



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA**

**MOACIRIA DE SOUZA LEMOS**

**ANÁLISE FILOGENÉTICA DA FAMÍLIA MULTIGÊNICA DA CATALASE EM  
ANGIOSPERMAS E EXPRESSÃO *IN SILICO* NA ESPÉCIE *ZEA MAYS***

**FORTALEZA**

**2019**

MOACIRIA DE SOUZA LEMOS

ANÁLISE FILOGENÉTICA DA FAMÍLIA MULTIGÊNICA DA CATALASE EM  
ANGIOSPERMAS E EXPRESSÃO *IN SILICO* NA ESPÉCIE *ZEA MAYS*

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestra em Bioquímica. Área de concentração: Ciências Biológicas II.

Orientador: Prof. Dr. Jose Helio Costa.

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

L578a Lemos, Moaciria de Souza.

Análise filogenética da família multigênica da catalase em angiospermas e expressão in silico na espécie Zea mays / Moaciria de Souza Lemos. – 2019.  
89 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Bioquímica, Fortaleza, 2019.  
Orientação: Prof. Dr. Jose Helio Costa.

1. Filogenia. 2. Expressão Gênica. 3. Zea mays. 4. Família multigênica. I. Título.

CDD 572

---

MOACIRIA DE SOUZA LEMOS

ANÁLISE FILOGENÉTICA DA FAMÍLIA MULTIGÊNICA DA CATALASE EM  
ANGIOSPERMAS E EXPRESSÃO *IN SILICO* NA ESPÉCIE *ZEA MAYS*

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestra em Bioquímica. Área de concentração: Ciências Biológicas II.

Aprovada em: \_\_\_/\_\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Jose Helio Costa (Orientador)  
Universidade Federal do Ceará (UFC)

---

Dr. Fabrício Eulálio Leite Carvalho  
Universidade Federal do Ceará (UFC)

---

Profª Drª Kátia Daniella da Cruz Saraiva  
Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB)

À minha família.

## AGRADECIMENTOS

A minha família pelo apoio incondicional e inesgotável fonte de inspiração.

Ao professor Jose Helio Costa, pela orientação.

Aos membros da banca examinadora pelas valiosas contribuições.

Aos meus amigos, Luciana, Paloma e João pela amizade sincera, apoio e momentos de diversão que foram de imensa ajuda durante esses dois anos de mestrado.

A Jéssica Moraes, pela amizade e apoio logo quando cheguei completamente perdida em Fortaleza.

Ao meu namorado Roberto Nascimento, pelo amizade, paciência e companheirismo, sempre me fazendo acreditar no meu potencial. Agradeço pelos momentos felizes que passamos. Os dias contigo foram mais agradáveis e as horas tristes e desafiadoras, mais suportáveis. Ao seu lado eu me tornei uma pessoa melhor e mais feliz. Te amo.

As minhas gatinhas, BiBi e Branquinha que me trazem tanta alegria.

As minhas colegas de laboratório, Thais, Lyndefânia e Dayane pelas contribuições e convívio agradável durante o mestrado.

Agradeço em especial a Clesivan Pereira, pela grande ajuda, paciência e pelos valiosos ensinamentos durante todo o tempo de mestrado. Grande parte dessa dissertação só foi possível graças a sua ajuda. Muito obrigada!

A Kátia Daniella, pela ajuda durante a classificação das sequências e também pela disponibilidade durante as correções iniciais da dissertação.

Ao professor Murilo, pela ajuda durante a realização da análise Bayesiana.

Ao Matheus, pela grande ajuda durante as análises de expressão *in silico*.

Ao professor Nicolas e ao Fabiano, por terem aceitado a participação na banca como membros suplentes.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa.

Enfim, agradeço a todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

“Aprender é a única coisa que a mente nunca se cansa, nunca tem medo e nunca se arrepende.”

