



BEATRIZ PANEGASSI DE SOUZA

**POLIADENILAÇÃO ALTERNATIVA EM ESPÉCIES DA
ALIANÇA *TABEBUIA* SOB SECA**

**LAVRAS – MG
2022**

BEATRIZ PANEGASSI DE SOUZA

**POLIADENILAÇÃO ALTERNATIVA EM ESPÉCIES DA ALIANÇA
*TABEBUIA SOB SECA***

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento de Plantas, área de concentração em genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

Prof. Dr. Evandro Novaes
Orientador

**LAVRAS-MG
2022**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Souza, Beatriz Panegassi de.

Poliadenilação alternativa em espécies da Aliança *Tabebuia*
sob seca / Beatriz Panegassi de Souza. – 2022.

58 p. : il.

Orientador(a): Evandro Novaes.

Dissertação (mestrado acadêmico) – Universidade Federal de
Lavras, 2022.

Bibliografia.

1. APAtrap. 2. Transcritos alternativos. 3. Seca. I. Novaes,
Evandro. II. Título.

BEATRIZ PANEGASSI DE SOUZA

**POLIADENILAÇÃO ALTERNATIVA EM ESPÉCIES DA ALIANÇA
TABEBUIA SOB SECA**

**ALTERNATIVE POLYADENYLATION IN *TABEBUIA* ALLIANCE SPECIES
UNDER DROUGHT**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

APROVADA em 15 de julho de 2022
Dr^a. Ana Maria Benko Iseppon - UFPE
Dr. Evandro Novaes – UFLA
Dr. Luciano Vilela Paiva - UFLA

Prof. Dr. Evandro Novaes
Orientador

**LAVRAS – MG
2022**

AGRADECIMENTOS

Agradeço aos meus pais, Patrícia e Wagner, por sempre me apoiarem em minhas escolhas e por todo amor que me dão. Todas as minhas conquistas, são de vocês também! Obrigada por tudo, eu amo vocês incondicionalmente.

Às minhas irmãs, Letícia e Lívia, por estarem sempre ao meu lado me dando tanto amor e carinho, vocês são incríveis e eu sou imensamente grata por tê-las em minha vida.

À Fran, que está sempre comigo nos melhores e piores momentos me dando forças, incentivo e abrigo todos os dias, eu amo você.

À Isa, minha amiga/irmã, pelos mais de 18 anos de amizade e por todos os outros que ainda estaremos compartilhando dos melhores sentimentos.

Ao Luiz e a Ana Luiza, vocês não imaginam o quanto sou grata por viver momentos com pessoas tão maravilhosas e admiráveis como vocês.

À minha família, em especial à minha Tia Val, por sempre se preocupar, me abrigar e cuidar de mim.

Agradeço aos amigos, por todas as conversas, risadas e apoio nos momentos complicados: Felipe, Léo, Andreza, Camila e Marcelo. O mestrado com certeza foi mais leve com vocês ao meu lado. Às meninas da República Tomara Que Caia, por serem minha segunda casa em Lavras.

À Universidade Federal de Lavras e ao Programa de Pós Graduação em Genética e Melhoramento de plantas. Ao professor Evandro, que topou me orientar durante o mestrado, por toda a confiança em mim e por todo crescimento profissional que me proporcionou.

O presente trabalho foi realizado com o apoio da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

RESUMO

A poliadenilação alternativa (APA) é um mecanismo co-transcricional que gera transcritos alternativos, podendo regular a expressão gênica das plantas em resposta a estresses bióticos/abióticos. Porém, em espécies vegetais, esse mecanismo permanece pouco explorado e conseqüentemente, pouco compreendido. Dessa forma, o objetivo do estudo foi analisar os padrões de APA em quatro espécies de ipês, duas de Cerrado (*Handroanthus ochraceus* e *Tabebuia aurea*) e duas de floresta (*Handroanthus impetiginosus* e *Handroanthus serratifolius*) submetidas à seca. Para isso, foram utilizados dados RNA-seq das quatro espécies sob dois tratamentos (seca e controle irrigado) como input na ferramenta APAtrap, que identifica potenciais sítios APA e suas dinâmicas de utilização entre tratamentos. Genes que apresentaram FDR (False Discovery Rate) < 0.05 e PD (Percentage Difference) > 0.2 foram considerados como genes com uso diferencial significativo de sítios de poliadenilação (DE-APA). *H. ochraceus* foi a espécie que apresentou o maior número de genes DE-APA (3,530) e *H. impetiginosus* o menor (120). *H. serratifolius* e *T. aurea* apresentaram números semelhantes de genes com DE-APA: 368 e 169, respectivamente. *H. ochraceus* compartilhou 102 e 103 DE-APA com *T. aurea* e *H. serratifolius*, respectivamente. Não houve genes DE-APA compartilhados entre as quatro espécies. Genes com DE-APA foram classificados em CR-APA, quando o sítio APA foi localizado em região codante; e em UTR-APA quando em região 3'UTR. *H. impetiginosus* apresentou 61 UTR-APA e 59 CR-APA; *H. serratifolius* teve 197 CR-APA e 171 UTR-APA; *T. aurea* com 228 CR-APA e 141 UTR-APA; por fim, *H. ochraceus* apresentou 1,089 CR-APA e 2,441 UTR-APA. Trabalhos prévios indicam que genes com menor extremidade 3'UTR tendem a ser mais expressos. Somente em *H. ochraceus* houve associação significativa entre o tamanho da 3'UTR e o nível de expressão gênica. Nesta espécie, entre os genes com 3'UTR curtas (APA proximal), 59% foram up-regulados, enquanto 62,5% dos genes UTR-APA com 3'UTR longas (APA distal) foram down-regulados. O teste exato de Fisher detectou categorias do Gene Ontology (GO) enriquecidas somente entre os genes com DE-APA de *H. ochraceus*. Essas categorias GO enriquecidas, estavam envolvidas em vias de regulação transcricional, transporte/localização de proteínas e processos catabólicos de macromoléculas. Esses resultados indicam que a resposta à seca pode ser regulada por APA em espécies de ipê. Dentre as espécies estudadas, *H. ochraceus* parece utilizar esse mecanismo de regulação co-transcricional de forma mais intensa. Porém, são necessários estudos adicionais para compreender as conseqüências da APA na expressão final dos genes nesta espécie.

Palavras-chave: APAtrap. Transcritos alternativos. Transcrição. Seca.

ABSTRACT

Alternative polyadenylation (APA) is a co-transcriptional mechanism that generates alternative transcripts, which can regulate plant gene expression in response to biotic/abiotic stresses. However, in plant species, this mechanism remains poorly understood. Thus, the objective of the study was to analyze APA patterns in four species of ipês, two from Cerrado (*Handroanthus ochraceus* and *Tabebuia aurea*) and two from forest (*Handroanthus impetiginosus* and *Handroanthus serratifolius*) submitted to drought. For this, RNA-seq data from the four species under two treatments (under drought and control) were used as input to the APATrap tool, which identifies potential APA sites and their utilization dynamics between treatments. Genes that presented FDR (False Discovery Rate) < 0.05 and PD (Percentage Difference) > 0.2 were considered as having significant differential use of polyadenylation sites (DE-APA). *H. ochraceus* was the species that presented the highest number of DE-APA genes (3,530) and *H. impetiginosus* the lowest (120). *H. serratifolius* and *T. aurea* had similar numbers of genes with DE-APA: 368 and 169, respectively. *H. ochraceus* shared 102 and 103 DE-APA with *T. aurea* and *H. serratifolius*, respectively. There were no DE-APA genes shared among all four species. Genes with DE-APA were classified as CR-APA, when APA was located in the coding region; and as UTR-APA when it was located in the 3'UTR of the gene. *H. impetiginosus* presented 61 UTR-APA and 59 CR-APA; *H. serratifolius* had 197 CR-APA and 171 UTR-APA; *T. aurea* with 228 CR-APA and 141 UTR-APA; finally, *H. ochraceus* presented 1,089 CR-APA and 2,441 UTR-APA. Previous work indicates that genes with shorter 3'UTR end tend to be more expressed. Only in *H. ochraceus* there was a significant association between the size of the 3'UTR and the level of gene expression. In this species, among genes with shorter 3'UTR (proximal APA), 59% were up-regulated, while 62.5% of the UTR-APA genes with longer 3'UTR (distal APA) were down-regulated. Fisher's exact test detected Gene Ontology (GO) categories enriched only among the DE-APA genes of *H. ochraceus*. These GO categories were involved in transcriptional regulatory pathways, protein transport/localization and macromolecule catabolic processes. These results indicate that the response to drought can be regulated by APA in ipê species. Among the studied species, *H. ochraceus* seems to use this co-transcriptional regulation mechanism more intensively. However, further studies are needed to understand the consequences of APA on the final expression of genes in this species.

Keywords: APATrap. Alternative transcripts. Transcription. Drought.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráfico demonstrando a ocorrência de espécies da família Bignoniaceae nas regiões brasileiras.....	15
Figura 2 - Filogenia de 12 dos 14 gêneros que compõem a Aliança Tabebuia. Em vermelho, as quatro espécies analisadas no presente estudo.	16
Figura 3 – Etapas em que a expressão gênica eucariótica pode ser controlada.....	18
Figura 4 - Diferentes formas de processamento de um RNA, por meio de splicing alternativo, gerando diversas isoformas (proteínas alternativas).....	20
Figura 5 - Sequências consenso e enzimas que determinam a clivagem e síntese da cauda poli-A na extremidade 3' dos transcritos.	22
Figura 6 - Representação esquemática das duas classes de poliadenilação alternativa: CR-APA, que altera sequências codantes; e UTR-APA, que altera o comprimento da região 3'UTR.	22
Figura 7 – Número de genes com diferentes quantidades de sítios de poliadenilação nas quatro espécies da Aliança Tabebuia.....	30
Figura 8 – Diagrama de Venn demonstrando o compartilhamento de genes com APA entre as quatro espécies da Aliança Tabebuia.....	31
Figura 9 - Classificação dos genes DE-APA quanto a preferência por sítios proximais e distais no tratamento sob seca nas quatro espécies da aliança Tabebuia.....	32
Figura 10 – Diagrama de Venn demonstrando o compartilhamento de genes DE-APA entre as quatro espécies da Aliança Tabebuia. Genes com DE-APA utilizam diferentes sítios de poliadenilação quando submetidas à seca.	33
Figura 11 - Comparação da densidade de sequências de RNA de <i>T. aurea</i> alinhadas na região 3' do gene CDL12_00024. A imagem mostra as sequências de RNA de duas amostras do tratamento de seca (roxo) e duas do controle (verde-água). O fim da região codante (CDS) do gene CDL12_00024 é representada pelo retângulo amarelo no canto inferior esquerdo. As linhas verticais representam a localização dos sítios APA, em azul os sítios proximais e em vermelho o sítio distal.....	34
Figura 12 - Proporções da localização dos sítios APA entre os genes DE-APA, sob seca, entre as quatro espécies da aliança Tabebuia.	35
Figura 13 - Comparação da densidade de sequências de RNA de <i>T. aurea</i> alinhadas na região 3' do gene CDL12_04454 (UTR-APA). A imagem mostra as sequências de RNA de duas amostras do tratamento seca (roxo) e duas do controle (verde-água). O fim da região codante	

(CDS) é representada pelo retângulo amarelo no canto inferior. As linhas verticais representam as localizações dos sítios APA, em azul o sítio proximal e em vermelho o sítio distal.	36
Figura 14 - Comparação da densidade de sequências de RNA de <i>H. ochraceus</i> alinhadas na região 3' do gene CDL12_00400 (UTR-APA). A imagem mostra as sequências de RNA de duas amostras do tratamento seca (roxo) e uma do controle (verde-água). O fim da região codante (CDS) é representada pelo retângulo amarelo no canto inferior. As linhas verticais representam as localizações dos sítios APA, em azul o sítio proximal e em vermelho o sítio distal.	36
Figura 15 – Comparação da densidade de sequências de RNA de <i>H. impetiginosus</i> alinhadas na região 3' do gene CDL12_00203 (UTR-APA). A imagem mostra as sequências de RNA de duas amostras do tratamento seca (roxo) e duas do controle (verde-água). O fim da região codante (CDS) é representada pelo retângulo amarelo no canto inferior. As linhas verticais representam as localizações dos sítios APA, em azul o sítio proximal e em vermelho o sítio distal.	37
Figura 16 - Verificação do número de genes DE-APA nas quatro espécies da aliança <i>Tabebuia</i> , considerando o mesmo número de réplicas de <i>H. ochraceus</i> . Essa análise verificou que a perda de réplicas em <i>H. ochraceus</i> não explica a maior quantidade de genes DE-APA nessa espécie.	38
Figura 17 - Diagrama de <i>Venn</i> demonstrando o compartilhamento de genes DE-APA, em azul, e genes diferencialmente expressos (DE), em vermelho, do trabalho de Sobreiro et al. (2021). A <i>T. aurea</i> . B <i>H. impetiginosus</i> . C <i>H. ochraceus</i> . D <i>H. serratifolius</i>	39
Figura 18 - Diferenças de expressão (Log_2 Fold Change) entre tratamento de seca e controle para os genes UTR-APA das quatro espécies da Aliança <i>Tabebuia</i> . Caixas vermelhas representam os UTR-APA com alongamento da 3'UTR (sítio distal). Caixas azuis representam os UTR-APA com encurtamento da 3'UTR (sítio proximal).....	41
Figura 19 - Diagrama de <i>Venn</i> demonstrando o compartilhamento das categorias GO dos genes DE-APA entre quatro espécies da Aliança <i>Tabebuia</i> sob seca. Os números indicados na imagem representam os três domínios GO (processo biológico, componente celular e função molecular), porém, somente as principais categorias de processo biológico são indicadas nas caixas.....	43

LISTA DE TABELAS

Tabela 1 - Vantagens da RNA-seq em comparação com outros métodos de transcritômica...	25
Tabela 2 – Número de genes com diferentes quantidades de sítios APA identificados nas quatro espécies da Aliança <i>Tabebuia</i>	31
Tabela 3 – Número de genes DE-APA de acordo com a quantidade de sítios APA identificados nas quatro espécies da Aliança <i>Tabebuia</i>	33
Tabela 4 - Tabela comparando os genes UTR-APA/GDE, quanto à utilização dos sítios (proximais ou distais) e o padrão de expressão destes genes (<i>up</i> ou <i>down</i>), nas quatro espécies da Aliança <i>Tabebuia</i>	40
Tabela 5 - Comparação entre a utilização de sítios proximais e distais, de genes UTR-APA, e o padrão de expressão destes genes (<i>UP</i> ou <i>down</i>), nas quatro espécies da Aliança <i>Tabebuia</i> ..	41
Tabela 6 - Número de genes com uso diferencial significativo ($PD \geq 0,2$ e $FDR < 0,05$) e não significativo ($PD < 0,2$ e $FDR \geq 0,05$) de APA, nas quatro espécies da Aliança <i>Tabebuia</i> sob seca.	45

SUMÁRIO

1 INTRODUÇÃO.....	12
2 REVISÃO BIBLIOGRÁFICA.....	15
2.1 Aliança Tabebuia.....	15
2.2 Regulação da expressão gênica.....	17
2.3 Poliadenilação alternativa.....	21
2.4 RNA-seq.....	24
3 MATERIAL E MÉTODOS.....	27
3.1 Dados RNA-seq de quatro espécies de ipês desafiados com a seca.....	27
3.2 Processamento dos dados RNA-seq.....	27
3.3 Identificação e classificação dos sítios APA.....	28
3.4 Análise de enriquecimento funcional dos genes DE-APA.....	29
4 RESULTADOS.....	30
4.1 Identificação e classificação dos sítios APA.....	30
4.2 Expressão de genes UTR-APA.....	38
4.3 Análise de enriquecimento funcional.....	42
5 DISCUSSÃO.....	44
6 CONCLUSÃO.....	49
REFERÊNCIAS BIBLIOGRÁFICAS.....	50
APÊNDICE A – Categorias GO enriquecidas em <i>H. ochraceus</i> (Processo biol.).....	57
APÊNDICE B – Categorias GO enriquecidas em <i>H. ochraceus</i> (Componente cel.).....	58

1 INTRODUÇÃO

Por se tratar de organismos sésseis, as plantas precisam adaptar rapidamente sua fisiologia em resposta a condições adversas, como estresses bióticos e abióticos. Nesse sentido, a regulação da expressão gênica é um importante mecanismo de resposta às alterações ambientais externas (TAIZ; ZEIGER, 2002). Este mecanismo pode ocorrer em diversos momentos, desde o início da transcrição do gene (controle transcricional) até após a tradução (controle pós-traducional), passando por diversas etapas, como o processamento do pré-mRNA (ALBERTS et al., 2017a). As etapas de processamento do mRNA ocorrem concomitantemente com o processo de transcrição e, portanto, as regulações durante esta etapa são conhecidas como mecanismos de controle co-transcricionais (PROUDFOOT et al., 2002).

Um desses mecanismos é a poliadenilação alternativa (APA, do inglês *alternative polyadenilation*) (PROUDFOOT et al., 2002). Este mecanismo culmina na formação de extremidades 3' alternativas. Isso ocorre devido à clivagem/poliadenilação em diferentes sítios durante o processamento do pré-mRNA (DI GIAMMARTINO; NISHIDA; MANLEY, 2011; LUTZ; MOREIRA, 2011). Quando os sítios poli-A alternativos se encontram em íntrons/exóns, o evento APA leva à produção de isoformas de proteínas a partir de um mesmo gene, sendo chamado de CR-APA. Porém, quando os sítios poli-A estão localizados em regiões 3' não traduzidas (3'UTR), a APA leva à formação de transcritos com diferentes comprimentos de 3'UTR, mas que codificam uma mesma proteína, sendo chamados de UTR-APA. Em CR-APA, a consequência pode ser a produção de proteínas com funções distintas ou até mesmo a alteração de uma proteína funcional para uma não funcional (e vice-versa). No caso de UTR-APA, o resultado pode ser uma mudança na localização, transporte, estabilidade e eficiência na tradução do mRNA (LEWIS; GUNDERSON; MATTAJ, 1995; COLGAN; MANLEY, 1997; WICKENS; ANDERSON; JACKSON, 1997). Dessa forma, APA tem recentemente ganhado atenção como mecanismo potencialmente importante para a regulação gênica co-transcricional (DI GIAMMARTINO; NISHIDA; MANLEY, 2011).

Tendo em vista o potencial regulatório da APA e que os mecanismos que controlam este evento ainda não são totalmente conhecidos, pesquisas vêm sendo desenvolvidas visando sua melhor compreensão. Sabe-se que eventos APA podem ser influenciados por diferentes condições fisiológicas, tais como crescimento, diferenciação e desenvolvimento celular, ou por patologias, como o câncer (DI GIAMMARTINO; NISHIDA; MANLEY, 2011). Mayr e Bartel (2009) demonstraram que a ocorrência de APA em células cancerosas, gerando transcritos com 3'UTR curtos, possui um papel generalizado na ativação da oncogênese. Em plantas, a APA

está começando a ser estudada, sendo poucos trabalhos já publicados com o tema. Hu et al. (2020) analisaram a ocorrência de poliadenilação alternativa em diferentes tecidos (folha, floema, xilema e raiz) de *Populus alba* e identificaram 10.213 sítios de poliadenilação em 6.988 genes, dos quais 32% (2.212) apresentaram dois ou mais sítios. Em outro estudo, Ye et al. (2019) demonstraram que eventos APA participam extensivamente da regulação de respostas a diferentes estresses (bióticos e abióticos) em arroz.

