGTTG	TGCAAAGAAACCAGACAACG
ART1[b]	CCAACCTTGGGACTTCTTGA	GCATAACAACGCTTCGATCA
ART2[b]	GGCTTGTCTTCTCGAACAGG	AATCTCGGGCAGATGTGTTC
GAPDH[c]	AGGCTGTTGGGAAAGTTCTTC	ACTGTTGGAACTCGGAATGC

[a]Primer used in PCR and [b]qPCR validation. [c]CRUZ (2009).

2.2.5 BIOLOGICAL SAMPLES OF COFFEE FRUIT RIPENING STAGES

Coffee fruit immature (for the G condition) and mature (for the R condition) of the C. arabica cv. Caturra Vermelho CIFC 19/1 were collected in greenhouse at the Universidade Federal de Viçosa and after frozen in liquid nitrogen. Total RNA was

extracted using 100 mg of the frozen tissue as proposed for SDS-LiCl method described by VENNAPUSA (2020). RNA quality was evaluated by agarose gel and quantification was performed by Qubit 3.0 Fluorometer (Invitrogen by Life technologies). A total of 1µg of RNA was treated with DNase (Promega – AM 6101) and the treatment was confirmed by amplification reactions with the constitutive gene GAPDH (CRUZ *et al.*, 2009) that had stable expression levels. cDNA synthesis was performed using the Synapse kit (S1402) according to the supplier's instructions.

2.2.6 EXPERIMENTAL VALIDATION OF AS EVENTS BY PCR

Conventional PCR was performed to validate the alternative splicing of apyrase and channel genes in G and R coffee fruit. The cDNA concentration was 20 ng and the primers 10 µM. Conditions used for the amplification were 40 cycles with 95 °C for 5 mi, denaturing at 95 °C for 30 s, annealing 50 °C for 30 s and extension at 72 °C for 2 min. Followed by a final extension at 72 °C for 10 min. The PCR product was visualized by electrophoresis on a 2.0% agarose gel with ethidium bromide.

Real-time PCR was performed to check the expression level of Potassium channel (AKT1) and Apyrase 7 (APY7) genes in immature and mature coffee fruits. The efficiency curve of the primers was performed through a serial dilution of genomic DNA, for the primers named apyrase 2, channel 1, 2 and 3, and cDNA, for the primers apyrase 1 and 3. The efficiency test was performed with serial dilutions of 1:2 ranging from 50ng to 0.78ng of cDNA or genomic DNA. The expression was detected by the GoTaq qPCR Master Mix (Promega - A6002) using the StepOnePlus™ Real-Time PCR System (Applied Biosystems) and the GAPDH as endogenous control. Two biological replicate experiments were conducted, and samples were tested in triplicate for each experiment. The conditions used for the amplification were 95 °C for 10 min, 40 denaturing cycles at 95 °C for 30 s, followed by annealing and extension at 60 °C for 1 min. After the 40 cycles, the temperature was gradually increased, by 1 °C increment, to 95 °C, resulting in the generation of a melting curve. The relative standard curve was used to calculate the relative quantity (Rq) values of each sample for each gene (DE OLIVEIRA MENDES *et al.*, 2013).

2.2.8 GRAPHICS AND ASSOCIATED STATISTICAL ANALYSIS

The graphs of Conventional and Real-time PCR validation were built on the GraphPad Prism 5.0 program (GraphPad Inc., San Diego, CA, United States) and statistical analysis between the experimental groups was performed by the t-test. The custom scripts filter_events.py, plot_events.py, annot_genes.py plotASdomains.py, extract_alternative_sequences.ipynb written for generate plot data to this analysis are available in the GitHub repository MiqueiasFernandes/bioinformatics. The Jupyter notebooks in PDF format to plot figures and extract alternative sequences are available on https://mikeias.net/coffee-notebook. The generic Galaxy workflow created for download reads by its accession (example “Run” column in 7.3 Supplementary Table S1) and quality control is available on https://usegalaxy.eu/u/miqueiasfernandes/w/coffee-qc.

2.3 RESULTS

2.3.1 ALTERNATIVE SPLICING EVENTS SCREENING DURING COFFEE FRUIT RIPENING

By the workflow, which is explained in the 7.10 Supplementary Fig. S5, mapped reads were utilized for discover AS with rMATS. Around ~98,2% of 57.6M reads were kept after quality control (7.3 Supplementary Table S1) applied to triplicates data of green (G), yellow (Y) and red (R) ripening stages. Considering only genome mapped reads, 80% stayed on sorted bam (7.5 Supplementary Fig. S2) of size about 1gb each of 9. for 55,174 mapped genes. Among them, 87,5% are in G, 96,5% in Y, and 85,8% in R experimental groups (Fig. 5a). Then the 9 BAM files were compared in rMATS in two runs: G versus Y, where 4,515 events were identified and R versus Y, where we found 4,941 events. Events with q-value < 0.05 obtained by false discovery rate (FDR) value were filtered to next steps. A total of 150 and 92 significant events were identified for GY and RY, respectively. The Venn diagram of 7.11 Supplementary Fig. S6 showed 18 shared genes in different AS events. In different conditions, a total of 7 genes were shared (Fig. 5b). The relation event/gene was different in conditions, but more expressive in A3SS as showed in relation on the 7.12 Supplementary Fig. S7 and in bars in Fig. 5c and Fig. 5d. In summary, 202 genes are undergone 241 significative events.

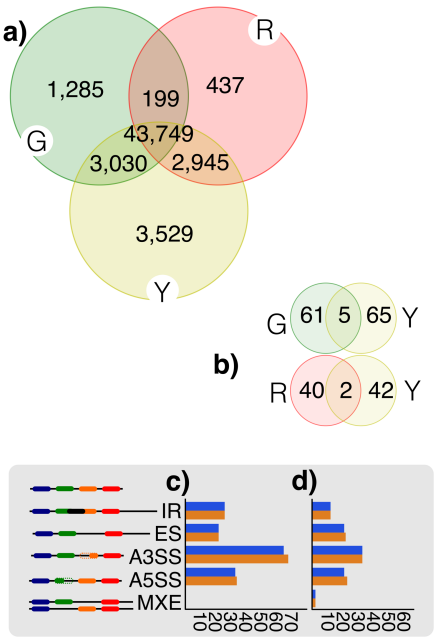


Fig. 5 Overview of alternative splicing events identified by rMATS during coffee fruit ripening. (a) The count of STAR genes with transcript reads mapped in each condition. (b) The number of genes with significative events identified by rMATS. The quantity of events (orange bars) and genes (blue bars) of each type of AS events is show in (c) analysis comparing G to Y and in (d) comparing R to Y. G, green fruit. Y, yellow fruit. R, red fruit. IR, intron retention. ES, exon skipping. A3SS, alternative acceptor 3' site. A5SS, alternative donor 5' site. MXE, mixed events.

In significant set of identified events, the most abundant are alternative acceptor (A3SS) and alternative donor (A5SS) mechanisms with 41,7% and 23,6% AS events respectively in GY and RY contrasts, as show in Fig. 6. The type with less significant events is MXE with 0,8%. MXE subtypes are rare in plants (HAAS, 2008). As gene differential expression is more intense in first stage of ripening (CHENG; FURTADO; HENRY, 2018), the relation 62% of AS events are in GY compared to 38% in RY. The 202 genes undergo AS are sizes between 1,157 and 32,164 base pair, with 23 introns. The CDS (coding sequence) region in these genes are about 59.4% of its size. About IR, the mean size of retained introns are 452.6 bp, and 165.1 bp, 267.4 bp is size of alternative region in A3SS and A5SS in mean, respectively.

The location of this genes, based on reference genome (NCBI accession GCF_003713225.1) are most frequents in chromosomes 1, 2, 6, and 7 (7.13 Supplementary Table S4). To verify gene duplications impact in AS analysis by allotetraploid nature of arabica specie the transcriptome was submitted to Orthovenn2 analysis, which a total of 26% of these genes are detected as inparalogs.

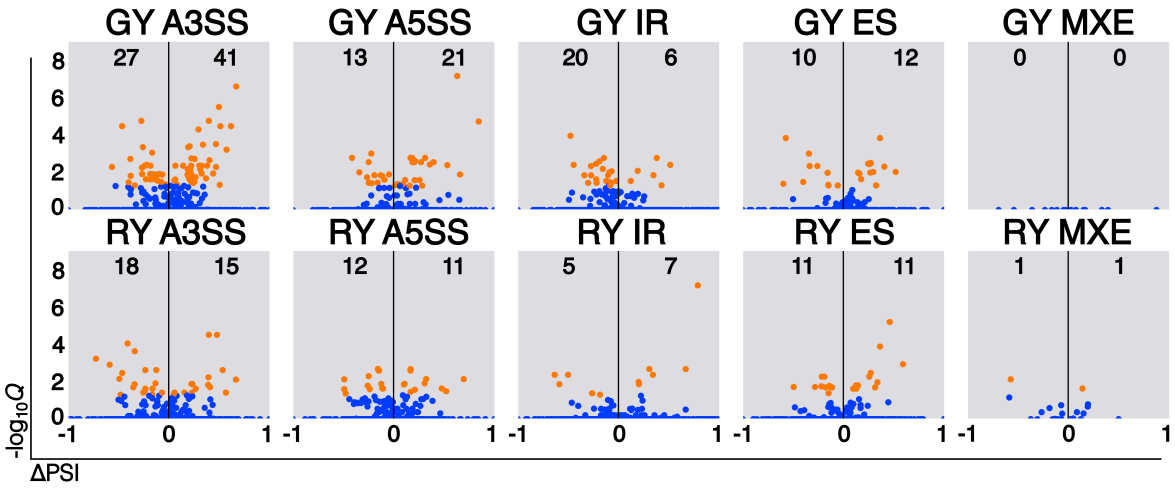


Fig. 6 Quantification of events identified by rMATS in three stages of arabica fruit ripening. In first row, the events A3SS, A5SS IR, ES and MXE identified in comparison between G with Y and in second row in R to Y comparison. In orange are the significative events and in X axis are the difference of percent spliced index (Δ PSI) for occurrence of alternative isoform in G/R to Y condition. The number annotated inside of each plot is the quantity of significative events identified. The top-right numbers represent the events with alternative exon in Y and in top-left the events with exon in the other compared condition (G or R). Q-values are FDR calculated by rMATS. G, green fruit. Y, yellow fruit. R, red fruit. IR, intron retention. ES, exon skipping. A3SS, alternative acceptor 3' site. A5SS, alternative donor 5' site. MXE, mixed events. PSI, percent spliced index of specific exon into the transcript population of a gene.

2.3.2 GENE FUNCTION RELATED WITH COFFEE FRUIT RIPENING IMPACTED BY AS

Because the genome available has no functional annotation, we used multiple databases annotation strategy (BOLGER, MARIE E.; ARSOVA; USADEL, 2018; LANTZ *et al.*, 2018), where 96% of 203 AS genes were annotated and 7 AS gene were identified with no annotation (7.15 Supplementary Table S6). In annotated gene set, the multicellular organism development (GO:0007275) and anatomical structure morphogenesis (GO:0009653) terms are the most abundant of biological process GO, with 27 and 11 genes annotated by these terms, respectively (7.14 Supplementary Table S5, and 7.17 Supplementary Table S8). The protein annotated names E3

ubiquitin-protein ligase PRT6 isoform (XP_027071496.1), sorting nexin 1-like isoform (XP_027072350.1) and copper-transporting ATPase RAN1 (XP_027087717.1) are the three most abundant in AS dataset with 14, 8 and 6 events, respectively. The top 10 most abundant Pfam domains annotated in AS genes include RNA recognition motif (PF00076) and Sugar (and other) transporter (PF00083) with 59 and 15 counts, respectively. The 101 (97%) of description pathways for the id from InterproScan5 and Eggnog-Mapper annotation were obtained from KEGG database to target selection. In 7.16 Supplementary Table S7, we defined by considering multiple annotations (like gene name, GO and involved pathways) and regulation in ripening stage the two-best target for experimental validation showed in two next figures.

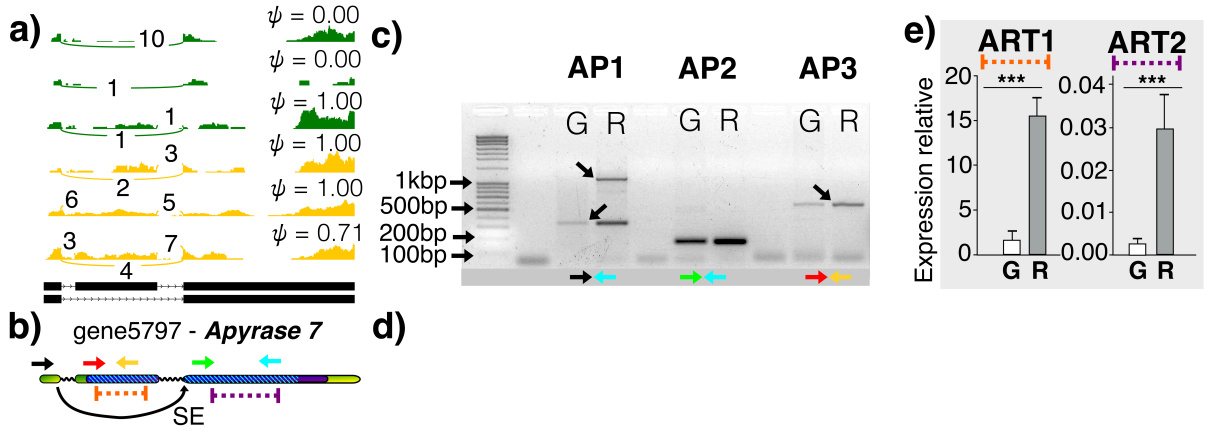


Fig. 7 AS validation in apyrase 7 during coffee fruit ripening. (a) The sashimi plot shown the read coverage in each replicate of case G and in intermediary stage Y. As the introns have big size, the plot is displayed in 100:1 scale for better visualization. (b) is shown the gene in graphic mode with black curved arrow indicates where occurs the exon skipping AS event, in purple is the CDS region and in dashed the PFam domains identified in functional annotation in isoform structure and where AS occurs, in green is showed the exon and introns is represented by black waves. This gene has three exons spanned by introns, the second exon is alternative and others two are constitutive. The colors arrow represented PCR primers used in the PCR and qPCR validation. (c) the results of PCR showed bands for canonical isoform (AP2) and AS isoform (AP1 and AP3) using primers combination forward green and reverse blue displayed in b. (e) Expression quantification of for canonical isoform (ART2) and AS isoform (ART1) by qPCR. The orange and purple doted lines represent the amplicon position in (b). The ladder used was HighRanger Plus 100 pb DNA (Norden Biotek Corporation – 12,000). G, sample of unrip fruit. R, sample of mature fruit.

The GO enrichment analysis of Orthovenn2 displayed in 7.18 Supplementary Fig. S8 showed 14 biological process terms enriched in AS genes. The term cellular protein modification process (GO:0006464), protoporphyrinogen IX biosynthetic process (GO:0006782), and tetrahydrofolate metabolic process (GO:0046653) were shared by more than one event (7.17 Supplementary Table S8). Expected enriched

terms associated to seedling development (GO:0090351) and myo-inositol transport (GO:0015798) term were also enriched in Y on both contrasts. The Green ripening stage (G) had the most represented enriched with 6 terms. The term regulation of translation (GO:0006417) of genes in G conditions undergo A3SS events is another interesting enriched biological process. Despite finding interesting results in the annotation the 7.15 Supplementary Table S6, and 7.16 Supplementary Table S7 was made to choose some targets for experimental validation. In addition to GO enrichment, other attributes like PSI, statistical significance (Q value) of the AS event was also used in this selection. Together these quantitative attributes of gene and event turned in the score of each event. Next, choosed tgenes Potassium channel (AKT1) and apyrase 7 (APY7) were validation using PCR and qPCR.

2.3.3 TARGET GENES AS EVENTS IN VITRO VALIDATION

To validate occurrence of AS events IR in gene AKT1 and SE in gene APY7 specific primers were used for PCR and qPCR. The Fig. 7 and Fig. 8 show for each gene alternative splicing predicted by RNASeq analysis, the genomic context of the gene, the primer designer strategies, and the validation of AS by alternative isoform detection by PCR and alternative isoform differential expression by qPCR. In Fig. 7b, the apyrase 7 transcript was predicted to have around half of its coding sequence (~1.6kbp) skipped by AS event in G condition. This alteration is very significative for translated protein excluding domain region important for protein activity. As the final product is involved in Pyrimidine and Purine metabolism pathway (Kegg accession id ko00230 and ko00240). This alteration is significative in the metabolic process.

For AKT1 gene the inclusion level in greater intron in Y condition (0.15, 0.07 and 0.04) is higher than inclusion level in G condition (0.0 in three samples). By the in-silico analysis the retained intron event inserts a PTC in coding sequence, what prevents its major part protein production, like a switch by degrading the transcripts by non-sense mediated decay (BAKER; PARKER, 2004). Because of these, the SE impact in product of apyrase 7 gene and PTC in IR event in Potassium channel AKT1 gene were selected to validation.

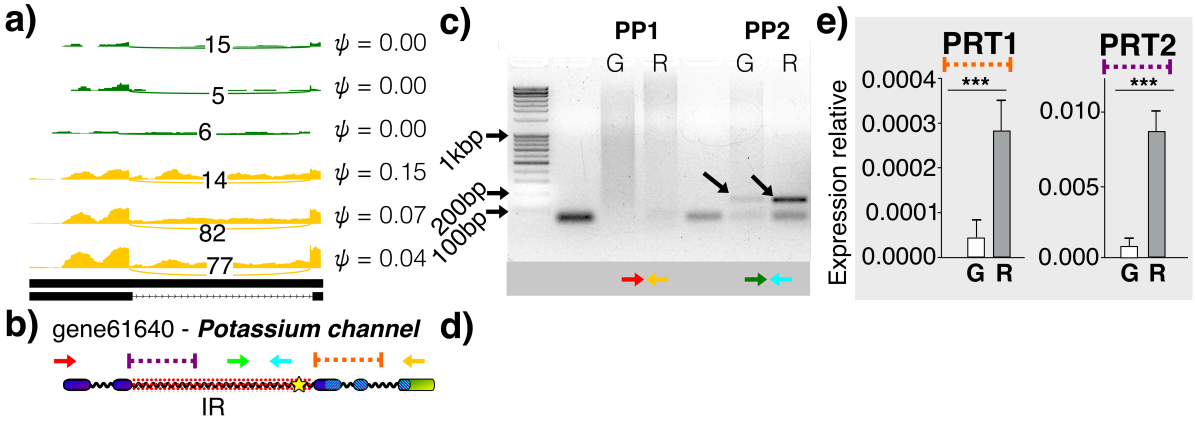


Fig. 8 AS validation in Potassium channel AKT1 during coffee fruit ripening. (a) The sashimi plot shown the read coverage in each replicate of case G and in intermediary stage Y. The reads coverage is expressively greater in retained intron region in case Y where occurs the AS event. (b) is showed the genomic context of gene with highlights in the retained intron of IR event represented by red doted region, the yellow star indicates that was identified premature terminator codon (PTC) inside retained intron on gene open read frame. The color arrows indicated primer sequence used to validation of AS event (c) conventional PCR validation using primers for canonical isoform (PP1) and AS isoform (PP2) (e) isoform specific expression by qPCR using primers for canonical isoform (PRT1) and AS isoform (PRT2). In both b and c, e the primer sequence is according with colored arrows showed in b. The orange and purple doted lines represent the amplicon position in (b). The ladder used was HighRanger Plus 100 pb DNA (Norden Biotek Corporation – 12,000). G, sample of unripe fruit. R, sample of mature fruit.

How the ψ calculation can be affected by read coverage according explained in 7.19 Supplementary Fig. S9, was needed to generate the sashimi plot stated in Fig. 7a to verify the read coverage in AS region where occurs the exon skipping (ES) event. In this gene in condition green (G) by in silico analysis the exon is not included in two replicates of this case, but in case intermediary (Y) this exon is included. The plotASdomains.py script allows to see the ES event overlap on coding sequence (CDS) and annotated functional domains what can be seen in Fig. 7b to identify targets where there is more AS impact on translated protein functional region with the manual curation of reads coverage on AS region and contextualization with functional region the APY7 gene. For the validation the designed primers on regions indicated by colored arrows in Fig. 7b was combinate in pairs, to PRC amplify of both alternative (around 1000 pb) and canonical isoform with different band size in agarose electrophorese Fig. 7c. The AP1 was designed to verify the amplicon size of PCR in immature (G) and mature (R) condition, where can see in G case 300bp band of amplicon of isoform without alternative exon, and in R case 1300bp amplicon of isoform what includes the

alternative exon. It is also possible to see the 300bp band in R case why this isoform is present in this case too. The AP2 primers was designed to validate gene gene5797 expression in both cases by 180bp band as validated, this is strategy i. explained in material and methods. Lastly the AP3 was designed to verify the alternative exon presence in R case, the presence is evident in Fig. 7e but also has a band for G case, wherever in qPCR we identified a greater expression in R case compared to G case showed in ART1 in Fig. 7e. The primers for real-time PCR validation were designed in regions indicated by orange (ART1) and purple (ART2) arrows in Fig. 7b. Then to verify difference in quantitative gene expression by qPCR the chosen regions were common (ART1) to both canonical and alternative isoforms and exclusive to alternative isoform (ART2) in alternative exon. By the ART2 did found differences with statistical significance by t-test.

The Potassium channel AKT1 gene also has read coverage in AS region and the AS impact in functional region as showed in Fig. 8a and Fig. 8b. Interesting in this, even the script identified premature terminator codon (PTC) inside retained intron annotated by yellow star in Fig. 8b. In conventional PCR validation that is showed in Fig. 8c is not possible to see the retained intron band for PP1 as expected by 680bp band in G and 1650bp in R condition, but in qPCR in Fig. 8e allow to see by PRT1 what have significative difference in relative expression in G and R cases, where have more expression of isoform with retained intron in mature sample. In case of PP2 PCR the band of R sample is more apparent in R case, here the idea was to validate which condition contained the intron what is according, in mature sample. This band is more intense in R sample confirming that AS retained intron region in mature fruit. This result was also corroborated by qPCR PRT2 what show significative difference about compared samples and yet with higher quantity of this intron region in mature condition.

Despite to ripening related genes, 55 proteins obtained from Uniprot database like 18 sucrose biosynthesis, 9 Cell wall degradation, 9 Ethylene-responsive element binding, 5 Ethylene biosynthesis and signaling, 4 Flavor volatiles biosynthesis, 2 Carotenoid biosynthesis obtained from ALEXANDER (2002) and CHENG (2016) gene list was aligned again protein set of AS genes (evalue 1e-5). Genes with expression changes involved in fruit ripening: Solyc05g012020 (Uniprot id Q8S4L3) MADS-box transcription factor MADS-rin RIN Transcription factor Ripening Inhibitor and Solyc06g069430 (Uniprot id A0A3Q7GYA) Uncharacterized protein FUL1

Transcription factor FRUITFULL-like MADS-box 1 was founded too, by align in gene37874 and gene70542 AS genes. Both genes are undergoing A3SS in GY comparison with PSI up in Y.

2.4 DISCUSSION

In this work was described and validate for the first time AS events in genes associated to coffee fruit ripening. The first step was the prediction of AS events using Illumina RNASeq data. The AS mechanism type proportion between IR, SE, A3SS, A5SS e MXE agrees with AS predicted for others plant models (CHAUDHARY; JABRE; *et al.*, 2019; HAAS, 2008; YAN *et al.*, 2021). Using this approach, new biological insights were obtained by integrated analyses of previously published data to predict the hundred events in dozens of AS genes involved in biological processes important in fruit ripening such as regulation of gene expression, response to endogenous stimulus, and signal transduction as detailed in 7.14 Supplementary Table S5-8. In significative events set, the number of IR and MXE events were identified in expected proportion, with IR in high abundance and MXE as rare as previously described for plants dataset (BEDRE *et al.*, 2019). The multi databases annotation strategy was efficient because just 7 genes had no annotation attributed and 87.6% were annotated for more than one database, including high number of genes annotated in curated Uniprotkb (83.7%), and 49% associated to metabolic pathways.

The relation gene to events is not the same in comparison GY and RY. In summary, in GY condition comparison, the genes are undergone to more AS events (131 genes) that observed to RY (84 genes) condition comparison. In annotated genes undergo AS, the mostly genes are related to ripening process (7.17 Supplementary Table S8), remain 8 genes associated to proteins with unknown function that can be studied to verify how is its involvement in ripening process. The reduction to raw data from rMATS to 2 target genes for experimental validation was based on 2 filters: i) statistical significance of FDR (97.5% removed) and ii) impact of AS in protein domain composition. The other genes included by these criteria are listed in 7.15 Supplementary Table S6. The enrichment analyses by orthoven2 and pathways annotation for AS genes are important to score genes by its involvement in ripping. The significant GO biologic identified processes include GO:0008380 RNA splicing (7 genes) and GO:0015031 protein transport (5 genes), MAPK signaling pathway occurs 3 times and can be used to define a group with 6,5% of total genes (6 genes). After this analysis, selected high confidence target is submitted to primer design. In this step, other precaution helps to avoid mistakes: in silico primer verification, to validate sequences and determine if in-paralogs are masking the results (HAAS, 2008). Here,

the NCBI blast primer tool was used to verify primers verification and NCBI arabica were confirmed. So, this work helps to understand this ripening process in coffee fruit, but still is need to better understand the relations of other genes with AS and ripening fruit.

The protein functional domain located on striped region draw in gene gene61640 on Fig. 8 identified by Interproscan5 as Ion channel (Pfam accession PF07885) and as Potassium channel domain (Interpro accession IPR013099) as the AS insert a PTC before this domain the gene becomes nonfunctional. So, the IR event observed in gene gene61640 (Potassium Chanel AKT1) set the AS with a fundamental role in ripening biological process by K+ control in the fruit. The K+ is an essential nutrient to growing and development of plants, what is uptake on high concentrations by potassium channels like AKT1 since roots (HIRSCH *et al.*, 1998). Because this cation is the most abundant in plant cells, it needs to be distributed from roots to fruits with help of enzymes translated by genes of Potassium Chanel AKT family collaboratively (DENNISON *et al.*, 2001; SUNG *et al.*, 2007). The potassium channels can be activated by phosphorylation and among its functions is a of help in transport of substances into the cell by plasma polarization (SUNG *et al.*, 2007). Like see in the plant where its grow depends on intracellular manutention of K+ (SÁNCHEZ-BARRENA *et al.*, 2020) here saw that when going from green to mature stage the channel is impeded by IR by control of transcripts of bulk K+ transport enzyme.

As we expected, during fruit development the gene gene61640 has its activity, but at a specific time when the fruit is ready to ripen the gene needs to interrupt its activity to control the K+ accumulation (since K+ has an impact on the chemical composition of the bean and consequently on its sensory properties) or even for having finished the fruit development who needs K+. To do this in this case the gene use of AS mechanism to produce truncated transcripts by IR event. In grapevine, as the K+ quantity take a negative impact in vine quality CUÉLLAR (2013) studied how is done this K+ transport then they observed that channel helps on rapid K+ transport on intermediary stage of ripening, so that the development and ripening consume K+ with a peak observed in maturation. While in grapes K+ alternates acidity, in coffee acidity is determined by chlorogenic acids (CHENG *et al.*, 2016), however the potassium has effects on the coffee bean chemical composition (CLEMENTE *et al.*, 2015).

It was observed that in the intermediary stage the first exon of gene gene5797 annotated as probable apyrase 7 undergo ES, as this exon have most of the functional domain on alternative site it means that AS event tends to affect the apyrase enzyme encoded by this gene. The apyrase (APY) wide exists in plants, metazoan, fungi, and bacteria and can regulate the extracellular ATP homeostasis (eATP) and intracellular (iATP) mediating biological process in plants. The intracellular ATP is a source of energy to support cellular metabolism, and the APYs has function to regulates the energy metabolism during development, ripening and senescence of fruits (SHAN *et al.*, 2022). The APY7 is expressed in some tissue on different development stages and has two trans-membrane domains and 5 highly conserved sequences (YANG *et al.*, 2013) in about 740aa peptide, here these domains are present in a 551aa protein. The plant apyrase enzyme (NTPDase) has its particularities and are recognized by its key roles if control the plant development. Has studies linking APY expression in grow control, probably by ethylene auxin crosstalk (CLARK, GREG *et al.*, 2021), this phytormones are associate to many biological process, inclusive ripening that is oriented in climacteric fruits like coffee by ethylene (GAO *et al.*, 2020). Thus, the AS can regulate ripening by fruit metabolism control by signal transduction by ethylene but need to experimental validation to confirm this link.

This first AS coffee results report are not deep. Recent studies in ripening process showed that this biological process is not all clean yet in its model: tomato (LI, ZHIWEI *et al.*, 2020; WANG, RUFANG *et al.*, 2020), but here with this first report the results show the AS like pos-transcriptional control mechanism, in special with IR as validated with real time PCR, this is large applied by plants to AS. To understand better AS influence in ripening, others RNA-seq datasets new research can use data available in SRA database for climacteric (guava: PRJNA472130, tomato: PRJNA391024, apple: PRJNA380154) and non-climacteric fruits (citrus: PRJEB12880, Grape: PRJNA394039,) to discover AS and possibly better understand this biological process. Finally, this work is particularly important to orient ripening in C. arabica research, to consider AS like cotranscriptional control in fruit ripening.

Capítulo 3

Análise de splicing alternativo no amadurecimento de tomate utilizando genoma e dados de RNASeq de banco de dados públicos, através de abordagem baseada em isoformas do software 3DRNASeq com eventos de DAS validados por PCR e qPCR.

3. SCREENING OF DIFFERENTIAL ALTERNATIVE SPLICING EVENTS IN TOMATO FRUIT RIPENING

Tomatoes are the second most consumed vegetable in the United States, with approximately 1.8 million tons produced annually for the fresh market. As a climacteric fruit, tomatoes undergo significant changes in sweetness, color, softness, and flavor during ripening. Alternative splicing (AS) is a common mechanism in plants, and recent research has shown that it is involved in ripening. In this study, we used the SL3.0 tomato genome and a high-resolution spatiotemporal RNA-seq dataset generated by Shinozaki to identify AS events. We analyzed data with 3DRNASeq software and found 6,950 differentially expressed (DE) genes, 100 differentially alternatively spliced (DAS) genes, 6,225 DE transcripts, and 201 differentially alternatively spliced transcripts (DTU transcripts). We developed a Python script to classify the identified events and found that A3SS and A5SS accounted for 56.3% of all events. We also identified 17 retained intron (RI) events in 15 genes, of which 47% had multiple premature termination codons (PTCs) in all three reading frames. Additionally, 48 genes were classified as exon skipping (SEs). The ripening process is still not fully understood, but understanding the complex biological process of ripening through the lens of AS can help the industry deliver fresh fruits with preserved nutritional characteristics and longer shelf lives.

Keywords: Tomato; Alternative splicing; ripening.

3.1 INTRODUCTION

Tomato is a member of the Solanum genus, the largest in the Solanaceae family which includes a variety of important crop plants such as eggplant, chili pepper and potato (WEESE; BOHS, 2007; WEI *et al.*, 2020). Like the potato, the tomato wild species originated in South America (MATA-NICOLÁS *et al.*, 2020), was spread to Europe to the world by early Spanish explorers. With 31.4 pounds of tomatoes available for consumption per person, this vegetable is the second commonly consumed in USA (USDA, 2019), with a world production that exceeds 180 million tons (FAO, 2017), Brazil is ninth of this result with about 4,2 mi tons produced (CONAB, 2019). It is a rich source of several vitamins, minerals, and fiber, as well as a dietary source of antioxidants 12 million tons are processed into various products such as soup, catsup, and salads, another 1.8 million tons are produced for the fresh market to consume ripening fruit (ZHENGFEI GUAN; TRINA BISWAS; FENG WU, 2017). These modern varieties to support this production is the result of intensive plant breeding programs since the beginning of the 20th century, and the natural biodiversity of tomato wild species has been key in this success (MATA-NICOLÁS *et al.*, 2020). By its importance it is between most studied organism with as be principal in project SOLNetwork (FERNANDEZ-POZO *et al.*, 2015), About 25,3k results on google scholar, 11 complete genome assembly on NCBI and Phytozome and 16,1k experiments on SRA where 562 of this data is of ripening analyzes.

Fruit ripening is a physiological process of Angiosperm species which is under developmental, hormonal, and epigenetic regulation driven by environmental stimuli (PALMA *et al.*, 2019). At mature green stage, the fruit undergoes important changes in sweetness, color, softness, and flavor to among other reasons become attractive to frugivores (LI, SHAN; CHEN; GRIERSON, 2021). This changes impact in fruit and its subproducts like in coffee (CHENG *et al.*, 2016) because this has been targeted to study to improve postharvest preservation of fruit (BOSE *et al.*, 2021).

Tomato is a perennial plant, although it is cultivated as an annual. Tomatoes require around 75 days from transplanting to first harvest. Considering ripening mechanism, tomato is categorized as climacteric by requiring ethylene for normal fruit ripening (ALEXANDER; GRIERSON, 2002), in this category it is a model for studying this process (SU *et al.*, 2021) its short time life and rich omics data like its genome available in public databases contribute for this. At transcriptomic level the peak of

ethylene is related to the regulatory network involving MADS, bZIP, NAC, MYB and other genes (ROHRMANN *et al.*, 2012). Recent epigenetic studies elucidate the methylation influence in this complex regulatory network (LI, ZHIWEI *et al.*, 2020). Today MADS-RIN, NAC-NOR and SPL-CNR transcription factor encoding genes are named master regulators and highlighted in studies for fruit ripening (WANG, RUFANG *et al.*, 2020).

Alternative pre-mRNA splicing AS is a common mechanism found in plants (LING *et al.*, 2019; SYED *et al.*, 2012) a important step between transcription and translation (KORNBLIHTT *et al.*, 2013) to increase protein diversity (CHAUDHARY; JABRE; *et al.*, 2019; KELEMEN *et al.*, 2013; NILSEN; GRAVELEY, 2010) then obtain different results from few genes (BARALLE; GIUDICE, 2017). In canonical splicing the nascent RNA harbors introns that are removed by the spliceosome (HERZEL *et al.*, 2017) by recognize branch point sequence (SHARP; ROBERTS; SHI, 2017), the splice site always spliced as exon or intron is called constitutive. In determinate context AS events like condition or tissue (FU, XIANG DONG; ARES, 2014) are regulated by Cis-acting RNA elements like sequence motifs, Trans-acting factors like SR1 proteins (ULE; BLENCOWE, 2019), and other RNA features, such as exon length and secondary structure (BARASH *et al.*, 2010) and others (BRAUNSCHWEIG *et al.*, 2013) and can be classified in exon skipping ES, alternative 5' splice site A5SS, alternative 3' splice site A3SS and intron retention RI which can be found combined in complex events (LI, YEYUN *et al.*, 2020; PARK *et al.*, 2018; VAQUERO-GARCIA *et al.*, 2016). RI events are more abundant in plants in animals SE is more common (CHAUDHARY; KHOKHAR; *et al.*, 2019). In special plant intron size long RI events undergo nonsense-mediated decay (BROGNA; WEN, 2009), recursive splicing and the inhibition of cryptic splice sites by the exon junction complex are two related processes with opposite outcome (GEHRING; ROIGNANT, 2021; SIBLEY; BLAZQUEZ; ULE, 2016). The alternative splice site ASS in mRNA can be quantified by index of splice in different samples then percent of inclusion of ASS in specific isoform at sample is denoted by percent sliced index PSI or Ψ and the different Ψ among samples is delta psi ΔPSI. Recent studies have elucidated the influence of AS in plant biologic processes (SZAKONYI; DUQUE, 2018; XIA *et al.*, 2020), stress response (FILICHKIN *et al.*, 2015; FU, LIANGBO *et al.*, 2019; LALOUM; MARTÍN; DUQUE, 2018; LEE; ADAMS, 2020), and global view like on tomato (CLARK, SARAH *et al.*, 2019).

Advances in High-Throughput Sequencing Technologies reavelated extend and complexity of AS with new analysis (BEDRE *et al.*, 2019). Next-generation platforms such as Illumina enabled researchers to submit a colossal volume of data of your data of transcriptomic analysis to public databases such as SRA which hosts more than 1k runs of RNASeq projects. Third generation sequencing as single-molecule real-time sequencing SMRT labeled an accurate pipeline Iso-seq (HU, HONGYIN *et al.*, 2020; WANG, MAOJUN *et al.*, 2018). In NGS pipelines is usual mapping of reads in genome (MERINO; CONESA; FERNÁNDEZ, 2019; TONG *et al.*, 2020), for now long reads have suggested quantifying it as sequence of isoform (ZHANG, CHI *et al.*, 2017). In this context some bioinformatic tools was written to discovery AS (REDDY *et al.*, 2012; WEN; MEAD; THONGJUEA, 2020a) which can be classified in three groups i) exon-based methods where read counts are assigned to features to find AS genes ii) Isoform-based methods which reconstruct full-length transcripts and estimate their relative abundances in each sample based on the sequencing reads and iii) event-based methods where splicing events themselves are quantified by calculating Ψ values for each event (Mehmood et al., 2020). With recent increment of IA applications new approaches has developed to analysis with deep learning (JAGANATHAN *et al.*, 2019; ZHANG, YI *et al.*, 2018; ZHANG, ZIJUN *et al.*, 2019).