Leonardo Da Vinci

## RESUMO

As plantas desenvolveram, ao longo da evolução, estratégias para mitigar os efeitos negativos dos estresses ambientais aos quais estão expostas. Dentre essas estratégias, o desenvolvimento de um sistema antioxidante desempenha uma função primordial em lidar com moléculas ou radicais livres nocivos à célula. A catalase é uma enzima que faz parte dessa linha de defesa, metabolizando o  $H_2O_2$ , que em concentrações elevadas causa efeitos danosos as células. Em plantas, geralmente a catalase é codificada por uma família multigênica de até três genes. Entretanto, a distribuição, bem como a função fisiológica de cada membro gênico ainda são temas pouco compreendidos. Dessa forma, esse trabalho objetivou analisar a estrutura e distribuição filogenética dos genes da catalase em angiospermas bem como a sua funcionalidade (expressão) em diferentes tecidos durante o desenvolvimento e condições de estresse em milho. Os genes da catalase foram anotados e as sequências de cDNA obtidas foram utilizadas para construção da árvore filogenética através da metodologia Bayesiana e também para determinação da estrutura gênica. A expressão gênica foi avaliada em dados transcriptômicos (*in silico*) de milho através da ferramenta HTSeq. A partir da análise filogenética quatro subfamílias da catalase foram identificadas em angiospermas (CAT2, CAT3, CAT1 e CAT4). As duas primeiras encontradas somente em espécies de monocotiledôneas, a terceira compartilhada por ambas as classes e a última encontrada somente em eudicotiledôneas. Os dados de estrutura gênica sugerem que a evolução dos genes ocorreu através de, predominantemente, eventos de deleção de íntrons e por meio de duplicação. A análise de expressão revelou que os genes *cat2* e *cat3* são expressos na raiz em condição de estresse osmótico e sob ataque de fungo, respectivamente. Ambos os genes apresentaram expressão no grão germinado e em tecido foliar maduro. Já o gene *cat1* foi expresso somente em tecidos reprodutivos. Os genes da catalase em milho possuem expressão dependente do tecido e do estágio de desenvolvimento da planta. Os dados obtidos nesse trabalho contribuem para elucidar a distribuição evolutiva dos genes da catalase em plantas, gerando informações para se estabelecer uma classificação mais ampla buscando inter-relação entre filogenia, estrutura e função gênica.

**Palavras-chave:** Filogenia. Expressão gênica. *Zea mays*. Família multigênica.



## ABSTRACT

Throughout evolution, plants have developed strategies to mitigate the negative effects of the environmental stresses to which they are exposed. Among these strategies, the development of an antioxidant system plays a primordial role in dealing with molecules or free radicals harmful to the cell. Catalase is an enzyme that is part of this line of defense, metabolizing  $H_2O_2$ , which at high concentrations causes harmful effects on cells. In plants, catalase is generally encoded by a multigenic family of up to three genes. However, the distribution as well as the physiological function of each gene member are still poorly understood subjects. This work aimed to analyze the structure and phylogenetic distribution of the catalase genes in angiosperms as well as their functionality (expression) in different tissues during development and stress conditions in corn. The catalase genes were annotated and the cDNA sequences obtained were used to construct the phylogenetic tree through Bayesian methodology and also to determine the gene structure. Gene expression was evaluated in corn transcriptomic data (in silico) using the HTSeq tool. From the phylogenetic analysis four catalase subfamilies were identified in angiosperms (CAT2, CAT3, CAT1 and CAT4). The first two found only in species of monocots, the third shared by both clades and the last one found only in eudicots. Gene structure data suggest that gene evolution occurred through predominantly introns deletion events and through duplication. Expression analysis revealed that the *cat2* and *cat3* genes are expressed in the root under dry condition and under fungus attack, respectively. Both genes had expression in the germinated grain and mature foliar tissue. The *cat1* gene was expressed only in reproductive tissues. All three genes studied have expression profile dependent on tissue and stage of development of the plant. The data obtained in this work contribute to elucidate the evolutionary distribution of the catalase genes in plants, generating information to establish a broader classification seeking interrelationship between phylogeny, structure and gene function.