O transcrito é o conjunto completo dos transcritos de uma célula em um dado estágio de desenvolvimento ou condição fisiológica. O estudo dos transcritomas pressupõe não só a identificação das sequências dos transcritos, como também suas quantificações. Dada a importância da regulação gênica no nível co-transcricional, é essencial que se conheça o transcrito para a elucidação destes mecanismos. (WANG; GERSTEIN; SNYDER, 2009; MARGUERAT; BÄHLER, 2010). O sequenciamento de RNAs em plataforma de alto desempenho (NGS), técnica conhecida como RNA-seq, possibilitou grandes avanços na análise do transcrito. A técnica consiste, basicamente, na conversão de RNAs em moléculas de cDNA, ligação com adaptadores e sequenciamento em plataforma NGS. Além de permitir o sequenciamento e a quantificação em alta resolução de transcritomas completos, a RNA-seq não necessita de um genoma/transcrito de referência, podendo ser aplicada nos mais diversos organismos (WANG; GERSTEIN; SNYDER, 2009; MARGUERAT; BÄHLER, 2010; MARTIN; WANG, 2011).

A RNA-seq tem sido amplamente utilizada em estudos envolvendo espécies florestais, com os mais diversos objetivos. Em *Tilia amaurensis* foi usada para a identificação e descrição de padrões de expressão de genes relacionados à embriogênese (KANG et al., 2021) e em *Pinus* para estudo de regulação de genes envolvidos em mecanismos de defesa ao *Fusarium circinatum* (ZAMORA-BALLESTEROS et al., 2021). Em diferentes gêneros (*Nothofagus*, *Populus*, *Eucalyptus*, *Handroanthus* e *Tabebuia*), a técnica foi utilizada para a análise de genes diferencialmente expressos sob condições de estresse biótico e abiótico (SOBREIRO et al., 2021; SANTOS et al., 2020; CHENG et al., 2021; ESTRAVIS-BARCALA et al., 2021). Além dessas aplicações, a RNA-seq pode ainda ser utilizada para o estudo e compreensão do processamento transcricional alternativo, como por exemplo os mecanismos de *splicing* alternativo e de poliadenilação alternativa (ELKON; UGALDE; AGAMI, 2013). Recentemente, Ye et al. (2018) desenvolveram uma ferramenta chamada APATrap, que detecta potenciais sítios APA e suas dinâmicas de utilização entre amostras com dados RNA-seq. O programa identifica diferentes sítios de finalização da transcrição por meio da profundidade de alinhamento de *reads* de cDNA em genes. Embora explorar a poliadenilação a partir de dados

RNA-seq apresente algumas complicações, o programa APATrap pode ser uma boa opção para estudos que visam à compreensão inicial de APA em espécies com genoma/transcritoma pouco explorado (SHAH et al., 2021).

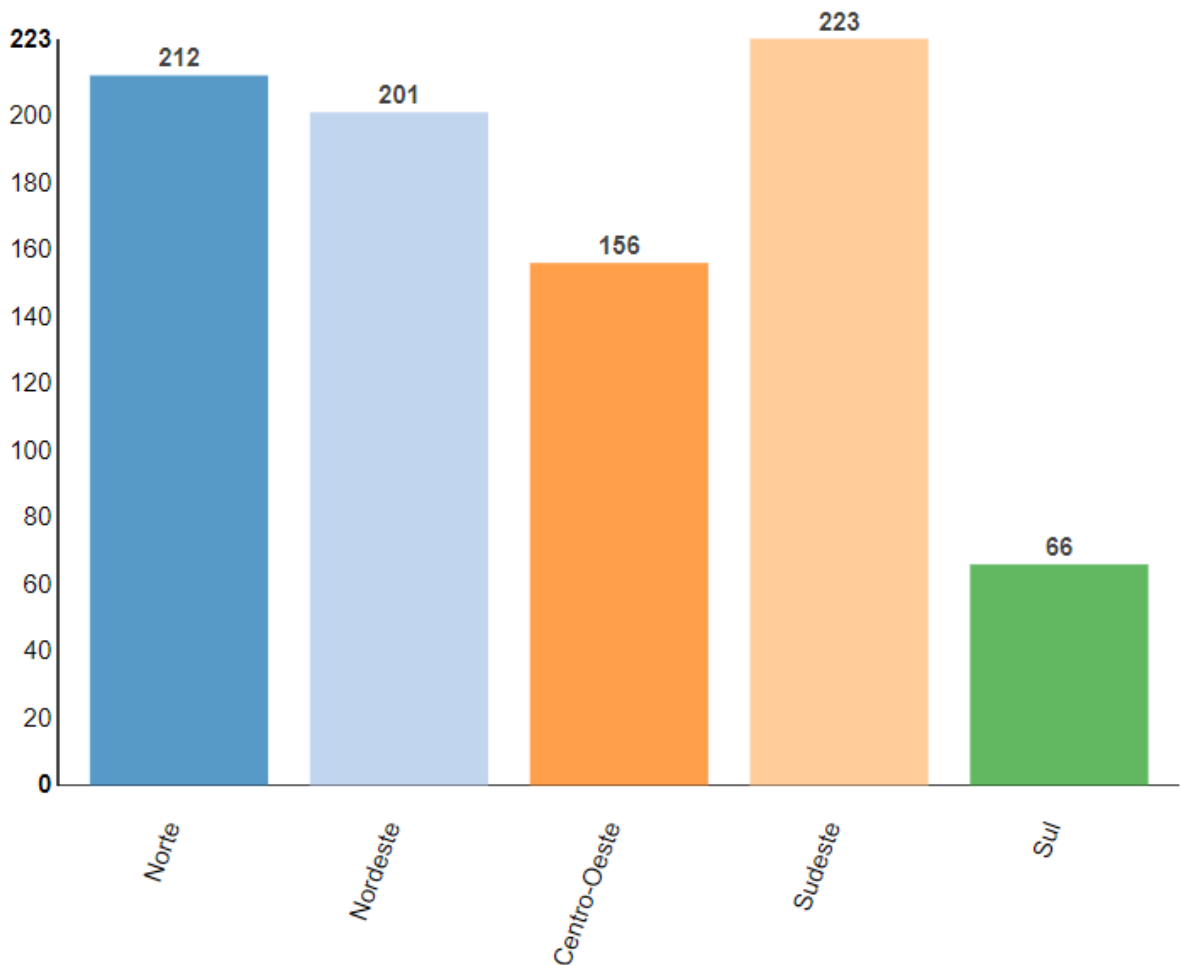
A escassez de estudos sobre a poliadenilação alternativa em espécies vegetais faz com que ainda não se saiba ao certo a importância desse mecanismo regulatório no controle da expressão dos genes de plantas. Portanto, o desenvolvimento de pesquisas sobre APA nessas espécies é um tema relevante para melhorar a compreensão dos mecanismos de controle de expressão gênica em vegetais. Além disso, estudos sobre APA em espécies de ipê, podem contribuir para a melhoria das informações genômicas/transcritômicas, que ainda são escassas nestas espécies.

2 REVISÃO BIBLIOGRÁFICA

2.1 Aliança *Tabebuia*

Bignoniaceae é uma família Pantropical com 104 gêneros e cerca de 860 espécies de árvores, lianas, por vezes arbustos e raramente ervas (FISCHER; THEISEN; LOHMANN, 2004). Apresenta ampla distribuição em regiões tropicais e subtropicais, com apenas algumas espécies ocorrendo em regiões de clima temperado. O centro de diversidade da família é a América do Sul, mais especificamente o Brasil, com 34 gêneros e 420 espécies distribuídas por todo o país, como pode ser observado na Figura 1 (GENTRY, 1980; FISCHER; THEISEN; LOHMANN, 2004; LOHMANN et al., 2020).

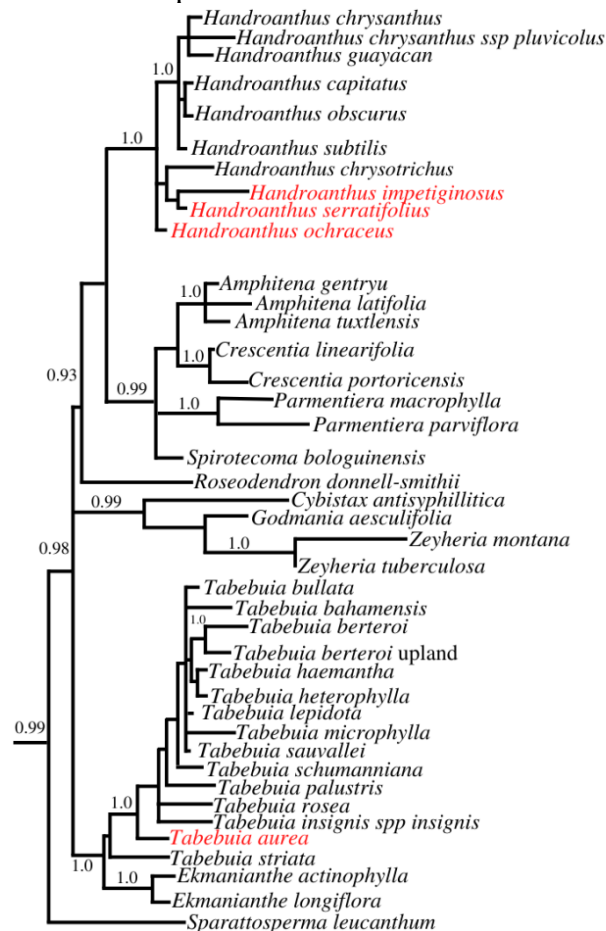
Figura 1 – Gráfico demonstrando a ocorrência de espécies da família Bignoniaceae nas regiões brasileiras.



Fonte: Lohmann et al. (2020)

De forma geral, a família apresenta hastes cilíndricas ou angulares, folhas opostas com lâminas compostas e flores de coloração variável: branca, creme, amarela, laranja, rosa, roxa, vermelha, verde ou azulada, dependendo da espécie (LOHMANN, 2004). Em um estudo molecular, Grose e Olmstead (2007a) identificaram um clado monofilético de Bignoniaceae composto por 14 gêneros neotropicais. Destes, apenas *Amphitecna* não apresenta folhas compostas digitadas, característica compartilhada por todos os outros 13 gêneros. *Tabebuia* e *Handroanthus* representam mais da metade do grupo que, por esse motivo, recebeu o nome de Aliança *Tabebuia*. Na Figura 2 é possível observar a relação filogenética dos gêneros, proposta por GROSE; OLMSTEAD (2007a) e revisada por GROSE; OLMSTEAD (2007b), sendo que os gêneros *Paratecoma* e *Romeroa* não foram incluídos nos estudos. Este é o segundo maior clado de Bignoniaceae e sete dos 14 gêneros são encontrados no Brasil: *Cybistax*, *Godmania*, *Handroanthus*, *Paratecoma*, *Sparattosperma*, *Tabebuia* e *Zeyheria* (GROSE; OLMSTEAD, 2007a, 2007b; OLMSTEAD et al., 2009; LOHMANN et al., 2020).

Figura 2 - Filogenia de 12 dos 14 gêneros que compõem a Aliança *Tabebuia*. Em vermelho, as quatro espécies analisadas no presente estudo.



Fonte: adaptado de GROSE; OLMSTEAD (2007b)

As espécies da família Bignoniaceae apresentam ampla utilização. Por exemplo, gêneros como *Jacaranda* e *Crescentia* são utilizados em ornamentações e, algumas espécies como *Jacaranda copaia*, são empregadas em projetos de reflorestamento. Gêneros como *Handroanthus* e *Tabebuia* são usados na indústria madeireira, devido à alta densidade e durabilidade de suas madeiras, que são muito resistentes ao ataque de fungos e insetos (GENTRY, 1980; FISCHER; THEISEN; LOHMANN, 2004; LOHMANN, 2004). Além disso, observa-se a importância etnobotânica e medicinal de diversas espécies dessa família (GENTRY, 1980). Recentemente, Ryan et al. (2021) demonstraram a ação anti-inflamatória da casca de *Handroanthus impetiginosus*.

Estudos citogenéticos têm sido realizados em espécies da Aliança *Tabebuia* (DE GUERRA, 2002; COLLEVATTI; DORNELAS, 2016; CORDEIRO et al., 2020), mas informações genômicas ainda possuem pouca disponibilidade. De todo o grupo, apenas os genomas de *H. guayacan* e *H. impetiginosus* foram sequenciados e disponibilizados no banco de dados NCBI mais recentemente (SILVA-JUNIOR et al., 2018; BURLEY et al., 2021). Além disso, em uma pesquisa realizada por Sobreiro et al. (2021), os autores analisaram os genes diferencialmente expressos (GDE) em quatro espécies deste grupo (*H. ochraceus*, *T. aurea*, *H. impetiginosus* e *H. serratifolius*) submetidas à seca. *H. ochraceus* e *T. aurea* são de ocorrência em cerrado *stricto sensu*, *H. impetiginosus* e *H. serratifolius* ocorrem em florestas estacional semidecídua. Os autores relataram que as espécies de cerrado foram menos responsivas à seca no nível transcricional, indicando uma possível tolerância a esse estresse, especialmente em *H. ochraceus*, que apresentou o menor número de GDE. Dessa forma, estudos genômicos e transcritômicos nestas espécies, contribuem tanto na elucidação dos mecanismos de regulação da expressão gênica em plantas, quanto na compreensão dos fatores que levam a essa regulação.

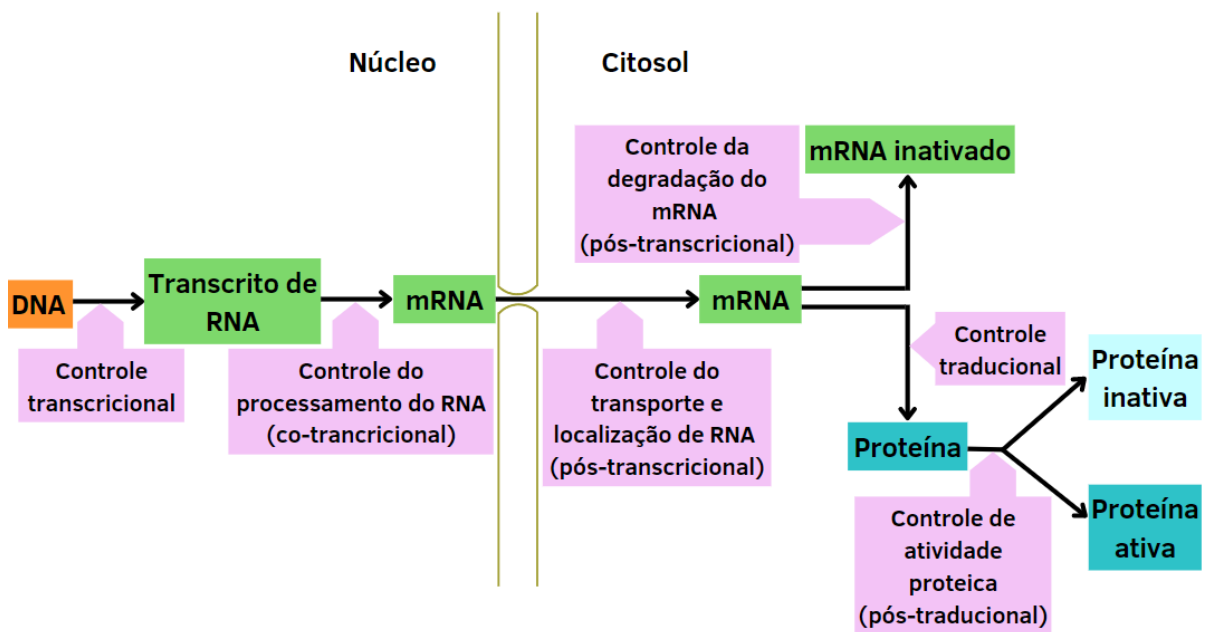
2.2 Regulação da expressão gênica

Por se tratar de organismos sésseis, que precisam perceber e responder rapidamente às mudanças no ambiente, as plantas contam com sensores moleculares, que reconhecem essas alterações ambientais e desencadeiam respostas fisiológicas. Por exemplo, espécies vegetais expostas a estresses abióticos (mudanças de temperatura, intensidade de luz, disponibilidade de nutrientes e água) podem ser prejudicadas caso não haja uma resposta fisiológica eficiente a essas condições. Desse modo, a regulação da expressão gênica apresenta-se como um

importante mecanismo de resposta às variáveis ambientais às quais as plantas estão sujeitas (TAIZ; ZEIGER, 2002; FLORIS et al., 2009).

A regulação da expressão gênica em eucariotos pode ser realizada em diferentes momentos (Figura 3), desde a transcrição (controles transcricional, co-transcricional e pós-transcricional) até durante e depois da tradução (controles traducional e pós-traducional).

Figura 3 – Etapas em que a expressão gênica eucariótica pode ser controlada.



Fonte: adaptado de Alberts et al. (2017a)

O nível de empacotamento da cromatina está diretamente relacionado ao acesso da maquinaria da transcrição ao DNA. Nesse sentido, o remodelamento da cromatina é um dos principais fatores de regulação transcricional. Os eventos mais importantes que alteram localmente a estrutura da cromatina são: a substituição de histonas e/ou modificações dessas por acetilação ou metilação, e a remodelagem/remoção de nucleossomos. Outra forma de regulação transcricional, é a especificação do momento e local em que cada gene deve ser transcrito. Os responsáveis por essa regulação são sequências específicas do DNA, chamadas de elementos *cis*-reguladores (como TATA-box), que geralmente se localizam no promotor do gene ou a montante dele. A esses *cis*-elementos ligam-se os *trans*-reguladores, que são proteínas de reconhecimento que regulam a transcrição. O TATA *Binding Protein* (TBP) e os fatores de transcrição são exemplos de *trans*-reguladores. Esses fatores de transcrição podem ativar ou reprimir a expressão gênica, auxiliando ou impedindo que a RNA-polimerase inicie a transcrição dos genes. Essa regulação ocorre como resposta das plantas a sinais ambientais

externos e, para isso, são necessários diversos sinais internos conhecidos como cascata de sinalização celular. A cascata, ou transdução de sinais, é uma série de reações bioquímicas sequenciais que culminam na alteração da expressão genes específicos. Dessa forma, o sinal inicial externo recebido pela planta é amplificado, resultando no controle transcricional da expressão de vários genes no núcleo da célula estimulada (TAIZ; ZEIGER, 2002; ARNHOLDT-SCHMITT, 2004; DUQUE et al., 2013; ALBERTS et al., 2017a).