Understand the complex ripening process enables industry to better use its protection in a sense to deliver fresh fruits with quality and nutritional characteristics preserved plus long shelf time. In transcriptional context, the AS mechanism can be useful to explain controls in a network of ripening regulators. In this scenery this work will use RNA-seq data of model for ripening studies, tomato, deposited on NCBI (SHINOZAKI *et al.*, 2018) yet not explored in ripening AS context, to find different AS genes among mature green and red ripe samples. For this we use exon-based 3D-Rnaseq (GUO, WENBIN *et al.*, 2021) R package to analyze transcript quantification by Salmon (PATRO *et al.*, 2017) again SL3.0 tomato genome (SATO *et al.*, 2012) then report experimental validated results.

3.2 MATERIAL E METHODS

The tomato genome assembly of cultivar Heinz 1706, annotation and transcriptomic SL3.0 was downloaded from RefSeq accession GCF_000188115.4 (SATO *et al.*, 2012) to analysis, this representative genome at Chromosome level was deposited by SOL Genomics Network (FERNANDEZ-POZO *et al.*, 2015). Reads of total pericarp tissue, in 4 replicates at mature green MG and red ripe stage RR of fruit development, deposited for project High-resolution spatiotemporal mapping of tomato fruit development (SHINOZAKI *et al.*, 2018) on NCBI on accession PRJNA391024. This samples listed on 7.20 Supplementary Table S1 were submitted to quality control on Galaxy.EU (COMMUNITY *et al.*, 2022) with tool Trim Galore! Then visualized on FASTQC (BROWN; PIRRUNG; MCCUE, 2017) and MULTIQC (EWELS *et al.*, 2016), the workflow and QC results are available on 7.21 Supplementary Fig. S1 and 7.22 Supplementary Fig. S2. Next the transcript quantification was performed by Salmon on Galxaxy.EU with quality reads and csv file of each sample.

To generate a table of gene-transcript mapping a custom python script was written. This table and Salmon (PATRO *et al.*, 2017) results were submitted to 3D-RNA seq APP (GUO, WENBIN *et al.*, 2021) web server https://3drnaseq.hutton.ac.uk. In 3D RNA-seq pipeline data preprocessing i) Low expressed transcripts and genes were filtered based on analyzing the data mean-variance trend when expressed transcripts were determined as which had 1 of the 16 samples with count per million reads (CPM) is 1 ii) The principal component analysis (PCA) plot showed the RNA-seq data did not have distinct batch effects and iii) The TMM method was used to normalize the gene and transcript read counts to -CPM. The generated plots of process i, ii, and iii are in 7.22 Supplementary Fig. S2-4. Limma (RITCHIE *et al.*, 2015) R package was used for 3D expression comparison. For DE genes/transcripts, P-values of multiple testing were adjusted with BH to correct false discovery rate (FDR). At the AS level, each individual transcript was compared to gene level, which was calculated as the weighted average of TPM of all transcripts of the gene. Then p-values of individual transcript comparison were summarized to a single gene level p-value with F-test. A gene was significantly DAS in a contrast group if it had an adjusted p-value < 0.01 and any of its transcript had a Ψ ratio 0.1. Then AS genes reported were downloaded to in house analysis, this full AS gene list is available on 7.25 Supplementary Table S2.

To assign gene, protein names and others annotation by NCBI and Uniprot (BATEMAN, 2019) in a table of differential AS genes the custom script anot_by_ncbi.py was written. Next the DE DAS and DTU plots was generated by de_das_dtu_genes.ipynb python notebook and events of differential AS genes was identified and classified by Identify_as_events_by_ss_vector.ipynb. The events found are manually curated. The custom scripts written for this analysis are available in the repository https://github.com/MiqueiasFernandes/bioinformatics.

To find GO enriched genes the proteins of DAS genes was submitted to Orthoven2 (XU *et al.*, 2019) webserver with default params. Selected targets to experimental validation by annotation was plotted on Genome Data Viewer (RANGWALA *et al.*, 2021) then annotated with domains based on InterproScan (JONES *et al.*, 2014). Selected targets based and annotation and impact of alteration of isoform was verified in previous plot and the CYP80E6, LOC101258920 and LOC104644837 genes are selected to validation.

For the experimental validation with PCR and qPCR (HARVEY; CHENG, 2016), first the alternative isoforms were obtained from NCBI to primer design in Primer3Plus (UNTERGASSER *et al.*, 2007) for under alternative region, spanning alternative region and 5' exon/intron junction. The primers for PCR conventional were designed with param "Product Size Ranges 200-1500" and "Product Size Ranges 150-300" for quantitative PCR. The primers sequences are available in 7.26 Supplementary Table S3.

The extraction of total RNA was made from macerated tissue, being 1 g of pericarp of green tomatoes equally divided in two tubes and 2 g of pericarp of mature tomatoes equally divided in four tubes followed by addition of 1 mL of TriReagent (Sigma–- T9424) into the macerated tissue, the next steps were made as proposed by MUOKI et al., (2012) with adaptations. The RNA quality was assessed by agarose gel and the quantification was performed by Qubit 3.0 Fluorometer (Invitrogen by Life technologies). The treatment with dNase (Promega – AM 6101) was from 1µg of RNA and the confirmation was made by PCR with the constitutive gene actin. cDNA synthesis was accomplished using the Applied Biosystems kit (4368814) according to the supplier's instructions.

The alternative splicing of the genes cytochrome P450 (A1PCR1 e A1PCR2), transcription factor (A2PCR1) and zinc finger (A3PCR1 e A3PCR2) was validated by

conventional PCR in green and mature tomatoes, using cDNA with the concentration of 50 ng and primers at 10μM. The amplification conditions were 95°C for 5 min, 35 denaturing cycles at 95 °C for 30 s, annealing was 50°C for A1PCR1 and A3PCR1, 55°C for A1PCR2 and 60°C for A2PCR1 and A3PCR2, all for 30 s and extension at 72°C for 1 min. Followed by a final extension at 72°C for 5 min. The PCR product of A1PCR2, A2PCR1 and A3PCR2 was visualized by electrophoresis on a 1.0% agarose gel with ethidium bromide, the amplicons of A1PCR1 and A3PCR1 were visualized on a 2.0% gel.

For confirmation of the expression level of the genes in green and mature tomatoes, real-time PCR was performed. The efficiency curve of the primers was performed through a serial dilution of a pool from cDNA of green and mature tomatoes for the endogenous, actin and ubiquitin, dilutions cDNA of green tomatoes for A2RTP1 and A3RTP2 and cDNA of mature tomatoes for A2RTP2 and A3RTP3. For the efficiency test, serial dilutions of 1:2 in the range of 50 ng to 0,78 ng of cDNA were made. The expression was detected by the GoTaq qPCR Master Mix (Promega–- A6002) using the StepOnePlus™ Real-Time PCR System (Applied Biosystems) and the actin and ubiquitin as endogenous controls. The experiments were conducted with five biological replicates in duplicate for each experiment. The conditions used for the amplification were 95 °C for 10 min, 40 denaturing cycles at 95 °C for 30 s, followed by annealing and extension at 60 °C for 1 min. After the 40 cycles, the temperature was gradually increased, by 1 °C increment, to 95 °C, resulting in the generation of a melting curve. The standard curve was used to calculate the relative quantity (Rq) values of each sample for each gene (DE OLIVEIRA MENDES *et al.*, 2013). The validation graphs of real-time PCR were built on the GraphPad Prism 5.0 program (GraphPad Inc., San Diego, CA, United States) and statistical analysis between the experimental groups was performed by the test-t.

3.3 RESULTS

To identify differential AS genes. the raw reads of tomato RNA-seq was submitted to quality control then to quantification to process on 3DRNASeq. The workflow until quantification performed on Galaxy.EU is show in 7.21 Supplementary Fig. S1. After, QC 97.3% of raw reads was used to quantify 35,895 transcripts of 23,726 genes, where 89,6% of this gene are protein coding and of this 26,5% has more than one annotated isoform. The TPM mean on MG samples is range between 34.9 and 38.9 and in RR in 38.0 to 40.2, where is 96,3% gene on MG and 91,4% gene on RR which TPM greater than zero. The master regulator genes LeSPL-CNR, MADS-MC, MADS-RIN, MADS-RIN-MADS-MC, NAC-NOR (WANG, RUFANG *et al.*, 2020; WANG, WEIHAO *et al.*, 2021) showed mean of TPM 17.8 in MG and 189.6 in RR. In MG samples the candidates to AS genes which show max TPM are LOC101249087, LOC101248825, LOC101247355, LOC101260734, LOC101257132 which range 3315.17 to 839.62 respectively, in RR the gene involved in fruit ripening ACO3 (LI, ZHIWEI *et al.*, 2020) was of the most TPM with 3027.8 followed by LOC101257132, ELIP, Psy1, LOC101255117 range until 1190.8 TPM. After 3DRNA-seq preprocessing step, where insignificant expressed transcripts and genes are filtered, the gene set was reduced to 17,265 Expressed genes with 26,459 Expressed transcripts. 3DRNA-seq analysis detected 6,950 DE genes, 100 DAS genes, 6,225 DE transcripts and 201 DTU transcripts like shown on Fig. 9 bellow.

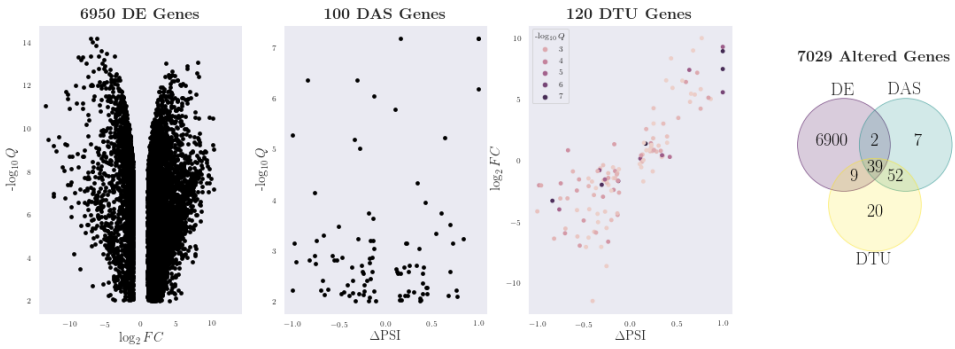


Fig. 9 3DRNA-Seq analysis DEG, DAS and DTU results. The DE genes UP regulated are more expressed in RR samples, same to DAS genes. In DTU genes the greater abs(ΔPSI) were selected to represent its gene; the statistical Q-value is reported by the DTU test. In the Venn graph 41% of DAS genes are differential expressed too. Only statistics significant with adjusted p values are shown and counted here.

In DE, gene set 4,467 are UP regulated and 2,483 are DOWN regulated, in this groups de top 10 more expressive in significance and fold change are LOC101248242, XTH7, LOC101266265, CYP77A19, LOC101264271 UP and PG2, LOC101253875, LOC101246642, LOC109119559, LOC112940921 DOWN regulated. The transcripts regulated only by transcription (DE), only by alternative splicing (DAS) and by both transcription and alternative splicing (DE & DAS) was 6,116, 92, 109 respectively. The pipeline found 13 Significant isoform switches between transcript isoforms of genes. Just two genes LOC101252559, LOC101253917 is in DE and DAS group and 39 is in set DE & DAS & DTU where CYP85A1, SlPer1, pds and others. In protein names of 7,029 DE or DAS or DTU gene set the keyword TRANSPORTER has frequency 2,3%, ZINC has 1.1% occurrence, OXIDASE occur in 1.2% and ETHYLENE is in 0.7% of protein names.

In top 5 gene with greater delta PSI and down delta PSI 2 gene with uncharacterized product and genes coding of protein DETOXIFICATION, kinesin-like protein KIN, probable LRR receptor-like serine/threonine-protein kinase RFK1, MORC family CW-type zinc finger, E3 ubiquitin-protein ligase TRAIP, acyl-lipid desaturase phosphodiesterase GDPD3 and TOM1-like protein. To classify events all DAS genes the upstream and downstream of alternative splice sites of DAS genes was classified as constitutive intron or exon, the quantification results are in next Fig. 10.b.

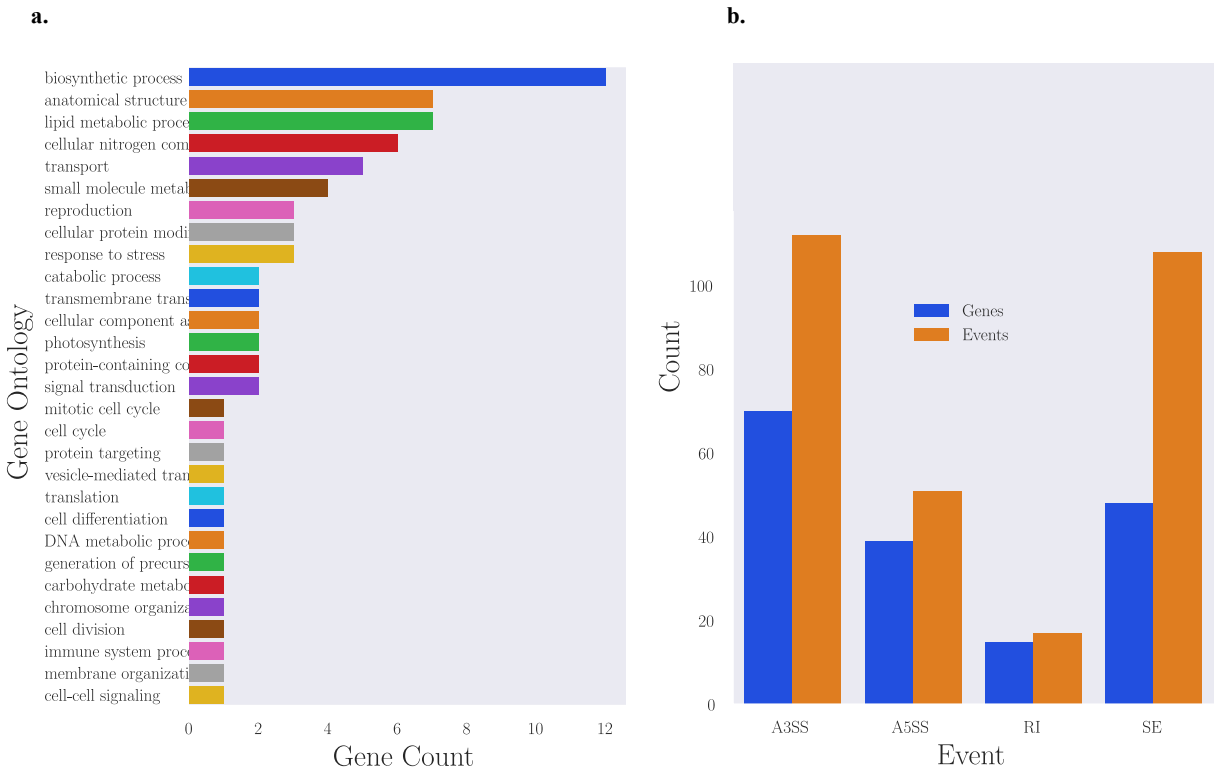


Fig. 10 AS events in DAS genes found by 3dRNA-seq analysis. a) count of genes by Gene ontology term annotation of AS genes. b) count of events found in genes that undergo AS classified by custom python script, the results were manually curated.

Alternative splice sites surrounded by constitutive intron and exon was classified by A3SS or A5SS, together are the most representative category with 163 events (56.3%) and alternative site size range between 1bp and 966bp with mean of 65.3bp. 17 Alternative splice sites surrounded only by constitutive exons was classified as RI in 15 genes. This event is common in plants, here in ripening contrast is less expressive, wherever its impact change phenotype with 47% of retained introns has multiple PTC found in three translate phases. The mean of retained intron size was 159bp. Alternative splice sites surrounded only by constitutive introns in 48 genes was classified as SE, the 108 events found has 68% of genes with impact in its coding region, considering all alternative sites 75% of genes has any AS event occur in CDS.

The protein names and go of DAS genes were annotated with Uniprot and orthovenn2 was used to perform go enrichment analysis. The GOS phosphorylation, protein autophosphorylation and glycerophospholipid catabolic process is found enriched. After processing with GO terms in GO Slim tool the biologic process GOs

biosynthetic process (GO 0009058) with 12 genes, anatomical structure development (GO 0048856) with 7 genes and lipid metabolic process (GO 0006629) with 7 genes was the most expressive as shown in Fig. 10.a.

The Interpro domains most frequent in annotated proteins was Cytochrome P450 (acc IPR001128) and Protein kinase domain (acc IPR000719) found in 8 proteins and AAA+ ATPase domain (acc IPR003593), Ankyrin repeat (acc IPR002110), Zinc finger RING-type (acc IPR001841), RNA recognition motif domain (acc IPR000504) found in 7 proteins, the first four domains is frequent in global analysis of AS genes in work of CLARK (2019). With this annotation and event annotation of impact of change in isoform to product the targets LOC101245278, LOC101258920 and LOC104644837 to experimental validation were defined as shown on next Fig. 11.

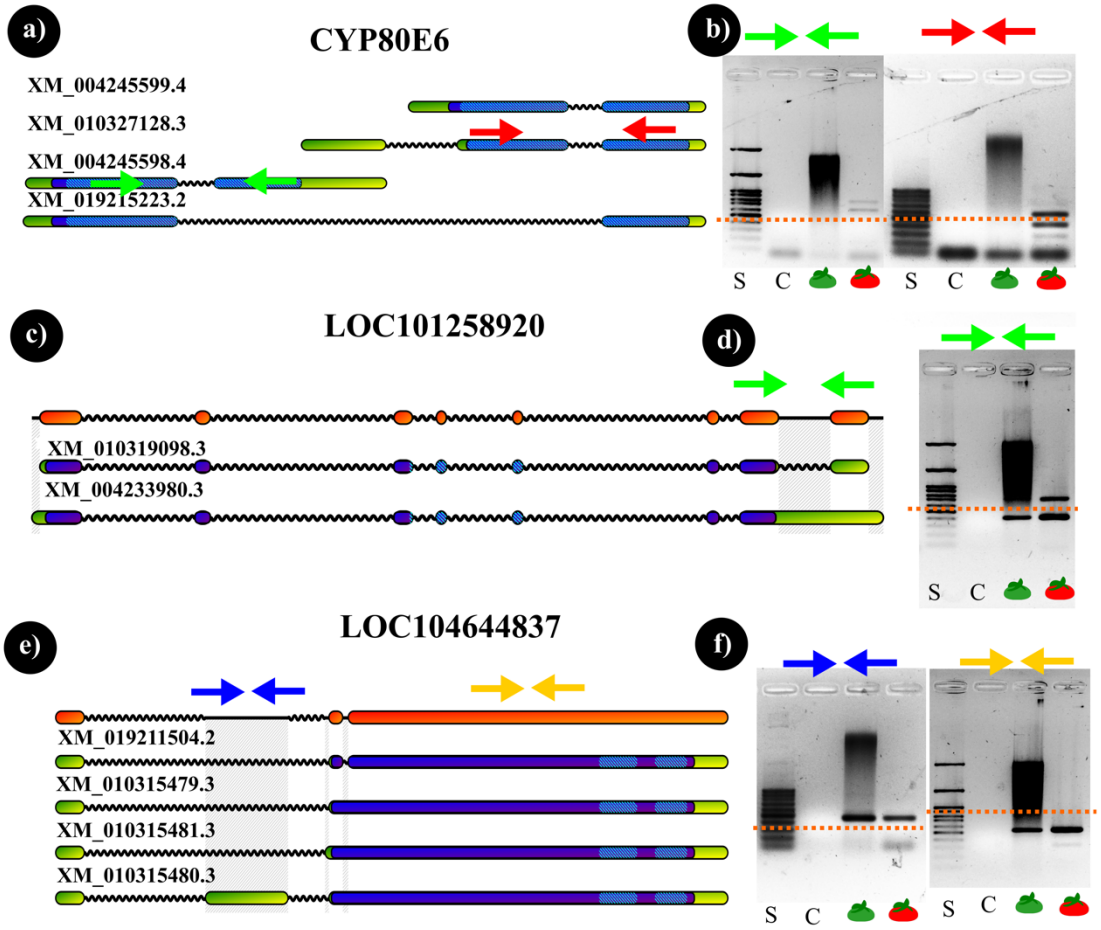


Fig. 11 Selected genes for experimental validation. In left is represented the genomic structure for three targets and in right PCR on unripe and mature tomato samples. The gene structure was obtained by Geneapp software, where constitutive exons is shown filled orange, alternative exons are in green rectangle, introns in black waved lines, purple coding region with functionals domains represented by dashed blue lines. The arrows show primers designed positions. S (scale): Ladder H3 RTU. C (control) water negative control. The dashed orange lines on b, d, f, represent region where be retained 500bp amplicons. In c) are validated an IR event, where on d) is observed in mature sample presence of two alternatives isoforms. Lastly in e) is shown a SE event where blue primer pair amplifying the second exon what is showed in f) evident and discrete bands in mature sample. The yellow primers are to validate presence in both samples.

3.4 DISCUSSION

Ripening is a complex biologic process performed by genomic, epigenomic, transcriptomic mechanism in molecular context (ALEXANDER; GRIERSON, 2002; LI, SHAN; CHEN; GRIERSON, 2021; PALMA *et al.*, 2019; WANG, RUFANG *et al.*, 2020). Tomato like the model to study this process has publications for this mechanism, each study increments more the understand of the participation of ripening actuators. In genomic scope the ripening is coordinated by RIN which has alternative isoforms, NAC-NOR and SplCNR genes (SLUGINA *et al.*, 2021). In epigenomic scope the overexpression of SlJMJ6 gene which encodes a histone lysine demethylase accelerates tomato fruit ripening (LI, ZHIWEI *et al.*, 2020). In transcriptomic, at transcriptional mechanisms the differential expression of ripening related genes (LI, JIAYIN *et al.*, 2016) to perform color, sweets, flavor among other characteristics coordinated by RIN perform ripening. Lastly in transcriptomic scope the pos transcriptional mechanism AS studied contemplated in this work showed 100 genes undergo AS by comparing samples in mature green and red rip stage, adding tree to envolved in this process: i) LOC101245278 coding of probable (S)-N-methylcoclaurine 3'-hydroxylase isozyme, ii) LOC101258920 coding of transcription factor UNE and iii) LOC104644837 coding of zinc finger BED domain-containing protein DAYSLEEPER-like with in vivo experimental validation.

The Zinc finger domain in proteins of AS genes and enriched GO phosphorylation, protein autophosphorylation found here also are related to fruit development regulation (WANG, KETAO *et al.*, 2016). Although RIN have 2 isoforms and the 2nd be related to ripening as role regulator (SLUGINA *et al.*, 2021), interesting here this master regulator not found it undergo AS. The phytoene desaturase (PDS) AS gene found here are related with ripening too (LI, SHAN *et al.*, 2020). LOC101251722 coding of potassium channel AKT2 are found undergo AS too in coffee analyze (FERNANDES, M; 2021). Is interesting evaluate involvement of uncharacterized 18 genes found undergo AS here: LOC101244484, LOC101245489, LOC101245961, LOC101247345, LOC101247498, LOC101248314, LOC101252360, LOC101253473, LOC101256296, LOC101256522, LOC101259888, LOC101262041, LOC101263285, LOC101263341, LOC101266909, LOC101266993, LOC101267838, LOC101257931. But to do this need to tool to facilitate the discover its relation based on its multi omics context.

Parallel to this work, the analyze with rMATS on Coffee (FERNANDES, M; 2021) plus this work serve to rise requisites to tool to sport alternative splicing analysis on screening of numerous results of genes or events in these pipelines. Considering pipeline base on rMATS (SHEN *et al.*, 2014) the differential is identification of novel events and classified results as type of event. Here the exon-based event has differential of be a method with perform generally better of event-based rMATS approach (MEHMOOD *et al.*, 2020). Wherever the raw results of this pipeline need to i) identification of occurred event, ii) classification of event, iii) extraction of alternative isoform with its alternative site highlighted, and iv) display the AS gene with its event, structure, proteins domains in its isoforms, expression, and its annotation. This is requisites to program a new tool to support AS analyses to better explore raw results of AS discover pipelines.

Tomato Ripening also use AS mechanisms to coordinate ripening process in the genes LOC101245278, LOC101258920 and LOC104644837 among others. By this AS is a crucial mechanism to plant development by its involvement in ripening process, but its full understand yet is unclear. To model ripening process the genomic, epigenomic and transcriptomic context must be considered and linked. In transcriptomic context the Galaxy-3DRnaSeq pipeline for find AS pos-transcriptional control is fast and accurate, but to explore full potential the raw results need additional customized tools. The requisites to tool for support AS analysis is show gene withs its classified AS events in details as PTC, expression profiles, domains, structure, and annotations.

Capítulo 4

Desenvolvimento de aplicativo web para a visualização de eventos de AS na estrutura dos genes tendo em vista as dificuldades encontradas em realizar as análises sob diferentes abordagens no café e no tomate.

4. NOTES OF USE OF WEB APPLICATION TO VISUALIZE ALTERNATIVE SPLICING FOR BIOMEDICINE

Alternative Splicing (AS) is an essential mechanism for eukaryotes. However, the consequences of deleting a single exon can be dramatic for the organism and can lead to cancer in humans. The protocol to study can be sampling followed by data analysis with bioinformatics to identify the different AS events in the control and case samples, data visualization for curation, and selection of relevant targets for experimental validation. The raw output of the analysis software does not favor the inspection of events by bioinformaticians requiring custom scripts for data visualization. In this work, we propose the Geneapp application with three modules: GeneappScript, GeneappServer, and GeneappExplorer. GeneappScript is a wrapper that assists in identifying AS in samples compared in two different approaches, while GeneappServer integrates data from AS analysis already performed by the user. In GeneappExplorer, the user visualizes the previous dataset by exploring AS events in genes with functional annotation. This screening that Geneapp allows to perform helps in the identification of targets for experimental validation to confirm the hypotheses under study. The Geneapp is freely available for non-commercial use at http://bioinfo.icb.ufmg.br/geneapp to facilitate the study of AS for bioinformatics.

Keywords: Alternative Splicing; Geneapp; webapp; data visualization

4.1 INTRODUCTION

Alternative Splicing (AS) is an essential mechanism as it allows the eukaryote to extend its proteome even with a limited number of genes (NILSEN; GRAVELEY, 2010). The cellular machinery accomplishes AS by combining alternative regions of the isoforms by the spliceosome, both in productive transcripts (SHARP; ROBERTS; SHI, 2017) and in non-productive ones (OUYANG *et al.*, 2021). In the literature, AS events are classified as exon skipping (ES), intron retention (IR), and alternative splicing site (A5SS and A3SS) (ZHANG, YUANJIAO *et al.*, 2021). The study of Chaudhary et al. points out that exon skipping (ES) is more frequent in animals, while (IR) occurs more in plants. IR events can have a regulatory effect by, for example, inserting a premature stop codon (PTC) that leads to transcript degradation via the nonsense-mediated decay (NMD) pathway (MONTEUUIS *et al.*, 2019) thereby potentially causing cancer in humans (FACKENTHAL; GODLEY, 2008). Thus, since dysregulation in AS may be associated with diseases such as cancer and diabetes, understanding how this mechanism works may help diagnose and treat cancer (ADUSUMALLI *et al.*, 2019; GIANNOPOULOU *et al.*, 2019; LORD; BARALLE, 2021; REN *et al.*, 2021) as well as in helping to produce solutions to treat these diseases (JI; MISHRA; DAVULURI, 2020; LE *et al.*, 2015; MARTINEZ-MONTIEL *et al.*, 2018). Given its importance in explaining human biological occurrences, AS research has become relevant in the medical scenario (BONNAL; LÓPEZ-OREJA; VALCÁRCEL, 2020; HU, WEI *et al.*, 2022; KAHRAMAN; BULJAN; VITTING-SEERUP, 2022; KIM, HYOUNG KYU *et al.*, 2018; LIU, QI; FANG; WU, 2022).

The traditional pipeline for studying biological phenomena in organisms by comparing transcriptomic profiling is accomplished by sampling the control and the different case, sequencing (RNASeq) the samples, analysis, and experimental validation. In the analysis steps, differentially expressed genes (DEGs) are identified from the count of reads that map to the genes (CONESA *et al.*, 2016). AS events, on the other hand, can be identified by the difference in expression at the isoform level (MERINO; CONESA; FERNÁNDEZ, 2019). If statistical significance exists, differential AS is defined (DAS) in the compared samples. DAS can be calculated by the difference between the inclusion ratios of the alternative exon (SHEN *et al.*, 2014) or the ratio between the expression rates of a given isoform compared to the gene expression rate (GUO, WENBIN *et al.*, 2021). Programs using the former approach are called event-

based, while the latter are exon-based (MEHMOOD *et al.*, 2020), and there may also be other approaches. In general, these programs generate tables that need to be curated by bioinformaticians to identify the real genes or isoforms that explain the biological phenomenon and then validate with PCR to close the study pipeline (WANG, ZHINING *et al.*, 2003) completely. This curation can be performed by visualizing the structure of the AS event, qualitatively and quantitatively inspecting the results, crossing structural and functional annotation data with the transcriptomic profile of the samples, extrapolating the DEG and DAS analyses with analyses at the pre-transcriptional or post-transcriptional level, and other methods can also be used.

In the case of visualizing the structure of AS events during curation, it is crucial to inspect how the coverage and depth of reads are mapped in the region where the event occurred and at the junction, as presented in the sashimi plots (KATZ *et al.*, 2015). Since visualizing the structure of the AS event is important, several software enables the graphical presentation of AS analysis, such as IGV (THORVALDSDÓTTIR; ROBINSON; MESIROV, 2013), *rmats2sashimiplot* (https://github.com/Xinglab/rmats2sashimiplot), *AS-Quant* (FAHMI *et al.*, 2021), *VALERIE* (WEN; MEAD; THONGJUEA, 2020b), *Vials* (STROBELT *et al.*, 2016), *Manananggal* (BARANN; ZIMMER; BIRZELE, 2017) and 3DRNASeq (GUO, WENBIN *et al.*, 2021). IGV is one of the most complete and extensible software for visualizing genomic data. Concerning AS, it allows visualization of the events in sashimi-plot format. However, it is limited to structural visualization, not exploring DAS analysis data that often includes volcanoes plot, for example. The rmats2sashimiplot only allows us to visualize the sashimi plot. AS-Quant shows good numbers compared to rMATS in the event identification part, but it generates only the event structure in the visualization part. VALERIE does not allow contextualizing the event with the structure of the protein-coding region and your functional domains. Vials and Manananggal provide extensive support for the AS event's structure but require software installation knowledge, which GeneAPP does not. 3DRNaseq allows the generation of several graphs to visualize DAS analysis through the web app. However, it does not display the event structure since it is isoform-based. The Geneapp application proposed in this work aims to meet these needs and provide resources for enriching DAS analyses by integrating the results with functional annotation from databases such as RefSeq

(O'LEARY *et al.*, 2016), Interpro (PAYSAN-LAFOSSE *et al.*, 2023), Uniprot (BATEMAN *et al.*, 2023), Gene Ontology (CONSORTIUM *et al.*, 2023) among others.

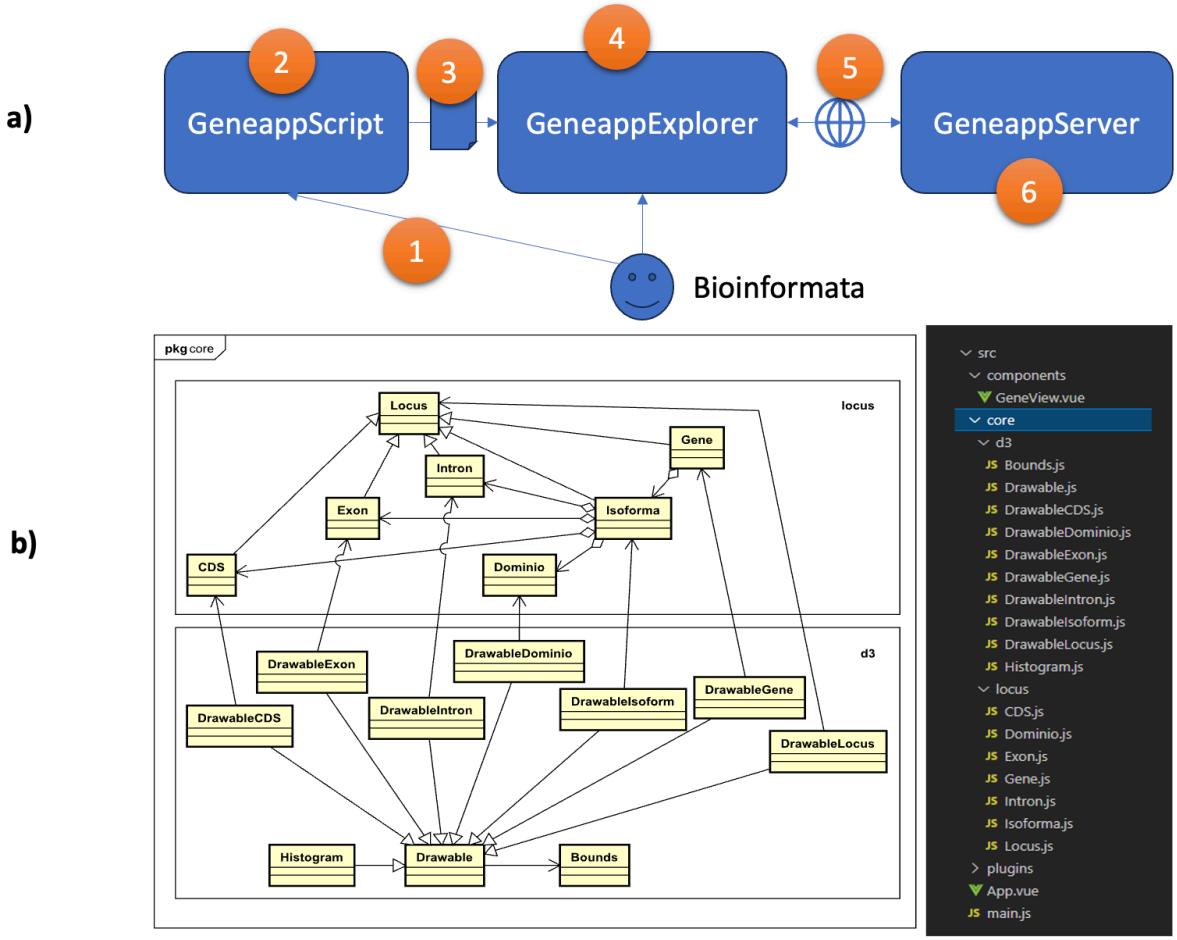


Fig. 12 Geneapp componentes overview. In top (a), its show how a bioinformatician interacts with the Geneapp components. The bioinformatician first uses the GeneappScript to generate data for analysis, which is then visualized in the GeneappExplorer. Alternatively, the bioinformatician can load existing data into the GeneappExplorer, which will then integrate the data and fill in any missing data by calling a REST API to the GeneappServer. The GeneappServer then returns the data needed for visualization in the GeneappExplorer. In bottom (b), we provide an overview of the core package, which is composed of the classes that implement the data visualization in the GeneappExplorer. The locus directory contains classes that abstract different types of loci in the genome. The Locus class is extended by the Gene, Isoforma, Exon, Intron, and CDS classes. The Gene class has an aggregation of Isoforma, which is aggregated by Exon, Intron, and CDS. The Dominio class abstracts conserved protein domains, which typically have coordinates based on the protein sequence, not the genome. The Isoforma class provides an interface between the Dominio and its genomic position as a set of Locus, since it may be partially located in the cds of the isoform. Each of these genomic structures has a corresponding C class in the d3 directory of the type Drawable for plotting. This class C inherits functionality implemented in the Drawable class, including square(), circle(), text(), line(), etc., which allow the structure to be drawn in SVG. The C class has an object to be plotted, which has the genomic coordinates that are converted to SVG coordinates using the scaleX() method available in the instance of Bounds stored in Drawable. The Histogram class contains functions for plotting the sequencing depth graph of gene regions.