**Keywords:** Phylogeny. Gene expression. *Zea mays*. Multigenic family.

## LISTA DE ILUSTRAÇÕES

Figura 1	- Distribuição dos genes das subfamílias da catalase em angiospermas.....	37
Figura 2	- Classificação proposta para espécies com classificação previamente definida no NCBI e/ou na literatura.....	38
Figura 3	- Árvore filogenética com as sequências de cDNAs da catalase, obtidos a partir da anotação em 110 genomas de plantas angiospermas.....	41
Figura 4	- Estrutura éxon-íntron dos genes pertencentes a subfamília CAT2.....	42
Figura 5	- Estrutura éxon-íntron dos genes pertencentes a subfamília CAT3.....	43
Figura 6	- Estrutura éxon-íntron dos genes pertencentes a subfamília CAT1.....	45
Figura 7	- Estrutura éxon-íntron dos genes pertencentes a subfamília CAT4.....	46
Figura 8	- Número de elementos <i>cis</i> regulatórios presentes na região promotora dos genes <i>cat1</i> , <i>cat2</i> e <i>cat3</i> de <i>Zea mays</i> .....	50
Figura 9	- Perfil de expressão dos genes <i>cat1</i> , <i>cat2</i> e <i>cat3</i> em diferentes partes da raiz primária obtidas de plantas de milho em condição de estresse osmótico.....	51
Figura 10	- Perfil de expressão dos genes <i>cat1</i> , <i>cat2</i> e <i>cat3</i> de <i>Zea mays</i> sob ataque pelo fungo <i>Cercospora zeae-maydis</i> .....	52
Figura 11	- Perfil de expressão dos genes <i>cat1</i> , <i>cat2</i> e <i>cat3</i> de <i>Zea mays</i> em diferentes tecidos reprodutivos.....	53
Figura 12	- Perfil de expressão dos genes <i>cat1</i> , <i>cat2</i> e <i>cat3</i> de <i>Zea mays</i> em diferentes tecidos.....	54

Figura 13 - Árvore filogenética baseada em sequências de aminoácidos gerada a partir do método da parcimônia..... 61

## LISTA DE TABELAS

Tabela 1 - Experimentos de RNA-Seq utilizados para se obter o perfil de expressão dos genes <i>cat1</i> , <i>cat2</i> e <i>cat3</i> em <i>Zea mays</i> .....	33
---	----

## LISTA DE ABREVIATURAS E SIGLAS

ABA	Ácido Abscísico
ACP	Análise de Componentes Principais
AIC	<i>Minimum Theoretical Information Criterion</i>
APX	Ascorbato peroxidase
AsA	Ascorbato
AsA-GSH	Enzimas do ciclo ascorbato-glutationa
BLAST	<i>Basic Local Alignment Search Tool</i>
CAT	Catalase
cDNA	DNA complementar
CPM	Contagem por milhão
DAP	Dias após a polinização
DHAR	Deidroascorbato redutase
EROs	Espécies Reativas de Oxigênio
EST	<i>Expressed Sequence Tags</i>
GDEs	Genes Diferencialmente Expressos
GLS	<i>Grey Leaf Spot</i>
PRX	Peroxirredoxinas
GSH	Glutationa reduzida
GTF	<i>General Transfer Format</i>
IPCC	<i>Intergovernmental Panel on Climate Change</i>
KcatG	Catalase bifuncional
MAPKs	<i>Mitogen Activated Protein Kinases</i>
MCMC	<i>Markov Chain Monte Carlo</i>
MDHAR	Monodeidroascorbato redutase
MeJA	Metil jasmonato
MnCATs	Catalases de Manganês
NCBI	<i>National Center for Biotechnology Information</i>
NR	<i>Nucleotide Collection</i>
PAMPS	<i>Pathogen-associated Molecular Pattern</i>
PEG	Polietilenoglicol
PTS	<i>Peroxisomal Targeting Signal</i>

Refseq_genomic	<i>Reference genomic sequences</i>
Refseq_rna	<i>Reference mRNA sequences</i>
SOD	<i>Superóxido Dismutase</i>
SOS	<i>Salt Overly Sensitive</i>
SRA	<i>Sequence Read Archive</i>
TSA	<i>Transcriptome Shotgun Assembly</i>
UPGMA	<i>Unweighted Pair Group Method using Arithmetic Averages</i>
USDA	<i>United States Department of Agriculture</i>
UTR	<i>Untranslated Region</i>
WGD	<i>Whole-genome Duplication</i>
WGS	<i>Whole-genomes Shotgun Reads</i>