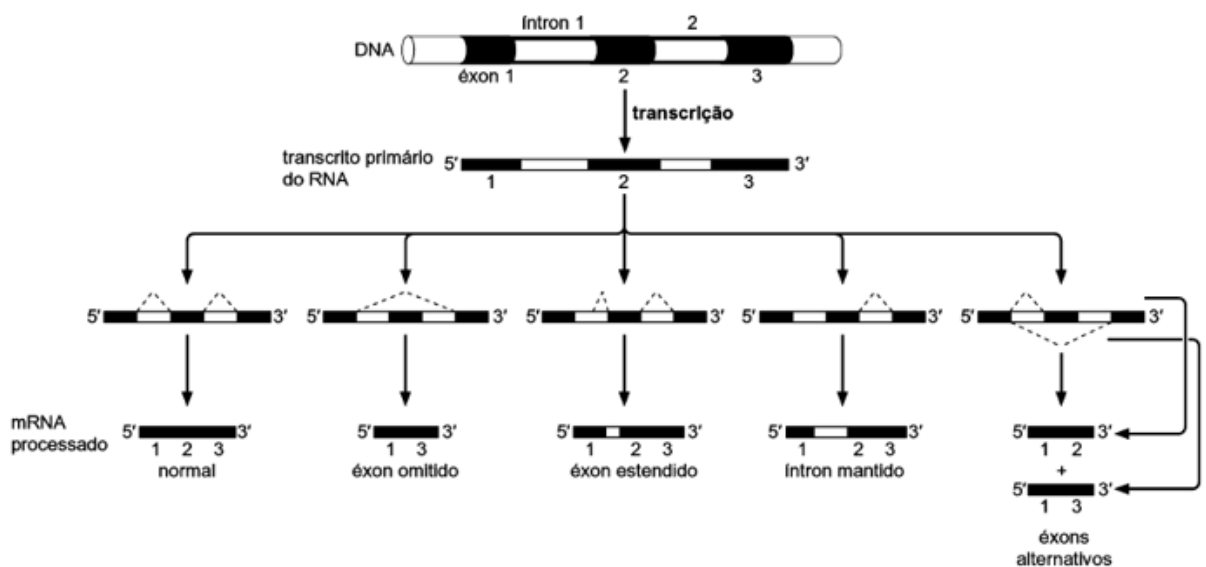
O controle co-transcricional pode ocorrer nos níveis de processamento do pré-mRNA (capeamento, *splicing* e poliadenilação) e na localização/estabilidade do mRNA. Uma vez iniciada a transcrição, a primeira etapa de modificação co-transcricional do pré-mRNA é a adição do *cap* na extremidade 5' (capeamento). O 5'-*cap* está envolvido na proteção do mRNA e na eficiência da tradução, uma vez que participa da associação com os ribossomos (STAATS, 2014). O *splicing* é a remoção dos íntrons e junção dos éxons, de forma ordenada, do transcrito primário do mRNA. Porém, existem algumas possibilidades de remoções e junções alternativas de exons e íntrons, nesse caso o *splicing* é denominado alternativo, e pode gerar transcritos alternativos (Figura 4). Sendo assim, um mesmo gene pode originar diferentes produtos polipeptídicos (isoformas) que podem, por exemplo, atuar em diferentes tecidos vegetais. Esse *splicing* alternativo também pode produzir uma proteína não funcional de acordo com a necessidade da célula (FLORIS et al., 2009; WATSON et al., 2015).

Após a adição do *cap-5'* e do *splicing*, ocorre a clivagem e poliadenilação (adição da cauda poli-A) da extremidade 3' do mRNA. Esta etapa está diretamente relacionada com o fim da transcrição. Duas proteínas de subunidades múltiplas (CstF e CPSF) atuam na clivagem da extremidade 3'UTR. Posteriormente, a enzima PAP (poli-A-polimerase) adiciona uma cadeia de adenosinas na extremidade 3' formada. Recentemente, a poliadenilação alternativa (APA) ganhou atenção como um fator de regulação co-transcricional da expressão gênica (DI GIAMMARTINO; NISHIDA; MANLEY, 2011; ALBERTS et al., 2017a; PROUDFOOT et al., 2002).

Embora os eventos acima citados ocorram concomitantemente com a transcrição, a regulação da expressão gênica nestes eventos ainda é denominada na literatura como pós-transcricionais, embora o termo co-transcricional seja mais adequado (PROUDFOOT et al., 2002). Um exemplo de regulação pós-transcricional, que ocorre efetivamente após o fim da transcrição, é a ação de micro-RNAs (miRNAs). Esses pequeno RNAs, com cerca de 22 pb, associam-se à proteína AGO (da subfamília *Argonauta*), formando o complexo miRISC (RNA - *induced silencing complex*). Além de AGO existem fatores específicos (da espécie; do mRNA-alvo; ou do próprio miRNA) que são necessários para o funcionamento correto do miRNA

(ZAHA, 2014). O RISC desempenha função de regulação negativa da expressão gênica, seja por meio da degradação dos mRNAs ou pela inibição da sua tradução. Mais especificamente, esse silenciamento da expressão gênica ocorre porque o complexo miRISC se liga a sequências complementares localizadas na região 3'UTR do mRNA-alvo (FLORIS et al., 2009; DUQUE et al., 2013; ALBERTS et al., 2017a).

Figura 4 - Diferentes formas de processamento de um RNA, por meio de splicing alternativo, gerando diversas isoformas (proteínas alternativas).



Fonte: Watson (2015)

O início da tradução se dá por meio da ligação entre os ribossomos e o mRNA. Diversas proteínas auxiliares, chamadas de fatores de iniciação (como eIF2), estão envolvidas neste processo. Dessa forma, a fosforilação de fatores de iniciação, atua como um importante controle traducional na regulação da expressão gênica. Por exemplo, eIF2 sofre uma mudança conformacional quando é fosforilado, não podendo mais exercer sua função e, conseqüentemente, não há o início da tradução. Outros fatores de iniciação também podem ser fosforilados, geralmente por consequência de vias de transdução de sinais (HERSHEY; SONENBERG; MATHEWS, 2012). Por fim, o controle pós-traducional pode ocorrer por meio de fosforilação e ubiquitinação de proteínas como respostas de plantas ao estresse abiótico. A fosforilação de proteínas pode estar relacionada à ativação ou inativação de sua função ou, em alguns casos, com a mudança funcional da mesma. A adição da ubiquitina está relacionada com a degradação da proteína em que este grupo químico é adicionado (DUQUE et al., 2013;

ALBERTS et al., 2017b). Nesse caso, as proteínas ubiquitinadas são direcionadas a complexos proteicos conhecidos como proteossomos, que realizam a degradação proteolítica.

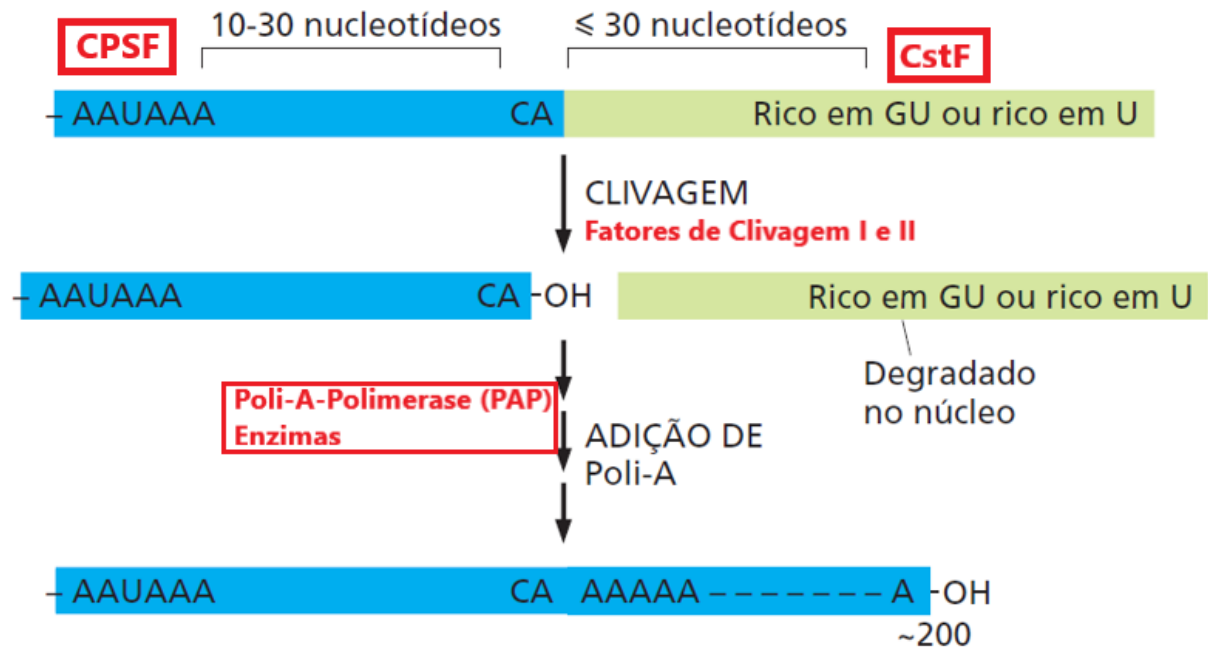
A partir disso, pode-se perceber que o controle da expressão gênica em eucariotos é complexo e envolve diversos mecanismos (MARTINS; FILHO, 2018). Embora muitos estudos tenham sido realizados visando à compreensão desses mecanismos em plantas, a poliadenilação alternativa permanece pouco explorada e, portanto, sua importância funcional ainda não foi bem determinada (SHEN et al., 2011). Dessa forma, é importante que sejam desenvolvidas pesquisas com o objetivo de compreender a ocorrência e importância de eventos APA no controle da expressão gênica em plantas.

2.3 Poliadenilação alternativa

Em quase todo mRNA eucariótico processado, encontra-se uma cauda poli-A na extremidade 3'. O processamento da extremidade 3' é uma etapa essencial, que envolve o reconhecimento de um sítio poli-A que sinaliza a necessidade de clivagem endonucleotídica, seguida da adição de uma cadeia de adenosinas (COLGAN; MANLEY, 1997; STAATS, 2014). A importância da cauda poli-A deve-se às suas funções de estabilizar, influenciar o transporte e a tradução do mRNA (LEWIS; GUNDERSON; MATTAJ, 1995; COLGAN; MANLEY, 1997; WICKENS; ANDERSON; JACKSON, 1997).

Em mamíferos, a sequência do sítio poli-A canônico, 5' AAUAAA 3' no mRNA, localiza-se de 10-30 bases a montante do local de clivagem/poliadenilação (Figura 5). Alguns mRNAs não contêm essa sequência consenso, diferindo em apenas uma base (sendo 5' AUUAAA 3' a mais comum). Quatro complexos de proteínas de subunidade múltiplas são responsáveis pela clivagem e poliadenilação da extremidade 3' do mRNA. O CPSF (fator de especificidade de clivagem) que reconhece 5' AAUAAA 3'; CstF (fator de estimulação de clivagem e poli-adenilação) que reconhece uma região *downstream* rica em U/GU; CFI e CFII (fatores de clivagem I e II) que realizam a clivagem (Figura 5). Além disso, são necessárias outras enzimas que determinam o tamanho da cauda e a presença da poli-A polimerase (PAP), responsável por adicionar a cadeia de poli-adeninas (COLGAN; MANLEY, 1997; DI GIAMMARTINO; NISHIDA; MANLEY, 2011; SHI, 2012).

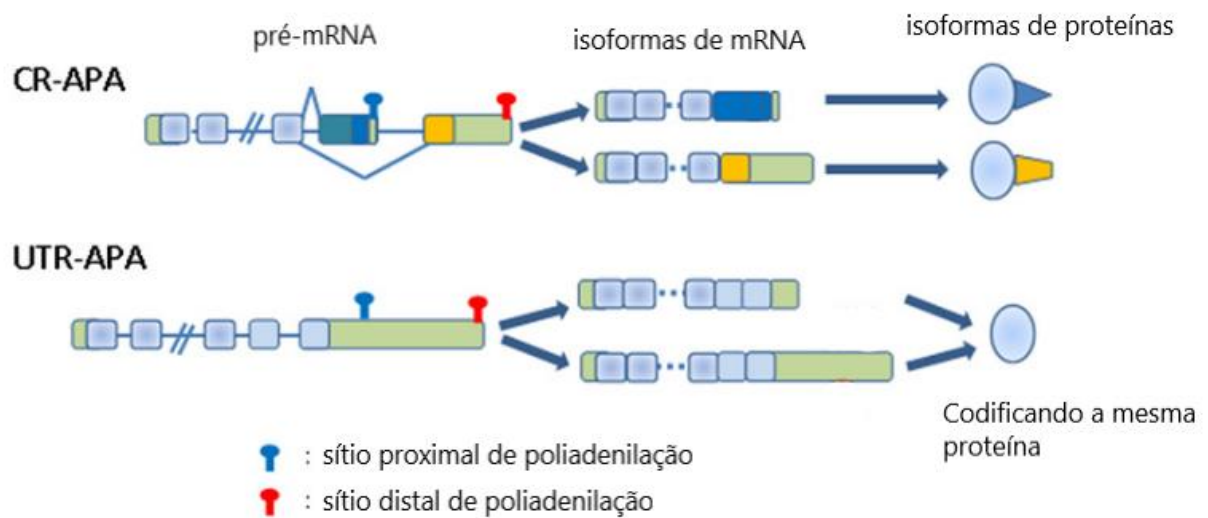
Figura 5 - Sequências consenso e enzimas que determinam a clivagem e síntese da cauda poli-A na extremidade 3' dos transcritos.



Fonte: adaptado de Alberts et al. (2017b)

Em eucariotos, os mRNAs de grande parte dos genes possuem extremidade 3' alternativas, ocasionadas pela clivagem/poliadenilação em diferentes sítios. Esse mecanismo é conhecido como poliadenilação alternativa (APA) (DI GIAMMARTINO; NISHIDA; MANLEY, 2011; LUTZ; MOREIRA, 2011). Os sítios APA podem estar localizados em região codante (CR-APA) e, neste caso, alteram a região codificadora de proteína levando a produção de isoformas de proteínas com terminais Carboxílicos distintos. Por esse motivo, a alteração da expressão gênica em genes com CR-APA, é dita qualitativa. Por outro lado, essa alteração pode ser quantitativa quando os sítios poli-A estão localizados na região 3' não traduzida (UTR-APA) e resultam transcritos com 3'UTR de diferentes comprimentos, mas que codificam uma mesma proteína (Figura 6) (COLGAN; MANLEY, 1997; SHI, 2012; CHEN et al., 2017). Como os elementos *cis* envolvidos no metabolismo do mRNA (como estabilidade, localização, transporte e tradução) estão presentes na 3'UTR, a ocorrência de APA nessas regiões pode efetivamente afetar a regulação do gene (TIAN; MANLEY, 2016). Além disso, regiões 3'UTR frequentemente contém sequências alvo de miRNAs e, portanto, UTR-APA podem adicionar ou remover essas sequências, levando a alterações na expressão gênica (MAYR, 2016; CHEN et al., 2017).

Figura 6 - Representação esquemática das duas classes de poliadenilação alternativa: CR-APA, que altera sequências codantes; e UTR-APA, que altera o comprimento da região 3'UTR.



Fonte: adaptado de Di Giammartino et al. (2011)

Sabe-se que fatores bióticos e abióticos podem influenciar a ocorrência de eventos APA, como exposição a patógenos à seca (YE et al., 2019). Porém, estes eventos ainda não são completamente compreendidos, apesar de diversas pesquisas estarem sendo realizadas a fim de entender as causas e mecanismos de regulação de APA em diversos organismos, por exemplo, humanos, ratos, arroz, *Populus*, entre outros (BEAUDOING et al., 2000; TIAN et al., 2005; MAYR; BARTEL, 2009; DI GIAMMARTINO; NISHIDA; MANLEY, 2011; SHEN et al., 2011; YE et al., 2019; HU et al., 2020). Ressalta-se que, em plantas, a APA ainda é pouco estudada e a sua compreensão pode fornecer informações importantes quanto ao controle da expressão de genes. Nesse sentido, YE et al. (2018) desenvolveram uma ferramenta de bioinformática, chamada APATrap, que identifica sinais de poliadenilação alternativa e detecta mudanças no uso desses sinais a partir de dados RNA-seq.

O passo inicial do APATrap é a identificação de regiões 3'UTR que são, posteriormente, comparadas com o genoma de referência para identificar novas 3'UTR ou extensões de 3'UTR. A partir disso, os sítios APA distais são definidos como as extremidades 3' dessas 3'UTRs. Para identificar potenciais sítios APA em cada gene, é utilizado um modelo baseado no erro quadrático médio. Para quantificar a diferença no uso dos sítios APA, entre duas amostras, o programa utiliza o índice de diferença percentual (PD – *percentage difference*). Para examinar mudanças significativas no uso desses sítios APA, o APATrap avalia a tendência linear por meio do coeficiente de correlação de Pearson. Para isso, o comprimento da 3'UTR e a expressão de um determinado gene, são comparados entre amostras. Por fim, o método de Benjamini-Hochberg é utilizado para estimar a taxa de falsa descoberta (FDR) (YE et al., 2018). Por se tratar de uma ferramenta nova, é importante que sejam desenvolvidos novos estudos que façam

uso de APAttrap. Os resultados podem contribuir para avaliar o funcionamento da ferramenta, além de contribuir com novas informações sobre a poliadenilação alternativa nas espécies em estudo.

2.4 RNA-seq

O conjunto completo de transcritos de uma célula, em um estágio de desenvolvimento e/ou condição fisiológica específica, é chamado de transcrito. Compreender o transcrito permite a identificação dos elementos funcionais do genoma e a realização de inferências sobre a constituição molecular de células e tecidos. A transcriptômica têm sido utilizada com o objetivo de avaliar o perfil de expressão de todos os genes; catalogar todas as espécies de transcritos (mRNAs e microRNAs, por exemplo); e, também, para identificar modificações co-transcripcionais, como *splicing* alternativo (MOROZOVA; MARRA, 2008; WANG; GERSTEIN; SNYDER, 2009).

Com o desenvolvimento do sequenciamento em plataformas de alto desempenho (NGS) (DIJK et al., 2014), um poderoso método surgiu para mapear e quantificar todos os transcritos de um transcrito. O método, denominado “RNA-seq”, envolve os seguintes passos: extração de RNA, uma etapa opcional de purificação caso haja interesse em apenas uma espécie de RNA (por exemplo mRNA ou miRNA), a fragmentação destes RNAs em tamanhos compatíveis com a plataforma de sequenciamento a ser utilizada (normalmente <500 pb), ligação com adaptadores, amplificação por PCR, sequenciamento em plataforma NGS e, por fim, a análise bioinformática dos dados (WANG; GERSTEIN; SNYDER, 2009; MARGUERAT; BÄHLER, 2010; HRDLICKOVA; TOLOUE; TIAN, 2017; STARK; GRZELAK; HADFIELD, 2019).

Em comparação com outras técnicas disponíveis para análise do transcrito, por exemplo o *microarray*, a RNA-seq apresenta vantagens (Tabela 1). Além de oferecer menor ruído de fundo advindo de outros transcritos e maior faixa dinâmica para detecção do nível de expressão de genes com expressão muito baixa ou muito alta, este método não necessita de um genoma de referência, podendo ser utilizado em organismos não-modelo. Outra vantagem, é a possibilidade de avaliar não somente o nível de expressão diferencial dos genes, mas também o nível de diferentes transcritos provenientes, por exemplo, de *splicing*. Adicionalmente, os custos da RNA-seq são relativamente menores aos de outras técnicas, tornando-a uma ótima escolha no estudo do transcrito (WANG; GERSTEIN; SNYDER, 2009; MARGUERAT;

BÄHLER, 2010; HRDLICKOVA; TOLOUE; TIAN, 2017; STARK; GRZELAK; HADFIELD, 2019).

Tabela 1 - Vantagens da RNA-seq em comparação com outros métodos de transcritômica.