Geneapp is a web application for integration, exploration, and visualization of AS data consisting of the pipeline-client-server suite, offering bioinformatician resources for the integration of raw outputs of DAS analyses from different tools with multiple biological databases, aiming to provide robust graphics and results guiding the selection of candidate targets for experimental validation. From our group's experience of DAS analyses in previous research, the identification of target genes by the output tables of DAS identification software through the application of filters often requires the researcher's expertise in computer programming techniques due to the need to implement customized scripts, since each data set has its peculiarity. On the other hand, curating results by visual inspection allows to detect spurious results when the data is contextualized, or even more important: the data in context becomes information (AMMARI *et al.*, 2018), allowing a semi-automated biodata curation (HIRSCHMAN *et al.*, 2012) which obtains insights that go unnoticed in purely methodical pipelines. Geneapp mitigates the need for advanced computer skills to explore the results by not downloading and installing software, with the researcher simply accessing the application through the browser, importing the data, and exporting the analyses. To provide the researcher with information during the visual inspection of AS events, Geneapp presents the genomic, transcriptomic, and proteomic context from the unrestricted data feed by the user, which does not limit the curation to these contexts, making it an important tool for bioinformatics.

4.2 RESULTS

The development of the application was carried out in 2022 on the GeneappScript, GeneappServer, and GeneappExplorer modules following the success of the web app prototype made available in 2021 for testing. The GeneappScript module is a robust wrapper that enables the bioinformatician to perform a complete DAS analysis, from the retrieval of network data to the functional annotation of genes under DAS. This module can be run on Debian distros, such as the Google Collaboratory (Colab) environment where it was developed. The GeneappServer module is a Flask backend that allows integrating outputs from different DAS analysis software that generate data in tabular outputs. Through GeneappExplorer, the bioinformatician can generate dozens of graphs to graphically visualize significant results implicit in the technical tables exported by DAS analysis software.

4.2.1 VISUALIZING AS EVENTS WITH GENEAPP

In fact, by accessing https://gene.mikeias.net/gene?id=836163, one can, for example, identify in silico the PTC caused by AS in the IR event in isoform NM_125434.4 when comparing with NM_203240.3. The implementation was carried out in 12 months. Following the development and testing of the application, it was made available to the community, hosted on Firebase (https://firebase.google.com), with code in a public repository on GitHub (https://github.com/MiqueiasFernandes/GeneAPP). Given the relevance of the software for research involving AS, their registration was carried out at the National Institute of Industrial Property - INPI.

4.2.2 FEATURES IMPLEMENTED AND OUTPUTS GENERATED BY GENEAPP

Geneapp proposes to support the end-to-end bioinformatician in DAS research by extending visualization to other analyses, such as DAS annotation and identification. In this sense, the user can start with the access of the SRA (RUN code) of the samples and finish with the graphics of the event structure and sequencing coverage, tables with statistics, and sequences of the isoforms with marked exons without the need for processing on his computer or in a high-performance computing environment (HPC). GeneappScript was developed in the Colab environment (BISONG, 2019), where this module can perform data download and processing. It is

important to note that this module triggers the processing of the analysis software using the default parameter to facilitate the user. However, the bioinformatician must review them and, if necessary, adjust them before processing the data. The GeneappServer module, in turn, is used by users who have already performed DAS analyses (already processed rMATS, for example) since it integrates, enriches, and filters the data. In the final step of the pipeline (Fig. 13a,b), the user imports the outputs obtained from the previous modules into the GeneappExplorer module to visualize, inspect, curate, and clean the identified AS events. In this last module, the user can also visualize the occurrence of AS among its isoforms only with the gene identifier - GeneID (MAGLOTT *et al.*, 2011) - deposited in NCBI.

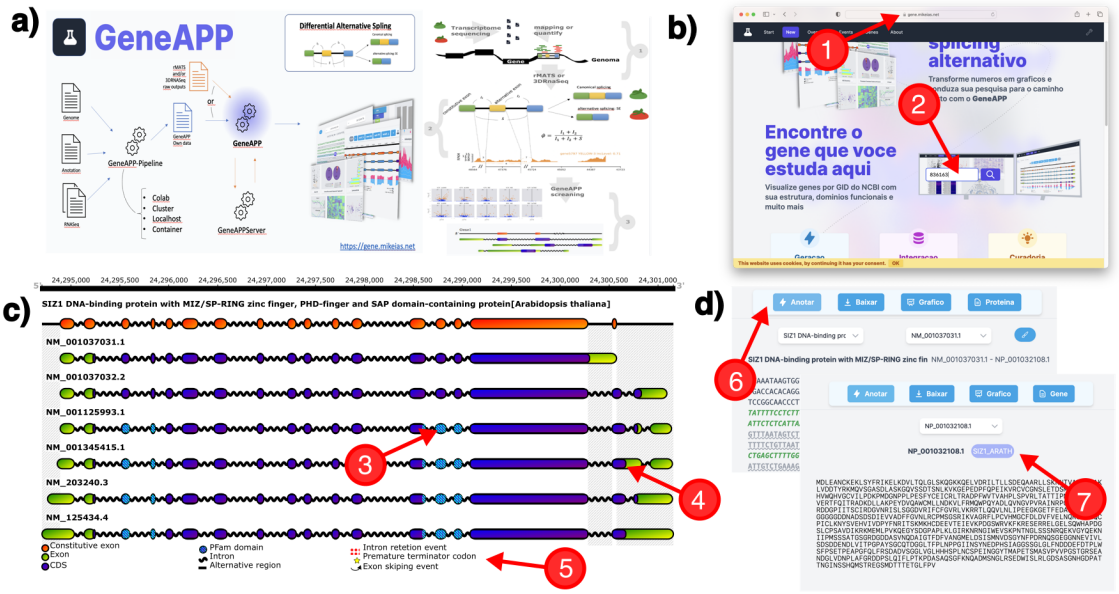


Fig. 13 Overview of the functionalities to visualize alternative splicing implemented in Geneapp. GeneAPP can be used together with the complete pipeline or accessed directly by http://bioinfo.icb.ufmg.br/geneapp, as indicated by arrow 1, to visualize genes with structure available at NCBI through GeneID insert the number in the field indicated by arrow 2. In a) is detailed the pipeline that uses the GeneappScript module to identify genes under DAS with rMATS and 3DRNaseq. The GeneappExplorer module (b) graphically displays the results of the identified DAS since the standard outputs of the identification programs are machine-friendly tables. The gene graph shown in c) displays the gene structure and its alternative regions and isoforms, if available, can be annotated domain family by InterproScan, shown in dashed blue as pointed by arrow 3. The functional annotation of the coding region is performed by the "Annotate" button indicated by arrow 6. The user can access the button at the top of the page displaying the gene diagram. In this example of GeneID 836163, the isoform retains, i.e., does not process an intronic region indicated by arrow 4, which causes a PTC, decreasing the size of the coding region that is colored purple. In the legend generated in the diagram, indicated by arrow 5, it can be seen that Geneapp can annotate AS events. However, it is necessary to load the files generated by GeneappScript. The "Graph" button switches to "Sequence" when selected, allowing the bioinformatician to copy the gene sequence with exon, intron, CDS, and start codon, among other annotated features. It is also possible to visualize the sequence of the translated protein of each isoform by clicking on the "Protein" button. Suppose the protein is annotated when your sequence is displayed. In that case, a link will be generated in the region pointed by arrow 7 that can lead the user to a similar protein with annotation available in UniProtKB. This cross-referencing of data helps the bioinformatician contextualize the AS event with information at the genomic, transcriptomic, and proteomic levels, assisting in the decision-making about whether a candidate target is involved in the biological process analyzed.

Among the analyses performed by the GeneappScript module, we can list: i) configuration of the environment with the installation of programs for mapping and quantification, DAS analysis and annotation with all their dependencies; ii) download of genomic, annotation, and transcriptomic data, iii) data pre-processing iv) mapping v) quantification vi) DAS analysis vii) annotation viii) clustering by sequence. The script is designed to continue the analyses where it left off as "divide and conquer", so the user can run each sample in parallel and then join these samples with one run. At the end of the analysis, the script generates a compressed file named import_geneapp.zip containing the 10 files to be imported into GeneappExplorer. The graphs and tables below explore this DAS analysis using the application developed in this work.

The GeneappServer module has a shell script that runs a Flask api to border the Axios client in GeneappExplorer. There is an instance of the module running on the SAGARANA server of the ICB lab headed by Professor Phd. José Miguel Ortega. At the beginning of the project, the module was provisioned for a queue of 10 users on the server. As the user enters his data, an observer in the module agent performs the necessary processing on the data and provides a link with the results that expire in 7 days to release resources on the server. A feature to evolve in this module is the implemented queuing which can be replaced by automated software allowing better management of the resources needed for network traffic and processing of a more significant number of users. The SAGARANA server is highly available and maintained with UFMG/CAPES resources.

The graph in Fig. 13.c is the main plot of the application and has been made available to display genes without the need to open files so that the user can know the structure of AS of genes available at NCBI. For example, the user can access the app through the link https://gene.mikeias.net/gene/?id=836163 and get a graph as presented in Fig. 13.c below. In this graph, the user can visualize where the AS event occurs and how large it is, the coverage of reads at the event site, and the structure of the gene and its isoforms. In this example, GeneID 39849 has an alternative region marked by the gray dashed line at the bottom, three constitutive exons, and two constitutive introns. It can be easily identified from the figure that the coding region has a distinct size between the isoforms and that the functional domain PF06495 annotated by the InterproScan API is only present in the second isoform that undergoes A3SS. This example does not show the events and reads coverage as per the figure in Figure

25 of 7.28 Supplementary Material 2 due to the user needing to enter the output from the DAS mapping and analysis software. However, it is a helpful graph for quickly visualizing genes and their structures with their functional domains.

GeneappExplorer can import DAS analysis data from GeneappScript or the user's outputs from manual execution, as shown in Fig. 13.a. After importing the files, the user can explore and download tables and graphs listed in Table 2. In this work, these outputs were generated based on the example data, and the tables were made available in 7.27 Supplementary Material 1and the graphs in 7.28 Supplementary Material 2. In the Pipeline_table, the user finds a quantitative view of the data used in the analysis, with metrics of the tool executions by the scripts and general metrics of the results. This table is essential for the user to check if all the steps of the analysis were performed as expected and where there was a failure in the execution if it occurred. In the Gene_table, the user will find a structure like a GFF to obtain machine-friendly access to structural information of genes and mRNAs. In the Annotation_table, it is possible to obtain the functional annotation of the genes containing ontology, InterPro description and identified Pathways. The Events_table contains the list of AS events with details. This table gathers the outputs of rMATS and 3DRNASeq into one. This last table has the columns Gene (which brings the gene identifier); Isoform1 (which brings the canonical isoform: of higher expression in the control sample); Isoform2 (which indicates the alternative isoform: of higher expression in the case sample); ΔPSI (in the case of 3DRNASeq registration the maximum of the identified events will be reported); FDR (false Discovery rate i.e. adjusted statistical significance); Δ Exonic Size (size of the alternative region) ; Δ CDS Size (size of the alternative region in the coding region of the transcript); Log2FC (fold-change transformed by logarithmic scale in base 2); Type (type of event reported by rMATS or inferred by GeneAPP; SE/RI/A3SS/A5SS); Maser (whether significant according to MASER analysis); Event (event identifier by rMATS); min mRNA TPM (smallest TPM calculated by 3DRNASeq for control and case); max mRNA TPM (similar to min mRNA TPM); min Protein Size (size of the most minor protein encoded by the gene); PTC (if PTC was identified in case of IR); etc; are presented.

Table 2 GeneAPP generated files available for download. The cited files are available in 7.27 Supplementary Material 1 (tables) and 7.28 Supplementary Material 2 (graphs), where the example data generated them.

Type	Archives	Description
Table	Experiment_table.csv Pipeline_table.csv Gene_table.csv Annotation_table.csv Events_table.csv	The experiment, gene, and pipeline tables are generated with the output data from the analyses. The annotation and event tables are generated by cross-referencing the annotation data with the DAS analysis. The user uses these tables to find specific information about a certain event or gene.
Descriptive graph	graphQc.svg graphRd.svg graphGc.svg graphUp.svg graphCv.svg graphAs.svg graphVen.svg graphMp.svg	These graphs are generated to give the user a global descriptive view of the analysis. In the funnel graph, the user can identify at which step of the pipeline the genes are retained. When GeneAPPScript obtains the results, the user can use these graphs to check: the quantity and size of total reads and quality of each sample, mapping in the genome, genes and transcripts, size and quantity of genes, transcripts, exons, introns, etc. It is also generated up SetPlot to check the number of genes expressed in each condition, the graph of sequencing coverage area in genes, the Venn diagram of DEG and DAS genes, and DAS of rMATS and 3DRNASeq.
Analysis chart	graphBar.svg graphTop.svg graphPie.svg graphCov.svg	These graphs have a more in-depth group view of DAS events. A bar chart shows how many genes were identified in DAS by event type and another by most relevant ΔPSI. About IR events, a pie chart is

	graphDea.svg graphFi.svg	presented with the PTC frequency slices in each retained region. For A3SS and A5SS, a line graph is generated of the coverage of reads in the upstream and downstream regions to the point where the event occurs. If the user has performed analysis by GeneAPPScript, a dendrogram of DAS genes based on their sequence similarity will be generated. A butterfly plot is also generated that relates DEG (log2FC) with DAS (ΔPSI) on the two Axes where each point is a gene, the size of the point is the impact of the DAS event, and the color is the statistical significance.
Volcanoplot	graphDe.svg graphScr.svg graphDa.svg graphDar.svg	Traditional volcano plots for differential expression analysis are generated. DAS events are plotted with shapes according to their type and size according to the event's impact on the gene.
Expression heatmap	graphHm.svg graphHm2.svg	In these graphs, the user can check the expression of each gene in each sample and of each transcript in each sample according to the data generated by 3DRNASeq.
Functional annotation chart	graphAn.svg graphLk.svg graphTr.svg graphWc.svg	These graphs help the user to understand the data according to the functional annotation.

AS event structure graph	gene-****.svg	This is the primary graphic of the app. It displays the gene structure with canonical and alternative regions, each isoform with annotated structural and protein domain regions, part of the gene in genomic context as miRNA, and the sequencing coverage in each sample. The expression of the gene and isoforms can be visualized in a summarized way concerning the gene or isoforms by the colored circle next to their name. The alternative regions have a dashed background, and the events are drawn at the coordinates identified by rMATS. In the case of an ES event, an arrow will be drawn indicating the alternative exon, and in the case of PTC found in an IR event, and a star will be noted.

Geneapp provides the user with integration with databases in some functional annotation figures through links or access codes of these databases. The Interpro database (PAYSAN-LAFOSSE *et al.*, 2023) provides annotations obtained by functional analysis of proteins by classifying them into families and predicting important sites and domains. The records of the PFam database (EL-GEBALI *et al.*, 2019) have grouped proteins into families determined by alignments using Markov chain models (HMM). The annotations of RefSeq (O'LEARY *et al.*, 2016) provide a comprehensive, integrated, and non-redundant collection of functionally annotated proteins by NCBI. The UniProtKB/Swiss-Prot bank (BATEMAN *et al.*, 2023) contains functional annotations obtained from the manual biocuration of computer-annotated records that often assign functional annotations by sequence similarity and database association. The Gene Ontology (GO) knowledge base is one of the largest sources of gene function information (CONSORTIUM *et al.*, 2023). The GO allows organ research through relationships between records by biological process, cellular components, and

molecular function. GeneAPP also integrates with other databases, such as SRA (sequence read the archive, part of NCBI), to obtain samples before annotation. In addition to databases, it also integrates with other web services such as, for example, to perform gene clustering based on sequence similarity, using the https://www.genome.jp/tools-bin/ete web service available on the GenomeNet infrastructure, which has several bioinformatics tools.

4.2.3 TUTORIAL FOR THE PRIMARY USE OF THE WEBAPP

The latest versions of the main browsers: Chrome, Firefox, and Safari, are allowed for use by GeneappExplorer. If they are running a compiled app on their machine, users with internet access can directly access the link https://gene.mikeias.net or http://localhost:8080. To compile on your machine, the user can clone the repository and compile their version or run a container from the docker image using the command "docker pull mikeiasfernandes93/geneapp," which provides an image that even has GeneappScript integrated. GeneappExplorer has also been tested in mobile browsers, as shown in Figure 26 of 7.28 Supplementary Material 2, where the intron indicated by the blue arrow participates in an IR event generating PTC pointed by the red arrow. The graphical user interface (GUI) provides a good usability experience by using responsive components of the Tailwind framework (https://tailwindcss.com).

Data import in Geneapp is simple since, to determine the transcriptomic profile, the user can inform the RUN access code that can be obtained by screening the experimental design in the Run Selector tool (https://www.ncbi.nlm.nih.gov/Traces/study/). Another important tip is that the scripts are optimized for NCBI file formats, so all the necessary ones can be accessed on the RefSeq page of the reference genome, example for humans is available at https://www.ncbi.nlm.nih.gov/genome/51?genome_assembly_id=1820449. In case of the need to import the data from another location, such as the cloud or Google Drive, for example, the user can edit GeneappScript to get the data from these locations as well. The format of the files from other software than rMATS and 3DRNASeq can be parsed to the format most adherent to rMATS available in the documentation https://github.com/Xinglab/rmats-turbo or 3Drnaseq defined from the format available at

https://github.com/wyguo/ThreeDRNAseq/blob/master/vignettes/user_manuals/3D_R

NA-seq_App_manual.md. If the mapping has already been performed, the "bam" file can be used directly by rMATS to identify the DAS or by Salmon (PATRO *et al.*, 2017) to quantify it without further processing. For GeneappServer, the data can be partially imported so that the user can build his dataset in the expected input pattern, which will automatically be validated for the expected format, adding missing data when possible. In the standard flow using the three Geneapp modules, the compressed files (zip) generated by GeneappScript contain the tables of DAS analysis and annotation, and the GeneappServer zip contains these integrated, filtered, annotated, and grouped files. In the case of gene visualization directly by GeneappExplorer, the necessary data are automatically imported by the NCBI and InterproScan API, requiring only that the user enter his email to use the APIs.

The files exported by Geneapp are simple. In the case of GeneappScript will be included quantitative data of the events identified and the logs of the executions of the programs. In this module, rMATS and 3DRNASeq analyses generate raw tabular outputs to perform functional annotation and cluster DAS genes. In the case of GeneappServer, the imported data are integrated by adding data for missing inputs for full use of GeneappExplorer. This module acts to filter the dataset and format it to divide it into smaller files facilitating the import into GeneappExplorer. GeneappExplorer outputs can be obtained through the "download" button at the app’s top or next to each graph or table. The images generated in this module are in svg format (vector), tables in csv format (columns separated by commas), and sequences in RichText format (rtf). The sequences can be used to design primers in specific regions for validation.

4.3 DISCUSSION

In RNA-Seq analyses, the transcript count, which can be performed by the metric TPM (ZHAO; YE; STANTON, 2020), can be used to determine gene expression. But before that, RNAseq mapping in the genome makes it possible to identify AS events (SHEN *et al.*, 2014), Transcript counts can indicate the differential use of isoforms (GUO, WENBIN *et al.*, 2021), or transcriptomic profiling (ZHANG, ZHONG FA *et al.*, 2013) which are additional to gene-level insights. As the cost of RNASeq sampling may make it unfeasible to sequence deeply enough to compare the expression of all isoforms, some studies whose samples are deposited in the SRA performed only the gene level verification since the sequencing depth of the gene is more significant since it is the sum of the expression of all isoforms. Since the Geneapp pipeline allows the bioinformatician to measure the sequencing coverage of each isoform, these public data may now be analyzed at the gene and isoform level, allowing to extract of even more knowledge from these studies by adding DAS analysis to DEG analysis. It is worth noting that it is necessary to perform experimental validation with PCR (to confirm the presence of alternative isoforms), qPCR (to quantitatively confirm the expression of alternative isoforms in the samples with statistical support), Sanger (to confirm the sequence of alternative isoforms) to validate the relationship of these changes in gene or isoform expression with the biological process under investigation.

The researcher can use the same RNA-Seq as the DEG analysis to study DAS. However, one must pay attention to the depth of the sequencing and the additional analyses to visualize the AS events structure. In developing DAS research to understand the ripening process of coffee beans, it was necessary to develop some scripts to filter and format the output files of rMATS and visualize the event's structure and data of the results. In a second work, we applied another pipeline, using quantification with Salmon and 3DRnaseq, which generated results in a different format than the previous one, requiring the implementation of new scripts for this new approach. The need to visualize the gene with mapping of reads, the structure of the event, and family protein domains annotated in the coding region added to what occurred in previous works justified the need for the development of Geneapp. Our research uses Geneapp in studies using orange data, already with results and grape data in the initial analysis. Now, with Geneapp, it will be possible to visualize in the same pipeline the two approaches allowing a more robust curation of DAS events,

whose effect is a more concise and accurate set of genes and isoforms regarding the impact of the biological process.

Thus, Geneapp was developed to integrate the same graph with different approaches to DAS analysis. As a result of this goal, several functions have been implemented to plot different graphs, allowing us to visualize the results by gene level or isoforms at rMATS or 3DRNASeq angle. Some graphs allow us to visualize the genes or isoforms or events in a group, like the funnel of Fig. 14.a which presents the group of genes flowing along the pipeline; in isoform group as the graph 28 in 7.28 Supplementary Material 2 generated by 3DRNASeq during GeneappScript execution; or even as in the graph in Fig. 14.e which shows the group of events by their type. Other graphs show each of these three in individual points on the same plot as the graph in Fig. 14.f where each node is a gene and the relationship between them shared annotation; in Fig. 14.d, which presents in the heatmap, the expression of each isoform in the samples; also in the case of individual events there is the Fig. 14.c which shows each identified event having on the axes PSI and FDR where the shapes ▲ means IR event, ■ are SE events and ● A3SS or A5SS events colored by significance, orange being the statistically significant ones. At the maximum level of granularity, the bioinformatician can visualize the gene, transcript, or event individually, as shown in the graph in Fig. 14.g or in Fig. 14.b, where each line is where the AS event occurs in the gene. The graphs in Fig. 14 and others are available in detail in 7.28 Supplementary Material 2, along with others that are variations of the examples cited.

The user can further understand their analysis with supporting graphs, such as these examples from 7.28 Supplementary Material 2: graph 4 which shows the amount of genes with TPM > 0 in each sample; graph 5 which shows the depth of reads along the normalized genes per sample; graph 21 which shows the amount of genes and annotations identified; graph 1 which shows the quality control of the samples with the inner bar being the % amount that did not pass quality control, the darker outer bar the amount of reads in the sample and the lighter one the average size of the reads in the sample both normalized; graph 3 which shows the % mapping of the reads in each sample in the genome, transcripts and genes with annotated AS; graph 14 that shows a grouping as an essential phylogeny to guide the user in groups of genes by the similarity of genomic sequence; graph 22 that shows words that appear more frequently in the annotation, the largest and darkest being the most frequent; graph 19

that shows the TPM of the gene in each sample; graph 23 that shows the total amount of amino acid of an Interpro annotation in the DAS proteome; graph 17 that presents the classic differential expression volcano plot with 3RNASeq data with DAS genes colored in red; among others.

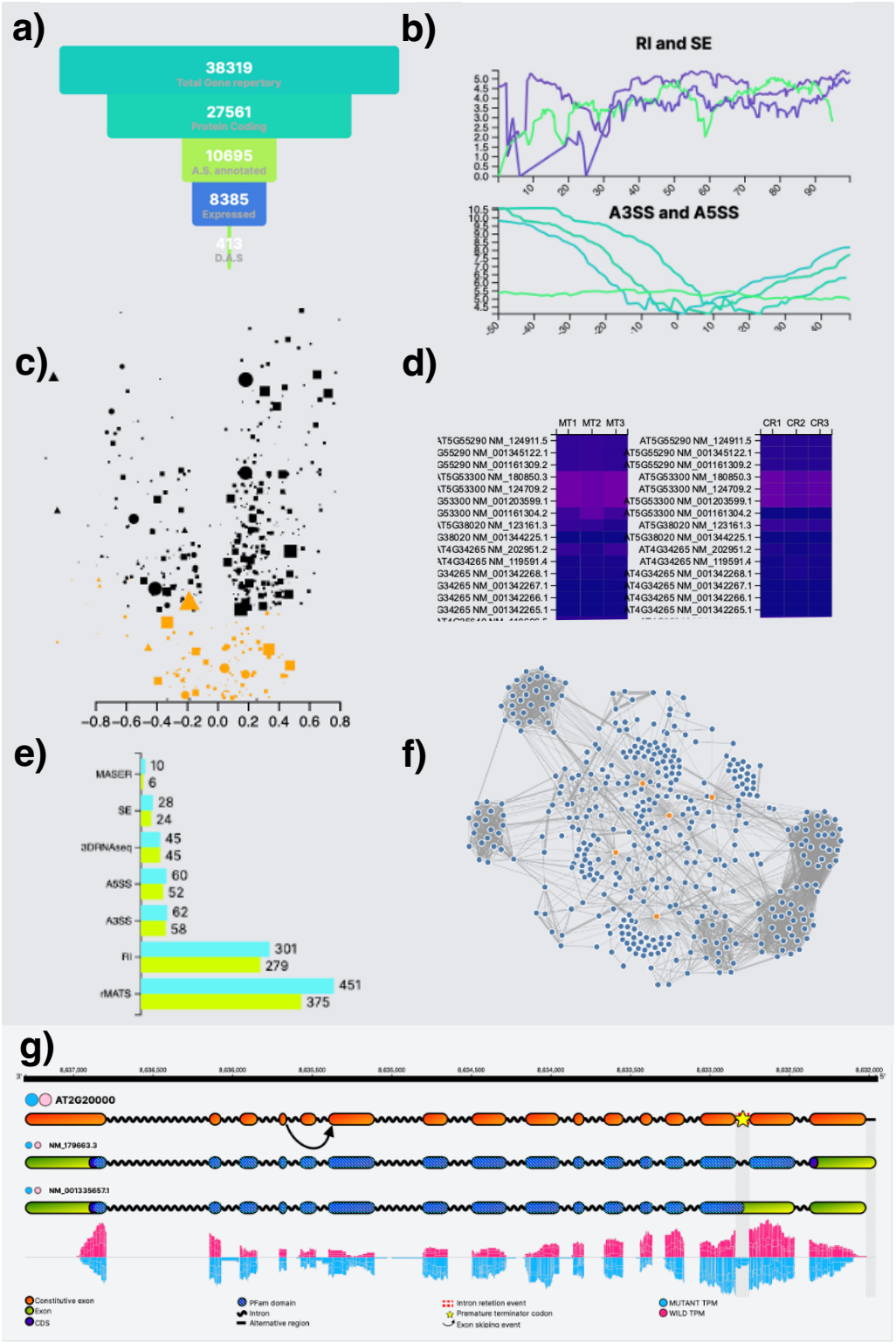


Fig. 14 Graphs generated by GeneAPP to assist the bioinformatician in DAS analysis. In the funnel of the figure in (a), the bioinformatician has access to the quantitative of the series of cleaning of the set of genes performed by GeneappScript. The result refers to the example analysis with accessions listed in 7.27 Supplementary Material 1. The lines of the upper graph in (b) show the coverage of reads along the alternative exon or intron of the genes that were selected to present in the figure. In the case of the inferred graph, the central point on the x-axis is the beginning of the alternative region, which in the case of A3SS and A5SS always occurs after an intron, generating the valley in the center of the graph. In (c) all DAS events identified by rMATS are plotted, the orange shapes being the events with FDR <0.005, the triangles RI events, the squares SE events, and the circles A3SS and A5SS events, the x-axis is the PSI. The heatmap (d) shows the TPM calculated by Salmon and provided by 3DRNaseq in the analyzed replicates grouped by control (CR1, CR2, and CR3) and treatment (MT1, MT2, MT3). In (e), the bioinformatician can find the number of genes in green and events in blue for the type of analysis or event, and MASER always presents itself as more restrictive by restricting the coverage of reads in the gene leading to the result with the support of more evidence of RNASeq. GeneappExplorer presents two groupings: the first as a circular dendrogram, according to the homology of the sequences by the analysis of GeneappScript (shown in Figure 14 of 7.28 Supplementary Material 2), and another based on the functional and structural annotation in the form of a graph that is shown in (f). In the graph, the nodes in orange are chromosomes, and in blue are the genes, the vertices between them refer to functional annotation sharing of the gene ontology or functional domain of a specific protein family, and the thickness of the edge presents the amount of annotation shared between the genes. Graph (g) shows the gene structure with annotated structural, functional, TPM, read coverage, and AS events. A further 18 types of graphs can be plotted, which are not shown in this figure due to space limitations. Disregard the numbers in the figure as the purpose of this figure is not to analyze the numbers but rather the meanings of the graphs. These graphs are available in 7.28 Supplementary Material 2 in detail.

Geneapp was designed to integrate data from specific formats of each software to facilitate the visualization of data in different formats in the graph of the software Fig. 14.g. GeneappServer, on the other hand, was developed using flexible input to allow integration with tables generated by other DAS analyzers. On the GeneappScript side, the bioinformatician can integrate other tools using, for example, the bam files generated by another mapping software of his choice or edit the shell script to invoke the execution of this other software. These integrations with other software allow you to extend your application. Another way to extend the scope of Geneapp is by providing links in the graphs where the user, as shown in Figure 23 of 7.28 Supplementary

Material 2, is taken to other databases where they can access more information. Thus, the purpose of Geneapp can be extended beyond the visualization of DAS results, such as the visualization of gene structure or the exploration of annotations of the dataset under study.

Geneapp was developed using free software to the maximum and is made up of libraries that have an open-source license. Gimp (https://www.gimp.org), Inkscape (https://inkscape.org), and VSCode (https://code.visualstudio.com) under Debian (https://www.debian.org) were part of the toolbox for APP development. The Vue framework (https://vuejs.org), the D3.Js data visualization library (https://d3js.org), and the Axios API client (https://axios-http.com) were used on the front and the Flask framework (https://flask.palletsprojects.com/en/latest) on the backend. In addition to these components, GeneappScript generates commands to execute, for example, STAR (DOBIN *et al.*, 2013), rMATS, HISAT (KIM, DAEHWAN *et al.*, 2019), 3DRNASeq and Salmon, which are software adhering to the open-source initiative or free for non-commercial use. In this sense, even with registered copyrights that restrict your use only for non-commercial purposes, Geneapp is available in the public repository https://github.com/MiqueiasFernandes/GeneAPP with open-source code. This strategy allows the Geneapp code to assist research with limited access to resources to be carried out.

The work cited as prompting the development of Geneapp was carried out to study AS events involved in the biological process of coffee bean ripening. To do this, we conducted a DAS study using RNASeq data obtained for differential expression experiment - these were generated by Cheng and collaborators in 2018 and published in SRA - and the genome made available by the WORLD COFFEE RESEARCH group (SCALABRIN *et al.*, 2020). In the in-silico analysis, we developed pipelines in Python to perform an automated curation of the 202 target genes identified with 241 DAS events by rMATS in the comparisons: green with the intermediate stage of ripeness (yellow) and ripe with yellow. We then performed a manual curation of the 241 events enriched with Interproscan5 annotation, which led to the interest in experimentally validating the Potassium channel AKT1 and Apyrase 7 genes with PCR and qPCR. After this work and with Geneapp, we are working with tomatoes in experimental validation. However, there are still studies with grape and orange data more advanced than others on the list of fruit RNAseq with data from experiments deposited in the SRA

suitable for DAS analysis to study the biological process of ripening by the transcriptomic profile of green and ripe fruit samples.

Geneapp was reviewed by three bioinformaticians who tested the operation and made suggestions to evolve the application. These notes constitute the project backlog that aims to benefit future work with the app. Among the corrective and incremental suggestions can be mentioned: i) Place a carousel component on the home page to display Geneapp information; ii) Create a download page to download the test files; iii) Host the download files on Github or similar; iv) Create pdf documentation explaining the results; v) Revise the texts in English; vi) Review responsive layout version for mobile devices with small screens; vii) Translate texts into another language; viii) Describe what the application does on the "Start" page; ix) Provide a flowchart with each step implemented in Geneapp to guide the user through the steps; x) Describe in detail each step of GeneappScript data acquisition, quality control, mapping, expression, DAS analysis; xi) Hide page names in the top navigation bar when unavailable xii) Make GeneappServer available in Docker; xiii) Homogenize GeneappServer in HPC/Cluster; xiv) Provide test files for other model organisms such as humans and fly; xv) Provide documentation to facilitate user execution of doker script; xvi) Adjust GeneAPP also to display Single-exon coding sequences (CDSs), and non-coding mRNAs; xvii) Provide a PayPal button for donations to support the project; xviii) Better explain the "divide and conquer" strategy on the homepage; xix) Increase the size of the legend displayed in the graphs; xx) Develop native import of DAS analysis outputs performed in other software xxi) Inform TPM metric through the y-axis in the individual gene graphs xxii) Implement functionality to select and name genes of interest in the graphs; xxiii) Provide a download button to download the list of elements (genes/transcripts/events) that appear in the graphs xxiv) Provide a button to download the list of applied filters; xxv) highlight the DAS event region on the page that displays the gene sequence. Even though this list is not exhaustive for points to work on Geneapp, the software in the current situation was thought of as MVP (minimum viable product), so it already allows to find, analyze, and visualize AS events in genes, for example, by accessing https://gene.mikeias.net/gene/?id=39849, as already discussed.

4.4 CONCLUSION

In this work, we present Geneapp, a suite of 3 modules: GeneappScript, GeneappServer, and GeneappExplorer, that can be used to study the differences in transcriptomic profiles at the isoform level. We show how easy it is to visualize the gene to identify AS events and explore results from DAS analyses obtained from rMATS and 3DRNAseq software. Geneapp facilitates data integration across multiple omics levels and biological databases. From the graphs generated by Geneapp, the bioinformatician discovers that not only are the numbers interesting but contextualizing them makes the result more confident. Finally, we list some expected evolution needs of the Geneapp to make it even more useful for transcriptome and bioinformatics analysis.

5. CONCLUSÃO E PERSPECTIVAS

Conforme análise in sílico realizada por Yan e colaboradores (2021), este trabalho encontrou AS ocorrendo durante o processo biológico de amadurecimento de frutos. O experimento com café corroborou ao identificar 241 eventos relacionados com amadurecimento. Os eventos de SE e RI confirmados na validação experimental com PCR indicam a necessidade de explorar mais dados públicos que possuem apenas análise transcriptômica em nível gênico. Na análise com tomate uma abordagem totalmente nova foi realizada para identificar DAS, oque permitiu levantar insights induzindo o desenvolvimento de ferramenta para integração de dados de AS.

Se por um lado a realização do experimento DAS em café e tomate buscou entender a biologia do amadurecimento, por outro as necessidades de automatizar o pipeline e integrar os resultados das diferentes abordagens despertou a contribuição computacional deste trabalho. O webapp desenvolvido para auxiliar na exploração de dados de AS apresenta-se como uma ferramenta para visualizar genes com foco em dados de DAS, diferenciando-o de outros softwares de visualização de dados biológicos. Nele há ainda implícito uma tentativa de compartilhar conceitos recentes da computação para a bioinformática; como framework Vue no aplicativo JS, visualização de dados com D3 e Tailwind no CSS. O GeneAPP evidenciou que a informática tem muito a apoiar na biologia, auxiliando no tratamento de dados para respaldar resposta a muitas perguntas que estão em aberto sobre os organismos e suas complexidades.

Mesmo que esta tese se apresente insipiente para esclarecer por completo o processo biológico de amadurecimento de frutos ela é vantajosa por contribuir no entendimento do processo no escopo de AS para café e tomate. Doutro modo, a experiência adquirida nesse trabalho é suficiente para conduzir outro experimento de estudo da expressão diferencial de isoformas de um conjunto de genes (DAS) nos moldes da metodologia científica, de modo a viabilizar a realização de pesquisa em outros cenários biológicos.

6. REFERÊNCIAS

ADUSUMALLI, Swarnaseetha *et al.* Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. *Aging Cell*, v. 18, n. 3, p. e12928, 1 jun. 2019. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/acel.12928>. Acesso em: 29 maio 2023.

ALEXANDER, Lucille; GRIERSON, Donald. 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. Disponível em: <https://academic.oup.com/jxb/article/53/377/2039/497226>. Acesso em: 15 jul. 2022.

AMMARI, Mais *et al.* Biocuration: Distilling data into knowledge. *PLoS Biology*, v. 16, n. 4, 16 abr. 2018. Disponível em: </pmc/articles/PMC5919672/>. Acesso em: 29 maio 2023.

ANTHONY, F. *et al.* The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics 2002 104:5*, v. 104, n. 5, p. 894–900, 2002. Disponível em: <https://link.springer.com/article/10.1007/s00122-001-0798-8>. Acesso em: 15 jul. 2022.