## SUMÁRIO

1	INTRODUÇÃO.....	13
2	FUNDAMENTAÇÃO TEÓRICA.....	16
2.1	Impactos dos estresses ambientais na agricultura .....	16
2.2	Respostas de plantas a estresses ambientais.....	18
2.3	Catalase em Plantas: origem, função e classificação.....	20
2.4	Duplicação gênica e a origem de famílias multigênicas.....	23
2.5	Estudos com a catalase na espécie <i>Zea mays</i> .....	25
3	OBJETIVOS.....	28
3.1	Objetivo Geral.....	28
3.2	Objetivos Específicos.....	28
4	MATERIAL E MÉTODOS.....	29
4.1	Identificação e anotação manual das sequências da catalase.....	29
4.2	Análise filogenética.....	29
4.3	Estrutura éxon-íntron dos genes da catalase.....	30
4.4	Análise das regiões promotoras dos genes da catalase do milho.....	31
4.5	Análise da expressão <i>in silico</i> utilizando dados de RNA-Seq para <i>Zea mays</i> .....	31
4.6	Análise estatística.....	36
5	RESULTADOS.....	36
5.1	Identificação, análise filogenética e classificação da catalase.....	36
5.2	Caracterização estrutural dos membros gênicos/proteínas da catalase de Angiospermas.....	42
5.3	Expressão <i>in silico</i> dos genes da catalase em <i>Zea mays</i> .....	50
5.3.1	<i>Expressão in silico dos genes da catalase em plantas de milho submetidas a estresse osmótico</i> .....	50
5.3.3	<i>Expressão in silico dos genes da catalase em plantas de milho em diferentes tecidos reprodutivos</i> .....	52

5.3.4	<i>Expressão in silico dos genes da catalase em plantas de milho em diferentes tecidos</i> .....	53
6	<b>DISCUSSÃO</b> .....	55
6.1	Distribuição das subfamílias da catalase em angiospermas.....	55
6.2	Catalase: sugestão de uma nova classificação em angiospermas.....	59
6.3	Estrutura éxon-íntron dos genes da catalase: um padrão evolutivo.....	62
6.4	História evolutiva da Família multigênica da catalase em Angiospermas.....	65
6.5	Papel fisiológico dos membros gênicos da catalase.....	67
7	<b>CONCLUSÃO</b> .....	72
	<b>REFERÊNCIAS</b> .....	73
	<b>ANEXO A - DADOS DE TAXONOMIA, NÚMERO DE GENES DA CATALASE POR ESPÉCIE E DISTRIBUIÇÃO DE CLADOS COM BASE NA FILOGENIA</b> .....	86



## 1 INTRODUÇÃO

Os estresses ambientais têm contribuído grandemente para perdas na agricultura nos últimos anos, o que afeta negativamente a produção de alimentos. Dentre esses estresses podemos citar os estresses abióticos, como seca, temperatura (alta e baixa) e salinidade, e os estresses bióticos, como o ataque de pragas e patógenos. Todavia, as plantas se adaptaram ao longo da evolução e como resultado disso, usam estratégias de defesa para mitigar os efeitos danosos dessas situações adversas (ZANDALINAS et al., 2017).

Sob uma situação de estresse, ocorre um desbalanço no metabolismo da planta, havendo a formação e o acúmulo de espécies reativas de oxigênio (EROs). Neste grupo, podemos destacar o peróxido de hidrogênio ( $H_2O_2$ ), que é considerado uma importante molécula sinalizadora, porém, quando em altas concentrações pode causar extensivo dano celular e inibição da fotossíntese (ZANDALINAS et al., 2017; TURKAN, 2017).

Dessa forma, em decorrência da toxicidade das EROs, as plantas desenvolveram mecanismos de defesa para lidar com essas moléculas e manter a homeostase celular, os quais incluem mecanismos não-enzimáticos e sistemas enzimáticos. Os primeiros contam com a atuação de moléculas antioxidantes, tais como glutathiona e carotenoides. Já os sistemas enzimáticos incluem enzimas como superóxido dismutase (SOD), peroxirredoxinas (PRX), ascorbato peroxidase (APX), catalase (CAT), dentre outras (SHARMA et al., 2012; SEWELAN; KAZAN; SCHENK et al., 2017).