Tecnologia	Tiling microarray	Sequenciamento de cDNA ou EST	RNA-seq
ESPECIFICAÇÕES DA TECNOLOGIA			
Princípio	Hibridização	Sequenciamento Sanger	Sequenciamento de alto desempenho
Resolução	De vários a 100 pb	Base única	Base única
Taxa de transferência	Alto	Baixo	Alto
Dependência da sequência genômica	Sim	Não	Não
Ruído de fundo	Alto	Baixo	Baixo
APLICAÇÕES			
Mapear simultaneamente regiões transcritas e expressão gênica	Sim	Limitada para expressão gênica	Sim
Faixa dinâmica para quantificar o nível de expressão gênica	Até algumas centenas de vezes	Não é prático	>8.000 vezes
Capacidade de distinguir diferentes isoformas	Limitado	Sim	Sim
Capacidade de distinguir a expressão alélica	Limitado	Sim	Sim
QUESTÕES PRÁTICAS			
Quantidade necessária de RNA	Alto	Alto	Baixo
Custo para mapear transcritomas de grandes genomas	Alto	Alto	Relativamente baixo

Fonte: Adaptado de Wang; Gerstein; Snyder (2009)

Devido às vantagens citadas, a utilização da RNA-seq em organismos não-modelo tem sido ampla. Em espécies florestais, por exemplo, a técnica tem sido utilizada para a identificação de genes relacionados à patogenicidade em *Eucalyptus* (SANTOS et al., 2020). Em *Populus trichocarpa*, a RNA-seq foi utilizada para a análise de expressão diferencial de genes em tecidos diferentes sob condição de estresse abiótico (CHENG et al., 2021) e, também,

na identificação de eventos de *splicing* e poliadenilação alternativa (HU et al., 2020). Dessa forma, a RNA-seq apresenta-se como uma técnica eficiente para o estudo do transcrito de qualquer espécie vegetal.

3 MATERIAL E MÉTODOS

3.1 Dados RNA-seq de quatro espécies de ipês desafiados com a seca

Foram utilizados os dados RNA-seq, obtidos por Sobreiro et al. (2021), de quatro espécies da Aliança *Tabebuia* submetidos a dois tratamentos (seca e controle). Das espécies, duas possuem ocorrência no Cerrado *stricto sensu* (*Tabebuia aurea* e *Handroanthus ochraceus*) e duas em florestas estacional semidecídua (*Handroanthus impetiginosus* e *Handroanthus serratifolius*).

Para obtenção do material biológico, Sobreiro et al. (2021) plantaram sementes das quatro espécies, em solo típico de Cerrado e sem adubação, em um viveiro da Universidade Federal de Goiás – UFG (Goiânia – GO, Brasil). O delineamento experimental utilizado foi o de blocos completos casualizados, com quatro espécies x dois tratamentos (seca e controle) x três blocos (totalizando 24 plantas). Durante 10 meses, as mudas foram regadas a cada 24 h e posteriormente, no pico da estação seca da região, iniciou-se o tratamento de seca. Para isso, os autores mantiveram o grupo controle com rega constante a cada 24h. Para o tratamento seca, a irrigação foi totalmente suspensa. Após 10 dias de tratamento, quando o solo do grupo submetido à seca atingiu 40% da capacidade de campo, os autores encerraram o experimento. As partes aéreas das plantas foram então coletadas para extração e sequenciamento de RNA. Durante a execução do experimento, duas plantas de *H. ochraceus* (uma de cada tratamento) morreram e, uma amostra do grupo controle desta espécie, foi perdida durante o controle de qualidade das sequências. Dessa forma, no presente estudo, essas amostras também foram excluídas das análises e testes foram realizados para verificar se isso influenciou nos resultados.

Sobreiro et al. (2021) extraíram o RNA total de toda a biomassa aérea das 21 plantas para posterior preparação de bibliotecas, que foram sequenciadas na BGI Genomics (Hong Kong, China). O sequenciamento foi realizado em plataforma Illumina HiSeq 2000, no modo de sequências pareadas 2x100pb, realizados em três canaletas (*lanes*). Os dados estão disponíveis no NCBI SRA, sob número de acesso: PRJNA608000, e foram utilizados para a realização do presente estudo.

3.2 Processamento dos dados RNA-seq

As sequências obtidas passaram por controle de qualidade, utilizando o FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Posteriormente, realizou-se a limpeza de sequências com o Trimmomatic (BOLGER; LOHSE; USADEL, 2014), de acordo com os parâmetros utilizados por Sobreiro et al. (2021). O genoma de *H. impetiginosus* (SILVA-JUNIOR et al., 2018) foi utilizado como referência e, para o mapeamento das leituras contra este genoma, utilizou-se o software STAR (DOBIN et al., 2013) com os seguintes parâmetros: `--runThreadN 18 --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNoverLmax 0.07 --outFilterMatchNmin 16 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --alignIntronMin 20 --alignMatesGapMax 1000 --alignIntronMax 10000`. Por fim, com o SAMStat (LASSMANN; HAYASHIZAKI; DAUB, 2011) realizou-se a análise de qualidade dos alinhamentos. Para a conversão dos resultados do alinhamento em formato Bed-Graph, requerido pelo programa APATrap, foram utilizados os programas SAMtools (LI et al., 2009) e BEDTools (QUINLAN; HALL, 2010).

3.3 Identificação e classificação dos sítios APA

Para a identificação de sítios de poliadenilação alternativa (APA), utilizamos o programa APATrap (YE et al., 2018) com parâmetros de *default* na linha de comandos do Linux. Esta ferramenta é capaz de identificar novas extensões de regiões 3' não traduzidas (3'UTR), contribuindo para a localização de potenciais sítios de poliadenilação alternativa. Os dados de RNA-seq de Sobreiro et al. (2021) (ver item 3.2) foram utilizados como *input* para as análises no APATrap.

O *output* de APATrap compreende um arquivo texto, que foi utilizado como *input* no pacote deAPA no programa R. Esse pacote detecta genes com mudanças significativas no uso de sítios APA entre as diferentes condições analisadas (seca e controle). A partir dos resultados obtidos, selecionamos somente os genes com uso diferencial significativo de APA (DE-APA). Para isso, o índice PD (*Percentage Difference*) deveria ser ≥ 0.2 e o FDR (*False Discovery Rate*) ou p-valor ajustado, deveria ser < 0.05 . O coeficiente de correlação de Pearson (*r*) foi utilizado para verificar se os genes com DE-APA significativos utilizavam sítios proximais ou distais no tratamento seca em relação ao controle (tratamento úmido). Para isso, $r > 0$ indicava

a utilização preferencial de sítios proximais, e $r < 0$ indicava maior utilização de sítios distais no tratamento com seca (YE et al, 2018).

Os DE-APA identificados foram classificados de acordo com a região de ocorrência de APA no gene. Essa classificação foi feita de acordo com Di Giammartino et al. (2011), em duas classes gerais: CR-APA (APA dentro de região codante) e UTR-APA (APA dentro de região 3' não traduzida – 3'UTR). Essa localização foi feita com o auxílio da ferramenta intersectBed do programa BEDTools (QUINLAN; HALL, 2010), utilizando as linhas indicativas de sequências codantes (CDS) do arquivo GFF3 com a anotação do genoma de *H. impetiginosus* (SILVA-JUNIOR et al., 2018). Sítios APA localizados dentro de sequências codantes e que, consequentemente alteram a sequência da proteína final, foram classificados como CR-APA. Por outro lado, os sítios APA localizados fora de sequências codantes e, portanto, em regiões 3'UTR, foram classificados como UTR-APA.

O programa IGV (*Integrative Genomics Viewer*) (ROBINSON et al. 2011) foi utilizado para visualização da quantidade de sequências de RNA, de cada biblioteca (espécie × tratamento × réplica), alinhadas na extremidade 3' dos genes com APA. Essa análise visual foi realizada para avaliar se os DE-APA possuíam, de fato, diferenças na profundidade de sequências alinhadas entre os tratamentos (seca × controle). Além disso, com o IGV também é possível verificar o local de ocorrência dos sítios APA (dentro de região codante ou em 3'UTR). Ou seja, com o programa IGV foi possível realizar um controle de qualidade em uma amostra de genes com APA e DE-APA significativos.

3.4 Análise de enriquecimento funcional dos genes DE-APA

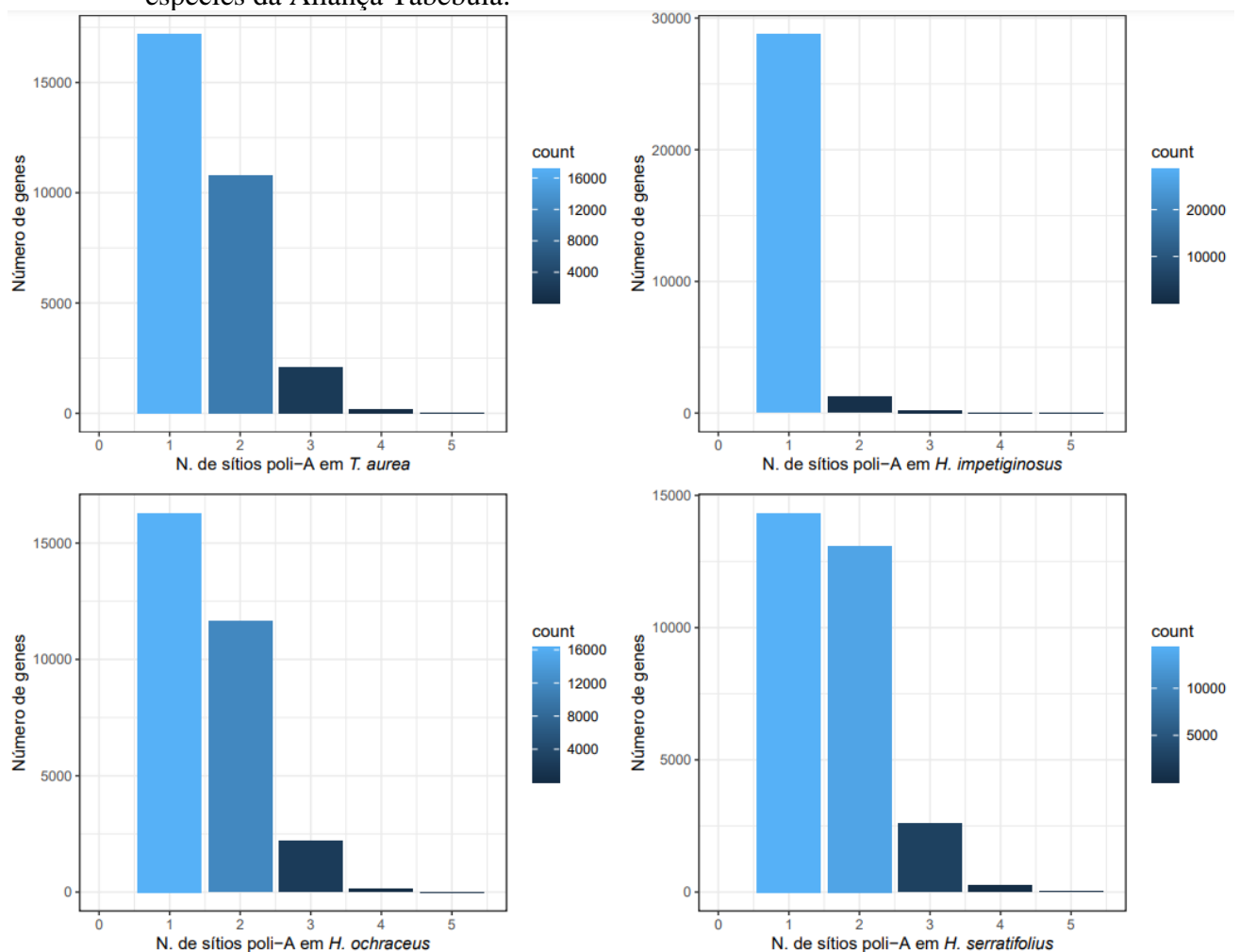
Nos genes DE-APA realizou-se análise de enriquecimento funcional com as categorias do *Gene Ontology* – GO (ASHBURNER et al., 2000). O objetivo foi identificar categorias GO mais frequentes entre os genes DE-APA do que o esperado em uma amostragem aleatória. Para isso, utilizou-se o teste exato de Fisher com $FDR < 0.05$ na ferramenta agriGO v2.0 (TIAN et al., 2017). As categorias GO anotadas para os genes de *H. impetiginosus* (SOBREIRO et al., 2021) foram utilizadas como referência. O pacote VennDiagram (CHEN, 2021) foi utilizado para analisar o compartilhamento de termos GO entre as espécies.

4 RESULTADOS

4.1 Identificação e classificação dos sítios APA

Em todas as espécies o número de genes com somente um sítio de poliadenilação foi superior ao número de genes com mais de um sítio (Figura 7). Quanto ao número de genes com eventos de poliadenilação alternativa, houve pouca variação entre as espécies, exceto por *H. impetiginosus*, em que a ocorrência desse evento foi drasticamente menor (Figura 7) (Tabela 2). Considerando somente os genes com APA, nas quatro espécies houve prevalência de dois sítios de poliadenilação. Além disso, um pequeno número de genes apresentou mais de cinco sítios APA (Tabela 2).

Figura 7 – Número de genes com diferentes quantidades de sítios de poliadenilação nas quatro espécies da Aliança Tabebuia.



Fonte: Da autora (2022)

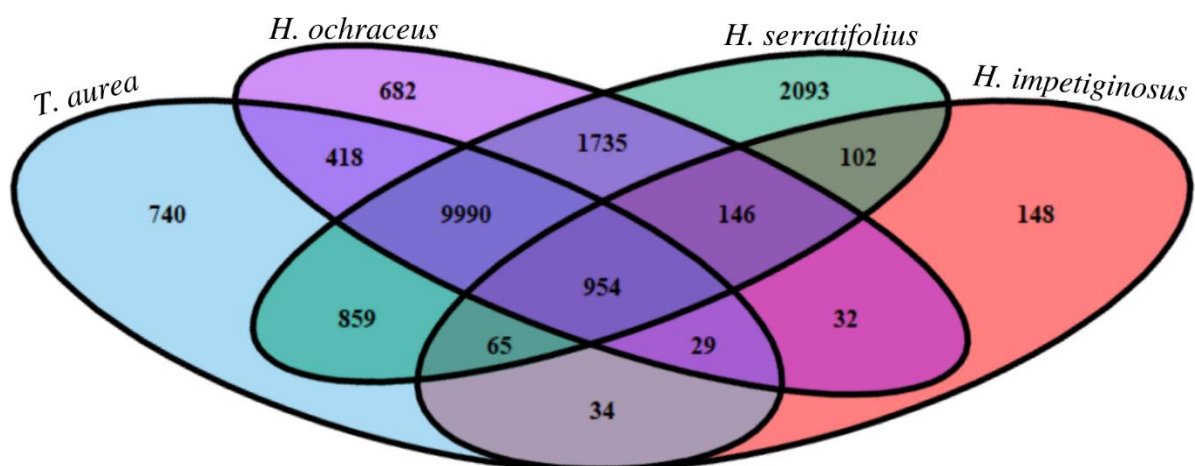
Tabela 2 – Número de genes com diferentes quantidades de sítios APA identificados nas quatro espécies da Aliança Tabebuia.

Número de sítios	<i>T. aurea</i>	<i>H. impetiginosus</i>	<i>H. ochraceus</i>	<i>H. serratifolius</i>
2	10.779	1.238	11.645	13.092
3	2.119	247	2.207	2.585
4	167	23	125	239
5	21	1	7	21
6-9	3	1	2	7
Total	13.089	1.510	13.986	15.944

Fonte: Da autora (2022)

A maioria dos genes com APA foram compartilhados entre as quatro espécies (Figura 8). Mesmo em *H. impetiginosus*, que apresentou menor número de genes com APA, a maioria destes também apresentou sítios APA nas outras três espécies (954). Além disso, *H. ochraceus*, *H. serratifolius* e *T. aurea* compartilharam entre si um número alto de genes com APA (9.990). *H. serratifolius* e *H. ochraceus* compartilharam um número relativamente alto de genes com APA (1.735) (Figura 8). Esse alto compartilhamento entre espécies, indica que a presença de sítios APA é conservada, pelo menos entre as espécies filogeneticamente próximas, como as da Aliança Tabebuia.

Figura 8 – Diagrama de Venn demonstrando o compartilhamento de genes com APA entre as quatro espécies da Aliança Tabebuia.

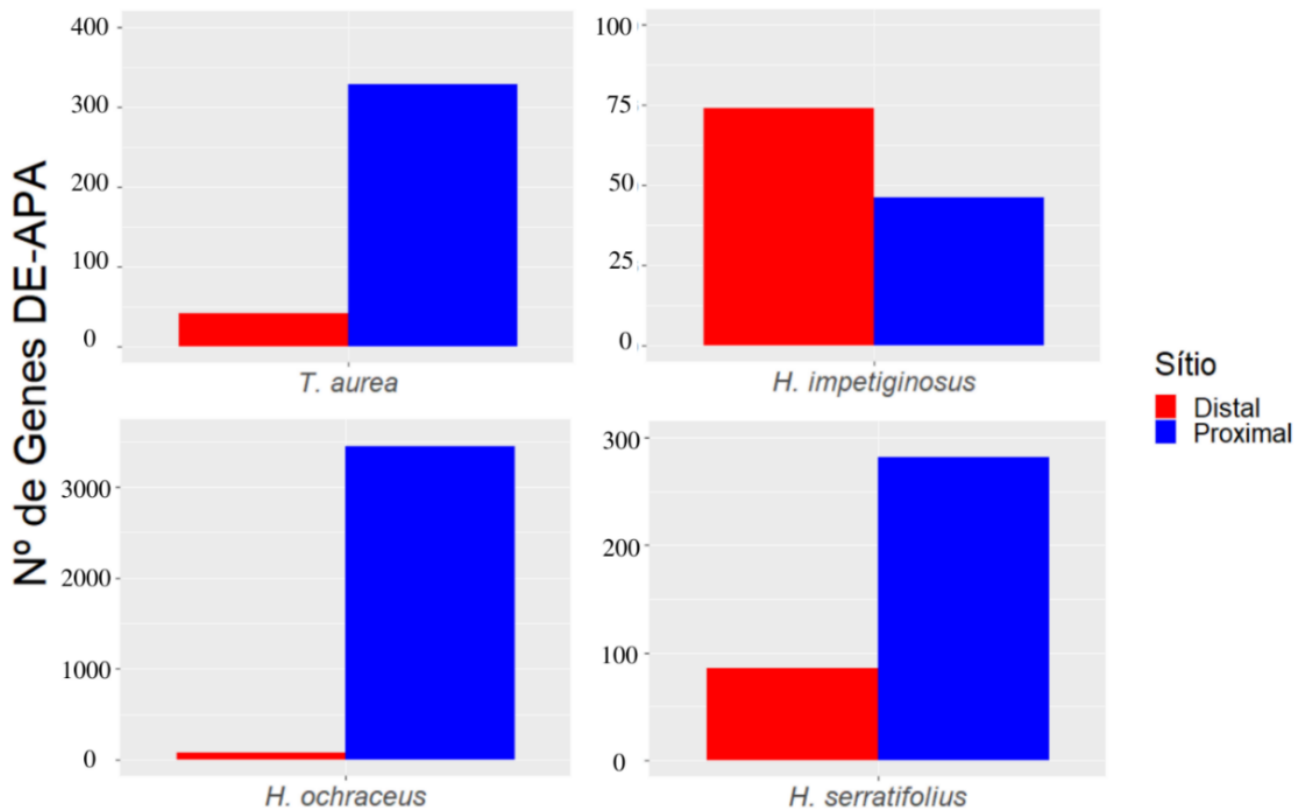


Fonte: Da autora (2022)

O número de genes com uso diferencial significativo de sítios APA (DE-APA) variou entre as espécies de 120 em *H. impetiginosus* a 3.530 em *H. ochraceus*. *T. aurea* e *H. serratifolius* apresentaram valores intermediários e próximos, com 369 e 368, respectivamente.