BAHARLOU HOUREH, Mandana *et al.* SpliceDetector: a software for detection of alternative splicing events in human and model organisms directly from transcript IDs. *Scientific Reports 2018 8:1*, v. 8, n. 1, p. 1–11, 22 mar. 2018. Disponível em: <https://www.nature.com/articles/s41598-018-23245-1>. Acesso em: 11 jul. 2023.

BAKER, Kristian E.; PARKER, Roy. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current Opinion in Cell Biology*, v. 16, n. 3, p. 293–299, 1 jun. 2004. Acesso em: 16 jul. 2022.

BAPAT, Vishwas A. *et al.* Ripening of fleshy fruit: Molecular insight and the role of ethylene. *Biotechnology Advances*, v. 28, n. 1, p. 94–107, 1 jan. 2010. Acesso em: 26 jun. 2023.

BARALLE, Francisco E.; GIUDICE, Jimena. Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology 2017 18:7*, v. 18, n. 7, p. 437–451, 10 maio 2017. Disponível em: <https://www.nature.com/articles/nrm.2017.27>. Acesso em: 15 jul. 2022.

BARANN, Matthias; ZIMMER, Ralf; BIRZELE, Fabian. Mananggal - a novel viewer for alternative splicing events. *BMC Bioinformatics*, v. 18, n. 1, p. 1–13, 21 fev. 2017. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1548-5>. Acesso em: 29 maio 2023.

BARASH, Yoseph *et al.* Deciphering the splicing code. *Nature 2010 465:7294*, v. 465, n. 7294, p. 53–59, 6 maio 2010. Disponível em: <https://www.nature.com/articles/nature09000>. Acesso em: 9 jul. 2023.

BATEMAN, Alex. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, v. 47, n. D1, p. D506–D515, 8 jan. 2019. Disponível em: <https://academic.oup.com/nar/article/47/D1/D506/5160987>. Acesso em: 15 jul. 2022.

BATEMAN, Alex *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, v. 51, n. D1, p. D523–D531, 6 jan. 2023. Disponível em: <https://academic.oup.com/nar/article/51/D1/D523/6835362>. Acesso em: 29 maio 2023.

BEDRE, Renesh *et al.* New era in plant alternative splicing analysis enabled by advances in high-throughput sequencing (HTS) technologies. *Frontiers in Plant Science*, v. 10, p. 740, 31 maio 2019. Acesso em: 15 jul. 2022.

BISONG, Ekaba. Google Colaboratory. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, p. 59–64, 2019. Disponível em: <https://link.springer.com/chapter/10.1007/978-1-4842-4470-8_7>. Acesso em: 29 maio 2023.

BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114–2120, 1 ago. 2014. Disponível em: <https://academic.oup.com/bioinformatics/article/30/15/2114/2390096>. Acesso em: 15 jul. 2022.

BOLGER, Marie E.; ARSOVA, Borjana; USADEL, Björn. Plant genome and transcriptome annotations: from misconceptions to simple solutions. *Briefings in Bioinformatics*, v. 19, n. 3, p. 437–449, 1 maio 2018. Disponível em: <https://academic.oup.com/bib/article/19/3/437/2843630>. Acesso em: 16 jul. 2022.

BONNAL, Sophie C.; LÓPEZ-OREJA, Irene; VALCÁRCEL, Juan. Roles and mechanisms of alternative splicing in cancer — implications for care. *Nature Reviews Clinical Oncology 2020 17:8*, v. 17, n. 8, p. 457–474, 17 abr. 2020. Disponível em: <https://www.nature.com/articles/s41571-020-0350-x>. Acesso em: 29 maio 2023.

BOSE, Santosh Kumar *et al.* Oligosaccharide is a promising natural preservative for improving postharvest preservation of fruit: A review. *Food Chemistry*, v. 341, p. 128178, 30 mar. 2021. Acesso em: 26 jun. 2023.

BRAUNSCHWEIG, Ulrich *et al.* Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell*, v. 152, n. 6, p. 1252–1269, 14 mar. 2013. Disponível em: <http://www.cell.com/article/S0092867413002663/fulltext>. Acesso em: 9 jul. 2023.

BROGNA, Saverio; WEN, Jikai. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology 2009 16:2*, v. 16, n. 2, p. 107–113, fev. 2009. Disponível em: <https://www.nature.com/articles/nsmb.1550>. Acesso em: 14 jul. 2023.

BROWN, Joseph; PIRRUNG, Meg; MCCUE, Lee Ann. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, v. 33, n. 19, p. 3137–3139, 1 out. 2017. Disponível em: <https://academic.oup.com/bioinformatics/article/33/19/3137/3865781>. Acesso em: 15 jul. 2022.

BYRNE, David H. Trends in fruit breeding. *Fruit Breeding*, p. 3–36, 1 jan. 2012. Disponível em: <https://link.springer.com/chapter/10.1007/978-1-4419-0763-9_1>. Acesso em: 26 jun. 2023.

CAMACHO, Christiam *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics*, v. 10, n. 1, p. 1–9, 15 dez. 2009. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-421>. Acesso em: 15 jul. 2022.

CENCI, Alberto; COMBES, Marie Christine; LASHERMES, Philippe. Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. *Plant Molecular Biology*, v. 78, n. 1–2, p. 135–145, 16 jan. 2012. Disponível em: <https://link.springer.com/article/10.1007/s11103-011-9852-3>. Acesso em: 15 jul. 2022.

CERRI, Martina; REALE, Lara. Anatomical traits of the principal fruits: An overview. *Scientia Horticulturae*, v. 270, p. 109390, 25 ago. 2020. Acesso em: 26 jun. 2023.

CHAUDHARY, Saurabh; KHOKHAR, Waqas; *et al.* Alternative splicing and protein diversity: Plants versus animals. *Frontiers in Plant Science*, v. 10, p. 708, 31 maio 2019. Acesso em: 15 jul. 2022.

CHAUDHARY, Saurabh; JABRE, Ibtissam; *et al.* Perspective on Alternative Splicing and Proteome Complexity in Plants. *Trends in Plant Science*, v. 24, n. 6, p. 496–506, 1 jun. 2019. Acesso em: 15 jul. 2022.

CHEN, Tong; QIN, Guozheng; TIAN, Shiping. Regulatory network of fruit ripening: current understanding and future challenges. *New Phytologist*, v. 228, n. 4, p. 1219–1226, 1 nov. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/nph.16822>. Acesso em: 26 jun. 2023.

CHENG, Bing *et al.* Influence of genotype and environment on coffee quality. *Trends in Food Science & Technology*, v. 57, p. 20–30, 1 nov. 2016. Acesso em: 15 jul. 2022.

CHENG, Bing *et al.* Slower development of lower canopy beans produces better coffee. *Journal of Experimental Botany*, v. 71, n. 14, p. 4201–4214, 6 jul. 2020. Disponível em: <https://academic.oup.com/jxb/article/71/14/4201/5811182>. Acesso em: 15 jul. 2022.

CHENG, Bing; FURTADO, Agnelo; HENRY, Robert J. The coffee bean transcriptome explains the accumulation of the major bean components through ripening. *Scientific Reports 2018 8:1*, v. 8, n. 1, p. 1–11, 30 jul. 2018. Disponível em: <https://www.nature.com/articles/s41598-018-29842-4>. Acesso em: 10 jul. 2022.

CLARK, Greg *et al.* Recent Advances Clarifying the Structure and Function of Plant Apyrases (Nucleoside Triphosphate Diphosphohydrolases). *International Journal of Molecular Sciences 2021, Vol. 22, Page 3283*, v. 22, n. 6, p. 3283, 23 mar. 2021. Disponível em: <https://www.mdpi.com/1422-0067/22/6/3283/htm>. Acesso em: 5 ago. 2022.

CLARK, Sarah *et al.* Expanding alternative splicing identification by integrating multiple sources of transcription data in tomato. *Frontiers in Plant Science*, v. 10, p. 689, 31 maio 2019. Acesso em: 15 jul. 2022.

CLEMENTE, Junia Maria *et al.* Effects of nitrogen and potassium on the chemical composition of coffee beans and on beverage quality. *Acta Scientiarum. Agronomy*, v. 37, n. 3, p. 297–305, 1 jul. 2015. Disponível em: <http://www.scielo.br/j/asagr/a/b69xxqtfwHwVDZQSPSXZDmk>. Acesso em: 4 ago. 2022.

COCK, Peter J.A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, v. 25, n. 11, p. 1422–1423, 1 jun. 2009. Disponível em: <https://academic.oup.com/bioinformatics/article/25/11/1422/330687>. Acesso em: 15 jul. 2022.

COLANERO, Sara *et al.* Alternative Splicing in the Anthocyanin Fruit Gene Encoding an R2R3 MYB Transcription Factor Affects Anthocyanin Biosynthesis in Tomato Fruits. *Plant Communications*, v. 1, n. 1, p. 100006, 13 jan. 2020. Acesso em: 11 jul. 2023.

COMMUNITY, The Galaxy *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, v. 50, n. W1, p. W345–W351, 5 jul. 2022. Disponível em: <https://academic.oup.com/nar/article/50/W1/W345/6572001>. Acesso em: 15 jul. 2022.

CONAB. *Tomate: Análise dos Indicadores da Produção e Comercialização no Mercado Mundial, Brasileiro e Catarinense*. . [S.l: s.n.], 2019. Disponível em: <http://www.conab.gov.br>.

CONESA, Ana *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biology 2016 17:1*, v. 17, n. 1, p. 1–19, 26 jan. 2016. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0881-8>. Acesso em: 29 maio 2023.

CONSORTIUM, The Gene Ontology *et al.* The Gene Ontology knowledgebase in 2023. *Genetics*, v. 224, n. 1, 4 maio 2023. Disponível em:

<https://academic.oup.com/genetics/article/224/1/iyad031/7068118>. Acesso em: 29 maio 2023.

CRUZ, Fernanda *et al.* Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. *Molecular Breeding*, v. 23, n. 4, p. 607–616, 3 maio 2009. Disponível em: <https://link.springer.com/article/10.1007/s11032-009-9259-x>. Acesso em: 13 set. 2022.

CUÉLLAR, Teresa *et al.* Potassium transport in developing fleshy fruits: the grapevine inward K+ channel VvK1.2 is activated by CIPK–CBL complexes and induced in ripening berry flesh cells. *The Plant Journal*, v. 73, n. 6, p. 1006–1018, 1 mar. 2013. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.12092>. Acesso em: 4 ago. 2022.

DALE, Ryan *et al.* Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nature Methods 2018 15:7*, v. 15, n. 7, p. 475–476, 2 jul. 2018. Disponível em: <https://www.nature.com/articles/s41592-018-0046-7>. Acesso em: 15 jul. 2022.

DANTAS, Luíza L.B. *et al.* Alternative Splicing of Circadian Clock Genes Correlates With Temperature in Field-Grown Sugarcane. *Frontiers in Plant Science*, v. 10, p. 488063, 23 dez. 2019. Acesso em: 9 jul. 2023.

DE OLIVEIRA MENDES, Tiago Antônio *et al.* Identification of Strain-Specific B-cell Epitopes in Trypanosoma cruzi Using Genome-Scale Epitope Prediction and High-Throughput Immunoscreening with Peptide Arrays. *PLOS Neglected Tropical Diseases*, v. 7, n. 10, p. e2524, 2013. Disponível em: <https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002524>. Acesso em: 30 jul. 2022.

DE SOUZA ROLIM, Glauco *et al.* Climate and natural quality of Coffea arabica L. drink. *Theoretical and Applied Climatology*, v. 141, n. 1–2, p. 87–98, 1 jul. 2020. Disponível em: <https://link.springer.com/article/10.1007/s00704-020-03117-3>. Acesso em: 15 jul. 2022.

DENG, Nan *et al.* A full-length transcriptome and gene expression analysis reveal genes and molecular elements expressed during seed development in Gnetum luofuense. *BMC Plant Biology*, v. 20, n. 1, p. 1–13, 1 dez. 2020. Disponível em:

<https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-020-02729-1>. Acesso em: 11 jul. 2023.

DENNISON, K. L. *et al.* Functions of AKT1 and AKT2 potassium channels determined by studies of single and double mutants of Arabidopsis. *Plant Physiology*, v. 127, n. 3, p. 1012–1019, 2001. Disponível em: </pmc/articles/PMC129271/>. Acesso em: 4 ago. 2022.

DENTI, Luca *et al.* ASGAL: Aligning RNA-Seq data to a splicing graph to detect novel alternative splicing events. *BMC Bioinformatics*, v. 19, n. 1, p. 1–21, 20 nov. 2018. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2436-3>. Acesso em: 11 jul. 2023.

DÍAZ RICCI, Juan C. *et al.* Editorial: Interplay Between Fungal Pathogens and Fruit Ripening. *Frontiers in Plant Science*, v. 11, p. 529163, 24 mar. 2020. Acesso em: 27 jun. 2023.

DING, Fangyuan; ELOWITZ, Michael B. Constitutive splicing and economies of scale in gene expression. *Nature Structural & Molecular Biology 2019 26:6*, v. 26, n. 6, p. 424–432, 27 maio 2019. Disponível em: <https://www.nature.com/articles/s41594-019-0226-x>. Acesso em: 9 jul. 2023.

DOBIN, Alexander *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, v. 29, n. 1, p. 15–21, 1 jan. 2013. Disponível em: <https://academic.oup.com/bioinformatics/article/29/1/15/272537>. Acesso em: 4 jun. 2023.

EL-GEBALI, Sara *et al.* The Pfam protein families database in 2019. *Nucleic Acids Research*, v. 47, n. D1, p. D427–D432, 8 jan. 2019. Disponível em: <https://academic.oup.com/nar/article/47/D1/D427/5144153>. Acesso em: 15 jul. 2022.

EWELS, Philip *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, v. 32, n. 19, p. 3047–3048, 1 out. 2016. Disponível em: <https://academic.oup.com/bioinformatics/article/32/19/3047/2196507>. Acesso em: 15 jul. 2022.

FACKENTHAL, James D.; GODLEY, Lucy A. Aberrant RNA splicing and its functional consequences in cancer cells. *Disease Models & Mechanisms*, v. 1, n. 1, p. 37–42, 1 jul. 2008. Disponível em: <https://journals.biologists.com/dmm/article/1/1/37/2098/Aberrant-RNA-splicing-and-its-functional>. Acesso em: 29 maio 2023.

FAHMI, Naima Ahmed *et al.* As-quant: Detection and visualization of alternative splicing events with rna-seq data. *International Journal of Molecular Sciences*, v. 22, n. 9, p. 4468, 1 maio 2021. Disponível em: <https://www.mdpi.com/1422-0067/22/9/4468/htm>. Acesso em: 29 maio 2023.

FAO. *FAOSTAT*. Disponível em: <https://www.fao.org/faostat/en/#home>. Acesso em: 17 jul. 2023.

FAO. *Land statistics and indicators*. Disponível em: <https://fenixservices.fao.org/faostat/static/documents/RL/cc0963en.pdf>. Acesso em: 26 jun. 2023.

FENG, Guizhi; WU, Juxun; YI, Hualin. Global tissue-specific transcriptome analysis of Citrus sinensis fruit across six developmental stages. *Scientific Data 2019 6:1*, v. 6, n. 1, p. 1–8, 21 ago. 2019. Disponível em: <https://www.nature.com/articles/s41597-019-0162-y>. Acesso em: 2 jul. 2023.

FERNANDEZ-POZO, Noe *et al.* The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Research*, v. 43, n. D1, p. D1036–D1041, 28 jan. 2015. Disponível em: <https://dx.doi.org/10.1093/nar/gku1195>. Acesso em: 14 jul. 2023.

FILICHKIN, Sergei *et al.* Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Current Opinion in Plant Biology*, v. 24, p. 125–135, 1 abr. 2015. Acesso em: 9 jul. 2023.

FORLANI, Sara; MASIERO, Simona; MIZZOTTI, Chiara. Fruit ripening: the role of hormones, cell wall modifications, and their relationship with pathogens. *Journal of Experimental Botany*, v. 70, n. 11, p. 2993–3006, 1 jun. 2019. Disponível em: <https://dx.doi.org/10.1093/jxb/erz112>. Acesso em: 26 jun. 2023.

FU, Liangbo *et al.* Transcriptomic and alternative splicing analyses reveal mechanisms of the difference in salt tolerance between barley and rice. *Environmental and Experimental Botany*, v. 166, p. 103810, 1 out. 2019. Acesso em: 11 jul. 2023.

FU, Xiang Dong; ARES, Manuel. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics 2014 15:10*, v. 15, n. 10, p. 689–701, 12 ago. 2014. Disponível em: <https://www.nature.com/articles/nrg3778>. Acesso em: 9 jul. 2023.

GAO, Jin *et al.* Role of ethylene response factors (ERFs) in fruit ripening. *Food Quality and Safety*, v. 4, n. 1, p. 15–20, 11 maio 2020. Disponível em: <https://dx.doi.org/10.1093/fqsafe/fyz042>. Acesso em: 2 jul. 2023.

GAPPER, Nigel E.; MCQUINN, Ryan P.; GIOVANNONI, James J. Molecular and genetic regulation of fruit ripening. *Plant Molecular Biology 2013 82:6*, v. 82, n. 6, p. 575–591, 13 abr. 2013. Disponível em: <https://link.springer.com/article/10.1007/s11103-013-0050-3>. Acesso em: 26 jun. 2023.

GEHRING, Niels H.; ROIGNANT, Jean Yves. Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends in Genetics*, v. 37, n. 4, p. 355–372, 1 abr. 2021. Disponível em: <http://www.cell.com/article/S0168952520302973/fulltext>. Acesso em: 14 jul. 2023.

GEORGOMANOLIS, Theodore; SOFIADIS, Konstantinos; PAPANTONIS, Argyris. Cutting a long intron short: Recursive splicing and its implications. *Frontiers in Physiology*, v. 7, n. NOV, p. 234102, 29 nov. 2016. Acesso em: 9 jul. 2023.

GIANNOPOULOU, Aikaterini F. *et al.* Gene-Specific Intron Retention Serves as Molecular Signature that Distinguishes Melanoma from Non-Melanoma Cancer Cells in Greek Patients. *International journal of molecular sciences*, v. 20, n. 4, 2 fev. 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30795533/>. Acesso em: 29 maio 2023.

GIOVANNONI, James J. Genetic Regulation of Fruit Development and Ripening. *The Plant Cell*, v. 16, n. suppl_1, p. S170–S180, 1 jun. 2004. Disponível em: <https://dx.doi.org/10.1105/tpc.019158>. Acesso em: 27 jun. 2023.

GUO, Da Long *et al.* Transcriptome analysis reveals mechanism of early ripening in Kyoho grape with hydrogen peroxide treatment. *BMC Genomics 2020 21:1*, v. 21, n. 1, p. 1–18, 11 nov. 2020. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07180-y>. Acesso em: 2 jul. 2023.

GUO, Wenbin *et al.* 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. *RNA Biology*, v. 18, n. 11, p. 1574–1587, 2021. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/15476286.2020.1858253>. Acesso em: 15 jul. 2022.

HAAS, B. J. Analysis of alternative splicing in plants with bioinformatics tools. *Current Topics in Microbiology and Immunology*, v. 326, p. 17–37, 2008. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-540-76776-3_2>. Acesso em: 16 jul. 2022.

HANDA, Avtar K.; TIZNADO-HERNÁNDEZ, Martín Ernesto; MATTOO, Autar K. Fruit development and ripening: A molecular perspective. *Plant Biotechnology and Agriculture: Prospects for the 21st Century*. [S.l.]: Elsevier, 2011. p. 405–424.

HARVEY, Samuel E.; CHENG, Chonghui. Methods for characterization of alternative RNA splicing. *Methods in Molecular Biology*, v. 1402, p. 229–241, 1 jan. 2016. Disponível em: <https://link.springer.com/protocol/10.1007/978-1-4939-3378-5_18>. Acesso em: 14 jul. 2023.

HERZEL, Lydia *et al.* Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology 2017 18:10*, v. 18, n. 10, p. 637–650, 9 ago. 2017. Disponível em: <https://www.nature.com/articles/nrm.2017.63>. Acesso em: 9 jul. 2023.

HIRSCH, Rebecca E. *et al.* A Role for the AKT1 Potassium Channel in Plant Nutrition. *Science*, v. 280, n. 5365, p. 918–921, 8 maio 1998. Disponível em: <https://www.science.org/doi/10.1126/science.280.5365.918>. Acesso em: 4 ago. 2022.

HIRSCHMAN, Lynette *et al.* Text mining for the biocuration workflow. *Database*, v. 2012, 1 jan. 2012. Disponível em:

<https://academic.oup.com/database/article/doi/10.1093/database/bas020/434747>. Acesso em: 29 maio 2023.

HOOPER, Joan E. A survey of software for genome-wide discovery of differential splicing in RNA-Seq data. *Human Genomics*, v. 8, n. 1, p. 1–6, 21 jan. 2014. Disponível em: <https://humgenomics.biomedcentral.com/articles/10.1186/1479-7364-8-3>. Acesso em: 11 jul. 2023.

HU, Hongyin *et al.* Analysis of Alternative Splicing and Alternative Polyadenylation in Populus alba var. pyramidalis by Single-Molecular Long-Read Sequencing. *Frontiers in Genetics*, v. 11, p. 498851, 7 fev. 2020. Acesso em: 11 jul. 2023.

HU, Wei *et al.* Systematic characterization of cancer transcriptome at transcript resolution. *Nature Communications 2022 13:1*, v. 13, n. 1, p. 1–16, 10 nov. 2022. Disponível em: <https://www.nature.com/articles/s41467-022-34568-z>. Acesso em: 29 maio 2023.

HUBER, Wolfgang *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods 2015 12:2*, v. 12, n. 2, p. 115–121, 29 jan. 2015. Disponível em: <https://www.nature.com/articles/nmeth.3252>. Acesso em: 15 jul. 2022.

HUERTA-CEPAS, Jaime *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, v. 34, n. 8, p. 2115–2122, 1 ago. 2017. Disponível em: <https://academic.oup.com/mbe/article/34/8/2115/3782716>. Acesso em: 15 jul. 2022.

ICO. *ICO Coffee Development Report 2020: "The value of coffee: Sustainability, inclusiveness and resilience of the coffee global value chain" Background*. . [S.l: s.n.], 2021. Disponível em: <https://www.internationalcoffeecouncil.com/cdr2020>. Acesso em: 10 jul. 2022.

JAGANATHAN, Kishore *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, v. 176, n. 3, p. 535- 548.e24, 24 jan. 2019. Disponível em: <http://www.cell.com/article/S0092867418316295/fulltext>. Acesso em: 14 jul. 2023.

JASSAL, Bijay *et al.* The reactome pathway knowledgebase. *Nucleic Acids Research*, v. 48, n. D1, p. D498–D503, 8 jan. 2020. Disponível em:

<https://academic.oup.com/nar/article/48/D1/D498/5613674>. Acesso em: 15 jul. 2022.

JI, Yanrong; MISHRA, Rama K.; DAVULURI, Ramana V. In silico analysis of alternative splicing on drug-target gene interactions. *Scientific Reports*, v. 10, n. 1, 1 dez. 2020. Disponível em: </pmc/articles/PMC6954184/>. Acesso em: 29 maio 2023.

JIANG, Guoxiang *et al.* Alternative splicing of MaMYB16L regulates starch degradation in banana fruit during ripening. *Journal of Integrative Plant Biology*, v. 63, n. 7, p. 1341–1352, 1 jul. 2021. Disponível em: <https://onlinelibrary-wiley.ez27.periodicos.capes.gov.br/doi/full/10.1111/jipb.13088>. Acesso em: 15 jul. 2022.

JOHN, Sheeba; OLAS, Justyna Jadwiga; MUELLER-ROEBER, Bernd. Regulation of alternative splicing in response to temperature variation in plants. *Journal of Experimental Botany*, v. 72, n. 18, p. 6150–6163, 30 set. 2021. Disponível em: <https://dx.doi.org/10.1093/jxb/erab232>. Acesso em: 9 jul. 2023.

JONES, Philip *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics*, v. 30, n. 9, p. 1236–1240, 1 maio 2014. Disponível em: <https://academic.oup.com/bioinformatics/article/30/9/1236/237988>. Acesso em: 15 jul. 2022.

KAHRAMAN, Abdullah; BULJAN, Marija; VITTING-SEERUP, Kristoffer. Editorial: Alternative Splicing in Health and Disease. *Frontiers in Molecular Biosciences*, v. 9, p. 265, 19 abr. 2022. Acesso em: 29 maio 2023.

KARASAWA, Marines Marli Gniech; MOHAN, Chakravarthi. Fruits as Prospective Reserves of bioactive Compounds: A Review. *Natural Products and Bioprospecting*, v. 8, n. 5, p. 335–346, 1 out. 2018. Disponível em: <https://link.springer.com/article/10.1007/s13659-018-0186-6>. Acesso em: 26 jun. 2023.

KARLOVA, Rumyana *et al.* Transcriptional control of fleshy fruit development and ripening. *Journal of Experimental Botany*, v. 65, n. 16, p. 4527–4541, 1 ago. 2014. Disponível em: <https://dx.doi.org/10.1093/jxb/eru316>. Acesso em: 26 jun. 2023.

KATZ, Yarden *et al.* Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics*, v. 31, n. 14, p. 2400–2402, 1 jul. 2015. Disponível

em: <https://academic.oup.com/bioinformatics/article/31/14/2400/254271>. Acesso em: 29 maio 2023.

KELEMEN, Olga *et al.* Function of alternative splicing. *Gene*, v. 514, n. 1, p. 1–30, 1 fev. 2013. Acesso em: 15 jul. 2022.

KIM, Daehwan *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8*, v. 37, n. 8, p. 907–915, 2 ago. 2019. Disponível em: <https://www.nature.com/articles/s41587-019-0201-4>. Acesso em: 4 jun. 2023.

KIM, Hyoung Kyu *et al.* Alternative splicing isoforms in health and disease. *Pflügers Archiv - European Journal of Physiology 2018 470:7*, v. 470, n. 7, p. 995–1016, 13 mar. 2018. Disponível em: <https://link.springer.com/article/10.1007/s00424-018-2136-x>. Acesso em: 29 maio 2023.

KONGOR, John Edem *et al.* Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile — A review. *Food Research International*, v. 82, p. 44–52, 1 abr. 2016. Acesso em: 27 jun. 2023.

KORNBLIHTT, Alberto R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature Reviews Molecular Cell Biology 2013 14:3*, v. 14, n. 3, p. 153–165, 6 fev. 2013. Disponível em: <https://www.nature.com/articles/nrm3525>. Acesso em: 15 jul. 2022.

LALOUM, Tom; MARTÍN, Guiomar; DUQUE, Paula. Alternative Splicing Control of Abiotic Stress Responses. *Trends in Plant Science*, v. 23, n. 2, p. 140–150, 1 fev. 2018. Acesso em: 15 jul. 2022.

LANTZ, Henrik *et al.* Ten steps to get started in Genome Assembly and Annotation. *F1000Research 2018 7:148*, v. 7, p. 148, 5 fev. 2018. Disponível em: <https://f1000research.com/articles/7-148>. Acesso em: 16 jul. 2022.

LAU, Edward *et al.* Splice-Junction-Based Mapping of Alternative Isoforms in the Human Proteome. *Cell Reports*, v. 29, n. 11, p. 3751- 3765.e5, 10 dez. 2019. Disponível em: <http://www.cell.com/article/S2211124719314974/fulltext>. Acesso em: 11 jul. 2023.

LE, Kai Qin *et al.* Alternative splicing as a biomarker and potential target for drug discovery. *Acta Pharmacologica Sinica 2015 36:10*, v. 36, n. 10, p. 1212–1218, 15 jun.

2015. Disponível em: <https://www.nature.com/articles/aps201543>. Acesso em: 29 maio 2023.

LEE, Joon Seon; ADAMS, Keith L. Global insights into duplicated gene expression and alternative splicing in polyploid Brassica napus under heat, cold, and drought stress. *The Plant Genome*, v. 13, n. 3, p. e20057, 1 nov. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/tpg2.20057>. Acesso em: 15 jul. 2022.

LI, Jiayin *et al.* Comprehensive RNA-Seq Analysis on the Regulation of Tomato Ripening by Exogenous Auxin. *PLOS ONE*, v. 11, n. 5, p. e0156453, 1 maio 2016. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0156453>. Acesso em: 14 jul. 2023.

LI, Shan *et al.* Roles of RIN and ethylene in tomato fruit ripening and ripening-associated traits. *New Phytologist*, v. 226, n. 2, p. 460–475, 1 abr. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/nph.16362>. Acesso em: 5 ago. 2022.

LI, Shan; CHEN, Kunsong; GRIERSON, Don. A critical evaluation of the role of ethylene and MADS transcription factors in the network controlling fleshy fruit ripening. *New Phytologist*, v. 221, n. 4, p. 1724–1741, 1 mar. 2019. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/nph.15545>. Acesso em: 27 jun. 2023.

LI, Shan; CHEN, Kunsong; GRIERSON, Donald. Molecular and Hormonal Mechanisms Regulating Fleshy Fruit Ripening. *Cells 2021, Vol. 10, Page 1136*, v. 10, n. 5, p. 1136, 8 maio 2021. Disponível em: <https://www.mdpi.com/2073-4409/10/5/1136/htm>. Acesso em: 27 jun. 2023.

LI, Yeyun *et al.* Comprehensive profiling of alternative splicing landscape during cold acclimation in tea plant. *BMC Genomics*, v. 21, n. 1, p. 1–16, 20 jan. 2020. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-6491-6>. Acesso em: 15 jul. 2022.

LI, Zhiwei *et al.* Histone demethylase SlJMJ6 promotes fruit ripening by removing H3K27 methylation of ripening-related genes in tomato. *New Phytologist*, v. 227, n. 4, p. 1138–1156, 1 ago. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/nph.16590>. Acesso em: 15 jul. 2022.

LING, Zhihao *et al.* Evolution of alternative splicing in eudicots. *Frontiers in Plant Science*, v. 10, p. 707, 31 maio 2019. Acesso em: 15 jul. 2022.

LIU, Mingchun *et al.* Ethylene Control of Fruit Ripening: Revisiting the Complex Network of Transcriptional Regulation. *Plant Physiology*, v. 169, n. 4, p. 2380–2390, 9 dez. 2015. Disponível em: <https://dx.doi.org/10.1104/pp.15.01361>. Acesso em: 27 jun. 2023.

LIU, Qi; FANG, Leiming; WU, Chengjun. Alternative Splicing and Isoforms: From Mechanisms to Diseases. *Genes*, v. 13, n. 3, 1 mar. 2022. Disponível em: </pmc/articles/PMC8951537/>. Acesso em: 29 maio 2023.

LORD, Jenny; BARALLE, Diana. Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Frontiers in Genetics*, v. 12, p. 1146, 1 jul. 2021. Acesso em: 29 maio 2023.

LYKKE-ANDERSEN, Søren; JENSEN, Torben Heick. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature Reviews Molecular Cell Biology 2015 16:11*, v. 16, n. 11, p. 665–677, 23 set. 2015. Disponível em: <https://www.nature.com/articles/nrm4063>. Acesso em: 9 jul. 2023.

MA, Tingyu *et al.* Genome-Wide Analysis of Light-Regulated Alternative Splicing in Artemisia annua L. *Frontiers in Plant Science*, v. 12, p. 733505, 29 set. 2021. Acesso em: 11 jul. 2023.

MAGLOTT, Donna *et al.* Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research*, v. 39, n. Database issue, p. D52, jan. 2011. Disponível em: </pmc/articles/PMC3013746/>. Acesso em: 29 maio 2023.

MARASCO, Luciano E.; KORNBLIHTT, Alberto R. The physiology of alternative splicing. *Nature Reviews Molecular Cell Biology 2022 24:4*, v. 24, n. 4, p. 242–254, 13 out. 2022. Disponível em: <https://www.nature.com/articles/s41580-022-00545-z>. Acesso em: 8 jul. 2023.

MARIE, Lison *et al.* G × E interactions on yield and quality in Coffea arabica: new F1 hybrids outperform American cultivars. *Euphytica*, v. 216, n. 5, 1 maio 2020. Acesso em: 10 jul. 2022.

MARTINEZ-MONTIEL, Nancy *et al.* Alternative Splicing as a Target for Cancer Treatment. *International Journal of Molecular Sciences*, v. 19, n. 2, 11 fev. 2018. Disponível em: </pmc/articles/PMC5855767/>. Acesso em: 29 maio 2023.

MATA-NICOLÁS, Estefanía *et al.* Exploiting the diversity of tomato: the development of a phenotypically and genetically detailed germplasm collection. *Horticulture Research 2020 7:1*, v. 7, n. 1, p. 1–14, 1 maio 2020. Disponível em: <https://www.nature.com/articles/s41438-020-0291-7>. Acesso em: 14 jul. 2023.

MCCARTHY, Fiona M. *et al.* AgBase: a unified resource for functional analysis in agriculture. *Nucleic Acids Research*, v. 35, n. suppl_1, p. D599–D603, 1 jan. 2007. Disponível em: <https://academic.oup.com/nar/article/35/suppl_1/D599/1111911>. Acesso em: 15 jul. 2022.

MCCOOK, Stuart; VANDERMEER, John. The Big Rust and the Red Queen: Long-term perspectives on coffee rust research. *Phytopathology*, v. 105, n. 9, p. 1164–1173, 1 set. 2015. Disponível em: <https://apsjournals.apsnet.org/doi/10.1094/PHYTO-04-15-0085-RVW>. Acesso em: 10 jul. 2022.

MEHMOOD, Arfa *et al.* Systematic evaluation of differential splicing tools for RNA-seq studies. *Briefings in Bioinformatics*, v. 21, n. 6, p. 2052–2065, 1 dez. 2020. Disponível em: <https://academic.oup.com/bib/article/21/6/2052/5648232>. Acesso em: 29 maio 2023.

MEHRABI, Zia; LASHERMES, Philippe. Protecting the origins of coffee to safeguard its future. *Nature Plants 2017 3:1*, v. 3, n. 1, p. 1–3, 6 jan. 2017. Disponível em: <https://www.nature.com/articles/nplants2016209>. Acesso em: 10 jul. 2022.

MELO, José Pedro; KALYNA, Maria; DUQUE, Paula. Current Challenges in Studying Alternative Splicing in Plants: The Case of Physcomitrella patens SR Proteins. *Frontiers in Plant Science*, v. 11, p. 512901, 24 mar. 2020. Acesso em: 9 jul. 2023.

MERINO, Gabriela A.; CONESA, Ana; FERNÁNDEZ, Elmer A. A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human RNA-seq studies. *Briefings in Bioinformatics*, v. 20, n. 2, p. 471–481, 25 mar. 2019. Disponível em: <https://academic.oup.com/bib/article/20/2/471/4524048>. Acesso em: 29 maio 2023.

MONTEUUIS, Geoffray *et al.* The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Research*, v. 47, n. 22, p. 11497–11513, 16 dez. 2019. Disponível em: <https://academic.oup.com/nar/article/47/22/11497/5625532>. Acesso em: 29 maio 2023.

MUJCIC, Redzo; OSWALD, Andrew J. Evolution of well-being and happiness after increases in consumption of fruit and vegetables. *American Journal of Public Health*, v. 106, n. 8, p. 1504–1510, 1 ago. 2016. Disponível em: <https://ajph.aphapublications.org/doi/10.2105/AJPH.2016.303260>. Acesso em: 26 jun. 2023.

MUOKI, Richard Chalo *et al.* An improved protocol for the isolation of RNA from roots of tea (Camellia sinensis (L.) O. Kuntze). *Molecular Biotechnology*, v. 52, n. 1, p. 82–88, 6 set. 2012. Disponível em: <https://link.springer.com/article/10.1007/s12033-011-9476-5>. Acesso em: 26 jul. 2023.

NAWAZ, Iqra *et al.* RNA-Seq profiling reveals the plant hormones and molecular mechanisms stimulating the early ripening in apple. *Genomics*, v. 113, n. 1, p. 493–502, 1 jan. 2021. Acesso em: 2 jul. 2023.

NILSEN, Timothy W.; GRAVELEY, Brenton R. Expansion of the eukaryotic proteome by alternative splicing. *Nature 2010 463:7280*, v. 463, n. 7280, p. 457–463, 27 jan. 2010. Disponível em: <https://www.nature.com/articles/nature08909>. Acesso em: 15 jul. 2022.

NOBLE, Jerald D. *et al.* The Genetic Regulation of Alternative Splicing in Populus deltoides. *Frontiers in Plant Science*, v. 11, p. 525233, 5 jun. 2020. Acesso em: 9 jul. 2023.

O'LEARY, Nuala A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, v. 44, n. D1, p. D733–D745, 4 jan. 2016. Disponível em: <https://academic.oup.com/nar/article/44/D1/D733/2502674>. Acesso em: 29 maio 2023.