A catalase (EC 1.11.1.6) é uma enzima encontrada predominantemente nos peroxissomos, catalisando uma reação de detoxificação do peróxido de hidrogênio, onde 2 moléculas de  $H_2O_2$  sofrem dismutação em água e  $O_2$  (MHAMDI et al., 2010).

Em plantas, a catalase já tem sido alvo de vários estudos, com trabalhos voltados para uma compreensão da sua função fisiológica. Contudo, apesar do grande enfoque dado a essa enzima, devido a ser altamente expressa e parte do sistema de defesa antioxidante, pouco ainda se conhece acerca da função fisiológica das diferentes isoformas, sua distribuição subcelular, os mecanismos que regulam essa distribuição e a verdadeira significância, do ponto de vista funcional, dessas enzimas em plantas (MHAMDI; NOCTOR; BAKER, 2012).

De uma forma geral, até o momento três genes codificando para a catalase são encontrados nas plantas. Geralmente, esses genes são nomeados como *cat1*, *cat2* e *cat3*. Essa nomenclatura é encontrada em milho e *Arabidopsis thaliana*. Contudo, em outras espécies, tais como arroz e *Nicotiana tabacum* outra nomenclatura é usada, não havendo correspondência entre as espécies, dificultando a interpretação dos dados em trabalhos científicos e demonstrando a necessidade de um sistema de classificação para os genes da catalase baseado nas relações filogenéticas, de forma a se homogeneizar a nomenclatura e a partir disso, criar hipóteses testáveis acerca da evolução, distribuição e função dessa enzima em plantas angiospermas.

Na espécie *Zea mays*, a família multigênica da catalase tem sido bastante estudada no decorrer dos últimos anos. Nessa espécie, 3 genes da catalase (*cat1*, *cat2* e *cat3*) codificam três isoformas distintas (CAT1, CAT2 e CAT3). Essas três isoformas exibem especificidade temporal e espacial em sua expressão e cada uma responde de forma variável a diferentes sinais ambientais.

O gene *cat1* em milho tem sido expresso durante os estágios iniciais do desenvolvimento da germinação da semente, endosperma leitoso (endosperma de sementes de milho no estágio R3 com grande acúmulo de amido) de sementes e no pólen após a antese. Esse gene também responde a condições de estresse, como salinidade, desidratação e elevadas temperaturas (SCANDALIOS; GUAN; POLIDOROS, 1997; GUAN; SCANDALIOS, 1998; SCANDALIOS; ACEVEDO; RUZCA, 2000; MENEZES-BENAVENTE et al., 2004).

O gene *cat2* pode ser encontrado na aleurona durante estágios mais tardios do desenvolvimento da semente e responde diferencialmente a ABA (ácido abscísico) (SCANDALIOS; GUAN; POLIDOROS, 1997). Esse gene também é expresso em folhas após exposição a luz (POLIDOROS; SCANDALIOS, 1997).

O gene *cat3* tem sido expresso no pericarpo durante o desenvolvimento do óvulo. A CAT3 é a isoenzima predominante em zimogramas do coleóptilo e em raízes (SCANDALIOS; GUAN; POLIDOROS, 1997). O gene *cat3* ainda é regulado por ritmo circadiano, e os transcritos são induzidos em embriões de milho sob tratamento com a toxina cercosporina purificada (WILLIAMSON; SCANDALIOS, 1992; POLIDOROS; SCANDALIOS, 1997).

As catalases de milho já vêm sendo estudadas com enfoque filogenético, junto com outras espécies, obtendo-se agrupamentos que nos permitem tirar informações de forma a se explorar e gerar uma classificação com base nessa espécie (GUAN;

SCANDALIOS, 1996; FRUGOLI et al., 1996; KLOTZ; KLASSEN; LOEWEN, 1997; BAGNOLI et al., 2004). Aliado a essa proposta de classificação, a inferência de uma função, relacionando as informações filogenéticas e de expressão, nos permitiria avançar numa compreensão mais detalhada acerca dessa família multigênica.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Impactos dos estresses ambientais na agricultura

As plantas são afetadas constantemente por fatores externos que causam efeitos negativos no crescimento e desenvolvimento vegetal. Esses fatores externos incluem estresses bióticos, como infecção por patógenos e ataques de herbívoros e estresses abióticos, como frio, deficiência nutricional, metais tóxicos no solo, intensidade luminosa, inundação, seca, salinidade e temperatura. Estes três últimos são os principais causadores de perdas na agricultura, sendo agravados principalmente pelas mudanças climáticas (ZHU, 2016; MEENA et al., 2017).