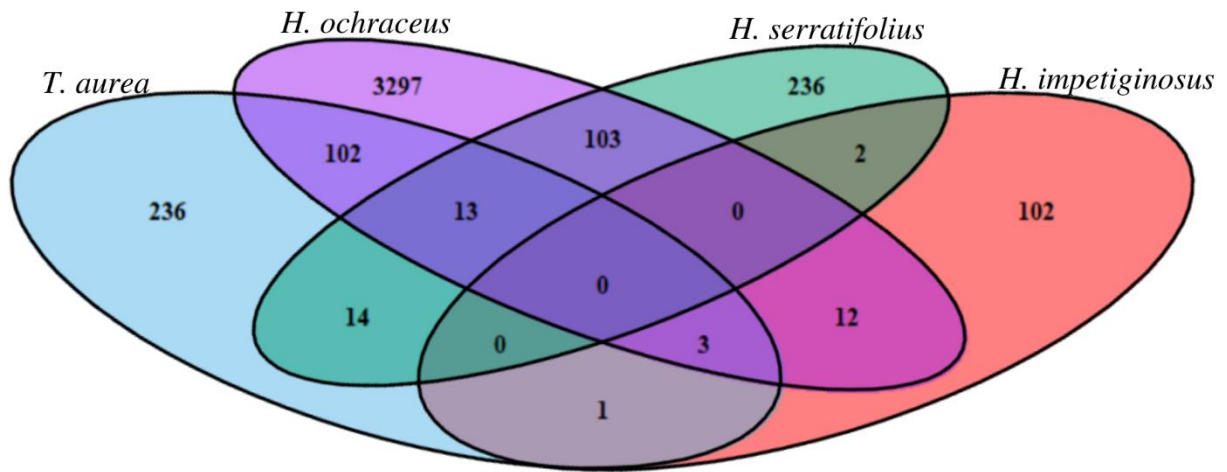
Quanto ao padrão de utilização dos sítios APA (distais ou proximais), as espécies mostraram comportamento semelhante, utilizando preferencialmente sítios proximais ($r > 0$) no grupo tratado em relação ao grupo controle. Somente *H. impetiginosus* comportou-se de forma oposta, em que o tratamento de seca utilizou preferencialmente sítios distais ($r < 0$) em relação ao controle (Figura 9). Não houve gene DE-APA compartilhado entre as quatro espécies. As espécies com os maiores números de compartilhamento, foram *H. ochraceus* com *T. aurea* (102 genes DE-APA comuns), e *H. ochraceus* com *H. serratifolius* (103). Entre essas três espécies, houve 13 genes DE-APA comuns (Figura 10). Esse baixo compartilhamento de genes DE-APA contrasta com o elevado compartilhamento de genes com APA (Figura 8), indicando que o uso dos sítios é menos conservado, mesmo entre espécies filogeneticamente próximas.

Figura 9 - Classificação dos genes DE-APA quanto a preferência por sítios proximais e distais no tratamento sob seca nas quatro espécies da aliança Tabebuia.



Fonte: Da autora (2022)

Figura 10 – Diagrama de Venn demonstrando o compartilhamento de genes DE-APA entre as quatro espécies da Aliança Tabebuia. Genes com DE-APA utilizam diferentes sítios de poliadenilação quando submetidas à seca.



Fonte: Da autora (2022)

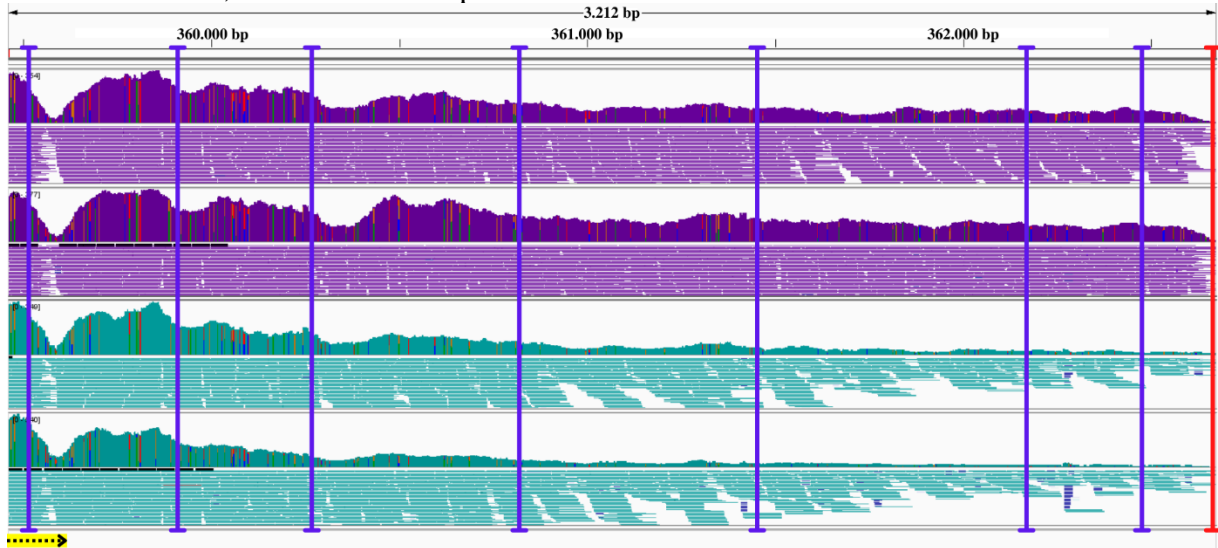
Os genes DE-APA foram analisados quanto ao número de sítios APA observados. A maior parte dos genes com DE-APA apresentaram somente dois sítios APA (um proximal e um distal). Somente *H. serratifolius* apresentou maior quantidade de genes DE-APA com 3 sítios de poliadenilação. Além disso, a presença de mais de cinco sítios APA foi observada somente em um gene com DE-APA de *T. aurea* (Tabela 3). Nessa espécie, o gene CDL12_00024 possui oito sítios APA e, quando desafiado com a seca, utilizou preferencialmente os sítios mais distais ($r < 0$) em comparação com o grupo controle. Dos oito sítios APA observados neste gene, somente um se encontrou em região CDS (Figura 11).

Tabela 3 – Número de genes DE-APA de acordo com a quantidade de sítios APA identificados nas quatro espécies da Aliança Tabebuia.

	<i>T. aurea</i>	<i>H. ochraceus</i>	<i>H. serratifolius</i>	<i>H. impetiginosus</i>
2 sítios	180 (48,78%)	2491 (70,57%)	159 (43%)	68 (56,7%)
3 sítios	155 (42,01%)	971 (27,51%)	170 (46%)	44 (36,7%)
4 sítios	29 (7,86%)	64 (1,81%)	36 (9%)	7 (5,8%)
5 sítios	4 (1,08%)	4 (0,11%)	4 (1%)	1 (0,8%)
8 sítios	1 (0,27%)	0	0	0
Total	369	3530	368	120

Fonte: Da autora (2022)

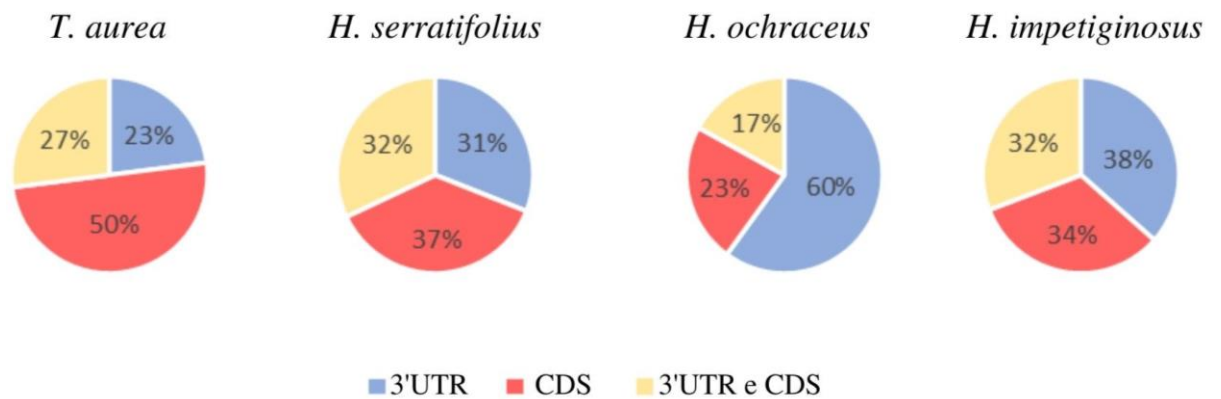
Figura 11 - Comparação da densidade de sequências de RNA de *T. aurea* alinhadas na região 3' do gene CDL12_00024. A imagem mostra as sequências de RNA de duas amostras do tratamento de seca (roxo) e duas do controle (verde-água). O fim da região codante (CDS) do ge CDL12_00024 é representada pelo retângulo amarelo no canto inferior esquerdo. As linhas verticais representam a localização dos sítios APA, em azul os sítios proximais e em vermelho o sítio distal.



Fonte: Gerada pela autora (2022) no programa IGV (ROBINSON et al. 2011)

Os genes DE-APA foram classificados quanto à região de ocorrência dos sítios APA, sendo CR-APA aqueles com ocorrência de sítios APA em região codante (CDS – *Coding DNA Sequence*) e UTR-APA os genes com sítios APA somente em região 3' não traduzida (3'UTR – *3' Untranslated Region*). *T. aurea* e *H. serratifolius* apresentaram os maiores números de genes CR-APA (50% e 37%, respectivamente) e menor número de genes UTR-APA. Além disso, ambas as espécies apresentaram uma proporção intermediária de genes DE-APA contendo tanto sítios APA em CDS quanto em 3'UTR, sendo 27% em *T. aurea* e 32% em *H. serratifolius*. Para *H. ochraceus* e *H. impetiginosus* foram encontrados resultados opostos, com maior número de genes UTR-APA: 60% e 38%, respectivamente. Genes DE-APA com sítios APA ocorrendo em ambas as regiões foram menos frequentes nessas espécies. *H. ochraceus* possui 23% de genes CR-APA, e *H. impetiginosus* apresenta 34% (Figura 12).

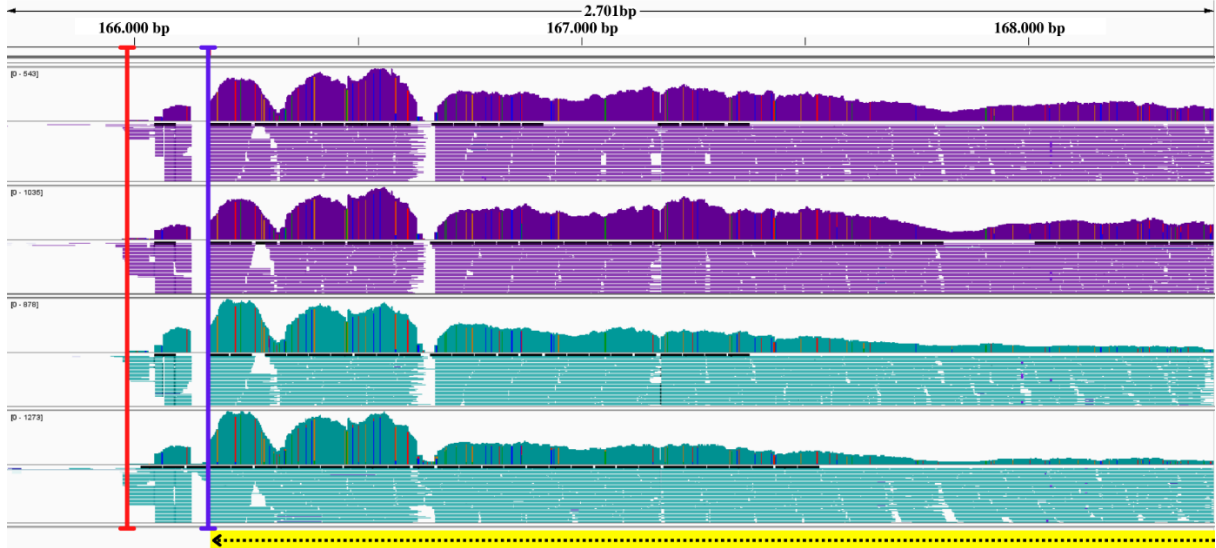
Figura 12 - Proporções da localização dos sítios APA entre os genes DE-APA, sob seca, entre as quatro espécies da aliança Tabebuia.



Fonte: Da autora (2022)

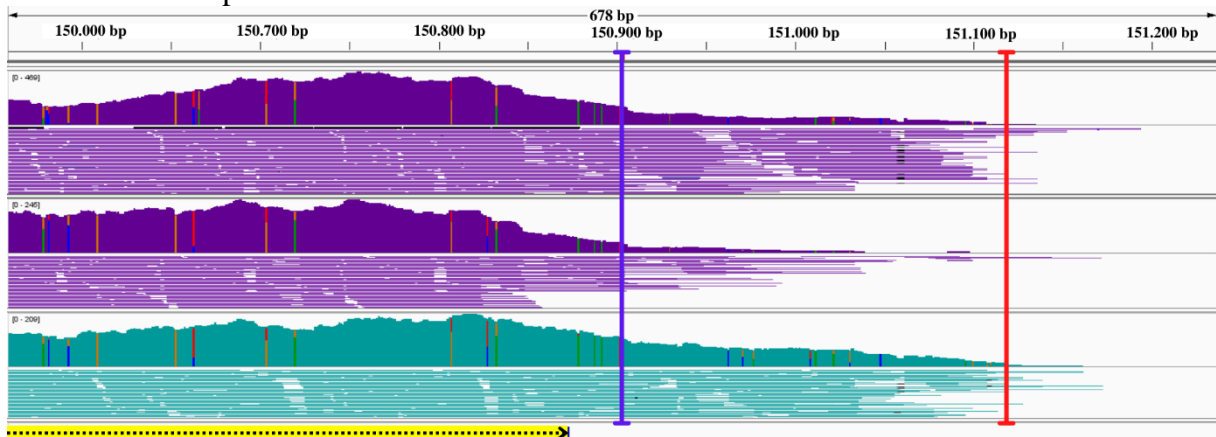
Genes DE-APA das quatro espécies foram selecionados para ilustrar a região de ocorrência dos sítios APA (CDS ou 3'UTR) e o padrão de utilização dos sítios (proximais ou distais). O gene CDL12_04454 em *T. aurea* e o CDL12_00400 de *H. ochraceus*, foram classificados como UTR-APA e apresentaram dois sítios APA, um proximal e um distal. Em ambos os casos, o sítio APA proximal foi preferencialmente utilizado no tratamento de seca, enquanto o sítio APA distal foi mais utilizado no controle ($r > 0$). Dessa forma, é possível visualizar uma quantidade maior de *reads* alinhadas, após o sítio proximal, no grupo controle (alongamento da 3'UTR) e uma menor quantidade no grupo seca (encurtamento da 3'UTR) (Figuras 13 e 14). Em *H. impetiginosus* o gene CDL12_00203 foi classificado como CR-APA. Neste caso, as amostras do grupo de seca utilizaram preferencialmente o sítio APA distal em comparação com o controle ($r < 0$), que utilizou mais o sítio APA proximal (Figura 15). Em genes CR-APA há alteração na sequência da proteína a ser sintetizada, uma vez que os sítios APA estão contidos em CDS.

Figura 13 - Comparação da densidade de sequências de RNA de *T. aurea* alinhadas na região 3' do gene CDL12_04454 (UTR-APA). A imagem mostra as sequências de RNA de duas amostras do tratamento seca (roxo) e duas do controle (verde-água). O fim da região codante (CDS) é representada pelo retângulo amarelo no canto inferior. As linhas verticais representam as localizações dos sítios APA, em azul o sítio proximal e em vermelho o sítio distal.



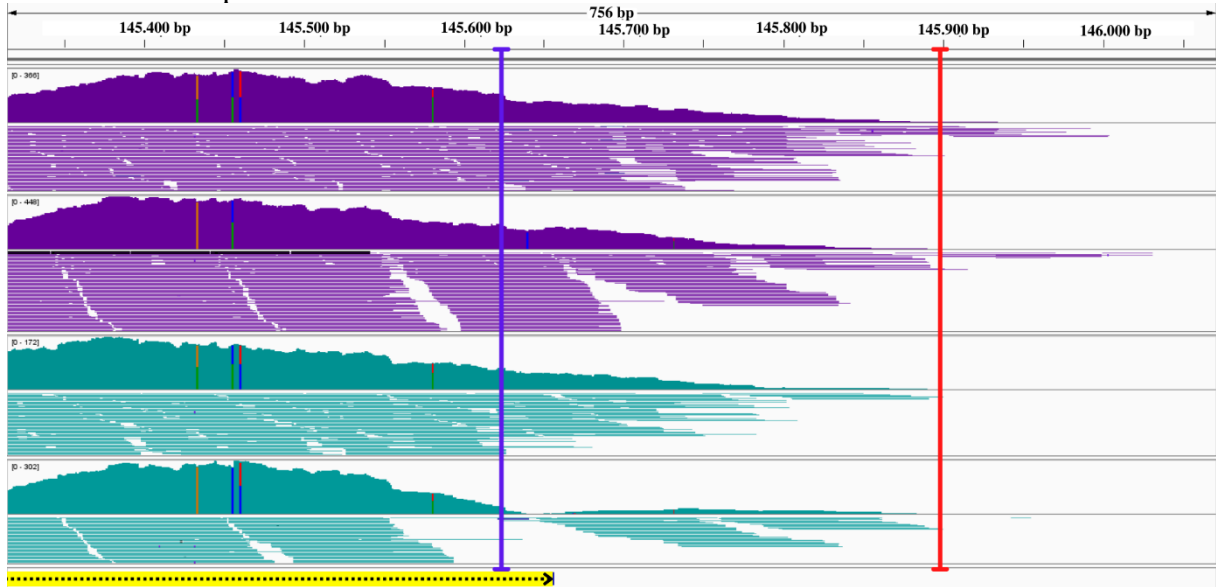
Fonte: Gerada pela autora (2022) no programa IGV (ROBINSON et al. 2011)

Figura 14 - Comparação da densidade de sequências de RNA de *H. ochraceus* alinhadas na região 3' do gene CDL12_00400 (UTR-APA). A imagem mostra as sequências de RNA de duas amostras do tratamento seca (roxo) e uma do controle (verde-água). O fim da região codante (CDS) é representada pelo retângulo amarelo no canto inferior. As linhas verticais representam as localizações dos sítios APA, em azul o sítio proximal e em vermelho o sítio distal.