OUYANG, Jiawei *et al.* Long non-coding RNAs are involved in alternative splicing and promote cancer progression. *British Journal of Cancer 2021 126:8*, v. 126,

n. 8, p. 1113–1124, 8 nov. 2021. Disponível em: <https://www.nature.com/articles/s41416-021-01600-w>. Acesso em: 29 maio 2023.

PALMA, José M. *et al.* Editorial: Fruit ripening: From present knowledge to future development. *Frontiers in Plant Science*, v. 10, p. 442947, 16 abr. 2019. Acesso em: 2 jul. 2023.

PARK, Eddie *et al.* The Expanding Landscape of Alternative Splicing Variation in Human Populations. *The American Journal of Human Genetics*, v. 102, n. 1, p. 11–26, 4 jan. 2018. Disponível em: <http://www.cell.com/article/S0002929717304548/fulltext>. Acesso em: 9 jul. 2023.

PATRO, Rob *et al.* Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods 2017 14:4*, v. 14, n. 4, p. 417–419, 6 mar. 2017. Disponível em: <https://www.nature.com/articles/nmeth.4197>. Acesso em: 2 jun. 2023.

PAUL, Vijay; PANDEY, Rakesh. Role of internal atmosphere on fruit ripening and storability - A review. *Journal of Food Science and Technology*, v. 51, n. 7, p. 1223–1250, 26 nov. 2014. Disponível em: <https://link.springer.com/article/10.1007/s13197-011-0583-x>. Acesso em: 26 jun. 2023.

PAYASI, A.; SANWAL, G. G. RIPENING OF CLIMACTERIC FRUITS AND THEIR CONTROL. *Journal of Food Biochemistry*, v. 34, n. 4, p. 679–710, 1 ago. 2010. Disponível em: <https://onlinelibrary-wiley.ez27.periodicos.capes.gov.br/doi/full/10.1111/j.1745-4514.2009.00307.x>. Acesso em: 26 jun. 2023.

PAYSAN-LAFOSSE, Typhaine *et al.* InterPro in 2022. *Nucleic Acids Research*, v. 51, n. D1, p. D418–D427, 6 jan. 2023. Disponível em: <https://academic.oup.com/nar/article/51/D1/D418/6814474>. Acesso em: 29 maio 2023.

PEREIRA, Lara *et al.* Genetic dissection of climacteric fruit ripening in a melon population segregating for ripening behavior. *Horticulture Research 2020 7:1*, v. 7, n. 1, p. 1–18, 1 nov. 2020. Disponível em: <https://www.nature.com/articles/s41438-020-00411-z>. Acesso em: 2 jul. 2023.

PESARESI, Paolo *et al.* Genetic regulation and structural changes during tomato fruit development and ripening. *Frontiers in Plant Science*, v. 5, n. APR, p. 82777, 23 abr. 2014. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpls.2014.00124/full>. Acesso em: 26 jun. 2023.

POHL, Martin *et al.* Alternative splicing of mutually exclusive exons—A review. *Biosystems*, v. 114, n. 1, p. 31–38, 1 out. 2013. Acesso em: 16 jul. 2022.

RANGWALA, Sanjida H. *et al.* Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Research*, v. 31, n. 1, p. 159–169, 1 jan. 2021. Disponível em: <https://genome.cshlp.org/content/31/1/159.full>. Acesso em: 14 jul. 2023.

REDDY, Anireddy S.N. *et al.* Deciphering the plant splicing code: Experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. *Frontiers in Plant Science*, v. 3, n. FEB, p. 18, 7 fev. 2012. Acesso em: 15 jul. 2022.

REN, Pingping *et al.* Alternative Splicing: A New Cause and Potential Therapeutic Target in Autoimmune Disease. *Frontiers in Immunology*, v. 12, p. 3284, 17 ago. 2021. Acesso em: 29 maio 2023.

RIGO, Richard *et al.* Alternative Splicing in the Regulation of Plant–Microbe Interactions. *Plant and Cell Physiology*, v. 60, n. 9, p. 1906–1916, 1 set. 2019. Disponível em: <https://dx.doi.org/10.1093/pcp/pcz086>. Acesso em: 9 jul. 2023.

RITCHIE, Matthew E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, v. 43, n. 7, p. e47–e47, 20 abr. 2015. Disponível em: <https://dx.doi.org/10.1093/nar/gkv007>. Acesso em: 11 jul. 2023.

ROHRMANN, Johannes *et al.* Tissue specificity and differential expression of transcription factors in tomato provide hints of unique regulatory networks during fruit ripening. *Plant Signaling and Behavior*, v. 7, n. 12, p. 1639–1647, 2012. Disponível em: <https://www.tandfonline.com/doi/abs/10.4161/psb.22264>. Acesso em: 2 jul. 2023.

SÁGIO, Solange Aparecida *et al.* Physiological and molecular analyses of early and late Coffea arabica cultivars at different stages of fruit ripening. *Acta Physiologiae*

Plantarum, v. 35, n. 11, p. 3091–3098, 17 nov. 2013. Disponível em: <https://link.springer.com/article/10.1007/s11738-013-1342-6>. Acesso em: 15 jul. 2022.

SAMMETH, Michael; FOISSAC, Sylvain; GUIGÓ, Roderic. A General Definition and Nomenclature for Alternative Splicing Events. *PLOS Computational Biology*, v. 4, n. 8, p. e1000147, 2008. Disponível em: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000147>. Acesso em: 10 jul. 2023.

SÁNCHEZ-BARRENA, María José *et al.* Recognition and Activation of the Plant AKT1 Potassium Channel by the Kinase CIPK23. *Plant Physiology*, v. 182, n. 4, p. 2143–2153, 6 abr. 2020. Disponível em: <https://academic.oup.com/plphys/article/182/4/2143/6116589>. Acesso em: 4 ago. 2022.

SATO, Shusei *et al.* The tomato genome sequence provides insights into fleshy fruit evolution. *Nature 2012 485:7400*, v. 485, n. 7400, p. 635–641, 30 maio 2012. Disponível em: <https://www.nature.com/articles/nature11119>. Acesso em: 14 jul. 2023.

SCALABRIN, Simone *et al.* A single polyploidization event at the origin of the tetraploid genome of Coffea arabica is responsible for the extremely low genetic variation in wild and cultivated germplasm. *Scientific Reports 2020 10:1*, v. 10, n. 1, p. 1–13, 13 mar. 2020. Disponível em: <https://www.nature.com/articles/s41598-020-61216-7>. Acesso em: 15 jul. 2022.

SCHAEFKE, Bernhard *et al.* The evolution of posttranscriptional regulation. *Wiley Interdisciplinary Reviews: RNA*, v. 9, n. 5, p. e1485, 1 set. 2018. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/wrna.1485>. Acesso em: 9 jul. 2023.

SEALE, Madeleine; NAKAYAMA, Naomi. From passive to informed: mechanical mechanisms of seed dispersal. *New Phytologist*, v. 225, n. 2, p. 653–658, 1 jan. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/nph.16110>. Acesso em: 26 jun. 2023.

SEYMOUR, Graham *et al. Genetics and epigenetics of fruit development and ripening*. *Current Opinion in Plant Biology*. [S.l: s.n.]. , fev. 2008

SHAN, Youxia *et al.* Role of apyrase-mediated eATP signal in chilling injury of postharvest banana fruit during storage. *Postharvest Biology and Technology*, v. 187, p. 111874, 1 maio 2022. Acesso em: 5 ago. 2022.

SHARP, Phillip; ROBERTS, Richard; SHI, Yigong. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nature Reviews Molecular Cell Biology 2017 18:11*, v. 18, n. 11, p. 655–670, 27 set. 2017. Disponível em: <https://www.nature.com/articles/nrm.2017.86>. Acesso em: 29 maio 2023.

SHEN, Shihao *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 51, p. E5593–E5601, 23 dez. 2014. Acesso em: 15 jul. 2022.

SHINOZAKI, Yoshihito *et al.* High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. *Nature Communications 2018 9:1*, v. 9, n. 1, p. 1–13, 25 jan. 2018. Disponível em: <https://www.nature.com/articles/s41467-017-02782-9>. Acesso em: 2 jul. 2023.

SIBLEY, Christopher R.; BLAZQUEZ, Lorea; ULE, Jernej. Lessons from non-canonical splicing. *Nature Reviews Genetics 2016 17:7*, v. 17, n. 7, p. 407–421, 31 maio 2016. Disponível em: <https://www.nature.com/articles/nrg.2016.46>. Acesso em: 15 jul. 2022.

SLUGINA, Maria A. *et al.* Characterization of RIN isoforms and their expression in tomato fruit ripening. *Cells*, v. 10, n. 7, p. 1739, 1 jul. 2021. Disponível em: <https://www.mdpi.com/2073-4409/10/7/1739/htm>. Acesso em: 2 jul. 2023.

SONG, Qi A. *et al.* Computational analysis of alternative splicing in plant genomes. *Gene*, v. 685, p. 186–195, 15 fev. 2019. Acesso em: 11 jul. 2023.

SOUSA, Tiago Vieira *et al.* Early selection enabled by the implementation of genomic selection in coffea arabica breeding. *Frontiers in Plant Science*, v. 9, p. 1934, 8 jan. 2019. Acesso em: 10 jul. 2022.

SPENGLER, Robert N. Anthropogenic Seed Dispersal: Rethinking the Origins of Plant Domestication. *Trends in Plant Science*, v. 25, n. 4, p. 340–348, 1 abr. 2020. Disponível em: <http://www.cell.com/article/S1360138520300224/fulltext>. Acesso em: 27 jun. 2023.

STROBELT, Hendrik *et al.* Vials: Visualizing Alternative Splicing of Genes. *IEEE Transactions on Visualization and Computer Graphics*, v. 22, n. 1, p. 399–408, 31 jan. 2016. Acesso em: 29 maio 2023.

SU, Xiao *et al.* A high-continuity and annotated tomato reference genome. *BMC Genomics*, v. 22, n. 1, p. 1–12, 1 dez. 2021. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-08212-x>. Acesso em: 14 jul. 2023.

SUN, Yuan; XIAO, Han. Identification of alternative splicing events by RNA sequencing in early growth tomato fruits. *BMC Genomics*, v. 16, n. 1, p. 1–13, 16 nov. 2015. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-2128-6>. Acesso em: 11 jul. 2023.

SUNG, Chul Lee *et al.* A protein phosphorylation/dephosphorylation network regulates a plant potassium channel. *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, n. 40, p. 15959–15964, 2 out. 2007. Disponível em: <https://www.pnas.org/doi/abs/10.1073/pnas.0707912104>. Acesso em: 4 ago. 2022.

SYED, Naeem H. *et al.* Alternative splicing in plants – coming of age. *Trends in Plant Science*, v. 17, n. 10, p. 616–623, 1 out. 2012. Disponível em: <http://www.cell.com/article/S1360138512001276/fulltext>. Acesso em: 15 jul. 2022.

SZAKONYI, Dóra; DUQUE, Paula. Alternative splicing as a regulator of early plant development. *Frontiers in Plant Science*, v. 9, p. 382146, 15 ago. 2018. Acesso em: 14 jul. 2023.

TALHINHAS, Pedro *et al.* The coffee leaf rust pathogen Hemileia vastatrix: one and a half centuries around the tropics. *Molecular Plant Pathology*, v. 18, n. 8, p. 1039–1051, 1 out. 2017. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/mpp.12512>. Acesso em: 10 jul. 2022.

TANG, Dengguo; GALLUSCI, Philippe; LANG, Zhaobo. Fruit development and epigenetic modifications. *New Phytologist*, v. 228, n. 3, p. 839–844, 1 nov. 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/nph.16724>. Acesso em: 27 jun. 2023.

TEROL, Javier *et al.* Transcriptomic analysis of Citrus clementina mandarin fruits maturation reveals a MADS-box transcription factor that might be involved in the regulation of earliness. *BMC Plant Biology*, v. 19, n. 1, p. 1–20, 31 jan. 2019. Disponível em: <https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-019-1651-z>. Acesso em: 2 jul. 2023.

THORVALDSDÓTTIR, Helga; ROBINSON, James T.; MESIROV, Jill P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, v. 14, n. 2, p. 178–192, 1 mar. 2013. Disponível em: <https://academic.oup.com/bib/article/14/2/178/208453>. Acesso em: 29 maio 2023.

TONG, Li *et al.* Impact of RNA-seq data analysis algorithms on gene expression estimation and downstream prediction. *Scientific Reports 2020 10:1*, v. 10, n. 1, p. 1–20, 21 out. 2020. Disponível em: <https://www.nature.com/articles/s41598-020-74567-y>. Acesso em: 14 jul. 2023.

TRESS, Michael L.; ABASCAL, Federico; VALENCIA, Alfonso. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends in Biochemical Sciences*, v. 42, n. 2, p. 98–110, 1 fev. 2017. Disponível em: <http://www.cell.com/article/S0968000416301189/fulltext>. Acesso em: 8 jul. 2023.

ULE, Jernej; BLENCOWE, Benjamin J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular Cell*, v. 76, n. 2, p. 329–345, 17 out. 2019. Disponível em: <http://www.cell.com/article/S1097276519307026/fulltext>. Acesso em: 15 jul. 2022.

UNTERGASSER, Andreas *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, v. 35, n. suppl_2, p. W71–W74, 1 jul. 2007. Disponível em: <https://academic.oup.com/nar/article/35/suppl_2/W71/2922185>. Acesso em: 15 jul. 2022.

USDA. *Potatoes and tomatoes are the most commonly consumed vegetables*. Disponível em: <https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=58340>. Acesso em: 17 jul. 2023.

VAQUERO-GARCIA, Jorge *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife*, v. 5, n. FEBRUARY2016, 1 fev. 2016. Acesso em: 10 jul. 2023.

VENNAPUSA, Amaranatha R. *et al.* A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber. *Scientific Reports 2020 10:1*, v. 10, n. 1, p. 1–13, 9 out. 2020. Disponível em: <https://www.nature.com/articles/s41598-020-73958-5>. Acesso em: 25 jul. 2022.

VUONG, Celine K.; BLACK, Douglas L.; ZHENG, Sika. The neurogenetics of alternative splicing. *Nature Reviews Neuroscience 2016 17:5*, v. 17, n. 5, p. 265–281, 20 abr. 2016. Disponível em: <https://www.nature.com/articles/nrn.2016.27>. Acesso em: 9 jul. 2023.

WANG, Ketao *et al.* Landscape and Fruit Developmental Regulation of Alternative Splicing in Tomato by Genome-Wide Analysis. *Horticultural Plant Journal*, v. 2, n. 6, p. 338–350, 1 nov. 2016. Acesso em: 11 jul. 2023.

WANG, Maojun *et al.* A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytologist*, v. 217, n. 1, p. 163–178, 1 jan. 2018. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/nph.14762>. Acesso em: 15 jul. 2022.

WANG, Rufang *et al.* Revisiting the Role of Master Regulators in Tomato Ripening. *Trends in Plant Science*, v. 25, n. 3, p. 291–301, 1 mar. 2020. Disponível em: <http://www.cell.com/article/S136013851930305X/fulltext>. Acesso em: 5 ago. 2022.

WANG, Weihao *et al.* Current insights into posttranscriptional regulation of fleshy fruit ripening. *Plant Physiology*, v. 192, n. 3, p. 1785–1798, 3 jul. 2023. Disponível em: <https://dx.doi.org/10.1093/plphys/kiac483>. Acesso em: 2 jul. 2023.

WANG, Weihao *et al.* The transcription factor SlHY5 regulates the ripening of tomato fruit at both the transcriptional and translational levels. *Horticulture Research 2021 8:1*, v. 8, n. 1, p. 1–15, 1 abr. 2021. Disponível em: <https://www.nature.com/articles/s41438-021-00523-0>. Acesso em: 2 jul. 2023.

WANG, Zhining *et al.* Computational Analysis and Experimental Validation of Tumor-associated Alternative RNA Splicing in Human Cancer. *CANCER RESEARCH*, v. 63, p. 655–657, 2003. Disponível em: <http://aacrjournals.org/cancerres/article-pdf/63/3/655/2512025/ch0303000655.pdf>. Acesso em: 15 jul. 2023.

WEESE, Terri L.; BOHS, Lynn. A three-gene phylogeny of the genus Solanum (Solanaceae). *Systematic Botany*, v. 32, n. 2, p. 445–463, abr. 2007. Acesso em: 14 jul. 2023.

WEI, Qingzhen *et al.* A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticulture Research 2020 7:1*, v. 7, n. 1, p. 1–15, 21 set. 2020. Disponível em: <https://www.nature.com/articles/s41438-020-00391-0>. Acesso em: 14 jul. 2023.

WEN, Wei Xiong; MEAD, Adam J.; THONGJUEA, Supat. Technological advances and computational approaches for alternative splicing analysis in single cells. *Computational and Structural Biotechnology Journal*, v. 18, p. 332–343, 1 jan. 2020a. Acesso em: 15 jul. 2022.

WEN, Wei Xiong; MEAD, Adam J.; THONGJUEA, Supat. VALERIE: Visual-based inspection of alternative splicing events at single-cell resolution. *PLOS Computational Biology*, v. 16, n. 9, p. e1008195, 1 set. 2020b. Disponível em: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008195>. Acesso em: 29 maio 2023.

WHITE, Philip J. Recent advances in fruit development and ripening: an overview. *Journal of Experimental Botany*, v. 53, n. 377, p. 1995–2000, 1 out. 2002. Disponível em: <https://dx.doi.org/10.1093/jxb/erf105>. Acesso em: 27 jun. 2023.

WHITEHEAD, Susan R. *et al.* Fruits, frugivores, and the evolution of phytochemical diversity. *Oikos*, v. 2022, n. 2, 1 fev. 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/oik.08332>. Acesso em: 26 jun. 2023.

WILKINSON, Max E.; CHARENTON, Clément; NAGAI, Kiyoshi. RNA Splicing by the Spliceosome. *https://doi.org/10.1146/annurev-biochem-091719-064225*, v. 89, p. 359–388, 22 jun. 2020. Disponível em: <https://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-091719-064225>. Acesso em: 8 jul. 2023.

XIA, Wei *et al.* Alternative splicing of flowering time gene FT is associated with halving of time to flowering in coconut. *Scientific Reports 2020 10:1*, v. 10, n. 1, p. 1–11, 15 jul. 2020. Disponível em: <https://www.nature.com/articles/s41598-020-68431-2>. Acesso em: 15 jul. 2022.

XU, Ling *et al.* OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, v. 47, n. W1, p. W52–W58, 2 jul. 2019. Disponível em: <https://academic.oup.com/nar/article/47/W1/W52/5485531>. Acesso em: 15 jul. 2022.

YAN, Xiaomin *et al.* Alternative splicing during fruit development among fleshy fruits. *BMC Genomics*, v. 22, n. 1, p. 1–14, 1 dez. 2021. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-08111-1>. Acesso em: 11 set. 2022.

YANG, Jian *et al.* Co-regulation of exine wall patterning, pollen fertility and anther dehiscence by Arabidopsis apyrases 6 and 7. *Plant Physiology and Biochemistry*, v. 69, p. 62–73, 1 ago. 2013. Acesso em: 5 ago. 2022.

YE, Jian *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, v. 13, n. 1, p. 134, 18 jun. 2012. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-134>. Acesso em: 15 jul. 2022.

YI, Gibum *et al.* Expanded transcriptomic view of strawberry fruit ripening through meta-analysis. *PLOS ONE*, v. 16, n. 6, p. e0252685, 1 jun. 2021. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0252685>. Acesso em: 2 jul. 2023.

ZHANG, Chi *et al.* Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics*, v. 18, n. 1, p. 1–11, 7 ago. 2017. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-4002-1>. Acesso em: 14 jul. 2023.

ZHANG, Yi *et al.* Discerning novel splice junctions derived from RNA-seq alignment: A deep learning approach. *BMC Genomics*, v. 19, n. 1, p. 1–13, 27 dez. 2018. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-5350-1>. Acesso em: 14 jul. 2023.

ZHANG, Yuanjiao *et al.* Alternative splicing and cancer: a systematic review. *Signal Transduction and Targeted Therapy 2021 6:1*, v. 6, n. 1, p. 1–14, 24 fev. 2021.

Disponível em: <https://www.nature.com/articles/s41392-021-00486-7>. Acesso em: 29 maio 2023.

ZHANG, Zhong Fa *et al.* Isoform level expression profiles provide better cancer signatures than gene level expression profiles. *Genome Medicine*, v. 5, n. 4, p. 1–13, 17 abr. 2013. Disponível em: <https://genomemedicine.biomedcentral.com/articles/10.1186/gm437>. Acesso em: 4 jun. 2023.

ZHANG, Zijun *et al.* Deep-learning augmented RNA-seq analysis of transcript splicing. *Nature Methods 2019 16:4*, v. 16, n. 4, p. 307–310, 25 mar. 2019. Disponível em: <https://www.nature.com/articles/s41592-019-0351-9>. Acesso em: 15 jul. 2022.

ZHAO, Shanrong; YE, Zhan; STANTON, Robert. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*, v. 26, n. 8, p. 903, 1 ago. 2020. Disponível em: </pmc/articles/PMC7373998/>. Acesso em: 4 jun. 2023.

ZHENGFEI GUAN; TRINA BISWAS; FENG WU. *The US Tomato Industry: An Overview of Production and Trade*. Disponível em: <https://edis.ifas.ufl.edu/publication/FE1027>. Acesso em: 17 jul. 2023.

ZHU, Qianglong *et al.* Comparative transcriptome analysis of two contrasting watermelon genotypes during fruit development and ripening. *BMC Genomics*, v. 18, n. 1, p. 1–20, 3 jan. 2017. Disponível em: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3442-3>. Acesso em: 2 jul. 2023.

ZWOLAK, Rafał; SIH, Andrew. Animal personalities and seed dispersal: A conceptual review. *Functional Ecology*, v. 34, n. 7, p. 1294–1310, 1 jul. 2020. Disponível em: <https://onlinelibrary-wiley.ez27.periodicos.capes.gov.br/doi/full/10.1111/1365-2435.13583>. Acesso em: 26 jun. 2023.

7. GLOSSÁRIO

DEG: Gene diferencialmente expresso, ou análise de expressão diferencial de genes, refere-se a uma análise de bioinformática que compara quantitativamente a expressão de genes em amostras sobre diferentes condições.

TPM: Transcritos por milhão é uma métrica para representar quantitativamente normalizada a expressão de um gene por meio de contagem de seus transcritos em uma amostra de RNASeq. Frequentemente é utilizada em tecnologias de nova geração como ilumina ou de terceira geração como PacBio. Consultar também RPM reads por milhão e FPKM fragmentos por milhão.

RNASeq: Tecnologia de sequenciamento de RNA em amostras biológicas frequentemente utilizada para quantificar a expressão de gene sob diferentes condições. Permite o bioinformata acessar as sequencias completas ou fragmentos de RNA representando cada par-de-base como ATCTG em um extenso arquivo de texto no formato FASTA contendo as leituras presentes nas amostras submetidas ao instrumento chamado sequenciador.

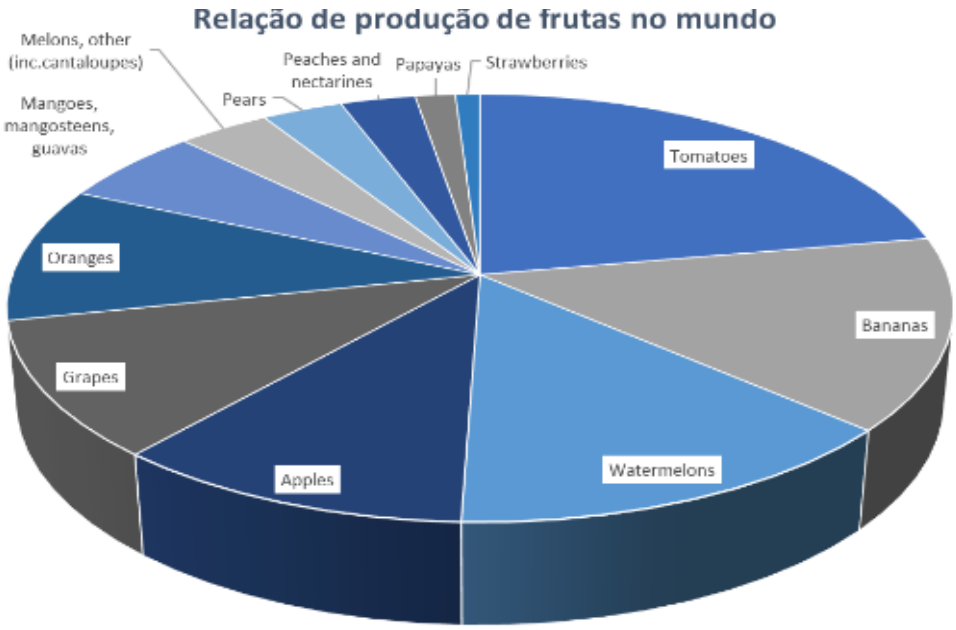
Volcano plot: Gráfico de representação de dados parecido com gráfico de dispersão duplicado em um único onde os pontos agrupam-se no centro inferior e se apresentam mais separados na parte superior à medida de avança no eixo x e no eixo y.

Flask: framework escrito na linguagem Python para construção de softwares de API http na arquitetura cliente-servidor.

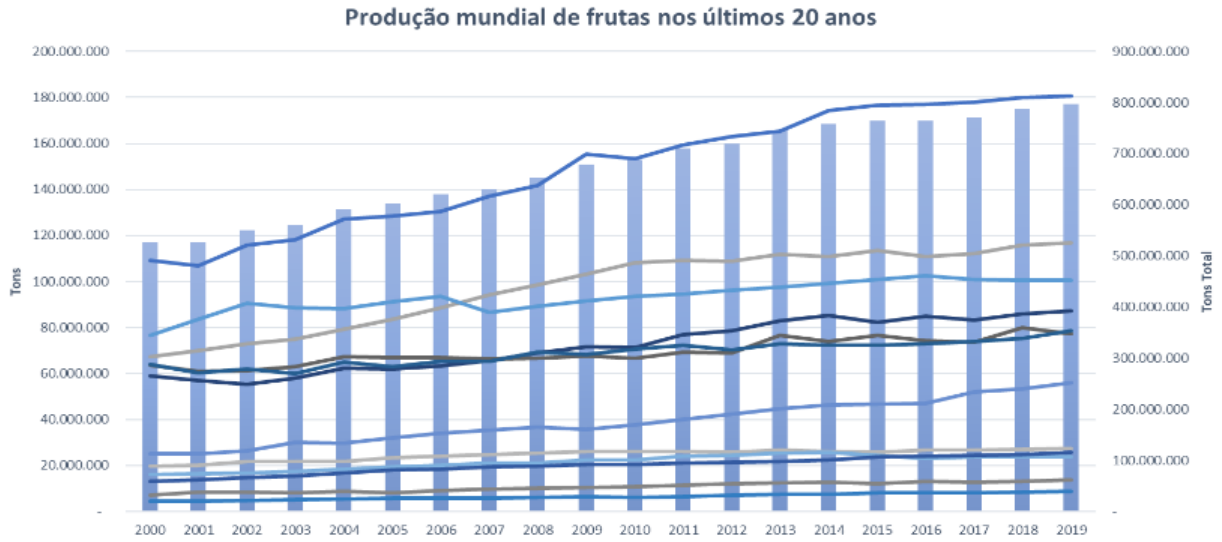
8. APÊNDICES

Tabelas e gráficos suplementares dos capítulos 1. Introdução, 2. RNASeq-based alternative splicing prediction reveals isoform alteration of potassium channel and apyrase associated with coffee fruit ripening, 3. Screening of Differential alternative splicing events in tomato fruit ripening e 4. Notes of use of web application to visualize alternative splicing for biomedicine.

7.1 PRODUÇÃO MUNDIAL DE FRUTOS



Produção mundial de frutos em toneladas conforme a FAO 2019. Os dados foram obtidos da ferramenta FAOStat http://www.fao.org/faostat filtrando pelos principais frutos produzidas no mundo: tomates, bananas, melancias, maçãs, uvas, laranjas, mangas, goiabas, melão, peras, pêssegos, nectarina, mamão e morangos. No gráfico de pizza a esquerda é apresentado a relação de cada fruto com relação a soma do total produzido ente 2000 e 2019 apresentado no gráfico a direita. No gráfico de barras é apresentado a soma da produção no ano em toneladas, as cores das linhas correspondem as cores do gráfico de pizza.



7.2 TABELA DE GENES ENVOLVIDOS NO PROCESSO DE AMADURECIMENTO

Lista de genes relacionados com o processo de amadurecimento dos frutos. Referencias [A] (LI, ZHIWEI *et al.*, 2020), [B] (LI, SHAN; CHEN; GRIERSON, 2021), [C] (LI, SHAN; CHEN; GRIERSON, 2019), [D] (ALEXANDER; GRIERSON, 2002), [E] (GAO *et al.*, 2020), [F] (KARLOVA *et al.*, 2014), [G] (SLUGINA *et al.*, 2021), [H] (GAPPER; MCQUINN; GIOVANNONI, 2013), [I] (LI, SHAN *et al.*, 2020), [J] (HANDA; TIZNADO-HERNÁNDEZ; MATTOO, 2011), [K] (GIOVANNONI, 2004), [L] (PAYASI; SANWAL, 2010), [M] (TANG; GALLUSCI; LANG, 2020), [N] (SEYMOUR *et al.*, 2008), [O] (WANG, RUFANG *et al.*, 2020). Os genes foram listados de tabelas e do corpo do texto das publicações, o nome é o radical do nome comum do gene. A coluna Uniprot é o identificador da proteína do gene no Uniprot identificado em alguma planta preferencialmente no tomate. Frutas onde o gene foi identificado: [1]Tomate, [2]Melão, [3]Banana, [4]Morango, [5]Maçã, [6]Pera, [7]Pessego, [8]Kiwi, [9]Arabidopsis, [10]Milho, [11]Pepino, [12]Uva, [13]Tangerina.

Gene	Descrição	Uniprot	Referência
Aroma			
AAT	Alcohol acyl transferase	Q6QLX4	A[1], B[2]
ADH	Alcohol dehydrogenase	P28032	A[1]
BCAT	Branched-chain-amino-acid aminotransferase 1	A0A3Q7JCI1	A[1]
LoxC	Lipoxygenase C	Q96573	A[1], B, C[1], D[1]
Sabor			
BHLH	Fator de transcrição basic helix–loop–helix	Q0JEB7	B[3], F
LIN	sucrose hydrolysis	Q5QER8	G[1]
EO / QR	2-methylene-furan-3-one reductase	K4BW79	E[4]
CHI	Chalcone-flavonone isomerase	A0A3Q7HFP3	C[1], H[1]
CHS	Chalcone synthase	P23419	C[1]
Cor			
4CL	4-coumarate CoA ligase	Q84P24	C[1]
ARF4	Auxin response factor	Q2LAI9	F[1]
CHY	Nonheme hydroxylases	Q0GGX1	C[1], I[1]
CRTISO	carotene isomerase	Q8S4R4	B, H, I[1]
GGPP	Geranylgeranyl pyrophosphate synthase	O04046	A[1], B, C[1], H
GLK2	Golden2-like	Q9FFH0	F[1]

	ISPE	4-diphos-phocytidyl-2-C-methyl-D-erythritol kinase	P93841	C[1]
	ISPF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	A0A3Q7HWC5	C[1]
	MYB	Fator de transcrição MYB	B5STZ8	B, F[1], H[1]
	NCED	9-cis-epoxycarotenoid dioxygenase	O24023	C[1]
	PDS	Phytoene desaturase	P28554	B, C[1], E[1], H, I[1]
	PSY	Phytoene synthase	P08196	A[1], B, C[1], G[1]
	PTOX	Alternative oxidase	Q9FEC9	C[1]
	ZDS	ζ-carotene desaturase	Q9SE20	B, G[1], H
	ZEP	Zeaxanthin epoxidase	P93236	C[1]
	ZISO	ζ-carotene isomerase	Q9SAC0	A[1], B, C[1], H
Textura				
	CEL	Endo-1,4-beta-glucanase	Q42871	A[1], C[1], J[1]
	EXP	Expansin	Q9C554	A[1], B[1], D[1], G[1], J[1]
	MAN	Endo-1,4-beta-mannanase	Q8L5J1	A[1], C[1]
	PG	Polygalacturonase	P05117	A[1], B, C[1], D, G[1], J[1], K
	PL	Pectate lyase	A0A3Q7FPN1	A[1], B, J[1]
	PME	Pectinesterase	P14280	C[1], D, J[1]
	SIXTH	Xyloglucan endotransglucosylase-hydrolase	Q9FZ05	C[1]
	TBG	β-galactosidase	A0A3Q7J3R3	A[1], C[1], J[1], K
Síntese de etileno				
	ACO	1-aminocyclopropane-1-carboxylate oxidase	P05116	A[1], C[1], D[1], E[5], J[1,2,3,5,6,8,7], L
	ACS	1-aminocyclopropane-1-carboxylate synthase	P18485	A[1], C[1], D[1], E[5], J[1,2,3,5,6,8,7], L[1]
	CTR	Ethylene-responsive protein kinase	A0A3Q7IMU1	C[1]
	EIL	ETHYLENE INSENSITIVE LIKE	Q76DI3	C[1], F[1]
	EIN	ETHYLENE INSENSITIVE	Q9S814	F, L[9]
	ETR	Ethylene receptor	Q41342	C[1], J[1]
	GR	Green ripe	Q1HTV5	C[1], J[1]
	SAM	S-adenosylmethionine synthase	P43281	C[1], J[1], L[1]
Outros				
	AP2	APETALA2-like ERF	B2ZP48	C[1], F[1,9]
	CNR	COLOURLESS NON-RIPENING	Q0PY35	C[1],F[1], J[10], O[1]
	ERF	Ethylene-responsive transcription factor	Q84XB3	D[1], E[1,3,4,5,7,8,11,12,13], F[1]
	FUL	FRUITFULL-like MADS-box	A0A3Q7GYA0	A[1], C[1], F[9], H[1]
	MADS	MADS-box transcription factor	Q8S4L3	F[1,4,5], J[1], O[1]
	NAC	plant-specific transcription factors	K4D6Q0	C[1], F[1], O[1]

NOR	NAC family transcription factor	Q56UP7	C[1], F[1]
MADS -RIN	MADS ripening inibithor	Q8S4L4	A[1], C[1], F[1], J[1], O[1]
DML2	DNA demethylase	Q9SR66	A[1], B[1], M[1,9], O

7.3 SUPPLEMENTARY TABLE S1

Table S1 Raw data used for analyses. In "Run" column has a identification to downlooad in Galaxy. This table was extracted from NCBI *Run Selector* tool.

Alias	Run	AvgSpotLen	Million Bases	BioSample	G. Bytes	Experiment	SRA_accession	SRA Study	Title	% after QC	Reads
SG1	ERR2231835	243	2.107	SAMEA104450747	0,77	ERX2285177	ERS2068689	ERP105943	Green stage coffee seed transcriptome replicate 1	98,9%	8.338.660
SG2	ERR2231836	249	3.770	SAMEA104450748	1,38	ERX2285178	ERS2068690	ERP105943	Green stage coffee seed transcriptome replicate 2	99,0%	14.515.144
SG3	ERR2231837	268	1.313	ERS2068691	0,48	ERX2285179	ERA1168391	ERP105943	Green stage coffee seed transcriptome replicate 3	99,8%	4.895.826
SY1	ERR2231838	244	993	SAMEA104450750	0,39	ERX2285180	ERS2068692	ERP105943	Yellow stage coffee seed transcriptome replicate 1	97,8%	3.882.179
SY2	ERR2231839	247	3.032	SAMEA104450751	1,22	ERX2285181	ERS2068693	ERP105943	Yellow stage coffee seed transcriptome replicate 2	97,6%	11.656.431
SY3	ERR2231840	247	1.568	SAMEA104450752	0,63	ERX2285182	ERS2068694	ERP105943	Yellow stage coffee seed transcriptome replicate 3	97,5%	6.009.040
SR1	ERR2231841	246	935	SAMEA104450753	0,37	ERX2285183	ERS2068695	ERP105943	Red stage coffee seed transcriptome replicate 1	97,7%	3.607.771
SR2	ERR2231842	244	575	SAMEA104450754	0,23	ERX2285184	ERS2068696	ERP105943	Red stage coffee seed transcriptome replicate 2	97,8%	2.247.207
SR3	ERR2231843	246	655	SAMEA104450755	0,26	ERX2285185	ERS2068697	ERP105943	Red stage coffee seed transcriptome replicate 3	97,5%	2.517.279

7.4 SUPPLEMENTARY FIG. S1

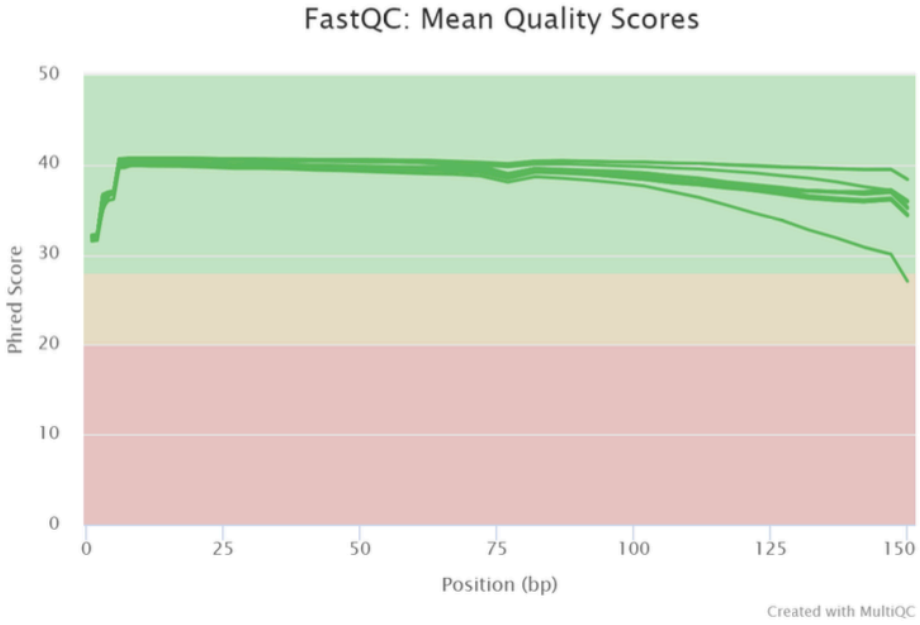


Fig. S1 RNA-seq Reads phred score reported by MultiQC after TRIMMOMATIC quality control.