Segundo estimativa do *International Panel for Climate Change* (IPCC) haverá um aumento na temperatura de 5°C no final deste século, o que pode ocasionar um aumento na frequência e severidade de eventos de climas extremos, afetando o rendimento das culturas, os ganhos do produtor, os preços dos alimentos, o fornecimento e a segurança alimentar (HALFORD et al., 2014).

Além do mais, a agricultura representa uma fonte de extrema importância para alimentação mundial. Ademais, com o crescimento populacional, um maior rendimento das culturas é necessário. Esse apelo por mais alimentos, por si só, já é difícil de se conseguir em condições agrícolas ótimas, porém esse quadro se modifica drasticamente quando se tem diferentes fatores a serem levados em consideração, tais como aumento da área disponível para cultivos, uso eficiente dos recursos hídricos, mudanças climáticas, estresses abióticos e estresses bióticos (LANDI et al., 2017).

Os estresses bióticos (ataques por pragas agrícolas e patógenos – fungos, nematoides, bactérias e vírus) afetam negativamente a segurança alimentar e são responsáveis por aproximadamente 30% das perdas das lavouras, tanto pré, como pós-colheita (ISLAM et al., 2018). Além disso, os métodos convencionais para controle de patógenos na agricultura fazem o uso de agrotóxicos. Se usado de forma indiscriminada, que é a forma mais frequente de uso, pode trazer prejuízos à saúde humana, assim também como impactos negativos ao ambiente, tais como a contaminação de reservatórios de águas, danos à fauna e a flora e o desenvolvimento de resistência por parte das pragas e patógenos (TOHIDFAR; KHOSRAVI, 2015).

Ademais, nas culturas em campo, os efeitos de ataques dos patógenos podem ser agravados diretamente por estresses abióticos, tornando as plantas mais suscetíveis a doenças. Aliado a tudo isso, as mudanças climáticas correntes influenciam o habitat de muitas pragas agrícolas – principalmente com o aumento da temperatura - promovendo a propagação de patógenos, acarretando prejuízos ao produtor (ATKINSON; URWIN, 2012; SUZUKI; KATANO, 2018).

Estresses abióticos como a seca já afetam 64% da área global terrestre, os efeitos da inundação (anoxia) já são percebidos em 13% da área terrestre, a salinidade em 6%, a deficiência de minerais em 9% e o frio em 57%. Em torno de 3,6 bilhões de hectares estão afetados por problemas de erosão, degradação e salinidade de solos, com um impacto em aproximadamente 50% da área total destinada a plantios. Toda essa problemática custa em torno de 12 bilhões de dólares para a economia mundial (MEENA et al., 2017).

Mariani e Ferrante (2017) relatam perdas de cerca de 40% e 42% aproximadamente (média global) na produção de milho causada por seca e altas temperaturas. Em culturas de trigo, esses valores chegam a 55% por água e 48% por estresse de temperatura. Esses autores destacam a forte limitação imposta pelos estresses abióticos sobre a produtividade de culturas comerciais, dentre elas o milho e o trigo, tratadas acima.

Dessa forma, futuramente torna-se imperativo a produção de plantas tolerantes e/ou resistentes a estresses ambientais no intuito de aumentar a produtividade e a sustentabilidade ambiental, pois plantas suscetíveis a estresses ambientais consomem mais água e fertilizantes, acarretando sobrecarga para o meio ambiente (ZHU, 2016).

Quando as plantas são expostas a estresses ambientais desencadeiam respostas, ativando mecanismos de defesa. Desse modo, a energia gerada pela fotossíntese na planta, ao invés de ser usada para a produção de biomassa, será alocada para vias metabólicas destinadas a defesa. Essa energia conhecida como “fitness cost” não contribui para a produtividade das culturas. Portanto, as plantas desenvolveram um sistema de balanceamento dos recursos entre o que é destinado para a produtividade e o que é determinado para a defesa, isso, a depender das necessidades fisiológicas (MARIANI; FERRANTI, 2017).