Fonte: Gerada pela autora (2022) no programa IGV (ROBINSON et al. 2011)

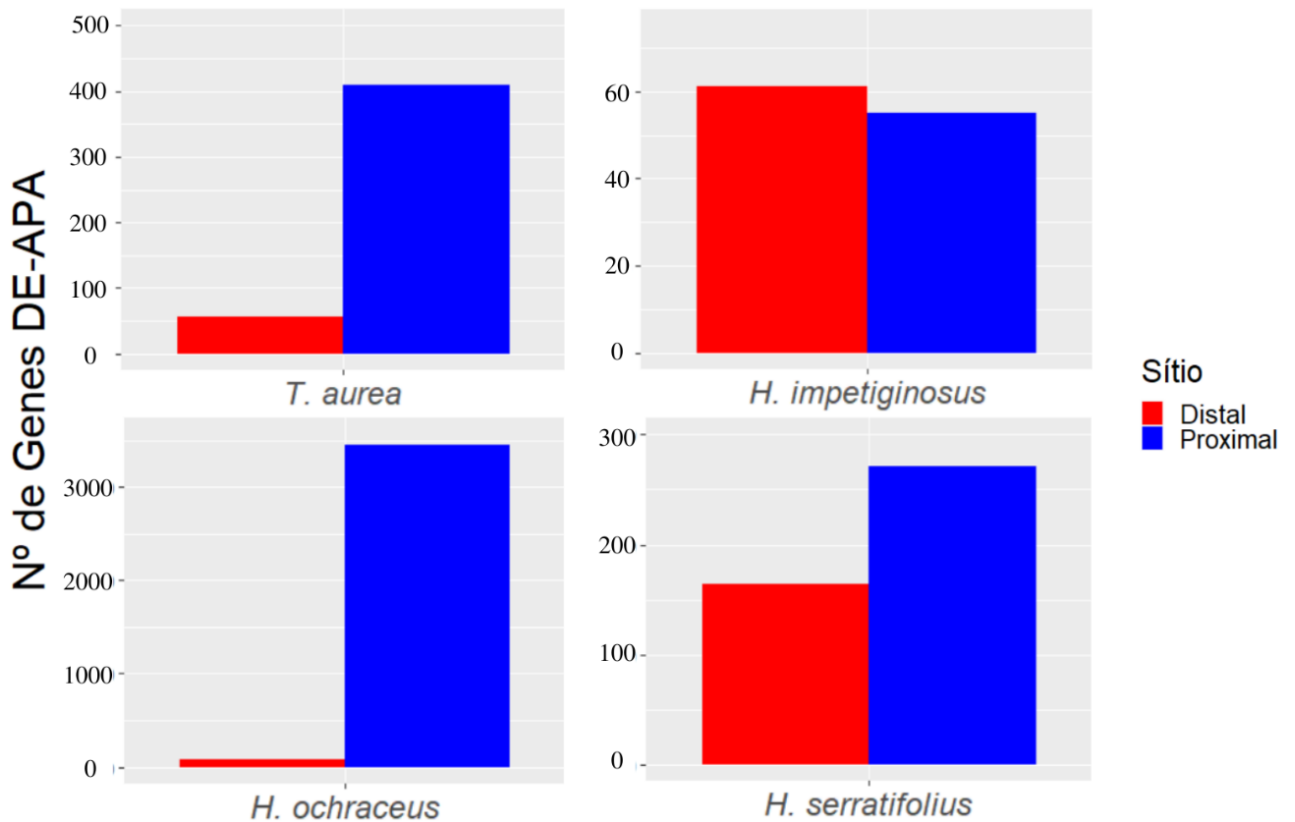
Figura 15 – Comparação da densidade de sequências de RNA de *H. impetiginosus* alinhadas na região 3' do gene CDL12_00203 (UTR-APA). A imagem mostra as sequências de RNA de duas amostras do tratamento seca (roxo) e duas do controle (verde-água). O fim da região codante (CDS) é representada pelo retângulo amarelo no canto inferior. As linhas verticais representam as localizações dos sítios APA, em azul o sítio proximal e em vermelho o sítio distal.



Fonte: Gerada pela autora (2022) no programa IGV (ROBINSON et al. 2011)

Como *H. ochraceus* foi a espécie com o maior número de genes DE-APA, refizemos as análises, para verificar se o número menor de réplicas nessa espécie influenciou esse resultado. Para isso, a identificação de APA e DE-APA foram refeitas selecionando-se aleatoriamente duas réplicas do grupo seca e uma do grupo controle para todas as demais espécies, igualando com *H. ochraceus*. O padrão dos genes DE-APA encontrados (Figura 16) foi muito semelhante ao encontrado quando todas as três réplicas biológicas foram consideradas em *H. impetiginosus*, *H. serratifolius* e *T. aurea* (Figura 9). Ou seja, *H. impetiginosus* manteve-se como a espécie com o menor número de genes DE-APA, *H. ochraceus* com o maior, e *T. aurea* e *H. serratifolius* com valores intermediários.

Figura 16 - Verificação do número de genes DE-APA nas quatro espécies da aliança Tabebuia, considerando o mesmo número de réplicas de *H. ochraceus*. Essa análise verificou que a perda de réplicas em *H. ochraceus* não explica a maior quantidade de genes DE-APA nessa espécie.



Fonte: Da autora (2022)

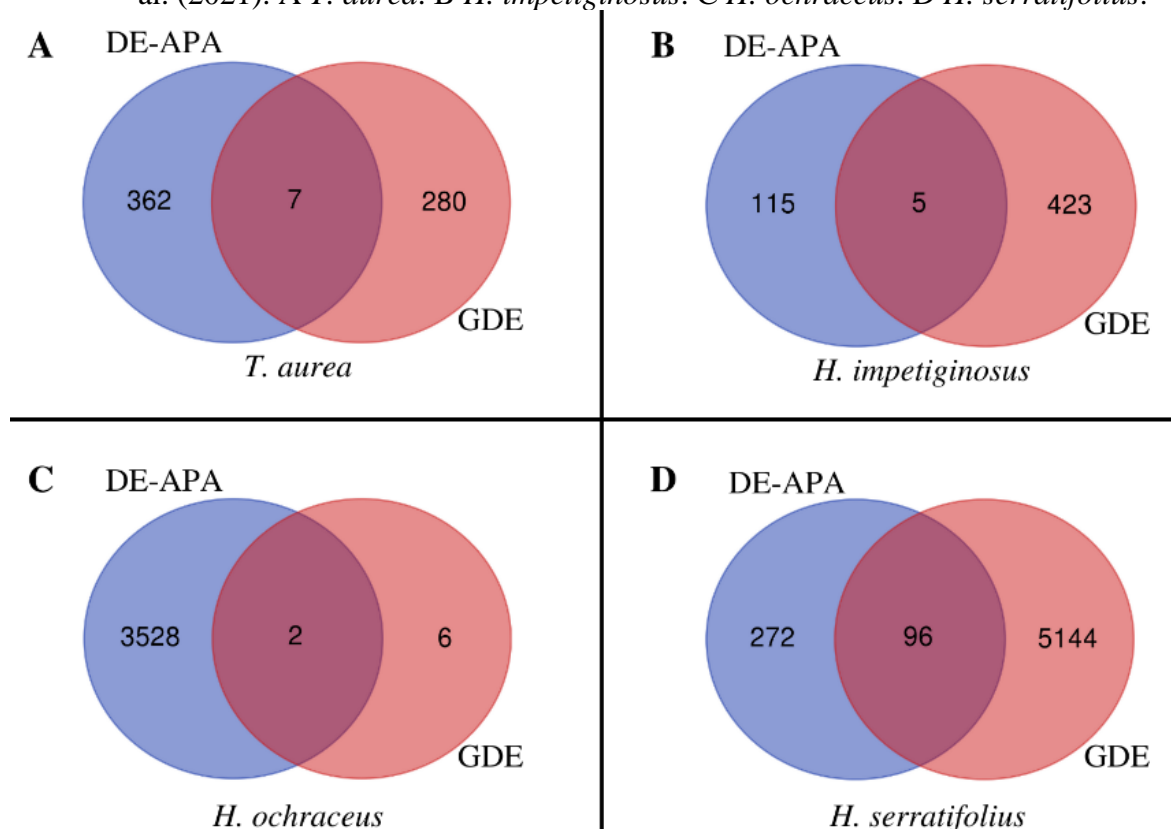
4.2 Expressão de genes UTR-APA

Os genes DE-APA identificados foram comparados com os genes diferencialmente expressos (DE) obtidos por Sobreiro et al. (2021), já que o encurtamento da região 3'UTR pode, por exemplo, remover sítios de ligação à miRNAs e, com isso, aumentar a expressão do gene (DI GIAMARTINO et al., 2011; YE et al., 2019). A espécie com o maior número de genes que compartilham esse padrão, isto é, com DE-APA e GDE, foi *H. serratifolius* (96). Por outro lado, *H. ochraceus* apresentou o menor número de compartilhamentos, com somente dois genes contendo DE-APA e expressão diferencial. *T. aurea* e *H. impetiginosus*, apresentaram respectivamente sete e cinco genes DE e DE-APA compartilhados (Figura 17). Esse baixo compartilhamento indica que o controle da poliadenilação alternativa e da expressão gênica, em resposta à seca, são mecanismos independentes nessas espécies.

Como visto, genes UTR-APA levam a alterações no tamanho da 3'UTR. Quando há o alongamento da 3'UTR (sítio distal) a expressão do gene tende a diminuir e, de forma oposta,

quando há o encurtamento da 3'UTR (sítio proximal) a expressão do gene tende a aumentar (DI GIAMARTINO et al., 2011; YE et al., 2019). Dentre os genes DE-APA e DE, selecionamos somente aqueles em que a APA ocorria exclusivamente em região 3'UTR (UTR-APA). Posteriormente, avaliamos se havia relação entre a utilização dos sítios APA (proximais ou distais) dos UTR-APA, com o padrão de expressão (*up* ou *down*) destes genes. Dentre os genes compartilhados em *H. serratifolius* foram identificados 31 UTR-APA, sendo três com 3'UTR longa e *down* regulados; dois com 3'UTR longa e *up* regulados; 11 com 3'UTR curta e *down* regulados; e 12 com 3'UTR curta e *up* regulados. O teste exato de Fisher não foi significativo (p-valor = 0.6358) e, portanto, não houve relação significativa entre o alongamento ou encurtamento da 3'UTR com o padrão de expressão dos genes. Em *H. ochraceus* e *H. impetiginosus*, não houve genes UTR-APA com expressão diferencial. Em *T. aurea* foram observados quatro genes UTR-APA que também foram DE, um com alongamento da 3'UTR e *down* regulado; um com encurtamento da 3'UTR e *down* regulado; e dois com encurtamento da 3'UTR e *up* regulados (Tabela 4). O teste de Fisher revelou que, assim como em *H. serratifolius*, em *T. aurea* não houve relação significativa (p-valor = 1).

Figura 17 - Diagrama de Venn demonstrando o compartilhamento de genes DE-APA, em azul, e genes diferencialmente expressos (DE), em vermelho, do trabalho de Sobreiro et al. (2021). A *T. aurea*. B *H. impetiginosus*. C *H. ochraceus*. D *H. serratifolius*.



Fonte: Da autora (2022)

Tabela 4 - Tabela comparando os genes UTR-APA/GDE, quanto à utilização dos sítios (proximais ou distais) e o padrão de expressão destes genes (*up* ou *down*), nas quatro espécies da Aliança *Tabebuia*.

Espécie	Expressão	Sítio APA do UTR-APA		P-valor (Teste exato de Fisher)
		Proximal	Distal	
<i>H. ochraceus</i>	<i>Up</i>	-	-	-
	<i>Down</i>	-	-	
<i>H. impetiginosus</i>	<i>Up</i>	-	-	-
	<i>Down</i>	-	-	
<i>H. serratifolius</i>	<i>Up</i>	15	2	0,6358
	<i>Down</i>	11	3	
<i>T. aurea</i>	<i>Up</i>	2	-	1
	<i>Down</i>	1	1	

Fonte: Da autora (2022)

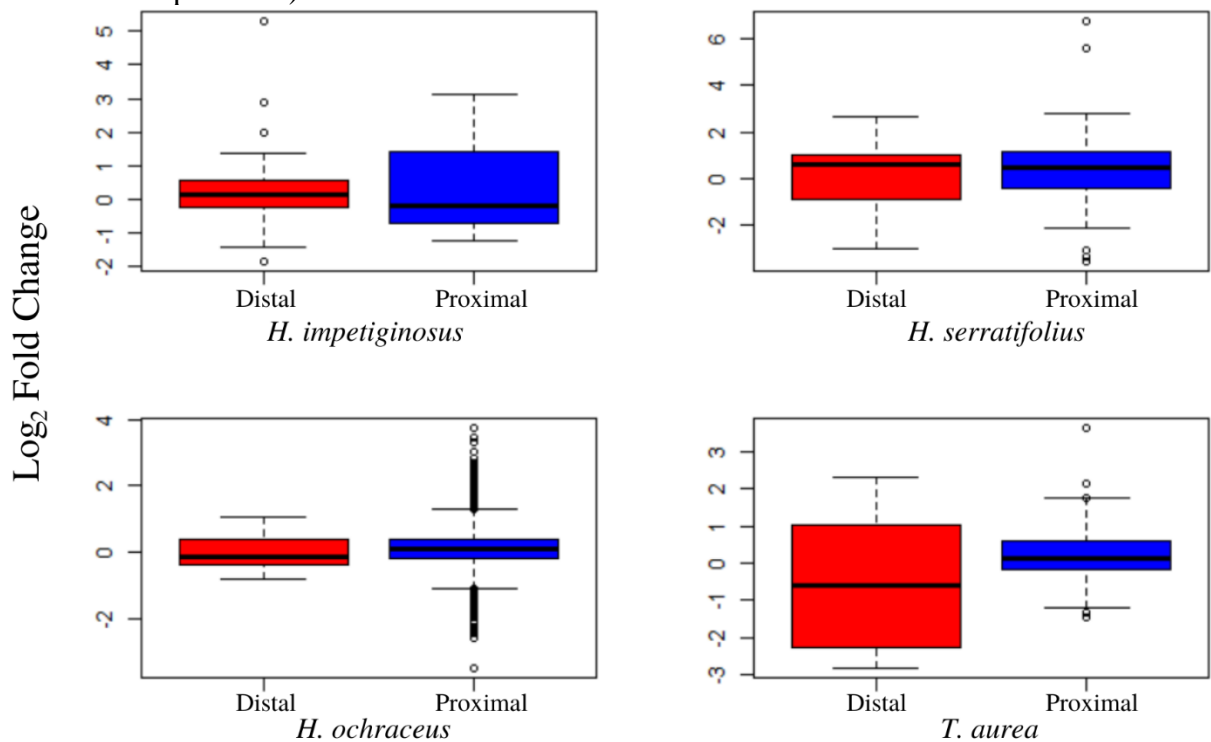
Posteriormente, realizou-se também uma análise comparativa dos genes UTR-APA com os dados disponíveis da expressão destes genes (SOBREIRO et al., 2021), buscando-se associação com a direção da expressão (*up* ou *down*-regulados na seca) independentemente de serem genes DE significativos ou não. *H. ochraceus* foi a única espécie que apresentou relação significativa, pelo teste de Fisher (p-valor = 0,027), entre a utilização dos sítios APA e a direção da expressão. Nessa espécie, isoformas com encurtamento da 3'UTR eram em sua maioria *up* regulados (1.258 ou 60,7%), e isoformas com alongamento da 3'UTR eram mais *down* regulados (14 ou 63,6 %) (Tabela 5) (Figura 18). Embora em *T. aurea* haja uma aparente relação, pois os transcritos com as 3'UTR curtas são mais *up* (51) do que *down* (27) regulados, e os transcritos com alongamento da 3'UTR são mais *down* (4) do que *up* (2) regulados (Tabela 5) (Figura 18), o teste exato de Fisher não foi significativo (p-valor = 0,19). Em *H. impetiginosus*, os transcritos com 3'UTR menores forma mais *down* regulados (10 ou 55,6%), e aqueles com 3'UTR maiores foram mais *up* regulados (17 ou 65,4%) (Tabela 5; Figura 18). Porém, em *H. impetiginosus* também não houve relação significativa (p-valor = 0,22, pelo teste exato Fisher) entre utilização do sítio APA e expressão. Para *H. serratifolius* não houve relação aparente (Figura 18), pois tanto transcritos com 3'UTR longas quanto curtas foram mais *up* regulados (Tabela 5).

Tabela 5 - Comparação entre a utilização de sítios proximais e distais, de genes UTR-APA, e o padrão de expressão destes genes (*UP* ou *down*), nas quatro espécies da Aliança *Tabebuia*.

Espécie	Expressão	Sítio APA do UTR-APA		P-valor (Teste exato de Fisher)
		Proximal	Distal	
<i>H. ochraceus</i>	<i>UP</i>	1.258	8	0,02688
	<i>Down</i>	813	14	
<i>H. impetiginosus</i>	<i>UP</i>	8	17	0,2215
	<i>Down</i>	10	9	
<i>H. serratifolius</i>	<i>UP</i>	58	11	1
	<i>Down</i>	34	7	
<i>T. aurea</i>	<i>UP</i>	51	2	0,1871
	<i>Down</i>	27	4	

Fonte: Da autora (2022)

Figura 18 - Diferenças de expressão (Log_2 Fold Change) entre tratamento de seca e controle para os genes UTR-APA das quatro espécies da Aliança *Tabebuia*. Caixas vermelhas representam os UTR-APA com alongamento da 3'UTR (sítio distal). Caixas azuis representam os UTR-APA com encurtamento da 3'UTR (sítio proximal).