7.5 SUPPLEMENTARY FIG. S2

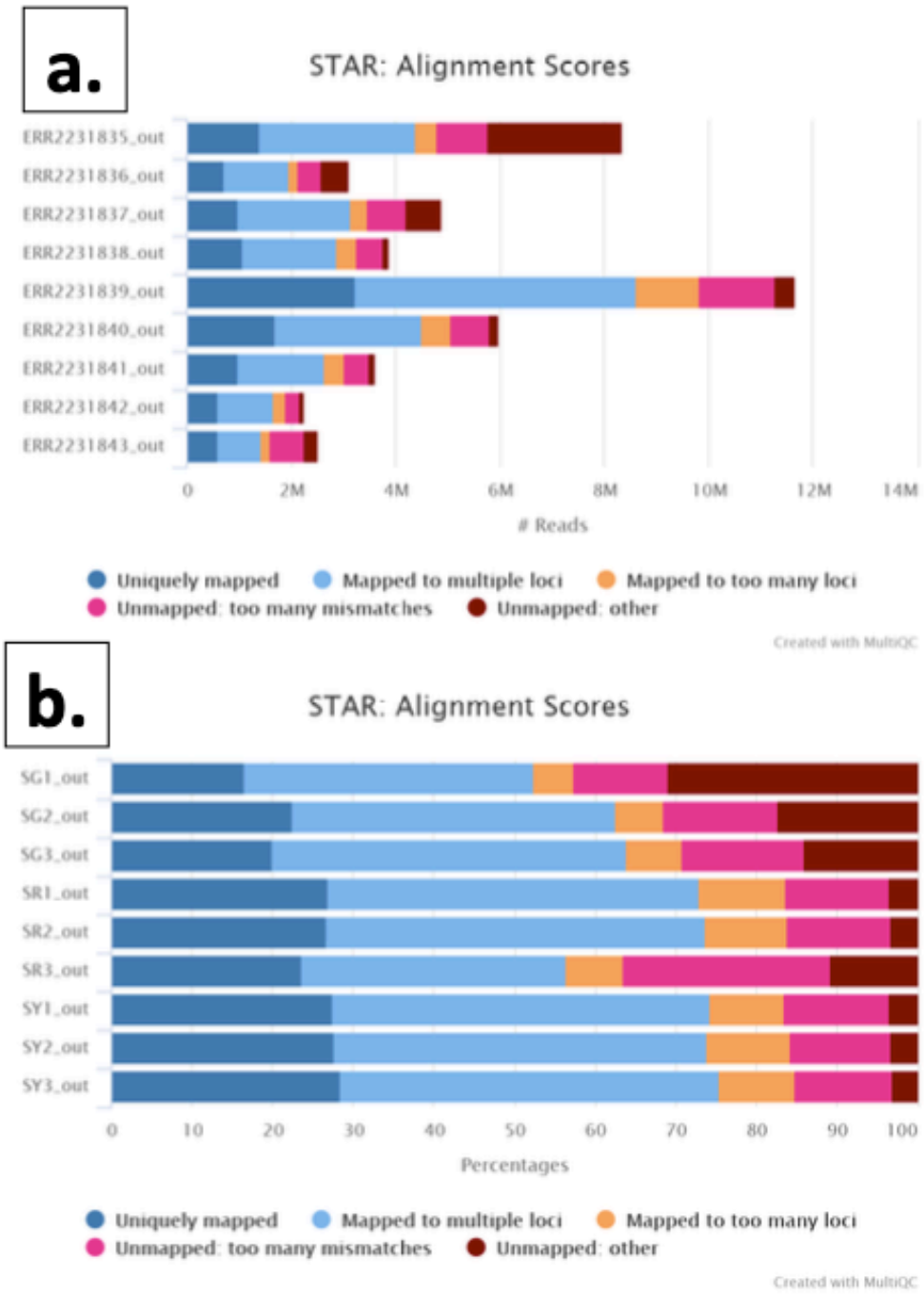
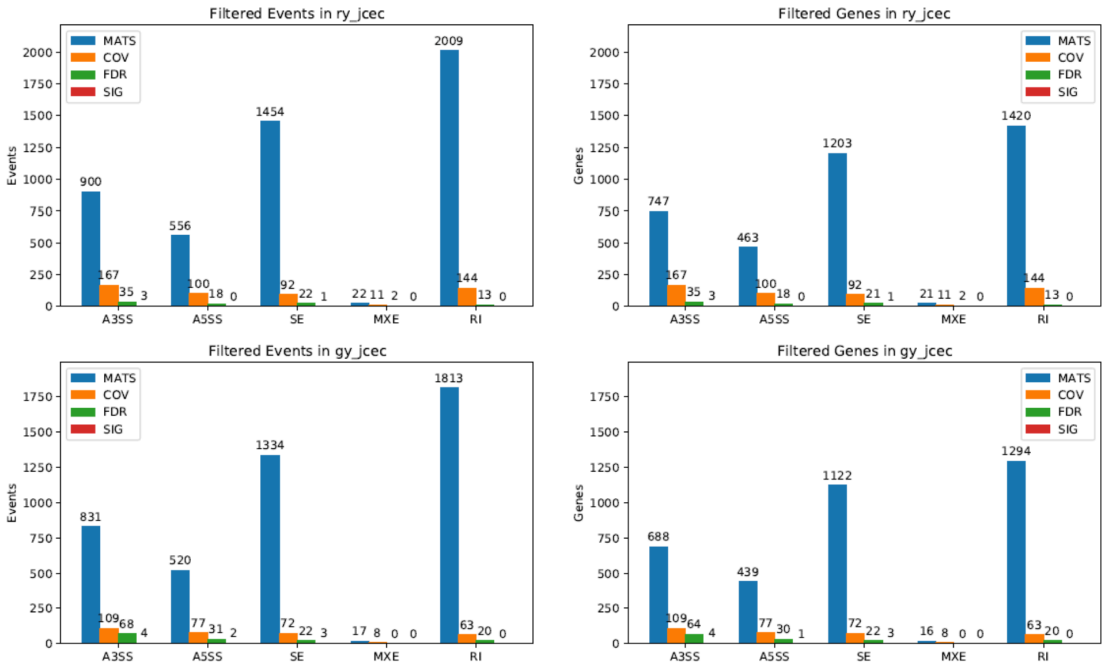


Fig. S2 Star scores of transcriptome alignment by genome reported by MultiQC. In a. is shown the aligned reads quantity and in b. the percentage of type of alignment by STAR.

7.6 SUPPLEMENTARY FIG. S3

Fig. S3 Events and gene quantification in GY and RY comparation analyses by rMATS obtained from plot_events.py after filter significative events from filter_events.py. MATS: quantity reported by rMATS, COV: value with reads coverage greater then 1, FDR: value with significance <= 0.05 FDR by rMATS statistic calc, SIG: quantity with coverage and statistical significance.



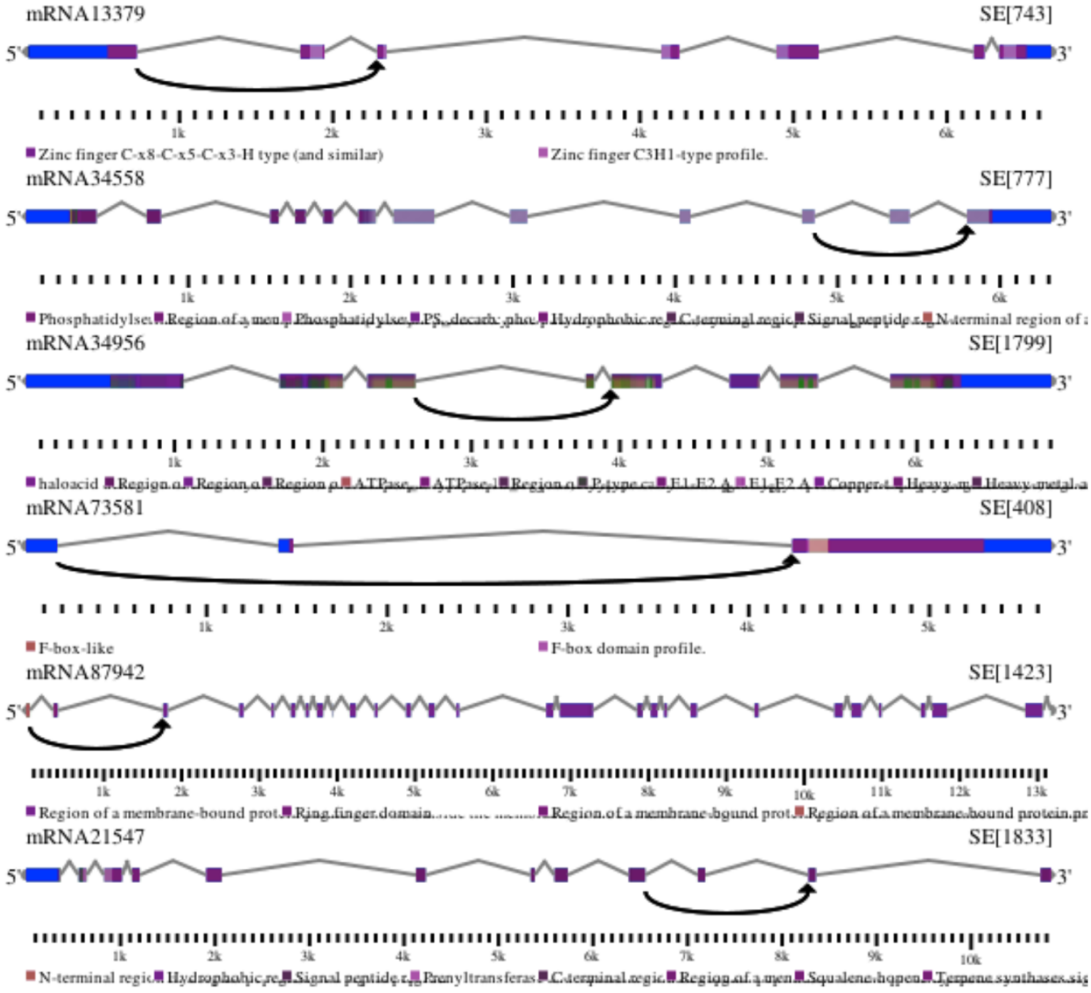
7.7 SUPPLEMENTARY TABLE S2

Table S2 GO Slim summary from anotatted genes undergo alternative splicing considering only Biologic Process aspect.

GO Term	Gene Ontology	Genes
GO:0009987	cellular process	111
GO:0008152	metabolic process	96
GO:0009058	biosynthetic process	45
GO:0006139	nucleobase-containing compound metabolic process	43
GO:0019538	protein metabolic process	37
GO:0006464	cellular protein modification process	32
GO:0016043	cellular component organization	32
GO:0007275	multicellular organism development	27
GO:0006950	response to stress	21
GO:0006810	transport	20
GO:0009791	post-embryonic development	18
GO:0000003	reproduction	17
GO:0007154	cell communication	16
GO:0006629	lipid metabolic process	15
GO:0009628	response to abiotic stimulus	15
GO:0007165	signal transduction	14
GO:0009056	catabolic process	13
GO:0009719	response to endogenous stimulus	13
GO:0009653	anatomical structure morphogenesis	11
GO:0009605	response to external stimulus	8
GO:0040007	growth	8
GO:0006259	DNA metabolic process	7
GO:0030154	cell differentiation	6
GO:0009908	flower development	5
GO:0040029	regulation of gene expression, epigenetic	5
GO:0005975	carbohydrate metabolic process	4
GO:0006412	translation	4
GO:0007049	cell cycle	4
GO:0009790	embryo development	4
GO:0009856	pollination	4
GO:0016049	cell growth	4
GO:0008219	cell death	3
GO:0009607	response to biotic stimulus	3
GO:0009991	response to extracellular stimulus	3
GO:0006091	generation of precursor metabolites and energy	1
GO:0009606	tropism	1
GO:0019725	cellular homeostasis	1

7.8 SUPPLEMENTARY FIG. S4

Fig. S4 Events of alternative splicing plotted by plotASdomains.py script. The back arrows indicate SE events identified by rMATS what generate alternative isoform of each gene. The blue region indicates an exon, gray an intron and purple coding region (CDS). The patterns in CDS indicate Pfam domains.



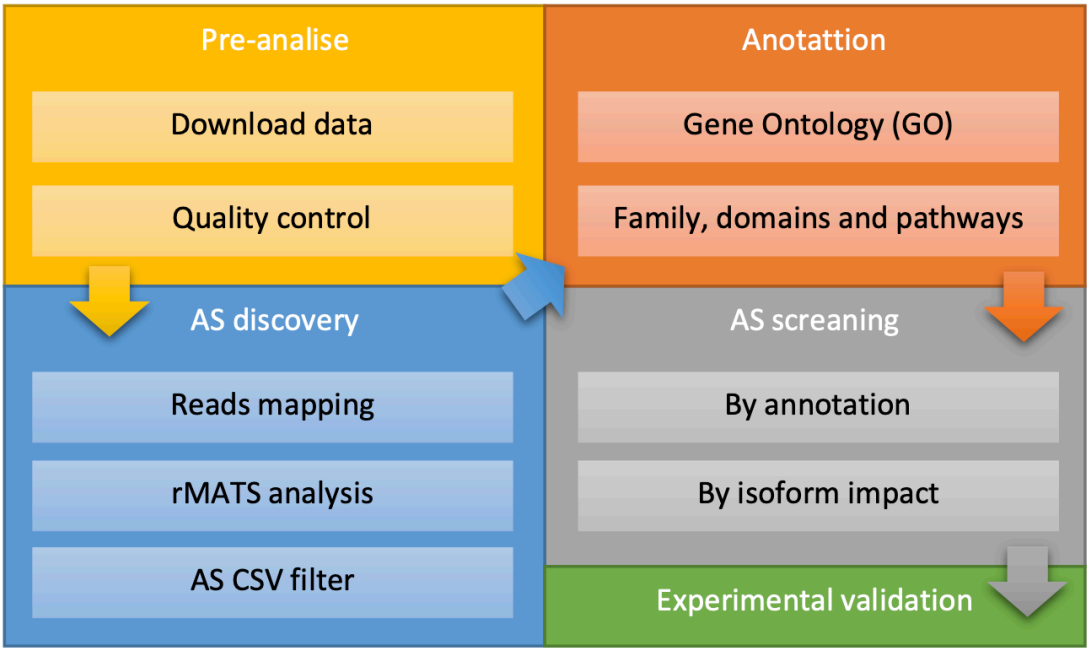
7.9 SUPPLEMENTARY TABLE S3

Table S3 Genes undergo AS anotattion.

EventID	Gene	FDR	Δ PSI	Event	Condition	mRNAs	Anotattion	Pathways	Dominios	GO	CLASS
451	gene14512	0.011787715715	-0.093	A3SS	GY	mRNA17334,mI	ubiquitin receptor	ko04141,ko0342	ProSiteProfiles: Ubiq	response to co	nucleobase-containing compound
1329	gene16741	0.020871054702	0.231	A3SS	GY	mRNA19988,mI	pre-mRNA-process	ko03040,map03	Pfam: FF domain|Pro	mRNA splicing,	nucleobase-containing compound
1339	gene2383	0.048486396056	-0.337	A3SS	GY	mRNA2842,mR	transcription initi	ko03022,map03	Phobius: Region of a	RNA polymeras	response to endogenous stimulus
1339	gene2383	0.048486396056	-0.337	A3SS	GY	mRNA2842,mR	transcription initi	ko03022,map03	Phobius: Region of a	RNA polymeras	signal transduction
1339	gene2383	0.048486396056	-0.337	A3SS	GY	mRNA2842,mR	transcription initi	ko03022,map03	Phobius: Region of a	RNA polymeras	nucleobase-containing compound
260	gene30238	0.000388050212	0.208	A3SS	GY	mRNA35932,mI	XP_027106644.1	ko01100,map00	Pfam: PIG-P|Phobius	GPI anchor bio	lipid metabolic process
1431	gene34245	0.004255948343	0.066	A3SS	GY	mRNA40710,mI	XP_027108022.1	map04510,mapC	ProSiteProfiles: Prot	protein serine/	anatomical structure morphogenese
1431	gene34245	0.004255948343	0.066	A3SS	GY	mRNA40710,mI	XP_027108022.1	map04510,mapC	ProSiteProfiles: Prot	protein serine/	response to abiotic stimulus
1054	gene40058	0.045055152058	0.506	A3SS	GY	mRNA47534,mI	carotene epsilon-c	map00906,ko01	Pfam: Cytochrome P4	carotenoid bio	lipid metabolic process
1201	gene43617	0.043761923102	-0.028	A3SS	GY	mRNA51872,mI	mitochondrial inn	ko03060,map03	Phobius: N-terminal	signal peptide	transport
1095	gene53362	0.006637490658	0.189	A3SS	GY	mRNA63425,mI	diacylglycerol kina	ko00561,ko0523	Phobius: Region of a	protein kinase	signal transduction
824	gene56392	0.032165311291	0.226	A3SS	GY	mRNA66952,mI	60S ribosomal pro	map03010,ko03	Pfam: Ribosomal L30	maturation of	nucleobase-containing compound
825	gene56392	0.017486245827	0.266	A3SS	GY	mRNA66952,mI	60S ribosomal pro	map03010,ko03	Pfam: Ribosomal L30	maturation of	nucleobase-containing compound
740	gene6505	0.004255948343	-0.237	A3SS	GY	mRNA7729,mR	probable 6-phosp	map00030,ko01	TIGRFAM: pgl: 6-pho	pentose-phosp	response to external stimulus
740	gene6505	0.004255948343	-0.237	A3SS	GY	mRNA7729,mR	probable 6-phosp	map00030,ko01	TIGRFAM: pgl: 6-pho	pentose-phosp	nucleobase-containing compound
1440	gene65950	0.022438062817	-0.152	A3SS	GY	mRNA78235,mI	biotin--protein lig	ko01100,ko0078	ProSiteProfiles: Bioti	cellular protein	response to external stimulus
773	gene66723	0.022438062817	-0.226	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to endogenous stimulus
774	gene66723	0.002150447229	0.405	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to endogenous stimulus
773	gene66723	0.022438062817	-0.226	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	signal transduction
774	gene66723	0.002150447229	0.405	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	signal transduction
773	gene66723	0.022438062817	-0.226	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to abiotic stimulus
774	gene66723	0.002150447229	0.405	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to abiotic stimulus
773	gene66723	0.022438062817	-0.226	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	multicellular organism development
774	gene66723	0.002150447229	0.405	A3SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	multicellular organism development
903	gene68399	0.023210165188	0.065	A3SS	GY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	response to external stimulus
903	gene68399	0.023210165188	0.065	A3SS	GY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	response to abiotic stimulus
903	gene68399	0.023210165188	0.065	A3SS	GY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	transport
903	gene68399	0.023210165188	0.065	A3SS	GY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	multicellular organism development
291	gene70542	0.043761923102	0.083	A3SS	GY	mRNA83738,mI	MADS-box transcri	map04022,mapC	ProSitePatterns: MAI	positive regula	response to abiotic stimulus
291	gene70542	0.043761923102	0.083	A3SS	GY	mRNA83738,mI	MADS-box transcri	map04022,mapC	ProSitePatterns: MAI	positive regula	multicellular organism development
291	gene70542	0.043761923102	0.083	A3SS	GY	mRNA83738,mI	MADS-box transcri	map04022,mapC	ProSitePatterns: MAI	positive regula	nucleobase-containing compound
1721	gene70831	0.020329334467	0.37	A3SS	GY	mRNA84070,mI	XP_027084320.1	map05131,ko03	ProSiteProfiles: Zinc	mRNA splicing,	nucleobase-containing compound
447	gene73859	0.025706061192	0.214	A3SS	GY	mRNA87679,mI	U4/U6 small nucle	ko03040,map03	Pfam: snoRNA bindin	spliceosomal t	nucleobase-containing compound
238	gene74509	0.011787715715	-0.264	A3SS	GY	mRNA88463,mI	small nuclear ribo	map05322,ko05	Pfam: LSM domain	spliceosomal s	nucleobase-containing compound
1053	gene76719	0.030699409002	-0.074	A3SS	GY	mRNA91127,mI	XP_027102156.1	ko01100,map00	PRINTS: Adrenodoxin	protoporphyri	multicellular organism development
1271	gene76790	0.012908968446	-0.069	A3SS	GY	mRNA91222,mI	splicing factor U2A	ko03040,map03	TIGRFAM: U2AF_lg: U	RNA splicing	nucleobase-containing compound
157	gene77426	0.012723425518	-0.063	A3SS	GY	mRNA91988,mI	serine/arginine-ric	map05168,ko03	Pfam: RNA recognitio	RNA splicing	regulation of gene expression
157	gene77426	0.012723425518	-0.063	A3SS	GY	mRNA91988,mI	serine/arginine-ric	map05168,ko03	Pfam: RNA recognitio	RNA splicing	nucleobase-containing compound
581	gene27402	0.049747953738	0.284	A5SS	GY	mRNA32602,mI	XP_027075628.1	ko00770,map04	Pfam: Cytidylyltransf	response to sal	response to abiotic stimulus
581	gene27402	0.049747953738	0.284	A5SS	GY	mRNA32602,mI	XP_027075628.1	ko00770,map04	Pfam: Cytidylyltransf	response to sal	lipid metabolic process
581	gene27402	0.049747953738	0.284	A5SS	GY	mRNA32602,mI	XP_027075628.1	ko00770,map04	Pfam: Cytidylyltransf	response to sal	transport
581	gene27402	0.049747953738	0.284	A5SS	GY	mRNA32602,mI	XP_027075628.1	ko00770,map04	Pfam: Cytidylyltransf	response to sal	nucleobase-containing compound
546	gene50370	0.013708140588	-0.19	A5SS	GY	mRNA59837,mI	protein RAE1-like	map05164,ko05	PRINTS: G protein be	transcription-c	transport
455	gene66723	0.025076680905	0.258	A5SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to endogenous stimulus
455	gene66723	0.025076680905	0.258	A5SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	signal transduction
455	gene66723	0.025076680905	0.258	A5SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to abiotic stimulus
455	gene66723	0.025076680905	0.258	A5SS	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	multicellular organism development
182	gene69372	0.003996912753	0.533	A5SS	GY	mRNA82389,mI	serine/arginine-ric	ko03040,map03	Pfam: RNA recognitio	RNA splicing	nucleobase-containing compound
313	gene6970	0.002359237925	0.212	A5SS	GY	mRNA8307,mR	SNF1-related prot	ko04151,map04	ProSiteProfiles: Kina	vegetative to re	signal transduction
846	gene71814	0.036915072968	-0.216	A5SS	GY	mRNA85236,mI	diacylglycerol kina	ko00561,ko0523	Pfam: Diacylglycerol	protein kinase	signal transduction
846	gene71814	0.036915072968	-0.216	A5SS	GY	mRNA85236,mI	diacylglycerol kina	ko00561,ko0523	Pfam: Diacylglycerol	protein kinase	lipid metabolic process
698	gene734	0.041075107049	0.092	A5SS	GY	mRNA907,mRN	WD40 repeat-cont	map04013,mapC	Pfam: LisH|Pfam: An	response to co	response to abiotic stimulus
698	gene734	0.041075107049	0.092	A5SS	GY	mRNA907,mRN	WD40 repeat-cont	map04013,mapC	Pfam: LisH|Pfam: An	response to co	nucleobase-containing compound
7386	gene44951	0.008480471128	-0.067	RI	GY	mRNA53481,mI	XP_027116294.1	ko00270,map00	ProSitePatterns: Cyst	cysteine biosyr	cell differentiation
7386	gene44951	0.008480471128	-0.067	RI	GY	mRNA53481,mI	XP_027116294.1	ko00270,map00	ProSitePatterns: Cyst	cysteine biosyr	anatomical structure morphogenese
186	gene6660	0.013759866139	0.347	RI	GY	mRNA7911,mR	serine/arginine-ric	ko03040,map03	Pfam: RNA recognitio	RNA splicing	nucleobase-containing compound
7314	gene73983	0.048441918892	-0.082	RI	GY	mRNA87834,mI	serine/threonine-p	ko04075,ko0415	ProSiteProfiles: Prot	protein phospl	response to endogenous stimulus
7314	gene73983	0.048441918892	-0.082	RI	GY	mRNA87834,mI	serine/threonine-p	ko04075,ko0415	ProSiteProfiles: Prot	protein phospl	signal transduction
7314	gene73983	0.048441918892	-0.082	RI	GY	mRNA87834,mI	serine/threonine-p	ko04075,ko0415	ProSiteProfiles: Prot	protein phospl	response to abiotic stimulus
7314	gene73983	0.048441918892	-0.082	RI	GY	mRNA87834,mI	serine/threonine-p	ko04075,ko0415	ProSiteProfiles: Prot	protein phospl	multicellular organism development
4309	gene76719	0.001632019622	-0.144	RI	GY	mRNA91127,mI	XP_027102156.1	ko01100,map00	PRINTS: Adrenodoxin	protoporphyri	multicellular organism development
908	gene1752	0.022499122155	0.177	SE	GY	mRNA2105,mR	serine/arginine-ric	map05168,map	Pfam: RNA recognitio	RNA splicing	nucleobase-containing compound
864	gene21193	0.03146027889	-0.398	SE	GY	mRNA25314	sphingosine-1-pho	ko00600,map04	Pfam: Pyridoxal-depe	sphingolipid m	lipid metabolic process
1735	gene29391	0.003503501681	0.258	SE	GY	mRNA34956	XP_027087717.1	ko01524,ko0401	Pfam: E1-E2 ATPase|	response to eth	response to endogenous stimulus
1735	gene29391	0.003503501681	0.258	SE	GY	mRNA34956	XP_027087717.1	ko01524,ko0401	Pfam: E1-E2 ATPase|	response to eth	signal transduction
1735	gene29391	0.003503501681	0.258	SE	GY	mRNA34956	XP_027087717.1	ko01524,ko0401	Pfam: E1-E2 ATPase|	response to eth	transport
1401	gene34245	0.0483941274	-0.146	SE	GY	mRNA40710,mI	XP_027108022.1	map04510,mapC	ProSiteProfiles: Prot	protein serine/	anatomical structure morphogenese
1401	gene34245	0.0483941274	-0.146	SE	GY	mRNA40710,mI	XP_027108022.1	map04510,mapC	ProSiteProfiles: Prot	protein serine/	response to abiotic stimulus
1003	gene5797	0.000140754386	-0.571	SE	GY	mRNA6900,mR	probable apyrase 7	map05169,ko05	Pfam: GDA1/CD39 (n	pollen exine fo	anatomical structure morphogenese
1003	gene5797	0.000140754386	-0.571	SE	GY	mRNA6900,mR	probable apyrase 7	map05169,ko05	Pfam: GDA1/CD39 (n	pollen exine fo	multicellular organism development
1003	gene5797	0.000140754386	-0.571	SE	GY	mRNA6900,mR	probable apyrase 7	map05169,ko05	Pfam: GDA1/CD39 (n	pollen exine fo	nucleobase-containing compound
1408	gene65950	0.010131205093	-0.156	SE	GY	mRNA78235,mI	biotin--protein lig	ko01100,ko0078	ProSiteProfiles: Bioti	cellular protein	response to external stimulus
766	gene66723	0.009222392237	0.296	SE	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to endogenous stimulus
766	gene66723	0.009222392237	0.296	SE	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	signal transduction
766	gene66723	0.009222392237	0.296	SE	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	response to abiotic stimulus
766	gene66723	0.009222392237	0.296	SE	GY	mRNA79172,mI	[Coffea arabica]	ko04075,ko0415	Pfam: CRM1 / Expor	ribosomal sma	multicellular organism development
451	gene14512	0.002356036343	-0.123	A3SS	RY	mRNA17334,mI	ubiquitin receptor	ko04141,ko0342	ProSiteProfiles: Ubiq	response to co	nucleobase-containing compound
927	gene29671	0.013130834165	0.226	A3SS	RY	mRNA35278,mI	XP_027072716.1	ko04712,map04	Pfam: CCT motif| Pro	rhythmic proc	signal transduction
927	gene29671	0.013130834165	0.226	A3SS	RY	mRNA35278,mI	XP_027072716.1	ko04712,map04	Pfam: CCT motif| Pro	rhythmic proc	response to abiotic stimulus
927	gene29671	0.013130834165	0.226	A3SS	RY	mRNA35278,mI	XP_027072716.1	ko04712,map04	Pfam: CCT motif| Pro	rhythmic proc	multicellular organism development
927	gene29671	0.013130834165	0.226	A3SS	RY	mRNA35278,mI	XP_027072716.1	ko04712,map04	Pfam: CCT motif| Pro	rhythmic proc	nucleobase-containing compound
1181	gene37851	0.037785503704	0.156	A3SS	RY	mRNA44932,mI	transcription initi	map05168,ko03	ProSiteProfiles: Trp-A	nucleus	nucleobase-containing compound
906	gene68399	0.044549725805	0.051	A3SS	RY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	response to external stimulus
906	gene68399	0.044549725805	0.051	A3SS	RY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	response to abiotic stimulus
906	gene68399	0.044549725805	0.051	A3SS	RY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	transport
906	gene68399	0.044549725805	0.051	A3SS	RY	mRNA81188,mI	sorting nexin 1-like	ko04144,map04	Pfam: Vps5 C termina	root developm	multicellular organism development
70	gene12648	0.002097093413	0.169	A5SS	RY	mRNA15053,mI	XP_027116294.1	ko00270,map00	ProSitePatterns: Cyst	cysteine biosyr	cell differentiation
70	gene12648	0.002097093413	0.169	A5SS	RY	mRNA15053,mI	XP_027116294.1	ko00270,map00	ProSitePatterns: Cyst	cysteine biosyr	anatomical structure morphogenese
778	gene60970	0.012924564941	0.11	A5SS	RY	mRNA72391,mI	protein kinase kin	ko04010,map04	ProSitePatterns: Prot	sugar mediated	response to endogenous stimulus
778	gene60970	0.012924564941	0.11	A5SS	RY	mRNA72391,mI	RecName: protein	ko04010,map04	ProSitePatterns: Prot	sugar mediated	signal transduction
778	gene60970	0.012924564941	0.11	A5SS	RY	mRNA72391,mI	RecName: protein	ko04010,map04	ProSitePatterns: Prot	sugar mediated	response to abiotic stimulus
778	gene60970	0.012924564941	0.11	A5SS	RY	mRNA72391,mRNA72392		ko04010,map04	ProSitePatterns: Prot	sugar mediated	lipid metabolic process
778	gene60970	0.012924564941	0.11	A5SS	RY	mRNA72391,mI	RecName: protein	ko04010,map04	ProSitePatterns: Prot	sugar mediated	multicellular organism development
4500	gene14520	0.002074227607	0.306	RI	RY	mRNA17343,mI	probable uridine r	ko00760,ko0024	Pfam: Inosine-uridin	purine nucleos	multicellular organism development
4500	gene14520	0.002074227607	0.306	RI	RY	mRNA17343,mI	probable uridine r	ko00760,ko0024	Pfam: Inosine-uridin	purine nucleos	nucleobase-containing compound
1833	gene18016	0.015558315575	0.145	SE	RY	mRNA21547	cycloartenol Synth	ko01100,ko0010	ProSitePatterns: Terp	cycloartenol sy	lipid metabolic process
1833	gene18016	0.015558315575	0.145	SE	RY	mRNA21547	cycloartenol Synth	ko01100,ko0010	ProSitePatterns: Terp	cycloartenol sy	multicellular organism development
1799	gene29391	0.005321540240	0.258	SE	RY	mRNA34956	XP_027087717.1	ko01524,ko0401	Pfam: E1-E2 ATPase|	response to eth	response to endogenous stimulus
1799	gene29391	0.005321540240	0.258	SE	RY	mRNA34956	XP_027087717.1	ko01524,ko0401	Pfam: E1-E2 ATPase|	response to eth	signal transduction
1799	gene29391	0.005321540240	0.258	SE	RY	mRNA34956	XP_027087717.1	ko01524,ko0401	Pfam: E1-E2 ATPase|	response to eth	transport

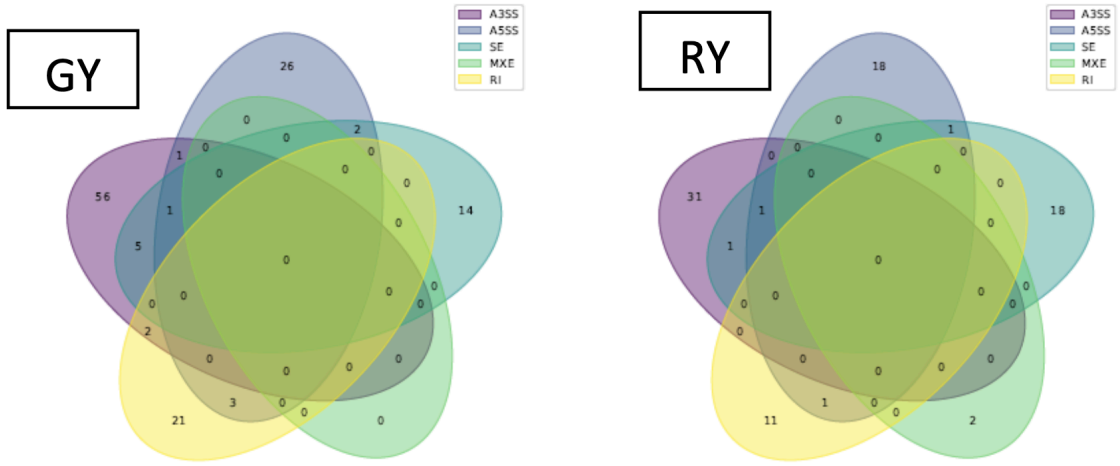
7.10 SUPPLEMENTARY FIG. S5

Fig. S5 Workflow used in this work. First, the fasta and gff3 genome files was downloaded, after the proteome was obtained with aux of gff-tools software. The RNASeq was downloaded from SRA and submitted to quality control for genome mapping. Next, the BAM files with mapping reads were submitted to rMATS analysis to get a csv file what was analyzed with python scripts to filter relevant events and plot quantitative graphics. The proteins of filtered events were annotated in EggNOG-mapper, NR, SwissProt and interproscan5 for determination of gene annotation and ontology, protein domain identification and associated pathways. With annotated proteins, AS events related to high interesse biological process was chosen with alteration happening in region with functional domain identified by InterproScan5. Between this, two events were selected of IR and SE types for experimental validation by PCR and real time-PCR.