## 2.2 Respostas de plantas a estresses ambientais

As plantas frequentemente respondem a estímulos do ambiente através da sua capacidade metabólica intrínseca. Normalmente, a planta está exposta a diversos estresses que podem prejudicar o seu crescimento e desenvolvimento, sendo assim, elas percebem o estresse e desencadeiam um arsenal de respostas, as quais podem envolver a ativação de diversas vias de defesa (PANDEY et al., 2017).

O termo estresse é definido como sendo uma condição fisiológica causada por alguma adversidade ambiental que limita o crescimento, o sucesso reprodutivo, o rendimento, a qualidade, dentre outras características desejáveis aos seres humanos (MHAMDI et al., 2010). O estresse vai ser qualquer fator ambiental que impeça que a planta manifeste todo o seu potencial genético, influenciando de forma adversa na produção da cultura (TAIZ et al., 2017).

Estresses em plantas podem ser desencadeados por fontes bióticas ou abióticas. Ambos compartilham a característica de grande formação e acúmulo, em níveis elevados, de EROs. No estresse biótico, as EROs são uma importante arma que pode ser produzida rapidamente em resposta a infecção, atuando através da ativação de linhas de defesa (reconhecimento dos patógenos via PAMPS, cascatas de MAPK, abertura de canais de cálcio, etc) e mediando respostas como moléculas de sinalização. Além do mais, EROs presentes no apoplasto geradas em plantas submetidas a infecção são suficientemente citotóxicas para causar a morte do patógeno (JWA; HWANG, 2017).

Assim como nos estresses bióticos, as EROs desempenham papéis cruciais em resposta a estresses abióticos. Nessas condições, ocorrem alterações na bioquímica e na fisiologia da planta, incluindo acúmulo de osmólitos, redução da fotossíntese, fechamento estomático e indução de genes responsivos a estresses. Em resposta a seca e salinidade, por exemplo, ocorre a ativação da via SOS (*Salt Overly Sensitive*), via das MAPKs, ácido abscísico, transportadores de íons, fatores de transcrição, etc. (LANDI et al., 2017).

As vias de sinalização em plantas responsáveis pela produção de EROs na célula, apesar de todo conhecimento já produzido, ainda permanecem desconhecidas em muitos aspectos. Sabe-se que há um grande número de eventos subjacentes aos mecanismos de sinalização com alto grau de interconectividade e que nem sempre

resultam de uma mudança da expressão de genes (SEWELAM; KAZAN; SCHENK, 2016).

De uma forma geral, as mudanças ambientais são sentidas pelas plantas através da formação de EROs, com conseqüente mudança na homeostase redox da célula. Tais EROs incluem moléculas consideravelmente reativas como os radicais hidroxila ( $\text{OH}^\bullet$ ), o oxigênio superóxido ( $\text{O}_2^{\bullet -}$ ) moléculas não-reativas como o oxigênio singleto ( $^1\text{O}_2$ ) e o peróxido de hidrogênio ( $\text{H}_2\text{O}_2$ ), derivados do  $\text{O}_2$  molecular, durante a respiração celular e a fotossíntese (VIGHI et al., 2016). Elas são geradas como conseqüência do metabolismo aeróbico normal da célula, bem como através de enzimas específicas que participam de determinados processos nas células, como NADPH-oxidases, peroxidases e superóxido-dismutase (INUPAKUTIKA et al., 2016).

Em plantas, os principais sítios de produção de EROs são o cloroplasto (gerando  $\text{O}_2^{\bullet -}$ ,  $\text{H}_2\text{O}_2$ , e  $^1\text{O}_2$ , produzidos principalmente como subprodutos da fotossíntese), a mitocôndria ( $\text{O}_2^{\bullet -}$  e  $\text{H}_2\text{O}_2$ , como subproduto da respiração) e peroxissomos (produz sobretudo  $\text{H}_2\text{O}_2$ , como subproduto da fotorrespiração). Até certo tempo, as EROs tinham sido sempre relacionadas a efeitos prejudiciais na célula, pois podem causar danos celulares, como oxidação da membrana, através da peroxidação de lipídeos, também podem ocasionar a oxidação de proteínas, de ácidos nucléicos, inibição enzimática, levando a danos irreversíveis na célula, com subsequente morte celular (INUPAKUTIKA et al., 2016; CHOUDHURY et al., 2017).