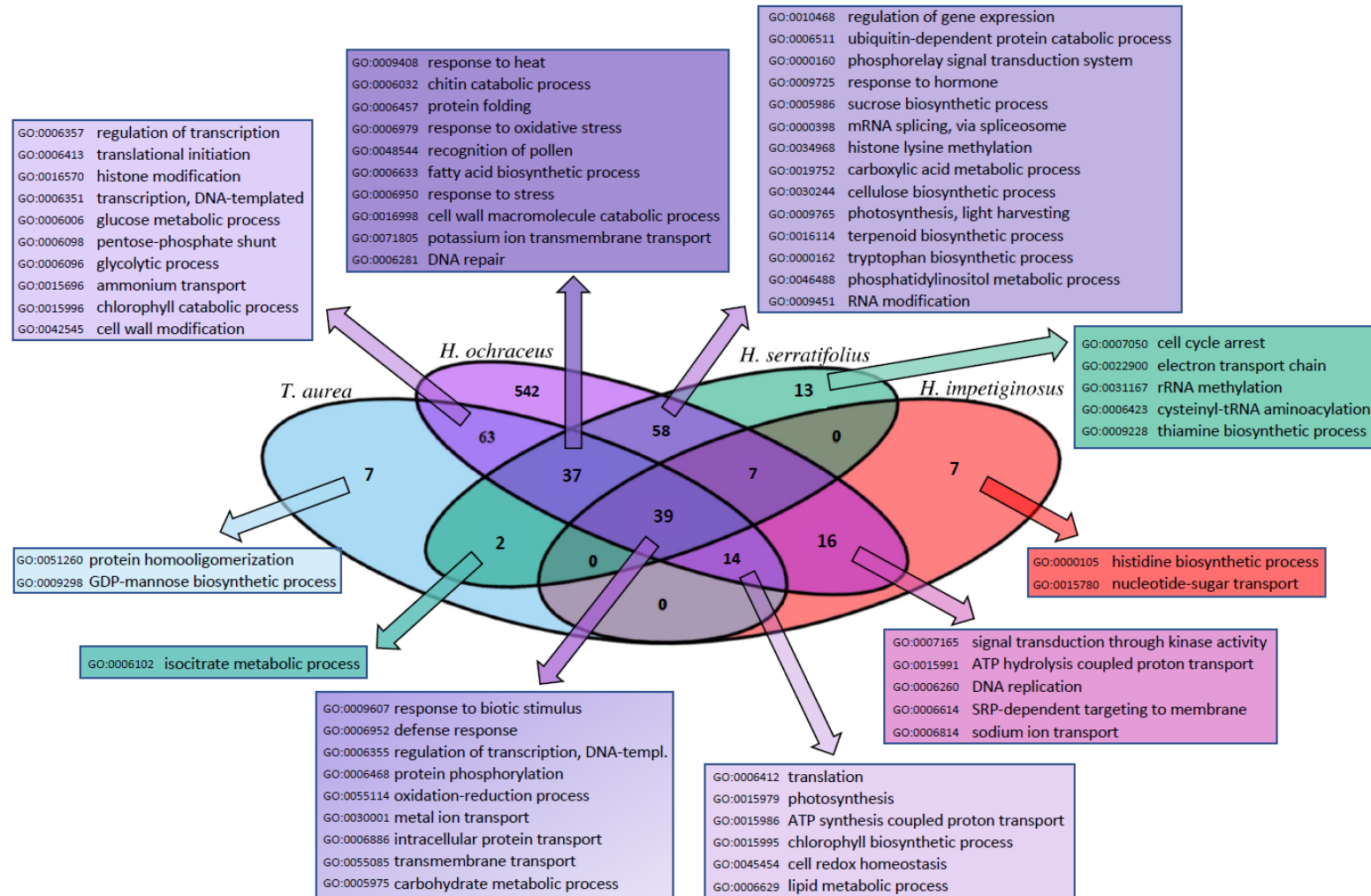
Fonte: Da autora (2022)

4.3 Análise de enriquecimento funcional

Analizamos inicialmente se haviam categorias GO (*Gene Ontology*) (ASHBURNER et al., 2000), dos genes DE-APA, compartilhadas entre as quatro espécies estudadas. *H. ochraceus* e *T. aurea* foram as espécies com o maior número de categorias GO compartilhadas (63), sendo as principais delas envolvidas na regulação da transcrição e tradução, “glucose metabolic process” e “chlorophyll catabolic process” (Figura 19). Entre *H. serratifolius* e *H. ochraceus*, houve 58 categorias GO coincidentes, das quais se destacam “regulation of gene expression.”, “ubiquitin-dependent protein catabolic process”, “phosphorelay signal transduction system”, “response to hormone” e “photosynthesis, light harvesting”. *T. aurea*, *H. serratifolius* e *H. ochraceus* apresentaram 37 categorias GO compartilhadas, sendo as principais “response to heat”, “response to oxidative stress”, “fatty acid biosynthetic process” e “response to stress”. Entre *H. ochraceus* e *H. impetiginosus*, das 16 categorias GO coincidente, pode-se destacar “signal transduction through kinase activity”, “DNA replication” e “SPR-dependent targeting to membrane”. Além disso, as quatro espécies compartilharam 39 categorias GO envolvidas em “response to biotic stimulus”, “defense response”, “protein phosphorylation” e “carbohydrate metabolic process”. Mais detalhes sobre o compartilhamento das categorias GO entre as espécies podem ser encontrados na Figura 19.

Por fim, o teste exato de Fisher detectou que somente *H. ochraceus* apresentou termos GO enriquecidos entre os genes DE-APA. Do total (40), 33 termos se enquadravam na categoria processo biológico (Apêndice A) e sete na categoria componente celular (Apêndice B). Pode-se destacar algumas categorias GO que estavam envolvidas em regulação da transcrição, processos catabólicos e regulação do transporte/localização de moléculas (Apêndice A). A falta de categorias GO enriquecidas nas demais espécies, pode ser resultado do menor número de genes com DE-APA, quando comparado com *H. ochraceus*. Um número menor de genes, diminui o poder estatístico no teste exato de Fisher, dificultando a identificação de categorias GO enriquecidas.

Figura 19 - Diagrama de Venn demonstrando o compartilhamento das categorias GO dos genes DE-APA entre quatro espécies da Aliança *Tabebuia* sob seca. Os números indicados na imagem representam os três domínios GO (processo biológico, componente celular e função molecular), porém, somente as principais categorias de processo biológico são indicadas nas caixas.



Fonte: Da autora (2022)

5 DISCUSSÃO

As alterações na expressão gênica estão diretamente relacionadas com a sobrevivência das plantas, em especial quando estas são submetidas a situações adversas e necessitam de rápidas mudanças em sua fisiologia (TAIZ; ZEIGER, 2002). Nesse sentido, pesquisas recentes têm demonstrado que a poliadenilação alternativa (APA) participa amplamente das respostas, em diversas espécies vegetais, a estresses bióticos e abióticos (LORENZO et al., 2017; YAN et al., 2021; YE et al., 2019). As análises realizadas neste estudo, revelaram que espécies de ipê, quando submetidas à seca, utilizam a poliadenilação alternativa em muitos genes (Figura 7). Das quatro espécies analisadas, somente *H. impetiginosus* apresentou um número relativamente pequeno (1.510) de genes com APA (Tabela 2) e destes, 954 foram compartilhados com as outras três espécies estudadas. Além disso, de todos os genes com APA identificados em *H. serratifolius*, *H. ochraceus* e *T. aurea*, a maioria (9.990) eram coincidentes nas três espécies (Figura 8), indicando que os sítios APA podem ser conservados pelo menos em espécies filogeneticamente próximas. Para a comprovação dessa hipótese, é interessante avaliar a sequência dos sítios canônicos de poliadenilação, dos não canônicos (sítios APA), e compará-los entre as espécies, a fim de constatar se eles são realmente conservados entre elas.

Além disso, nosso estudo demonstra que muitos desses sítios APA tiveram uso diferencial significativo (DE-APA) quando as plantas de ipê foram submetidas à seca. Dessa forma, nossos resultados sugerem que APA pode estar envolvida no controle co-transcricional da expressão em resposta à seca, de muitos genes de ipê, especialmente em *H. serratifolius*, *H. ochraceus* e *T. aurea*. Porém, cabe ressaltar que, ao contrário do grande compartilhamento de genes com múltiplos sítios APA, o uso diferencial significativo, entre tratamento de seca × controle, não foi conservado entre as espécies (Figura 10). Por outro lado, o grande compartilhamento de categorias funcionais Gene Ontology (GO) (Figura 19), indica que apesar dos genes/isoformas com DE-APA nas diferentes espécies não serem os mesmos, eles podem participar dos mesmo processos biológicos. Esse é um sinal biológico importante nos resultados, pois indica que alguns processos são, possivelmente, mais controlados por APA do que outros.

Outra constatação importante em nossos resultados, é que a maioria dos genes com APA não fizeram uso diferencial significativo (DE-APA) desses sítios entre tratamento de seca × controle (Tabela 6). Esse resultado pode ser devido ao fato de que a utilização de dados RNA-seq de leitura curta não é o melhor método para identificação de APA e, talvez, por possuírem

quantidade significativa de sequências com sobreposição, as leituras podem ter sido atribuídas a isoformas incorretas (SHAH et al., 2021). Análises com metodologias mais adequadas à identificação de APA, como por exemplo PAS-Seq, 3P-Seq, Iso-Seq e 3'-Seq, podem talvez gerar resultados diferentes. Contudo, dado o alto custo dessas metodologias, a ampla disponibilidade de dados RNA-seq pode ser aproveitada para fornecer informações iniciais relevantes sobre o uso de APA em plantas (CHEN et al., 2017; SHAH et al., 2021).

Tabela 6 - Número de genes com uso diferencial significativo ($PD \geq 0,2$ e $FDR < 0,05$) e não significativo ($PD < 0,2$ e $FDR \geq 0,05$) de APA, nas quatro espécies da Aliança *Tabebuia* sob seca.

Uso diferencial	<i>T. aurea</i>	<i>H. impetiginosus</i>	<i>H. ochraceus</i>	<i>H. serratifolius</i>
Significativo (DE-APA)	369 (3%)	120 (8%)	3530 (25%)	368 (2%)
Não significativo	12.720 (97%)	1.390 (92%)	10.456 (75%)	15.579 (98%)
Total	13.089	1.510	13.986	15.944

Fonte: Da autora (2022)

APA pode alterar a expressão de genes de formas diferente, de acordo com o local de ocorrência dos sítios nos genes. Nesse estudo, os genes DE-APA foram classificados em CR-APA, quando os sítios estavam localizados em região codante, e em UTR-APA, quando os sítios estavam em região 3'UTR. CR-APA pode levar à produção de isoformas de proteínas, que podem ter funções distintas ou até mesmo perda de função (CHEN et al., 2017; DI GIAMMARTINO; NISHIDA; MANLEY, 2011). Em nosso estudo, encontramos uma maior proporção de CR-APA em espécies de ipê sob seca (Figura 12) quando comparado ao que tem sido descrito para outras espécies vegetais em condições normais, como *Arabidopsis* (17%) e *Medicago truncatula* (10%) (WU et al., 2011, 2014). Porém, semelhante aos nossos resultados, em *Arabidopsis thaliana* sob condição de hipoxia, também há relato do aumento de genes com APA em região codante, bem como a associação destes eventos com a instabilidade das isoformas de mRNA geradas (LORENZO et al., 2017). Portanto, nossos resultados podem indicar que a seca à qual as espécies de ipê foram submetidas, levaram a um possível aumento da poliadenilação alternativa em região codante. Embora CR-APA tenha sido observada em algumas espécie de plantas, sua funcionalidade ainda é pouco explorada e, portanto, pouco compreendida (HUNT, 2012; XING; LI, 2011). Dessa forma, os genes CR-APA identificados no presente estudo requerem mais atenção, abrindo a possibilidade de novas pesquisas que visem compreender (em nível de proteína) as consequências deste evento.

Em *H. ochraceus*, *T. aurea* e *H. serratifolius*, identificamos que a maioria dos genes UTR-APA apresentaram encurtamento da 3'UTR, devido à utilização preferencial de sítios proximais na condição de seca. Por outro lado, *H. impetiginosus* quando submetido a esse estresse, apresentou um alongamento da 3'UTR na maioria dos genes UTR-APA (Figura 10). O alongamento ou encurtamento da 3'UTR tem sido frequentemente associado, respectivamente, com a diminuição ou aumento da expressão dos genes (SANDBERG et al., 2008; YE et al., 2019). Essa associação tem sido feita devido ao fato de que, na região 3'UTR encontram-se sequências-alvo de reguladores (por exemplo, miRNAs e RBPs), que podem levar à desestabilização/degradação do mRNA (DI GIAMMARTINO; NISHIDA; MANLEY, 2011; JI et al., 2011; SANDBERG et al, 2008; YE et al., 2019). Em nossos resultados, somente *H. ochraceus* apresentou uma relação significativa entre o aumento da expressão de genes com 3'UTR curtas, e a diminuição da expressão de genes com 3'UTR longas (Tabela 5). Esse resultado pode indicar que a escolha dos sítios proximais ou distais nos genes desta espécie, talvez estejam excluindo ou incluindo sequências alvo de fatores repressivos dos mRNAs (YE et al., 2019). Esse resultado é particularmente interessante, pois *H. ochraceus* foi a única espécie que praticamente não teve o transcrito modulado (somente 14 genes DE) quando desafiado com a seca (SOBREIRO et al., 2021). Esse resultado abre a possibilidade de que, nessa espécie, a regulação a expressão gênica em resposta a seca, pode se dar via uso diferencial de sítios APA, ao invés de uma alteração na quantidade total de mRNA, conforme mensurado via RNA-Seq por Sobreiro et al. (2021).

Nas outras três espécies, não houve relação aparente entre o nível de expressão e o comprimento da 3'UTR (Figura 18), indicando que os eventos de APA nos genes destas espécies não alteraram de forma significativa a estabilidade dos mRNAs. Uma hipótese que pode ser levantada para esse resultado é que, nestas espécies, somente poucos genes UTR-APA contém sítios que são reconhecidos por elementos de regulação do mRNA e, portanto, independente do comprimento da 3'UTR, não sofrem os efeitos destes reguladores (ELKON; UGALDE; AGAMI, 2013).

Apesar de nossos resultados serem promissores e inovadores, é importante ressaltar que os dados RNA-seq não são os mais indicados para a identificação de APA. Isso porque com RNA-Seq, geralmente, a cobertura de sequências nas extremidades 5' e 3' dos genes são relativamente baixas. Uma possibilidade de correção para este problema, é a utilização da RNA-seq enriquecida em 3' (CHEN et al., 2017). Além disso, uma metodologia que pode superar os problemas da RNA-seq, é o sequenciamento de molécula longa única, com sequenciadores da empresa PacBio (RHOADS; AU, 2015). Esse método é conhecido como Iso-seq e permite a

identificação e quantificação mais precisa de APA, já que sequencia toda a extensão dos transcritos. Porém, essas ainda são metodologias onerosas. Além disso, têm-se hoje uma ampla disponibilidade de dados RNA-seq (CHEN et al., 2017; SHAH et al., 2021). Dessa forma, para lidar com a enorme quantidade de dados gerados por análises RNA-seq, é importante que se explore as ferramentas já disponíveis (ZHAO et al., 2014). Aqui optamos por utilizar a ferramenta APATrap (YE et al., 2018), porém, existem outras ferramentas disponíveis com o mesmo propósito, como TAPAS (AREFEEN et al., 2018), QAPA (HA; BLENCOWE; MORRIS, 2018), DaPars (XIA et al., 2014), entre outras. Dessa forma, é interessante que sejam realizadas análises comparativas dos nossos resultados, com outras ferramentas disponíveis, para que a caracterização de APA em ipê seja o mais precisa possível.

A função dos genes de ipês, com uso diferencial de sítios APA (genes DE-APA) sob seca, foi analisada por meio de categorias GO. Uma das principais resposta à seca nas plantas, é a regulação da atividade fotossintética. A redução da fotossíntese pode se dar, por exemplo, pela alteração da pressão de turgescência nas células guarda (pelo transporte de íons e água através das membranas) de forma a estimular o fechamento estomático (BASU et al., 2016; OSAKABE et al., 2014). Em todas as espécies analisadas, encontramos diversas categorias GO, entre os genes DE-APA, envolvidas em fotossíntese (Figura 19), como por exemplo, “processo catabólico da clorofila”, “processo biossintético da clorofila”, “transporte transmembrana do íon potássio”, “cadeia transportadora de elétron”, e muitas outras. Esse resultado indica que APA pode estar relacionada com a regulação co-transcricional da fotossíntese em resposta à seca em ipê.

Além disso, compartilhadas entre as quatro espécies, foram observadas categorias GO envolvidas em “resposta de defesa” e “resposta à estímulo biótico” (Figura 19). Entre *H. ochraceus*, *H. serratifolius* e *T. aurea*, categorias como “resposta ao calor”, “resposta ao estresse oxidativo” e “resposta ao estresse”, também foram comuns. Portanto, alguns genes de ipês que foram regulados por APA após a exposição à seca, são responsáveis por resposta a estresses. Um ponto interessante é que, genes DE-APA com essas categorias funcionais, foram identificados em mais de uma das quatro espécies de ipê (Figura 19), indicando que esse método de regulação co-transcricional pode ser particularmente importante na resposta dessas espécies arbóreas à seca.

Vias de regulação do metabolismo do RNA, bem como de regulação direta da expressão gênica, também foram observadas tanto individualmente nas espécies, como compartilhadas entre elas (Figura 19). Genes com poliadenilação alternativa, envolvidos em termos GO de metabolismo do RNA, também foram observados em *Populus trichocarpa* na resposta a

estresses de temperatura (YAN et al., 2021). A existência de categorias envolvidas em tradução e catabolismo de proteínas pela via das ubiquitinas, indicam que APA também pode estar relacionada com a regulação de vias de controle pós-transcricional de genes. Dessa forma, a APA em ipê sob seca, parece regular genes que atuam sobre a própria modulação da expressão gênica. O grande compartilhamento de categorias GO entre os genes com DE-APA das diferentes espécies, demonstra que esse mecanismo regulatório possivelmente incide com maior frequência em genes envolvidos em processos biológicos e funções moleculares específicas. Como já mencionado, muitas das categorias compartilhadas estão envolvidas em respostas a mudanças no ambiente, como mecanismos de defesa, resposta a estresses, transdução de sinais, regulação da transcrição, degradação de proteínas por meio de ubiquinas, entre outros. Em contraste com esse grande compartilhamento de categorias GO (Figura 19), tem-se um baixo número de genes DE-APA coincidentes entre as espécies (Figura 10). Isso provavelmente ocorre devido ao modo como as plantas evoluíram, em que múltiplas rodadas de poliploidização geraram famílias multigênicas em que os genes podem assumir novas ou sub funcionalidades (“neo or subfunctionalization”) em diferentes linhagens evolutivas (VAN de PEER; MIZRACHI; MARCHAL, 2017).

Sabe-se que a poliadenilação alternativa, em conjunto com outros mecanismos de regulação da expressão gênica, permite o redirecionamento da produção proteica, em células vegetais submetidas a condições de estresse (CHAKRABARTI et al., 2020). Portanto, nossos resultados tomados em conjunto, sugerem que APA atua como um importante mecanismo de regulação co-transcricional, em resposta a seca, em espécies de ipê. Como as espécies de ipês, particularmente as de Cerrado (*H. ochraceus* e *T. aurea*), apresentam alta tolerância à seca, o estudo dessas espécies pode revelar quais são os mecanismos que as tornam mais adaptadas à seca. A identificação de muitos genes DE-APA (3.530) em *H. ochraceus* pode, por exemplo, explicar porque essa, que é uma espécie altamente adaptada à seca, praticamente não apresentou genes diferencialmente expressos sob seca (somente 14 em Sobreiro et al., 2021).

6 CONCLUSÃO

Utilizando a ferramenta APATrap, identificamos diversos sítios de poliadenilação alternativa em *H. ochraceus*, *H. impetiginosus*, *H. serratifolius* e *T. aurea*. Embora muitos dos genes com APA sejam comuns nas quatro espécies, genes com uso diferencial significativo dos sítios de APA (DE-APA), entre os tratamentos de seca e controle, não apresentaram alto índice de compartilhamento entre elas. Dessa forma, os sítios APA parecem ser conservados entre as espécies, porém, a alteração significativa no uso desses sítios não parece ocorrer de forma semelhante entre elas. Além disso, nossas análises revelaram que as espécies de ipê, quando submetidas à seca, aumentam a ocorrência de APA em região codante (CR-APA), contrastando com o que tem sido descrito na literatura para outras espécies vegetais em condições normais. Em relação a APA ocorrendo em região 3'UTR (UTR-APA), somente *H. ochraceus* apresentou relação significativa entre a utilização de sítios proximais com o aumento da expressão dos genes, e a utilização de sítios distais com a diminuição da expressão gênica. Nas demais espécies, a UTR-APA não parece ter grande impacto na expressão dos genes.