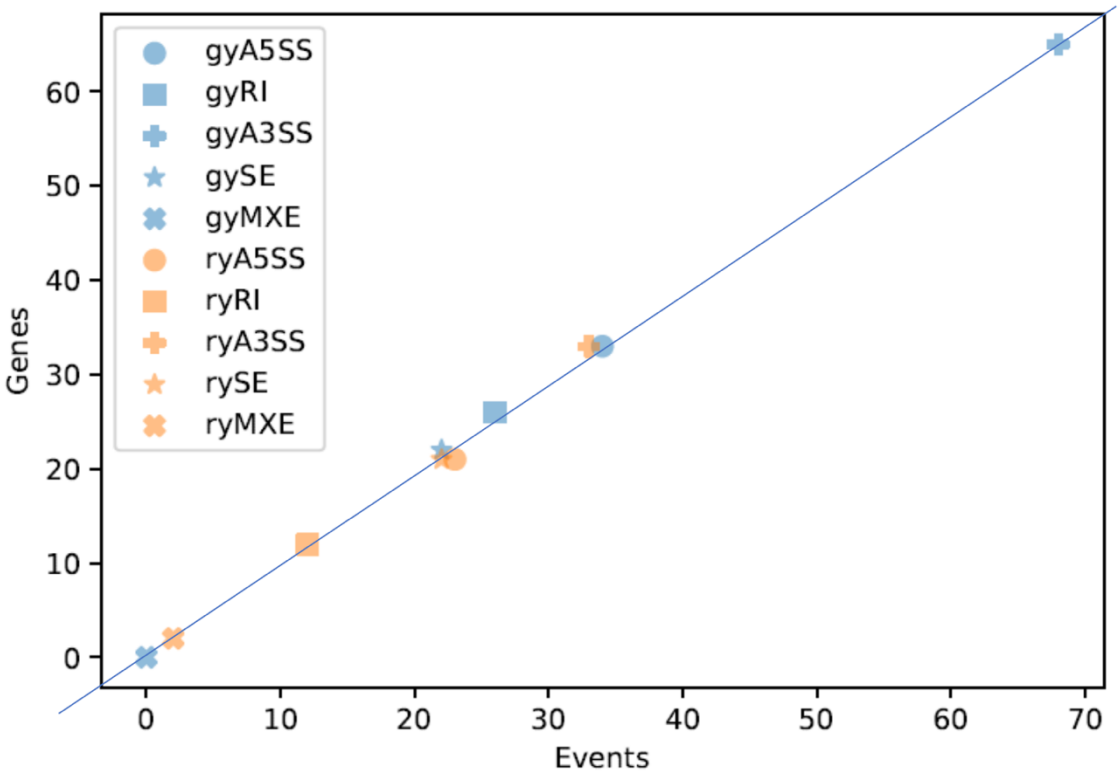
7.11 SUPPLEMENTARY FIG. S6

Fig. S6 Shared genes undergo AS in different event types in two compared conditions: Green x Yellow and Red x Yellow.



7.12 SUPPLEMENTARY FIG. S7

Fig. S7 Quantity of gene *versus* events relation.



7.13 SUPPLEMENTARY TABLE S4

Table S4 Distribution of AS genes into *C. arabica* genome.

Chr	Genes
1c	38
1e	38
2c	50
2e	52
3c	31
3e	33
4c	31
4e	38
5c	30
5e	27
6c	50
6e	52
7c	41
7e	40
8c	28
8e	28
9c	20
9e	25
10c	30
10e	33
11c	32
11e	24
Unknown	18

7.14 SUPPLEMENTARY TABLE S5

Table S5 Gene ontology annotation by Orthovenn2 analysis

Gene	GO Aspect	GO term	Description
gene32750	F	GO:0017112	Rab guanyl-nucleotide exchange factor activity
gene22710	P	GO:0046777	protein autophosphorylation
gene60403	P	GO:0015031	protein transport
gene56006	P	GO:0015031	protein transport
gene47482	P	GO:0015031	protein transport
gene43160	P	GO:0015031	protein transport
gene22839	P	GO:0015031	protein transport
gene43153	P	GO:0015031	protein transport
gene32818	P	GO:0015031	protein transport
gene38154	P	GO:0015031	protein transport
gene23340	P	GO:0015031	protein transport
gene51461	P	GO:0009908	flower development
gene43183	C	GO:0005886	plasma membrane
gene42154	C	GO:0005886	plasma membrane
gene8729	P	GO:0055114	oxidation-reduction process
gene70956	F	GO:0004722	protein serine/threonine phosphatase activity
gene55742	P	GO:0046653	tetrahydrofolate metabolic process
gene76795	P	GO:0010029	regulation of seed germination
gene65931	P	GO:0008380	RNA splicing
gene76790	P	GO:0008380	RNA splicing
gene14345	P	GO:0008380	RNA splicing
gene75730	P	GO:0010119	regulation of stomatal movement
gene15529	F	GO:0004842	ubiquitin-protein transferase activity
gene64643	F	GO:0016597	amino acid binding
gene21492	P	GO:0071076	RNA 3' uridylation
gene69654	F	GO:0015267	channel activity
gene65950	P	GO:0006464	cellular protein modification process
gene65679	F	GO:0046872	metal ion binding
gene11260	F	GO:0046872	metal ion binding
gene41374	F	GO:0046872	metal ion binding
gene58866	P	GO:0009409	response to cold
gene71814	P	GO:0007205	protein kinase C-activating G protein-coupled receptor signaling pathway
gene53362	P	GO:0007205	protein kinase C-activating G protein-coupled receptor signaling pathway
gene14123	P	GO:0016255	attachment of GPI anchor to protein
gene16068	F	GO:0004523	RNA-DNA hybrid ribonuclease activity
gene62527	P	GO:0048364	root development
gene68399	P	GO:0048364	root development
gene53439	P	GO:0006535	cysteine biosynthetic process from serine
gene44951	P	GO:0006535	cysteine biosynthetic process from serine
gene12648	P	GO:0006535	cysteine biosynthetic process from serine
gene77388	P	GO:2001142	nicotinate transport
gene56392	P	GO:0000463	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 58S rRNA, LSU-rRNA)

gene77426	P	GO:0008380	RNA splicing
gene6660	P	GO:0008380	RNA splicing
gene1752	P	GO:0008380	RNA splicing
gene12324	P	GO:0008380	RNA splicing
gene69372	P	GO:0008380	RNA splicing
gene67924	F	GO:0004725	protein tyrosine phosphatase activity
gene41208	P	GO:0009737	response to abscisic acid
gene74040	P	GO:0006511	ubiquitin-dependent protein catabolic process
gene21015	P	GO:0006511	ubiquitin-dependent protein catabolic process
gene37751	P	GO:0010114	response to red light
gene36349	P	GO:0010182	sugar mediated signaling pathway
gene53254	C	GO:0016021	integral component of membrane
gene36792	C	GO:0016021	integral component of membrane
gene57633	C	GO:0016021	integral component of membrane
gene33659	C	GO:0016021	integral component of membrane
gene29208	C	GO:0016021	integral component of membrane
gene76625	P	GO:0090065	regulation of production of siRNA involved in RNA interference
gene40487	P	GO:0043666	regulation of phosphoprotein phosphatase activity
gene10428	P	GO:0016567	protein ubiquitination
gene9026	P	GO:0016567	protein ubiquitination
gene30238	P	GO:0006506	GPI anchor biosynthetic process
gene56884	P	GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline
gene39887	P	GO:0072583	clathrin-dependent endocytosis
gene39635	P	GO:0015798	myo-inositol transport
gene16741	P	GO:0000398	mRNA splicing, via spliceosome
gene70831	P	GO:0000398	mRNA splicing, via spliceosome
gene61405	P	GO:0000398	mRNA splicing, via spliceosome
gene6505	P	GO:0006098	pentose-phosphate shunt
gene50370	P	GO:0000972	transcription-dependent tethering of RNA polymerase II gene DNA at nuclear periphery
gene29671	P	GO:0048511	rhythmic process
gene60945	P	GO:0016567	protein ubiquitination
gene61490	F	GO:0019905	syntaxin binding
gene7901	P	GO:0009308	amine metabolic process
gene8217	P	GO:0019252	starch biosynthetic process
gene28924	P	GO:0016226	iron-sulfur cluster assembly
gene67353	C	GO:0005739	mitochondrion
gene23862	P	GO:0019252	starch biosynthetic process
gene71329	F	GO:0004691	cAMP-dependent protein kinase activity
gene71768	P	GO:0007021	tubulin complex assembly
gene23573	P	GO:0006952	defense response
gene2010	P	GO:0006874	cellular calcium ion homeostasis
gene18780	F	GO:0043022	ribosome binding
gene34291	P	GO:0055085	transmembrane transport
gene6970	P	GO:0016032	viral process
gene63905	P	GO:0016032	viral process
gene58119	P	GO:0009820	alkaloid metabolic process

gene8619	C	GO:0005794	Golgi apparatus
gene37874	P	GO:0045944	positive regulation of transcription by RNA polymerase II
gene70542	P	GO:0045944	positive regulation of transcription by RNA polymerase II
gene35221	P	GO:0000038	very long-chain fatty acid metabolic process
gene48904	P	GO:0009733	response to auxin
gene59092	P	GO:0010090	trichome morphogenesis
gene14512	P	GO:0009409	response to cold
gene27402	P	GO:0009651	response to salt stress
gene34245	P	GO:0009651	response to salt stress
gene13351	P	GO:0006433	prolyl-tRNA aminoacylation
gene28716	F	GO:0016279	protein-lysine N-methyltransferase activity
gene48468	P	GO:0006486	protein glycosylation
gene59501	P	GO:0016485	protein processing
gene27502	P	GO:0006457	protein folding
gene63086	P	GO:0009911	positive regulation of flower development
gene38562	P	GO:0090351	seedling development
gene73983	P	GO:0090351	seedling development
gene40058	P	GO:0016117	carotenoid biosynthetic process
gene66723	P	GO:0006468	protein phosphorylation
gene73983	P	GO:0006468	protein phosphorylation
gene76719	P	GO:0006782	protoporphyrinogen IX biosynthetic process
gene60970	P	GO:0010182	sugar mediated signaling pathway
gene49121	P	GO:0009651	response to salt stress
gene73859	P	GO:0000244	spliceosomal tri-snRNP complex assembly
gene74509	P	GO:0000387	spliceosomal snRNP assembly
gene670	P	GO:0071596	ubiquitin-dependent protein catabolic process via the N-end rule pathway
gene34245	F	GO:0004674	protein serine/threonine kinase activity
gene43617	P	GO:0006465	signal peptide processing
gene66169	P	GO:0045927	positive regulation of growth
gene21956	P	GO:0010118	stomatal movement
gene61640	P	GO:0048767	root hair elongation
gene39886	P	GO:0006355	regulation of transcription, DNA-templated
gene42453	P	GO:0006355	regulation of transcription, DNA-templated
gene44862	P	GO:0006355	regulation of transcription, DNA-templated
gene59397	P	GO:0006355	regulation of transcription, DNA-templated
gene37851	C	GO:0005634	nucleus
gene51028	F	GO:0017112	Rab guanyl-nucleotide exchange factor activity
gene56882	P	GO:0009408	response to heat
gene11437	P	GO:0009744	response to sucrose
gene56302	P	GO:0031297	replication fork processing
gene70928	F	GO:0051015	actin filament binding
gene49283	P	GO:0042127	regulation of cell proliferation
gene32	P	GO:0071900	regulation of protein serine/threonine kinase activity
gene24835	P	GO:0006511	ubiquitin-dependent protein catabolic process
gene59392	F	GO:0016829	lyase activity
gene14123	P	GO:0016255	attachment of GPI anchor to protein

gene7100	P	GO:0006351	transcription, DNA-templated
gene67940	P	GO:0007275	multicellular organism development
gene66007	P	GO:0007275	multicellular organism development
gene20213	F	GO:0004672	protein kinase activity
gene66129	P	GO:0006890	retrograde vesicle-mediated transport, Golgi to ER
gene12895	P	GO:0006625	protein targeting to peroxisome
gene58475	P	GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline
gene31058	P	GO:0000387	spliceosomal snRNP assembly
gene72622	P	GO:0009626	plant-type hypersensitive response
gene22342	P	GO:0006629	lipid metabolic process
gene2776	F	GO:0015098	molybdate ion transmembrane transporter activity
gene60453	P	GO:0072593	reactive oxygen species metabolic process
gene64875	P	GO:0045039	protein import into mitochondrial inner membrane
gene29208	P	GO:0071786	endoplasmic reticulum tubular network organization
gene49302	P	GO:0051603	proteolysis involved in cellular protein catabolic process
gene2383	P	GO:0051123	RNA polymerase II preinitiation complex assembly
gene9537	P	GO:0001666	response to hypoxia
gene21657	P	GO:0009259	ribonucleotide metabolic process
gene1651	F	GO:0003723	RNA binding
gene59434	F	GO:0005227	calcium activated cation channel activity
gene34262	P	GO:0000050	urea cycle
gene66723	P	GO:0000056	ribosomal small subunit export from nucleus
gene29050	P	GO:0016540	protein autoprocessing
gene34225	P	GO:0000381	regulation of alternative mRNA splicing, via spliceosome
gene25024	P	GO:0042176	regulation of protein catabolic process
gene6732	P	GO:0000162	tryptophan biosynthetic process
gene42848	P	GO:0010030	positive regulation of seed germination
gene76371	P	GO:0006486	protein glycosylation
gene45233	P	GO:0045944	positive regulation of transcription by RNA polymerase II
gene5797	P	GO:0010584	pollen exine formation
gene4598	P	GO:0006417	regulation of translation
gene734	P	GO:0009409	response to cold
gene69689	P	GO:0009416	response to light stimulus
gene14135	P	GO:0009909	regulation of flower development
gene14520	P	GO:0006152	purine nucleoside catabolic process
gene21193	P	GO:0006665	sphingolipid metabolic process
gene29391	P	GO:0009723	response to ethylene
gene6970	P	GO:0010228	vegetative to reproductive phase transition of meristem
gene18016	F	GO:0016871	cycloartenol synthase activity
gene74066	P	GO:0010212	response to ionizing radiation
gene34578	F	GO:0001078	proximal promoter DNA-binding transcription repressor activity, RNA polymerase II-specific

7.15 SUPPLEMENTARY TABLE S6

Table S6 Table to mannualy choose best targets to validation considering genotipe influence.

Event ID	Gene	Via	mRNA	Tipo	SMP	Fenótipo ocorre em	Alteração	Fenótipo	GO	GO Slim
773	gene66723	**Thermogenesis|MAPK|Ribosome|Plant hormone signal transduction|PI3K-Akt signaling pathway**	mRNA79173	A3SS	GY	UP Y	1	ribosomal small subunit export from nucleus	?	response to endogenous stimulus
1721	gene70831	Spliceosome	mRNA84070	A3SS	GY	UP G	1	splicing factor U2af small subunit B-like	mRNA splicing, via spliceosome	nucleobase-containing compound
1339	gene2383	Basal transcription factors	mRNA2843	A3SS	GY	UP Y	1	transcription initiation factor TFIID subunit 12b-like isoform	RNA polymerase II preinitiation complex assembly	response to endogenous stimulus
455	gene66723	**Thermogenesis|MAPK|Ribosome|Plant hormone signal transduction|PI3K-Akt signaling pathway**	mRNA79175	A5SS	GY	UP G	1	ribosomal small subunit export from nucleus	?	response to endogenous stimulus
581	gene27402	SNARE|Pantothenate and CoA biosynthesis	mRNA32605	A5SS	GY	UP G	1	syntaxin-112-like	response to salt stress	response to abiotic stimulus
186	gene6660	Spliceosome	mRNA7912	RI	GY	UP G	1	serine/arginine-rich SC35-like splicing factor SCL33	RNA splicing	nucleobase-containing compound
1146	**gene42848**		**mRNA50894**	**RI**	**GY**	**UP Y**	**2**	**sugar transporter ERD6-like 6**	**positive regulation of seed germination**	**transport**
7314	**gene73983**	**Thermogenesis|MAPK|Ribosome|Plant hormone signal transduction|PI3K-Akt signaling pathway**	**mRNA87834**	**RI**	**GY**	**UP Y**	**2**	**serine/threonine-protein kinase SAPK2-like**	**protein phosphorylation**	**response to endogenous stimulus**
5968	**gene14135**		**mRNA16860**	**RI**	**GY**	**UP Y**	**2**	**ubiquitin-like-specific protease ESD4**	**regulation of flower development**	**multicellular organism development**
667	***gene61640***		***mRNA73174***	***RI***	***GY***	***UP Y***	***3***	***potassium channel AKT1-like***	***root hair elongation***	***cell differentiation***
4798	gene67924		mRNA80619	RI	GY	UP Y	1	tyrosine-protein phosphatase isoform	protein tyrosine phosphatase activity	response to external stimulus
1735	gene29391	**MAPK signaling pathway - plant**	mRNA34956	SE	GY	UP G	1	copper-transporting ATPase RAN1	response to ethylene	response to endogenous stimulus
1401	gene34245	Focal adhesion|Chemokine signaling pathway|PI3K-Akt signaling pathway|B cell receptor signaling pathway|Cell cycle	mRNA40712	SE	GY	UP Y	1	shaggy-related protein kinase alpha	protein serine/threonine kinase activity	anatomical structure morphogenese
1377	gene74066		mRNA87942	SE	GY	UP Y	1	E3 ubiquitin-protein ligase RFWD3-like	?	response to abiotic stimulus
1253	gene64875		mRNA76983	SE	GY	UP G	1	mitochondrial import inner membrane translocase subunit TIM10-like	protein import into mitochondrial inner membrane	transport
1003	***gene5797***	***Pyrimidine|Purine metabolism***	***mRNA6900***	***SE***	***GY***	***UP Y***	***3***	***probable apyrase 7***	***pollen exine formation***	***anatomical structure morphogenese***
202	gene36792		mRNA43686	SE	GY	UP G	1	protein YIF1B-like	integral component of membrane	transport
244	gene58278		mRNA69205	A3SS	RY	UP R	1	RNA polymerase Rpb4	?	nucleobase-containing compound
1113	gene38545		mRNA45758	A3SS	RY	UP Y	1	Pfam: Lipase (class 3)	?	lipid metabolic process
319	gene670		mRNA823	A5SS	RY	UP Y	1	E3 ubiquitin-protein ligase PRT6 isoform	ubiquitin-dependent protein catabolic process via the N-end rule pathway	regulation of gene expression
677	gene39635		mRNA47016	A5SS	RY	UP Y	1	inositol transporter 1-like isoform	myo-inositol transport	transport
778	gene60970	**MAPK signaling pathway**	mRNA72392	A5SS	RY	UP R	1	protein kinase	sugar mediated signaling pathway	response to endogenous stimulus
197	gene21015		mRNA25113	A5SS	RY	UP R	1	E3 ubiquitin-protein ligase Praja-2-like isoform	ubiquitin-dependent protein catabolic process	multicellular organism development
4805	gene67924		mRNA80619	RI	RY	UP Y	2	tyrosine-protein phosphatase isoform	protein tyrosine phosphatase activity	response to external stimulus

1782	*gene56302*		*mRNA66853*	*RI*	*RY*	*UP Y*	*3*	*SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like*	*replication fork processing*	*response to abiotic stimulus*
229	*gene75730*		*mRNA89907*	*RI*	*RY*	*UP R*	*3*	*mitochondrial pyruvate carrier 1-like*	*?*	*transport*
743	gene11260		mRNA13379	SE	RY	UP R	1	zinc finger CCCH domain-containing protein ZFN-like isoform	metal ion binding	nucleobase-containing compound
777	**gene29050**		**mRNA34558**	**SE**	**RY**	**UP Y**	**2**	**phosphatidylserine decarboxylase proenzyme 1, mitochondrial-like isoform**		**anatomical structure morphogenese**
1799	gene29391	**MAPK signaling pathway - plant**	mRNA34956	SE	RY	UP R	1	copper-transporting ATPase RAN1	response to ethylene	response to endogenous stimulus
1833	gene18016	Steroid biosynthesis	mRNA21547	SE	RY	UP R	1	cycloartenol Synthase isoform	cycloartenol synthase activity	lipid metabolic process

7.16 SUPPLEMENTARY TABLE S7

Table S7 GO enrichiment by Orthoven2 analysis.

GO	annotattion	evalue	Label	Condition	mRNA	Gene	ID evento	FDR	PSI	qVALUE	Significative
GO:0000162	tryptophan biosynthetic process	0.0263777706648816	A3SS	GY	mRNA8026	gene6732	12	0.0204487737644	-0.127	SIG	NO
GO:0000463	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	3,25E+10	A3SS	GY	mRNA66954	gene56392	824	0.0321653112911	0.226	SIG	NO
GO:0001666	response to hypoxia	1,67E+11	A3SS	GY	mRNA11319	gene9537	1713	0.0199864664836	0.201	SIG	NO
GO:0006098	pentose-phosphate shunt	0.0097997658971618	A3SS	GY	mRNA7730	gene6505	740	0.00425594834338	-0.237	SIG	NO
GO:0006355	regulation of transcription, DNA-templated	0.0004189803684544	A3SS	GY	mRNA53376	gene44862	954	0.000298002372356	0.469	SIG	NO
GO:0006417	regulation of translation	0.0001417440018311	A3SS	GY	mRNA5458	gene4598	153	0.000432653488097	-0.256	SIG	NO
GO:0006464	cellular protein modification process	7,96E+07	A3SS	GY	mRNA78241	gene65950	1440	0.0224380628178	-0.152	SIG	YES
GO:0006625	protein targeting to peroxisome	1,54E+10	A3SS	GY	mRNA15362	gene12895	1047	0.0127925699295	-0.128	SIG	YES
GO:0006629	lipid metabolic process	0.0002064231063006	A3SS	GY	mRNA26646	gene22342	489	0.0328981089518	-0.397	SIG	NO
GO:0006782	protoporphyrinogen IX biosynthetic process	2,55E+04	A3SS	GY	mRNA91131	gene76719	1053	0.0306994090023	-0.074	SIG	YES
GO:0007021	tubulin complex assembly	0.0097997658971618	A3SS	GY	mRNA85168	gene71768	841	0.00425594834338	0.223	SIG	NO
GO:0007205	protein kinase C-activating G-protein coupled receptor signaling pathway	1,95E+11	A3SS	GY	mRNA63426	gene53362	1095	0.00663749065835	0.189	SIG	NO
GO:0007275	multicellular organismal development	0.0200985865485847	A3SS	GY	mRNA78309	gene66007	1343	0.000317018847658	0.34	SIG	NO
GO:0008380	RNA splicing	0.0008892113859861	A3SS	GY	mRNA91990	gene77426	157	0.0127234255189	-0.063	SIG	NO
GO:0010090	trichome morphogenesis	0.0003956735037218	A3SS	GY	mRNA70186	gene59092	625	0.0230754563129	-0.184	SIG	NO
GO:0015031	protein transport	0.005352719878102	A3SS	GY	mRNA38996	gene32818	1177	4,88E+06	0.295	SIG	NO
GO:0015798	myo-inositol transport	2,51E+09	A3SS	GY	mRNA47022	gene39635	1093	0.00425594834338	0.325	SIG	YES
GO:0016117	carotenoid biosynthetic process	0.0001843657863878	A3SS	GY	mRNA47537	gene40058	1054	0.0450551520587	0.506	SIG	NO
GO:0016255	attachment of GPI anchor to protein	0.0007280966819781	A3SS	GY	mRNA16846	gene14123	1028	0.00172611765676	0.226	SIG	NO
GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline	0.0009523444949986	A3SS	GY	mRNA67541	gene56884	912	2,45E+04	0.67	SIG	NO
GO:0042780	tRNA 3'-end processing	0.0021259221806578	A3SS	GY	mRNA82380	gene69365	1234	0.0271276060069	-0.11	SIG	NO
GO:0043666	regulation of phosphoprotein phosphatase activity	0.0021259221806578	A3SS	GY	mRNA48045	gene40487	959	0.00721882091079	0.213	SIG	NO
GO:0045927	positive regulation of growth	0.0001843657863878	A3SS	GY	mRNA78499	gene66169	147	0.0111157244464	0.221	SIG	NO
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	0.0105984780443909	A3SS	GY	mRNA44967	gene37874	270	0.0259552829784	0.056	SIG	NO
GO:0051123	RNA polymerase II transcriptional preinitiation complex assembly	5,37E+09	A3SS	GY	mRNA2846	gene2383	1339	0.0484863960566	-0.337	SIG	YES
GO:0051603	proteolysis involved in cellular protein catabolic process	0.0076766010983047	A3SS	GY	mRNA58621	gene49302	518	0.0306994090023	0.255	SIG	NO

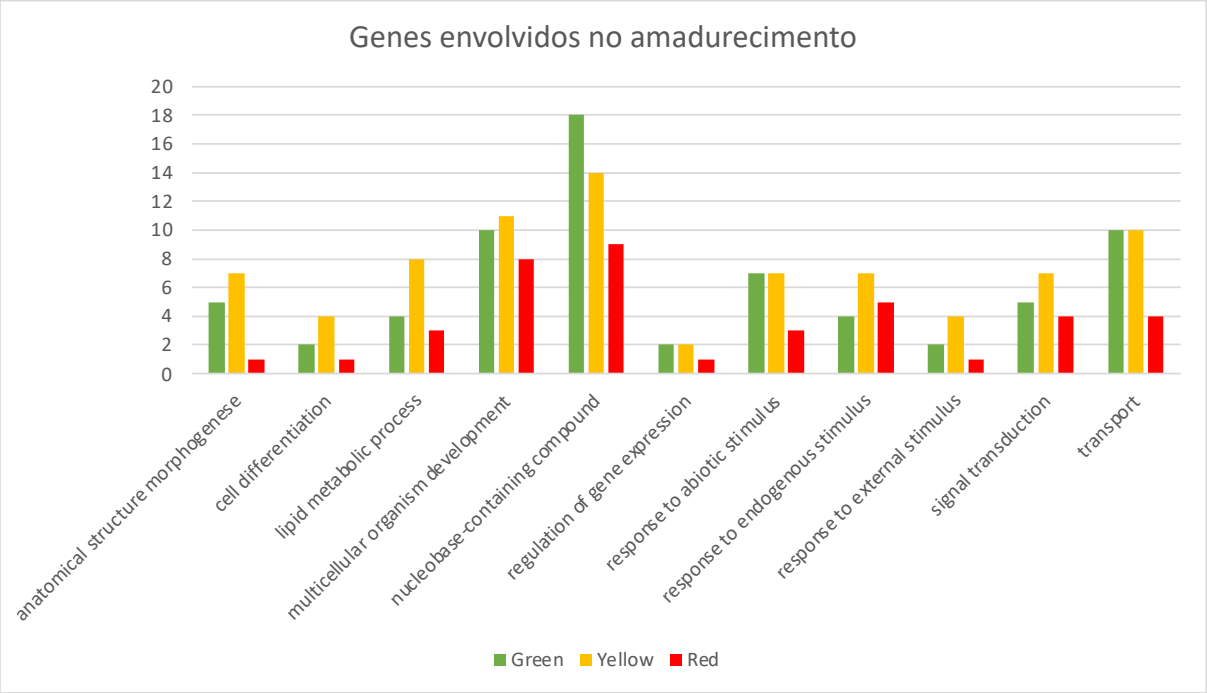
GO:0055085	transmembrane transport	0.0122506778364924	A3SS	GY	mRNA40774	gene34291	1445	0.0174862458276	0.252	SIG	NO
GO:0071786	endoplasmic reticulum tubular network organization	0.0003589393113826	A3SS	GY	mRNA34746	gene29208	1675	0.00658573079186	0.313	SIG	NO
GO:0000387	spliceosomal snRNP assembly	0.0060846502232073	A5SS	GY	mRNA36906	gene31058	665	0.046926321215	0.011	SIG	NO
GO:0006874	cellular calcium ion homeostasis	0.0002562196637923	A5SS	GY	mRNA2414	gene2010	180	0.0137081405885	-0.128	SIG	NO
GO:0007205	protein kinase C-activating G-protein coupled receptor signaling pathway	0.000507821058871	A5SS	GY	mRNA85237	gene71814	846	0.036915072968	-0.216	SIG	NO
GO:0008380	RNA splicing	0.0100151535712257	A5SS	GY	mRNA78211	gene65931	844	0.0416416201133	0.04	SIG	NO
GO:0009409	response to cold	0.0114059124115961	A5SS	GY	mRNA908	gene734	698	0.041075107049	0.092	SIG	NO
GO:0009651	response to salt stress	0.0128189058794498	A5SS	GY	mRNA32602	gene27402	581	0.0497479537384	0.284	SIG	NO
GO:0010029	regulation of seed germination	0.0002562196637923	A5SS	GY	mRNA91231	gene76795	216	0.00366387074714	0.36	SIG	NO
GO:0010114	response to red light	0.0001193531561871	A5SS	GY	mRNA44812	gene37751	769	0.0249577588804	0.114	SIG	NO
GO:0010228	vegetative to reproductive phase transition of meristem	0.0024801249048809	A5SS	GY	mRNA8308	gene6970	313	0.00235923792573	0.212	SIG	NO
GO:0015031	protein transport	0.0002538217399123	A5SS	GY	mRNA66499	gene56006	512	0.0150899207582	-0.112	SIG	NO
GO:0046777	protein autophosphorylation	0.000317635971034	A5SS	GY	mRNA27077	gene22710	353	0.00274083230755	0.188	SIG	NO
GO:0006782	protoporphyrinogen IX biosynthetic process	6,81E-02	RI	GY	mRNA91131	gene76719	4309	0.00163201962221	-0.144	SIG	YES
GO:0008380	RNA splicing	0.0127238047838268	RI	GY	mRNA78211	gene65931	5945	0.00384128621994	0.514	SIG	NO
GO:0009308	amine metabolic process	8,97E+08	RI	GY	mRNA9407	gene7901	6369	0.0277647490619	0.122	SIG	YES
GO:0009909	regulation of flower development	0.0032302822085562	RI	GY	mRNA16860	gene14135	5968	0.00384128621994	-0.444	SIG	NO
GO:0010030	positive regulation of seed germination	0.0005647865203837	RI	GY	mRNA50895	gene42848	1146	0.0228347985606	-0.221	SIG	NO
GO:0042127	regulation of cell proliferation	2,91E+07	RI	GY	mRNA58597	gene49283	6698	0.0352778508359	-0.301	SIG	YES
GO:0046777	protein autophosphorylation	0.0003714508289444	RI	GY	mRNA27077	gene22710	2420	0.00800691462275	0.191	SIG	NO
GO:0048767	root hair elongation	0.0005647865203837	RI	GY	mRNA73176	gene61640	667	0.0326059578288	-0.086	SIG	NO
GO:0090351	seedling development	4,88E+10	RI	GY	mRNA87838	gene73983	7314	0.0484419188925	-0.082	SIG	NO
GO:0000038	very long-chain fatty acid metabolic process	1,59E+08	SE	GY	mRNA41839	gene35221	1697	0.0483941274	-0.055	SIG	YES
GO:0006464	cellular protein modification process	3,13E+05	SE	GY	mRNA78241	gene65950	1408	0.0101312050932	-0.156	SIG	YES
GO:0006468	protein phosphorylation	0.0085318313434596	SE	GY	mRNA79172	gene66723	766	0.0092223922371	0.296	SIG	NO
GO:0006874	cellular calcium ion homeostasis	0.0001514520994892	SE	GY	mRNA2414	gene2010	234	0.000959587529935	-0.341	SIG	NO
GO:0008380	RNA splicing	0.0074263956570201	SE	GY	mRNA2105	gene1752	908	0.0224991221552	0.177	SIG	NO
GO:0009626	plant-type hypersensitive response	0.0026278473755156	SE	GY	mRNA86236	gene72622	1071	0.00303577677073	0.272	SIG	NO
GO:0010029	regulation of seed germination	0.0001514520994892	SE	GY	mRNA91231	gene76795	309	0.0483941274	0.325	SIG	NO
GO:0015031	protein transport	0.0074263956570201	SE	GY	mRNA45292	gene38154	1196	0.0113966927581	0.159	SIG	NO

GO:0045039	protein import into mitochondrial inner membrane	0.0007417896173718	SE	GY	mRNA76984	gene64875	1253	0.00915236304346	0.519	SIG	NO
GO:0006351	transcription, DNA-templated	0.0337416844247951	A3SS	RY	mRNA8473	gene7100	721	0.0135353860338	-0.338	SIG	NO
GO:0006417	regulation of translation	2,75E+09	A3SS	RY	mRNA5460	gene4598	154	0.049508122571	-0.164	SIG	YES
GO:0006486	protein glycosylation	0.0001291834225313	A3SS	RY	mRNA57574	gene48468	1621	0.023472429078	0.268	SIG	NO
GO:0007275	multicellular organismal development	0.0022437669514699	A3SS	RY	mRNA80651	gene67940	801	0.00121325010921	-0.586	SIG	NO
GO:0016032	viral process	1,22E+03	A3SS	RY	mRNA75851	gene63905	1150	0.0194107930084	0.41	SIG	YES
GO:0046777	protein autophosphorylation	0.0004499840243142	A3SS	RY	mRNA27077	gene22710	600	0.0494631805218	-0.478	SIG	NO
GO:0048364	root development	0.0130320872649912	A3SS	RY	mRNA81189	gene68399	906	0.044549725805	0.051	SIG	NO
GO:0048364	root development	0.0130320872649912	A3SS	RY	mRNA81188	gene68399	906	0.044549725805	0.051	SIG	NO
GO:0048511	rhythmic process	0.0052226910283397	A3SS	RY	mRNA35282	gene29671	927	0.0131308341653	0.226	SIG	NO
GO:0090351	seedling development	7,33E+05	A3SS	RY	mRNA45786	gene38562	1186	0.0377855037042	0.568	SIG	YES
GO:0006535	cysteine biosynthetic process from serine	1,08E+11	A5SS	RY	mRNA63516	gene53439	347	0.022850308837	0.478	SIG	NO
GO:0015031	protein transport	6,84E+10	A5SS	RY	mRNA66503	gene56006	515	0.0225635305701	-0.112	SIG	NO
GO:0015798	myo-inositol transport	1,02E+08	A5SS	RY	mRNA47022	gene39635	677	0.00251138690838	-0.148	SIG	YES
GO:0016226	iron-sulfur cluster assembly	1,39E+09	A5SS	RY	mRNA34400	gene28924	425	0.026759345337	0.211	SIG	YES
GO:0046653	tetrahydrofolate metabolic process	9,79E+06	A5SS	RY	mRNA66188	gene55742	992	0.0129245649415	-0.122	SIG	YES
GO:0071596	ubiquitin-dependent protein catabolic process via the N-end rule pathway	4,79E+10	A5SS	RY	mRNA824	gene670	318	0.0247891728858	-0.49	SIG	NO
GO:0006152	purine nucleoside catabolic process	0.0001434002042833	RI	RY	mRNA17343	gene14520	4500	0.00207422760786	0.306	SIG	NO
GO:0006486	protein glycosylation	0.0002938146636846	RI	RY	mRNA90654	gene76371	3073	0.0135588013568	-0.586	SIG	NO
GO:0006511	ubiquitin-dependent protein catabolic process	0.0050794091154707	RI	RY	mRNA29584	gene24835	1136	0.00423740409117	-0.5	SIG	NO
GO:0009308	amine metabolic process	1,11E+08	RI	RY	mRNA9407	gene7901	6376	0.0102011620957	0.201	SIG	YES
GO:0010119	regulation of stomatal movement	4,45E+08	RI	RY	mRNA89913	gene75730	229	3,50E+01	0.765	SIG	YES
GO:0015031	protein transport	0.0335106265229928	RI	RY	mRNA56409	gene47482	1372	0.0133544917839	0.203	SIG	NO
GO:0009626	plant-type hypersensitive response	0.0011161460560915	SE	RY	mRNA86236	gene72622	1096	0.00335267448925	0.272	SIG	NO
GO:0009820	alkaloid metabolic process	1,97E+10	SE	RY	mRNA69006	gene58119	1468	0.0175237444089	-0.218	SIG	NO
GO:0016540	protein autoprocessing	7,70E+10	SE	RY	mRNA34559	gene29050	777	0.0189570164559	-0.492	SIG	NO
GO:0019252	starch biosynthetic process	0.0013576175360553	SE	RY	mRNA28462	gene23862	4	0.00532154024096	-0.188	SIG	NO
GO:0046653	tetrahydrofolate metabolic process	1,95E+06	SE	RY	mRNA66190	gene55742	1748	0.0165549615968	-0.139	SIG	YES

7.17 SUPPLEMENTARY TABLE S8

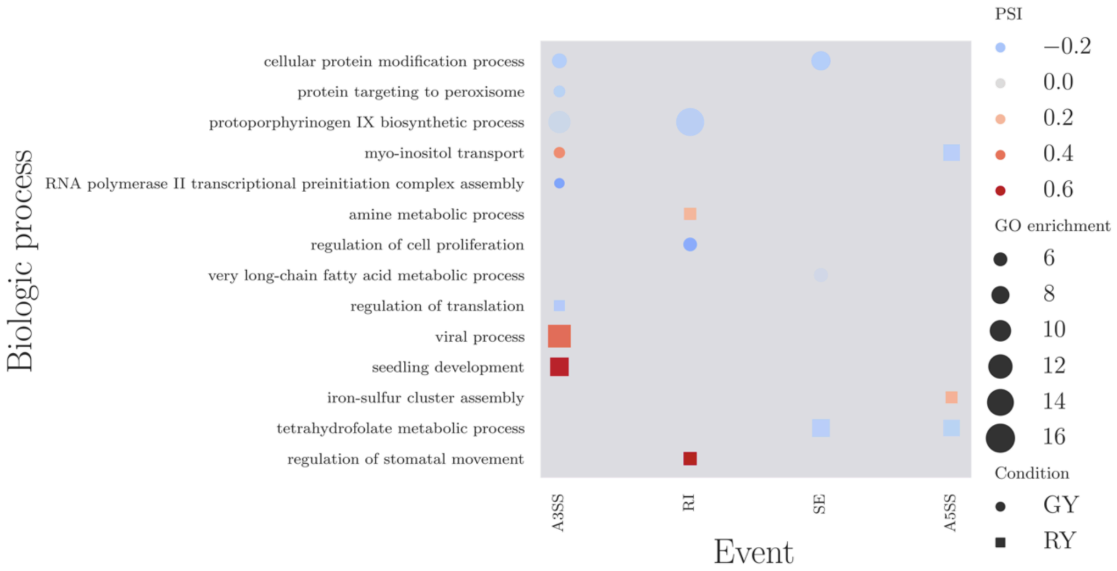
Table S8 GO grouped AS genes.