Embora os genes DE-APA não tenham sido compartilhados entre as espécies, a análise dos termos *Gene Ontology* (GO), revelou que estes genes estavam envolvidos em processos biológicos semelhantes em todas as espécies. Das categorias GO identificadas, muitas estavam envolvidas em vias importantes de resposta a estresses e de regulação da transcrição. Esses resultados indicam que, possivelmente, as espécies de ipê submetidas à seca, utilizam a APA como um mecanismo co-transcricional tanto na regulação de vias de resposta ao estresse, quanto na regulação da própria transcrição. *H. ochraceus*, em especial, apresentou um número muito superior de genes DE-APA (3.530) e, conseqüentemente, foi a única espécie com termos GO enriquecidos. Esses foram resultados particularmente importantes, visto que *H. ochraceus* é uma espécie altamente tolerante à seca, portanto, a APA é possivelmente um dos mecanismos que medeia tal tolerância.

Por fim, nossos resultados fornecem informações iniciais importantes sobre a poliadenilação alternativa em espécies de ipê submetidas à seca. Os resultados aqui presentes, podem auxiliar no desenvolvimento de futuras pesquisas que visem compreender, de forma mais concisa, o complexo mecanismo de resposta à seca em ipês.

REFERÊNCIAS

- ALBERTS, B. et al. Controle da expressão gênica. *In*: ALBERTS, B. *Biologia molecular da célula*. 5. ed. **Artmed**, 2017a. cap. 7. p. 369–438.
- ALBERTS, B. et al. Como as células leem o genoma: do DNA à proteína. *In*: ALBERTS, B. *Biologia molecular da célula*. 5. ed. **Artmed**, 2017b. cap. 6. p. 299–368.
- AREFEEN A. et al. TAPAS: tool for alternative polyadenylation site analysis. **Bioinformatics**. 2018. Disponível em: <<https://doi.org/10.1093/bioinformatics/bty110>>.
- ARNHOLDT-SCHMITT, B. Stress-induced cell reprogramming. A role for global genome regulation? **Plant Physiology**, v. 136, n. 1, p. 2579–2586, 2004.
- ASHBURNER, M. et al. Gene ontology: Tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25–29, maio 2000.
- BASU S. et al. Plant adaptation to drought stress. **F1000Research**. 2016. Disponível em: <<https://doi.org/10.12688/f1000research.7678.1>>.
- BEAUDOING, E. et al. Patterns of variant polyadenylation signal usage in human genes. **Genome Research**, v. 10, n. 7, p. 1001–1010, 2000.
- BOLGER, M. A.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114–2120, agosto 2014. Disponível em: <<https://doi.org/10.1093/bioinformatics/btu170>>.
- BU, D. et al. KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. **Nucleic Acids Research**, v. 49, n. 1, p. 317–325, 2021.
- BURLEY, J. T. et al. Genome assemblies for two Neotropical trees: Jacaranda copaia and Handroanthus guayacan. **G3: Genes, Genomes, Genetics**, v. 11, n. 2, 2021.
- CHAKRABARTI, M. et al. Wide-ranging transcriptome remodelling mediated by alternative polyadenylation in response to abiotic stresses in Sorghum. **The Plant Journal**. 2020. Disponível em: <<https://doi.org/10.1111/tpj.14671>>.
- CHEN, H. VennDiagram: Generate high-resolution Venn and Euler plots. R package version 1.7.1. 2021. Disponível em: <<https://CRAN.R-project.org/package=VennDiagram>>.
- Chen M. et al. A survey on identification and quantification of alternative polyadenylation sites from RNA-seq data. **Brief Bioinform**. 2020. Disponível em: <<https://doi.org/10.1093/bib/bbz068>>.
- CHEN, W. et al. Alternative Polyadenylation: Methods, Findings, and Impacts Methods and Findings of Alternative Polyadenylation. **Genomics, Proteomics and Bioinformatics**, v. 15, n. 5, p. 287–300, 2017. Disponível em: <<https://doi.org/10.1016/j.gpb.2017.06.001>>.

CHENG, Z. et al. Genome-wide search and structural and functional analyses for late embryogenesis-abundant (LEA) gene family in poplar. 2021. Disponible em: <<https://doi.org/10.1186/s12870-021-02872-3>>.

COLGAN, D. F.; MANLEY, J. L. Mechanism and regulation of mRNA polyadenylation. **Genes and Development**, v. 11, n. 21, p. 2755–2766, 1997.

COLLEVATTI, R. G.; DORNELAS, M. C. Clues to the evolution of genome size and chromosome number in *Tabebuia* alliance (Bignoniaceae). **Plant Systematics and Evolution**, v. 302, n. 5, p. 601–607, 2016.

CORDEIRO, J. M. P. et al. Heterochromatin and numeric chromosome evolution in bignoniaceae, with emphasis on the neotropical clade *Tabebuia* alliance. **Genetics and Molecular Biology**, v. 43, n. 1, p. 1–11, 2020.

DE GUERRA, A. Cariología de Dos Especies del Género *Tabebuia* Gomes (Bignoniaceae) Karyology of Two Species of Genus *Tabebuia* Gomes (Bignoniaceae). **Revista UDO Agrícola**, v. 2, n. 1, p. 14–21, 2002.

DI GIAMMARTINO, D. C.; NISHIDA, K.; MANLEY, J. L. Mechanisms and Consequences of Alternative Polyadenylation. **Molecular Cell**, v. 43, n. 6, p. 853–866, 2011. Disponible em: <<http://dx.doi.org/10.1016/j.molcel.2011.08.017>>.

DIJK E. L. V. et al. Ten years of next-generation sequencing technology. **Trends Genet.** 2014. Disponible em: <<https://doi.org/10.1016/j.tig.2014.07.001>>.

DOBIN A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, v. 29, n 1, p. 15–21. Janeiro, 2013. Disponible em: <<https://doi.org/10.1093/bioinformatics/bts635>>.

DUQUE, A. S. et al. Abiotic Stress Responses in Plants: Unraveling the Complexity of Genes and Networks to Survive. In: *Abiotic Stress - Plant Responses and Applications in Agriculture*. **IntechOpen**, 2013. p. 49–101.

ELKON, R.; UGALDE, A; AGAMI, R. Alternative cleavage and polyadenylation: extent, regulation and function. **Nature Review Genetics** 14, 496–506. 2013. Disponible em: <<https://doi.org/10.1038/nrg3482>>.

ESTRAVIS-BARCALA, M. et al. Deciphering the transcriptomic regulation of heat stress responses in *Nothofagus pumilio*. **PLoS ONE**, v. 16, n. 3 March, p. 1–22, 2021.

FISCHER, E.; THEISEN, I.; LOHMANN, L. G. Bignoniaceae. In: *Flowering Plants. Dicotyledons*. **Springer**, 2004. v. 7. p. 9–38.

FLORIS, M. et al. Post-transcriptional Regulation of Gene Expression in Plants during Abiotic Stress. **OPEN ACCESS Int. J. Mol. Sci**, v. 10, p. 10, 2009. Disponible em: <www.mdpi.com/journal/ijms>.

FRANCO OSPINA, L. A. et al. Actividad antiinflamatoria, antioxidante y antibacteriana de dos especies del género *Tabebuia*. **Revista Cubana de Plantas Medicinales**, v. 18, n. 1, p. 34–46, 2013.

GENTRY, H. A. Organization for flora neotropica. In: **Brittonia**. GENTRY, 1980: New York Botanical Garden Press on behalf of Organization for Flora Neotropica, 1980. 25p. 225–228.

GROSE, S. O.; OLMSTEAD, R. G. Evolution of a charismatic neotropical clade: Molecular phylogeny of *Tabebuia* s. l., Crescentieae, and allied genera (Bignoniaceae). **Systematic Botany**, v. 32, n. 3, p. 650–659, 2007a.

GROSE, S. O.; OLMSTEAD, R. G. Taxonomic revisions in the polyphyletic genus *Tabebuia* s. l. (Bignoniaceae). **Systematic Botany**, v. 32, n. 3, p. 660–670, 2007b.

HA K.C.; BLENCOWE B. J.; MORRIS Q. QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. **Genome Biol.** 2018. Disponível em: <<https://doi.org/10.1186/s13059-018-1414-4>>.

HERSHEY, J. W. B.; SONENBERG, N.; MATHEWS, M. B. Principles of Translational Control : An Overview. 2012.

HRDLICKOVA, R.; TOLOUE, M.; TIAN, B. RNA-Seq methods for transcriptome analysis. **Wiley Interdisciplinary Reviews: RNA**, v. 8, n. 1, 2017.

HU, H. 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. 1–10. February, 2020.

HUNT A. G. RNA regulatory elements and polyadenylation in plants. **Front Plant Sci.** 2012. Disponível em: < <https://doi.org/10.3389/fpls.2011.00109>>.

JI Z., et al. Transcriptional activity regulates alternative cleavage and polyadenylation. **Molecular Systems Biology.** 2011. Disponível em: <<https://doi.org/10.1038/msb.2011.69>>

KANEHISA, M.; GOTO, S. **KEGG: Kyoto Encyclopedia of Genes and Genomes Nucleic Acids Research** Oxford University Press, , 1 jan. 2000. . Disponível em: <<http://www.genome.ad.jp/kegg/>>. Acesso em: 19 jun. 2021.

KANG, H. I. et al. Comparative transcriptome analysis during developmental stages of direct somatic embryogenesis in *Tilia amurensis* Rupr. **Scientific Reports**, v. 11, n. 1, p. 6359, 1 dez. 2021. Disponível em: <<https://doi.org/10.1038/s41598-021-85886-z>>. Acesso em: 21 jun. 2021.

LASSMANN, T.; HAYASHIZAKI, Y.; DAUB, O. C. SAMStat: monitoring biases in next generation sequencing data. **Bioinformatics.** v. 27, n. 1, p. 130–131. Janeiro, 2011. Disponível em: <<https://doi.org/10.1093/bioinformatics/btq614>>.

LEWIS, J. D.; GUNDERSON, S. I.; MATTAJ, I. W. The influence of 5' and 3' end structures on pre-mRNA metabolism. **Journal of Cell Science**, v. 108, n. SUPPL. 19, p. 13–19, 1995.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics.** v. 25, n. 16, p. 2078–2079. Agosto, 2009. Disponível em:

<<https://doi.org/10.1093/bioinformatics/btp352>>.

LOHMANN, L. G. Bignoniaceae. In: Flowering Plants of the Neotropics. **Princeton University Press**, 2004. p. 51–53.

LOHMANN, L. G. et al. 2020. Bignoniaceae in Flora do Brasil 2020. **Jardim Botânico do Rio de Janeiro**. Disponível em: <<http://reflora.jbrj.gov.br/reflora/floradobrasil/FB112305>>. Acesso em: 14 jun. 2021.

LORENZO L. et al. Noncanonical Alternative Polyadenylation Contributes to Gene Regulation in Response to Hypoxia. **Plant Cell**. 2017. Disponível em: <<https://doi.org/10.1105/tpc.16.00746>>.

LUTZ, C. S.; MOREIRA, A. Alternative mRNA polyadenylation in eukaryotes: An effective regulator of gene expression. **Wiley Interdisciplinary Reviews: RNA**, v. 2, n. 1, p. 22–31, 2011.

MARGUERAT, S.; BÄHLER, J. RNA-seq: From technology to biology. **Cellular and Molecular Life Sciences**, v. 67, n. 4, p. 569–579, 2010.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature Reviews Genetics**, v. 12, n. 10, p. 671–682, 2011. Disponível em: <<http://dx.doi.org/10.1038/nrg3068>>.

MARTINS, E. A. C.; Filho, P. R. M. Mecanismos de expressão gênica em Eucariotos. **Revista Da Biologia**. 2018. Disponível em: <<https://doi.org/10.7594/revbio.04.01>>.

MAYR, C. Evolution and Biological Roles of Alternative 3'UTRs. **Trends in Cell Biology**, v. 26, n. 3, p. 227–237, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.tcb.2015.10.012>>.

MAYR, C.; BARTEL, D. P. Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. **Cell**, v. 138, n. 4, p. 673–684, 2009. Disponível em: <<http://dx.doi.org/10.1016/j.cell.2009.06.016>>.

MOORE, M. J. From birth to death: The complex lives of eukaryotic mRNAs. **Science**, v. 309, n. 5740, p. 1514–1518, 2005.

MOROZOVA, O.; MARRA, M. A. Applications of next-generation sequencing technologies in functional genomics. **Genomics**, v. 92, n. 5, p. 255–264, 2008. Disponível em: <<http://dx.doi.org/10.1016/j.ygeno.2008.07.001>>.

OLMSTEAD, R. G. et al. A molecular phylogeny and classification of Bignoniaceae. **American Journal of Botany**, v. 96, n. 9, p. 1731–1743, 2009.

OSAKABE Y. et al. Response of plants to water stress. **Front Plant Sci**. 2014. Disponível em: <<https://doi.org/10.3389/fpls.2014.00086>>.

VAN de PEER Y.; MIZRACHI E.; MARCHAL K. The evolutionary significance of polyploidy. **Nat Rev Genet**. 2017. Disponível em: <<https://doi.org/10.1038/nrg.2017.26>>.

PROUDFOOT N. J. et al. Integrating mRNA processing with transcription. **Cell**. 2002. Disponível em: < [https://doi.org/10.1016/S0092-8674\(02\)00617-7](https://doi.org/10.1016/S0092-8674(02)00617-7)>.

QUINLAN, R. A.; HALL M. I. BEDTools: a flexible suite of utilities for comparing genomic features. **Bioinformatics**. v. 26, n. 6, p. 841–842. Março, 2010. Disponível em: <<https://doi.org/10.1093/bioinformatics/btq033>>.

RHOADS A.; AU K. F. PacBio sequencing and its applications. **Genomics Proteomics Bioinformatics**. 2015. Disponível em: <<https://doi.org/10.1016/j.gpb.2015.08.002>>.

ROBINSON, J. T. et al. Integrative genomics viewer. **Nature biotechnology**. v. 29, p. 24-6. 2011. Disponível em: <[doi:10.1038/nbt.1754](https://doi.org/10.1038/nbt.1754)>.

RYAN, R. Y. M. et al. The medicinal plant *Tabebuia impetiginosa* potently reduces pro-inflammatory cytokine responses in primary human lymphocytes. **Scientific Reports**, v. 11, n. 1, p. 1–12, 2021. Disponível em: <<https://doi.org/10.1038/s41598-021-85211-8>>.

SANDBERG R. et al. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. **Science**. 2008. Disponível em: <<https://www.science.org/doi/10.1126/science.1155390>>.

SANTOS, S. A. et al. Comparative genomic and transcriptomic analyses reveal different pathogenicity-related genes among three eucalyptus fungal pathogens. **Fungal Genetics and Biology**, v. 137, 2020. Disponível em: <<https://doi.org/10.1016/j.fgb.2019.103332>>.

SHAH A. et al. Benchmarking sequencing methods and tools that facilitate the study of alternative polyadenylation. **Genome Biol**. 2021. Disponível em: < [doi:10.1186/s13059-021-02502-z](https://doi.org/10.1186/s13059-021-02502-z) >.

SHEN, Y. et al. Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing. **Genome Research**, v. 21, n. 9, p. 1478–1486, 2011.

SHI, Y. Alternative polyadenylation: New insights from global analyses. **Rna**, v. 18, n. 12, p. 2105–2117, 2012.

SILVA-JUNIOR, O. B. et al. Genome assembly of the Pink Ipê (*Handroanthus impetiginosus*, Bignoniaceae), a highly valued, ecologically keystone Neotropical timber forest tree. **Giga Science**, v. 7, n. 1, p. 1–16, 2018.

SOBREIRO, M. B. et al. RNA-Seq reveals different responses to drought in Neotropical trees from savannas and seasonally dry forests. **BMC Plant Biol** 21, 463 (2021). Disponível em: <<https://doi.org/10.1186/s12870-021-03244-7>>.

STAATS, C. C. Controle da expressão gênica em eucariotos. *In*: ZAHA, A. **Biologia Molecular Básica**. 5. ed. **Artmed**, 2014, cap. 11, p. 233-254.

STARK, R.; GRZELAK, M.; HADFIELD, J. RNA sequencing: the teenage years. **Nature Reviews Genetics**, v. 20, n. 11, p. 631–656, 2019. Disponível em: <[http://dx.doi.org/10.1038/s41576-019-0150-2](https://doi.org/10.1038/s41576-019-0150-2)>.

TAIZ, L.; ZEIGER, E. *Plant Physiology*. 3. ed. **Sinauer Associates**, 2002. cap. 14, p. 309-339.

TIAN, B. et al. A large-scale analysis of mRNA polyadenylation of human and mouse genes. **Nucleic Acids Research**, v. 33, n. 1, p. 201–212, 2005.

TIAN, B.; MANLEY, J. L. Alternative polyadenylation of mRNA precursors. **Nature Publishing Group**, v. 18, n. 1, p. 18–30, 2016. Disponível em: <<http://dx.doi.org/10.1038/nrm.2016.116>>.

TIAN, T. et al. AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. **Nucleic Acids Research**, v. 45, n. 1, p. 122–129, 2017.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57–63, 2009. Disponível em: <<http://doi.org/nrg2484>>.

WATSON, J. D. et al. Processamento do RNA. In: *Biologia Molecular do Gene*. 7. ed. **Artmed**, 2015. p. 467–507

WICKENS, M.; ANDERSON, P.; JACKSON, R. J. Life and death in the cytoplasm: Messages from the 3' end. **Current Opinion in Genetics and Development**, v. 7, n. 2, p. 220–232, 1997.

WU X. et al. Genome-wide determination of poly(A) sites in *Medicago truncatula*: evolutionary conservation of alternative poly(A) site choice. **BMC Genomics**. 2014. Disponível em: <<https://doi.org/10.1186/1471-2164-15-615>>.

WU X. et al. Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. **PNAS**. 2011. Disponível em: <<https://doi.org/10.1073/pnas.1019732108>>.

XIA, Z. et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. **Nat Commun**. 2014 Disponível em: <<https://doi.org/10.1038/ncomms6274>>.

XING, D.; LI, Q.Q. Alternative polyadenylation and gene expression regulation in plants. **WIREs RNA**. 2011. Disponível em: <<https://doi.org/10.1002/wrna.59>>.

YAN, C. et al. Alternative Polyadenylation in response to temperature stress contributes to gene regulation in *Populus trichocarpa*. **BMC Genomics**, 2021. Disponível em: <<https://doi.org/10.1186/s12864-020-07353-9>>.

YE, C. et al. APATrap: Identification and quantification of alternative polyadenylation sites from RNA-seq data. **Bioinformatics**, v. 34, n. 11, p. 1841–1849, 2018.

YE, C. et al. Genome-wide alternative polyadenylation dynamics in response to biotic and abiotic stresses in rice. **Ecotoxicology and Environmental Safety**, v. 183, 2019. Disponível em: <<https://doi.org/10.1016/j.ecoenv.2019.109485>>.

ZAHA, A. Controle da expressão gênica em eucariotos. *In*: ZAHA, A. *Biologia Molecular Básica*. 5. ed. **Artmed**, 2014, cap. 14, p. 301-318.

ZAMORA-BALLESTEROS, C. et al. Dual rna-sequencing analysis of resistant (*Pinus pinea*) and susceptible (*pinus radiata*) hosts during *fusarium circinatum* challenge. **International Journal of Molecular Sciences**, v. 22, 2021. Disponível em: </pmc/articles/PMC8156185/>. Acesso em: 21 jun. 2021.

ZHAO S. et al. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. **PLoS One**. 2014. Disponível em: <<https://doi.org/10.1371/journal.pone.0078644>>.

APÊNDICE B – Categorias GO enriquecidas em *H. ochraceus* (Componente cel.)