Ontology	Green	Yellow	Red
anatomical structure morphogenese	5	7	1
cell differentiation	2	4	1
lipid metabolic process	4	8	3
multicellular organism development	10	11	8
nucleobase-containing compound	18	14	9
regulation of gene expression	2	2	1
response to abiotic stimulus	7	7	3
response to endogenous stimulus	4	7	5
response to external stimulus	2	4	1
signal transduction	5	7	4
transport	10	10	4
Total redundant	69	81	40



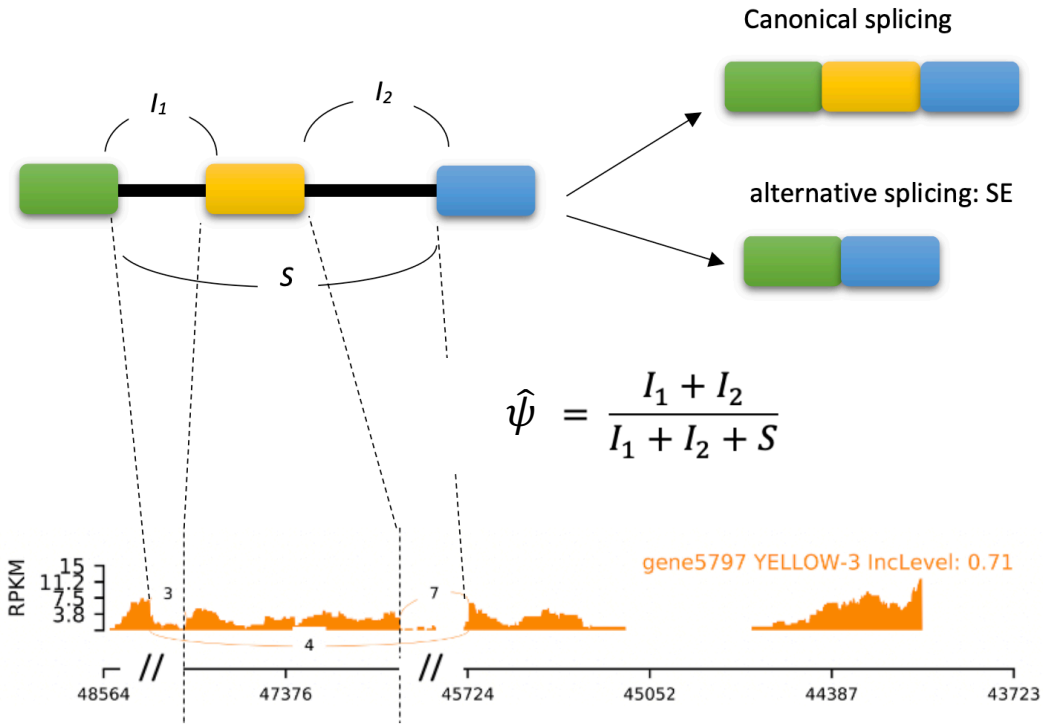
7.18 SUPPLEMENTARY FIG. S8

Fig. S8 GOs of biological process enriched in AS genes of GY and RY comparisons. The quantity greater of ΔPSI (red) indicate the alternative exon inclusion index in yellow condition. The size of form is the -log10(evalue) of enrichment of GO term calculated by Orthovenn2, only result with e-value < 0.05 are show.



7.19 SUPPLEMENTARY FIG. S9

Fig. S9 Sample of rMATS method to estimate inclusion level by reads count. First in SE alternative splice event reads count for exon alternative of isoform what includes the exon (I_1 and I_2 counts) are in relation to the total reads (I_1, I_2 and S counts) for estimate the percent spliced in $\hat{\psi}$. In the bottom image are 3 + 7 reads what supports exon inclusion, and 4 what support exon skipping, with 10 as total, then (3+7)/(3+7+4) = 0.71 of inclusion level.



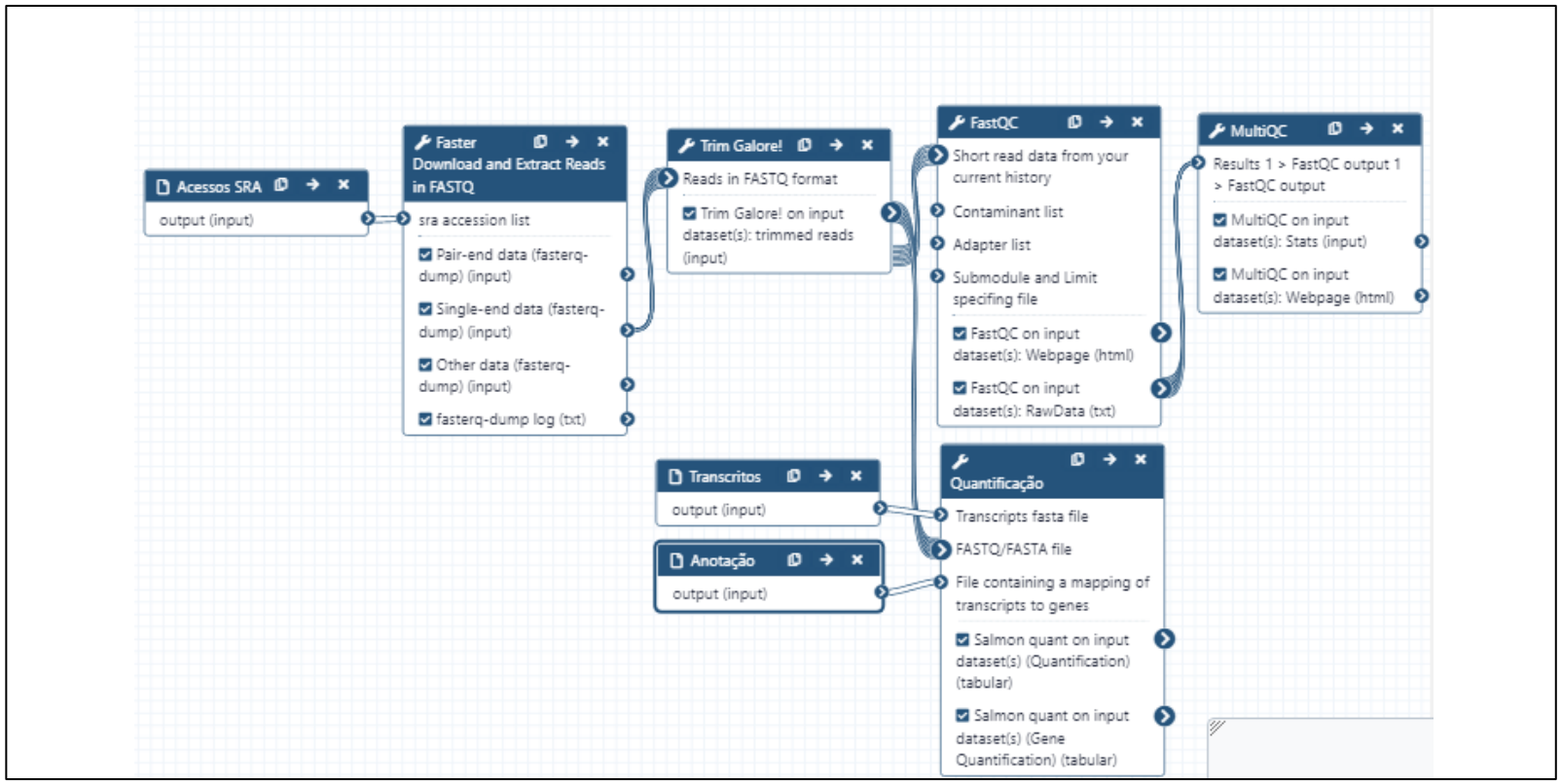
7.20 SUPPLEMENTARY TABLE S1

Suplemmentary Table 1. Raw data

Run	Bases	Bytes	dev_stage	Experiment	Library Name	Sample Name	Tissue	Sample	M Seqs	M Seqs after QC %
SRR5724093	692.032.608	290.319.665	MG-equatorial	SRX2940204	J35-1	J35	Total pericarp	MGE1	6,9	97,1%
SRR5724092	1.642.706.521	691.536.511	MG-equatorial	SRX2940205	J35-2	J35	Total pericarp	MGE2	16,3	97,5%
SRR5724091	633.155.769	264.629.044	MG-equatorial	SRX2940206	J35-3	J35	Total pericarp	MGE3	6,3	98,4%
SRR5724090	1.163.476.772	486.158.348	MG-equatorial	SRX2940207	J35-4	J35	Total pericarp	MGE4	11,5	98,3%
SRR5724049	1.316.924.254	545.372.095	RR	SRX2940248	J90-1	J90	Total pericarp	RR1	13,0	98,5%
SRR5724050	856.762.800	354.937.684	RR	SRX2940247	J90-2	J90	Total pericarp	RR2	8,5	97,6%
SRR5724042	889.503.162	371.457.495	RR	SRX2940255	J90-3	J90	Total pericarp	RR3	8,8	94,3%
SRR5724043	837.698.141	349.341.931	RR	SRX2940254	J90-4	J90	Total pericarp	RR4	8,3	96,4%

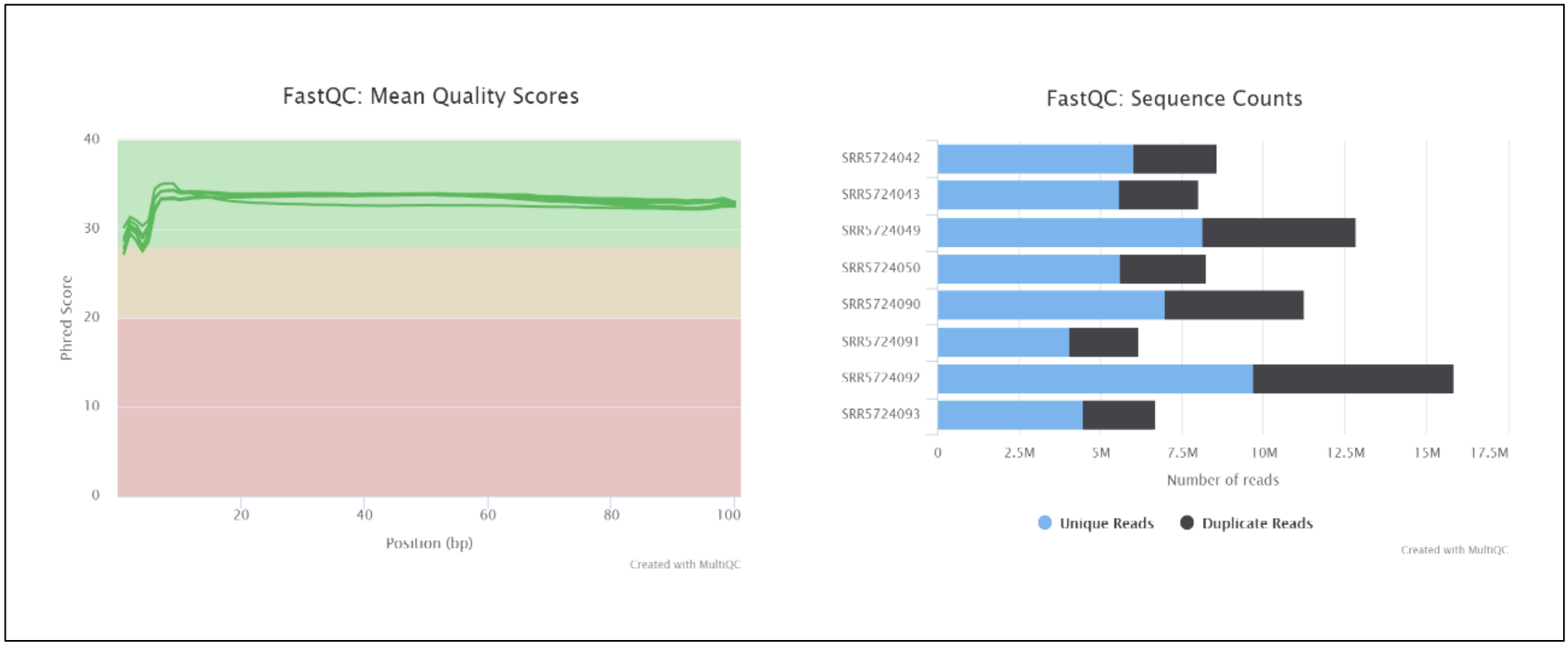
7.21 SUPPLEMENTARY FIG. S1

Supplementary figure 1. Galaxy workflow to perform and report automated quality control and Salmon transcripts quantify for reuse users need to input list of SRA accessions and transcript as fasta and table gene to transcript mapping, it is available on web server with GUI by Galaxy.EU



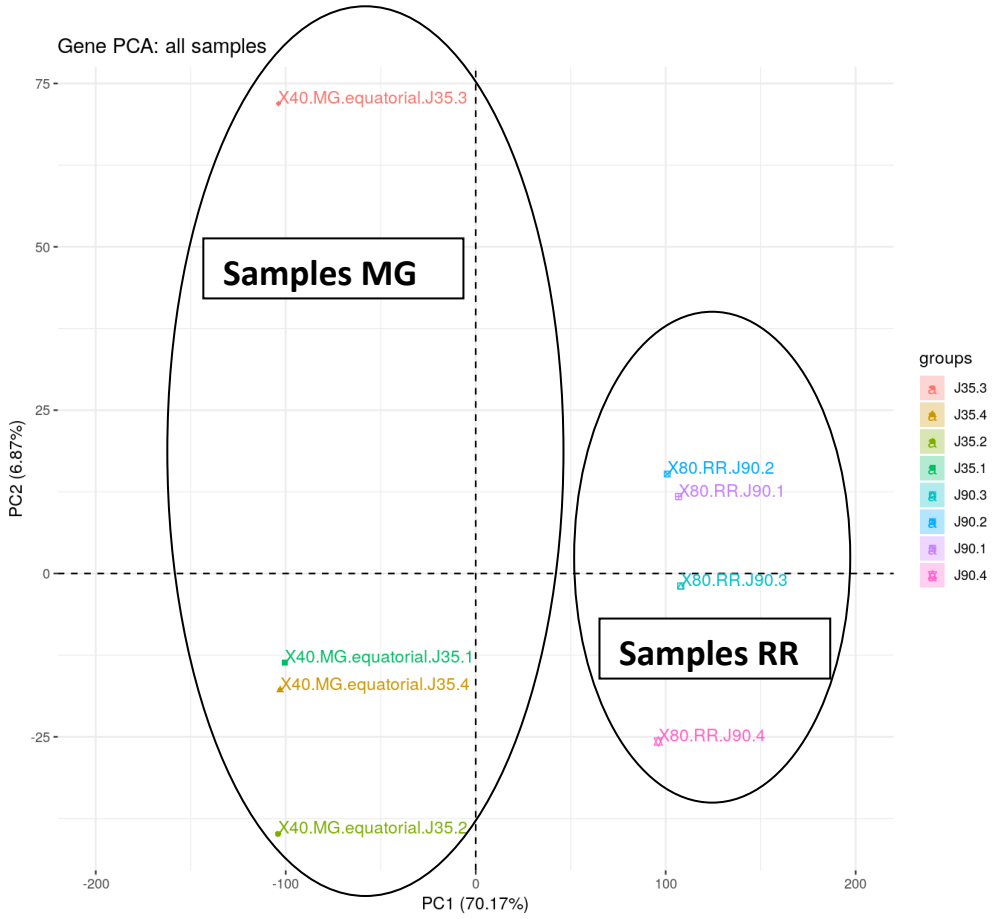
7.22 SUPPLEMENTARY FIG. S2

Supplementary figure 2. QC control reported by MULTIQC. On left Main quality scores reported in MultiQC after low quality reads removed by Trim Galore! In right Quantity of high-quality reads per sample.



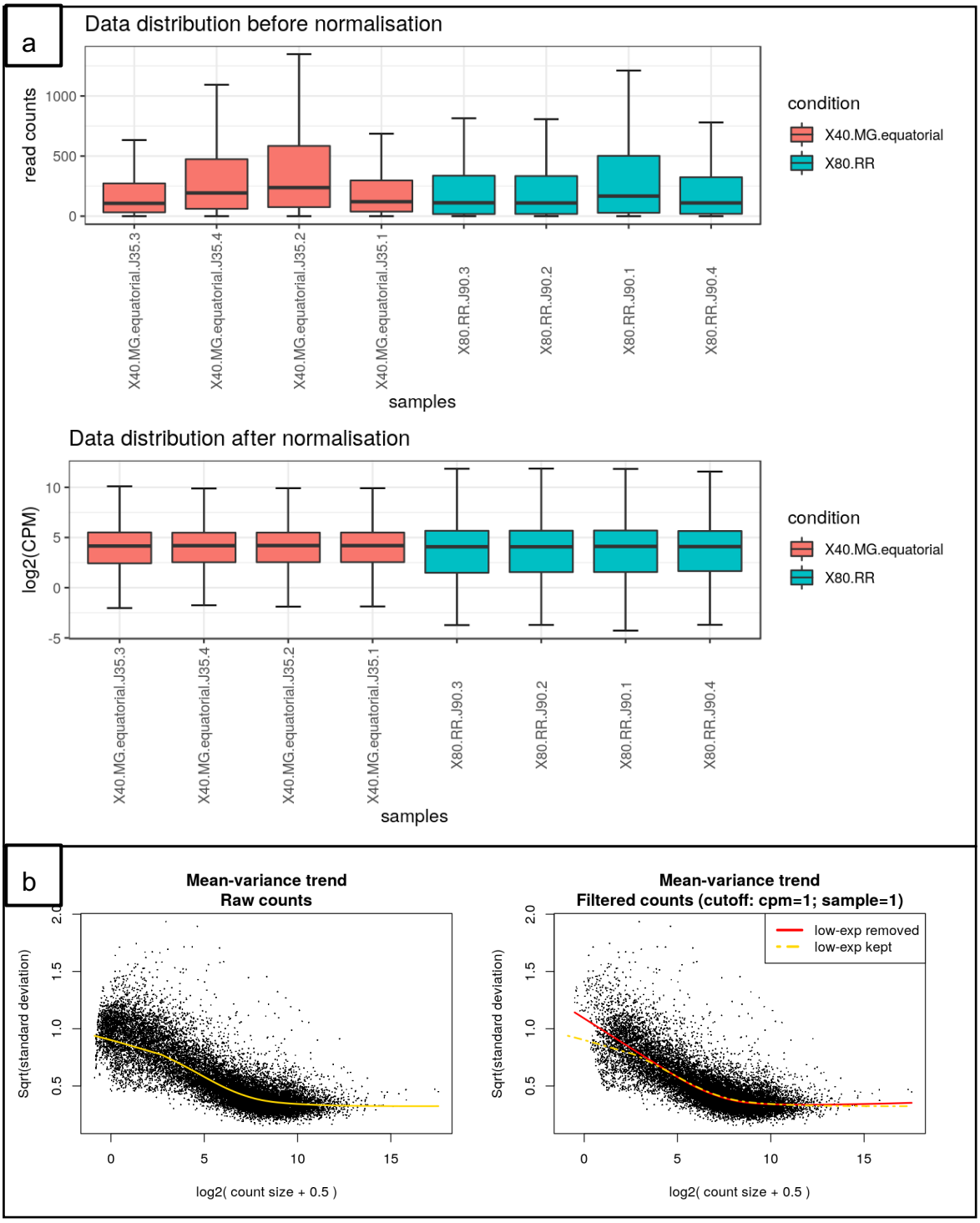
7.23 SUPPLEMENTARY FIG. S3

Supplementary figure 3. PCA of gene expression reported by 3DRnaseq.



7.24 SUPPLEMENTARY FIG. S4

Supplementary figure 4. (**a**) 3DRNASeq Gene expression normalization and (**b**) Gene mean-variance trend filter.



7.25 SUPPLEMENTARY TABLE S2

TABLE S2. Significant DAS genes list and statistics.

target	contrast	adj.pval	maxdeltaPS	log2FC	up.down
CYP85A1	X40.MG.equatorial-X80.RR	0.00132802759820421	-0.467350274057055	-7,39576E+14	down-regulated
LOC101244006	X40.MG.equatorial-X80.RR	0.000329921737183554	-0.5	3,47288E+14	up-regulated
LOC101244484	X40.MG.equatorial-X80.RR	0.00730132794334482	-0.657561541668782	-1,36351E+14	down-regulated
LOC101245201	X40.MG.equatorial-X80.RR	0.000876173399892903	-0.118840768571667	0.199916384343435	up-regulated
LOC101245278	X40.MG.equatorial-X80.RR	6995040213	-0.763258908578169	-1,22422E+14	down-regulated
LOC101245489	X40.MG.equatorial-X80.RR	63375109,62	1	0.60478455656748	up-regulated
LOC101245882	X40.MG.equatorial-X80.RR	0.00684575842170949	0.447685949080311	-0.214847647951949	down-regulated
LOC101245961	X40.MG.equatorial-X80.RR	0.00941284449697878	-0.32136012143764	0.221089693184056	up-regulated
LOC101246025	X40.MG.equatorial-X80.RR	44999736,86	0.344720444512384	-3,57029E+11	down-regulated
LOC101246155	X40.MG.equatorial-X80.RR	0.00816463341942904	-0.173432423570242	-0.632856974306865	down-regulated
LOC101246357	X40.MG.equatorial-X80.RR	0.000231695705219765	-0.134288606017892	1,39517E+14	up-regulated
LOC101246361	X40.MG.equatorial-X80.RR	0.00168998553850982	-0.170065978986949	2,91457E+14	up-regulated
LOC101246893	X40.MG.equatorial-X80.RR	529274414,2	-1	3,72838E+14	up-regulated
LOC101247345	X40.MG.equatorial-X80.RR	0.0091238858839879	0.176064078525636	0.304490728116299	up-regulated
LOC101247498	X40.MG.equatorial-X80.RR	636976678,6	-0.329964998995063	0.0451054993243565	up-regulated
LOC101247787	X40.MG.equatorial-X80.RR	0.00150158594279069	-0.818937648027599	-2,43926E+14	down-regulated
LOC101247954	X40.MG.equatorial-X80.RR	0.00740986239304256	0.241587620311712	-0.678234701442156	down-regulated
LOC101248222	X40.MG.equatorial-X80.RR	0.00206612169539083	0.411138123422376	5,00933E+14	up-regulated
LOC101248300	X40.MG.equatorial-X80.RR	0.000574826487663374	-0.291466994202044	4,50969E+13	up-regulated
LOC101248314	X40.MG.equatorial-X80.RR	0.000710311505791379	0.724192891810424	0.785496636300131	up-regulated
LOC101248986	X40.MG.equatorial-X80.RR	0.00207030724525686	0.20137936848199	-0.78847753092556	down-regulated
LOC101249078	X40.MG.equatorial-X80.RR	0.000619617787874801	-0.161141367494131	-0.252850671557649	down-regulated
LOC101249485	X40.MG.equatorial-X80.RR	6435257,639	1	2,70377E+14	up-regulated
LOC101251219	X40.MG.equatorial-X80.RR	0.00253496003831736	-0.137467665826225	-0.416051963750162	down-regulated
LOC101251722	X40.MG.equatorial-X80.RR	0.000715419264679302	0.224569420665278	-0.0762679257876089	down-regulated
LOC101252029	X40.MG.equatorial-X80.RR	0.00646870259082405	0.310414176340981	-1,28241E+14	down-regulated
LOC101252258	X40.MG.equatorial-X80.RR	0.00736555358944551	0.679884222987321	7,52718E+14	up-regulated
LOC101252360	X40.MG.equatorial-X80.RR	0.00176437570798233	-0.751899934650576	4,03768E+14	up-regulated
LOC101252559	X40.MG.equatorial-X80.RR	0.00916264630546761	0.387676413589818	-2,77317E+13	down-regulated
LOC101252600	X40.MG.equatorial-X80.RR	0.000176210688822232	0.602486585111368	-0.511358011359515	down-regulated
LOC101252721	X40.MG.equatorial-X80.RR	0.00141379560825733	-0.17415558358922	-0.0786748913775801	down-regulated
LOC101253077	X40.MG.equatorial-X80.RR	87224843,41	-0.12661739751613	-0.358476243263497	down-regulated
LOC101253473	X40.MG.equatorial-X80.RR	0.00512730094189678	0.44015088049505	1,08456E+14	up-regulated
LOC101253757	X40.MG.equatorial-X80.RR	0.00588987219277493	-0.69782194802896	2,03212E+14	up-regulated
LOC101253775	X40.MG.equatorial-X80.RR	43111952,58	-0.837360049006395	0.0391201958293479	up-regulated
LOC101253816	X40.MG.equatorial-X80.RR	0.000110690729865618	0.428374454814057	-1,62739E+14	down-regulated
LOC101253917	X40.MG.equatorial-X80.RR	0.00960729076861937	-0.106588020742381	-1,10346E+13	down-regulated
LOC101254006	X40.MG.equatorial-X80.RR	0.00684575842170949	0.399570286986169	0.206244225988753	up-regulated

LOC101254143	X40.MG.equatorial-X80.RR	59692127,5	0.638827638294371	0.10879555708721	up-regulated
LOC101254150	X40.MG.equatorial-X80.RR	0.00927760676771596	0.150154846272414	-0.34986988064342	down-regulated
LOC101254156	X40.MG.equatorial-X80.RR	0.00681376802549168	-0.477725745045094	-0.958683569088813	down-regulated
LOC101254260	X40.MG.equatorial-X80.RR	0.00469375256374101	-0.4106131700716	-0.482112031505362	down-regulated
LOC101255628	X40.MG.equatorial-X80.RR	0.00587070902408609	-1	0.251966803848006	up-regulated
LOC101255908	X40.MG.equatorial-X80.RR	0.000449330337287761	0.542413343141484	0.203166875885976	up-regulated
LOC101256051	X40.MG.equatorial-X80.RR	0.00719501086113107	-0.233118421713499	-0.477783221242567	down-regulated
LOC101256296	X40.MG.equatorial-X80.RR	0.00941284449697878	0.375587690993891	0.898021075923987	up-regulated
LOC101256522	X40.MG.equatorial-X80.RR	6435257,639	0.168022910772664	-1,14899E+13	down-regulated
LOC101256738	X40.MG.equatorial-X80.RR	162217365,4	0.111285031924536	-0.0665277984071562	down-regulated
LOC101256831	X40.MG.equatorial-X80.RR	0.00471157271959003	-0.228121749063029	0.289657541447692	up-regulated
LOC101257064	X40.MG.equatorial-X80.RR	0.00122981494091921	-0.760877380174328	1,00212E+14	up-regulated
LOC101257084	X40.MG.equatorial-X80.RR	0.0066151377101069	-0.350468321095158	-1,25936E+14	down-regulated
LOC101257100	X40.MG.equatorial-X80.RR	0.00608143485319102	-0.230651509010127	0.805843193994791	up-regulated
LOC101257542	X40.MG.equatorial-X80.RR	0.00681376802549168	-0.285796829033057	-0.813027225797378	down-regulated
LOC101257882	X40.MG.equatorial-X80.RR	0.000710311505791379	-0.981201042779773	5,47957E+14	up-regulated
LOC101257931	X40.MG.equatorial-X80.RR	0.0012481017115551	-0.573459705502606	-0.120943911343084	down-regulated
LOC101257984	X40.MG.equatorial-X80.RR	0.00962460536718531	-0.183204432809703	0.665908158798676	up-regulated
LOC101258125	X40.MG.equatorial-X80.RR	0.000574826487663374	0.845460526186015	-0.276278009227217	down-regulated
LOC101258184	X40.MG.equatorial-X80.RR	0.00191935328943645	0.258069145642991	-0.662610955085229	down-regulated
LOC101258920	X40.MG.equatorial-X80.RR	0.00255701167430188	0.250329847758866	-1,04266E+14	down-regulated
LOC101259103	X40.MG.equatorial-X80.RR	0.00253496003831736	0.129027527536056	-0.384524267986511	down-regulated
LOC101259292	X40.MG.equatorial-X80.RR	0.00515257429199856	0.448104110497895	-0.240339393058034	down-regulated
LOC101259772	X40.MG.equatorial-X80.RR	0.00188100362320535	-0.25383552859972	-0.946081369059318	down-regulated
LOC101259814	X40.MG.equatorial-X80.RR	0.00604180848386638	0.254103377577568	-0.692851075165508	down-regulated
LOC101259888	X40.MG.equatorial-X80.RR	0.00150158594279069	-0.549923306990873	-0.203108682159741	down-regulated
LOC101260211	X40.MG.equatorial-X80.RR	0.00206612169539083	-0.358911287230927	2,20834E+14	up-regulated
LOC101260719	X40.MG.equatorial-X80.RR	0.00756764644629146	0.208268332042893	0.311529329406455	up-regulated
LOC101261457	X40.MG.equatorial-X80.RR	0.00836879349949388	0.376615733776447	-0.92519820094466	down-regulated
LOC101261639	X40.MG.equatorial-X80.RR	0.00377303026749139	-0.674990143780122	3,09339E+14	up-regulated
LOC101261960	X40.MG.equatorial-X80.RR	0.00626234586017524	-0.120189023331666	-0.371199043635042	down-regulated
LOC101262041	X40.MG.equatorial-X80.RR	43111952,58	-0.305377744151863	-1,43325E+14	down-regulated
LOC101262589	X40.MG.equatorial-X80.RR	0.00765030771228298	-0.389475301506252	1,50155E+14	up-regulated
LOC101263245	X40.MG.equatorial-X80.RR	0.00762348470128317	0.149988508555946	1,06474E+14	up-regulated
LOC101263285	X40.MG.equatorial-X80.RR	0.00354423214533009	-0.163976315105552	0.210766247552121	up-regulated
LOC101263341	X40.MG.equatorial-X80.RR	0.00588987219277493	0.761996200599632	2,06564E+13	up-regulated
LOC101263397	X40.MG.equatorial-X80.RR	0.00255701167430188	0.696702623467185	-2,29796E+14	down-regulated
LOC101263627	X40.MG.equatorial-X80.RR	0.000710311505791379	0.234656678479418	-0.967589332483271	down-regulated
LOC101263952	X40.MG.equatorial-X80.RR	963275880,6	-0.27129250668663	-0.459957604710128	down-regulated
LOC101264109	X40.MG.equatorial-X80.RR	0.00280534656703734	0.160407166734939	-0.76064688290788	down-regulated
LOC101264171	X40.MG.equatorial-X80.RR	0.00278047273787419	0.399086274778784	0.822930101209898	up-regulated
LOC101264545	X40.MG.equatorial-X80.RR	0.0016613900218444	-0.958815727834968	5,61124E+14	up-regulated
LOC101264788	X40.MG.equatorial-X80.RR	0.00188100362320535	0.331263215315599	-0.280766999674712	down-regulated

LOC101265213	X40.MG.equatorial-X80.RR	0.000876173399892903	0.366638746321176	0.320997976090004	up-regulated
LOC101266673	X40.MG.equatorial-X80.RR	0.00424770541383831	0.403856307138071	-1,06682E+14	down-regulated
LOC101266909	X40.MG.equatorial-X80.RR	0.00569569061042193	-0.616175764623917	0.851788858861276	up-regulated
LOC101266993	X40.MG.equatorial-X80.RR	0.0001801999443547	-0.177403516305815	-2,13536E+14	down-regulated
LOC101267163	X40.MG.equatorial-X80.RR	0.00273397041045293	-0.23520784211676	-0.316548065425775	down-regulated
LOC101267654	X40.MG.equatorial-X80.RR	0.00132802759820421	-0.142685280584753	-0.740916609403538	down-regulated
LOC101267684	X40.MG.equatorial-X80.RR	0.000298366103422698	0.700960318895274	-0.513423654959467	down-regulated
LOC101267838	X40.MG.equatorial-X80.RR	0.00773261498252611	0.772798454729431	7,56712E+14	up-regulated
LOC101268021	X40.MG.equatorial-X80.RR	0.00226815704831765	0.119701917672642	0.173217201934546	up-regulated
LOC101268268	X40.MG.equatorial-X80.RR	0.00278047273787419	0.136962828578897	-0.761093200411056	down-regulated
LOC101268326	X40.MG.equatorial-X80.RR	0.000619617787874801	-0.13030736997932	-1,46852E+14	down-regulated
LOC101268345	X40.MG.equatorial-X80.RR	0.00049405085979606	-0.66912279286183	3,80127E+14	up-regulated
LOC101268591	X40.MG.equatorial-X80.RR	0.00152137760515526	-0.520525671305539	0.220803183676444	up-regulated
LOC101268806	X40.MG.equatorial-X80.RR	0.00627937037654833	-0.446252503847462	-0.379643888285541	down-regulated
LOC104644837	X40.MG.equatorial-X80.RR	0.00377303026749139	-0.349927763549033	0.103924029202328	up-regulated
LOC104645871	X40.MG.equatorial-X80.RR	6435257,639	1	3,57433E+13	up-regulated
LOC543863	X40.MG.equatorial-X80.RR	0.00253496003831736	-0.249361317399971	-0.00399078316357482	down-regulated
pds	X40.MG.equatorial-X80.RR	0.00377303026749139	-0.20175491930753	-2,02217E+14	down-regulated
SlPer1	X40.MG.equatorial-X80.RR	0.000619617787874801	-0.747236328376515	2,02726E+13	up-regulated

7.26 SUPPLEMENTARY TABLE S3

TABLE S3. Primer design to experimental validation in tomato with PCR and qPCR.

Nome	Alvo	Tipo	Verde	Maduro	Foward	Reverse
A1PCR1	Alvo 1 / LOC101245278 / CYP80E6	PCR	560	0	ATCAATGGGCTGGACTTCTG	TGGTGTCTGAACCAGCAGAG
A1PCR2	Alvo 1 / LOC101245278 / CYP80E6	PCR	0	797	GACATCACTGGGCTGGAGAT	AGATTTGCGCGTCTTTAGGA
A2PCR1	Alvo 2 / LOC101258920	PCR	552	419	AGCAAAGCTCATGGAGGAAG	AAAGAAATGAAGAAGCTCCTAATCC
A2RTP1	Alvo 2 / LOC101258920	rtPCR	193	0	mesmo que A2PCR1	AGGAGTTCTTGATTGCACCTG
A2RTP2	Alvo 2 / LOC101258920	rtPCR	141	141	mesmo que A2PCR1	TGTTCTGGCTTGACCATTGA
A3PCR1	Alvo 3 / LOC104644837 / zinc finger	PCR	240	870	TCTCTCTCCTCCAGCTACGG	GGCACCATCTCATTATTCTCAG
A3PCR2	Alvo 3 / LOC104644837 / zinc finger	PCR	542	542	TTCAAGTGCCAAGCACAAAG	GGTCAGAGGCAGAGGAAGTG
A3RTP1	Alvo 3 / LOC104644837 / zinc finger	rtPCR	0	191	ATGCGTAGTGAGCGGAGAGT	CTGCAACTCCATTGCTTCAA
A3RTP2	Alvo 3 / LOC104644837 / zinc finger	rtPCR	100	0	ATTCAAAATTTCCAACAGCC	GGCACCATCTCATTATTCTCAG
A3RTP3	Alvo 3 / LOC104644837 / zinc finger	rtPCR	157	157	ATCCACGAGCTCTTCCTTGA	TTGACTGCTGGCTTGTTGTC

7.27 SUPPLEMENTARY MATERIAL 1

A planilha com as tabelas esta disponível para download no XLSX pelo link:

https://zenodo.org/record/8159389/files/Supplementary_Material_1.xlsx?download=1

7.28 SUPPLEMENTARY MATERIAL 2

As imagens estão agrupadas em único arquivo PDF disponível pelo link:

https://zenodo.org/record/8159389/files/Supplementary_Material_2.pdf?download=1