As papéis das EROs, ainda são amplamente discutidos, sendo que elas podem atuar como sinalizadores para funções ligadas ao crescimento e desenvolvimento celular, como também podem atuar como sinais que desencadeiem morte celular, processo central pelo qual ocorre a renovação celular e que também está ligado a respostas de defesa. Se as EROs irão atuar como um sinal de “vida” ou de “morte”, isso vai depender do contexto molecular e celular no qual elas acumulam (FOYER; RUBAN; NOCTOR, 2017).

Tem sido considerado que, se mantido em níveis basais, as EROs atuam como mensageiros secundários nas vias de transdução de sinais moleculares, incluindo a regulação de vias metabólicas, processos relacionados ao desenvolvimento, aclimatação a estresses abióticos e defesa contra patógenos. No entanto, para que se mantenha o nível de EROs em equilíbrio para a execução ótima de suas funções é necessário um ajuste fino das taxas de produção e degradação (INUPAKUTIKA et al., 2016).

No decorrer da evolução, as plantas desenvolveram mecanismos de defesa para lidar com os danos advindos das EROs. Esses incluem mecanismos não-enzimáticos e sistemas enzimáticos. Os primeiros contam com a atuação de moléculas antioxidantes, tais como glutathiona, ácido ascórbico, carotenoides e  $\alpha$ -tocoferol. Já os sistemas enzimáticos incluem enzimas como superóxido dismutase (SOD), peroxirredoxinas (PRX), enzimas do ciclo ascorbato-glutathiona (AsA-GSH), ascorbato peroxidase (APX), monodeidroascorbato redutase (MDHAR), deidroascorbato redutase (DHAR), glutathiona redutase e catalase (CAT) (SHARMA et al., 2012; INUPAKUTIKA et al., 2016; SEWELAN; KAZAN; SCHENK, 2017).

### **2.3 Catalase em Plantas: origem, função e classificação**

A cerca de 3.8 bilhões de anos atrás, baixos níveis de oxigênio passaram a ser emitidos para a atmosfera, através do surgimento de organismos fotossintetizantes. A partir desse momento, os demais organismos que foram surgindo necessitavam do desenvolvimento de mecanismos que pudessem sobreviver em um ambiente na presença de oxigênio e também que pudessem lidar com as EROs geradas por causa das reações químicas sofridas pela molécula de oxigênio (INUPAKUTIKA et al., 2016).

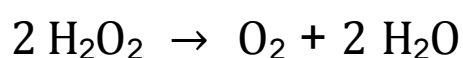
A partir de então, mecanismos foram surgindo para proteger os organismos contra os efeitos danosos de tais moléculas reativas, propiciando um ambiente favorável para o estabelecimento da vida na terra. Tais mecanismos incluem sistemas enzimáticos como a enzima catalase, que segundo estudo realizado por Inupakutika et al. (2016) foi encontrada em organismos que surgiram a cerca de 4.1-3.5 bilhões de anos atrás. Isto vai de encontro com a estimativa de 3.8 bilhões de anos, onde acredita-se que houve o aparecimento das primeiras formas de vida fotossintetizantes. Dessa forma, a catalase desde cedo desempenhou um papel notável nas formas de vida ancestrais dos organismos atuais.

Esta enzima é encontrada em praticamente todos os organismos vivos existentes, com variações apenas no número de genes encontrados. A catalase se apresenta diferente de outras enzimas antioxidantes pela habilidade de não requerer um composto redutor para auxiliar na catálise da reação de dismutação. Ela efetua uma reação de detoxificação do peróxido de hidrogênio gerado na célula por diversas atividades celulares, apresentando uma alta taxa de *turnover*, com uma baixa afinidade pelo  $H_2O_2$  (se comparado a outras enzimas, como APX) e um valor de  $K_M$



abaixo de 100  $\mu\text{M}$ . Em condições *in vivo*, essa enzima funciona bem abaixo da sua capacidade, mas esta situação muda quando ocorre incrementos das concentrações de  $\text{H}_2\text{O}_2$  para níveis mais elevados (MHAMDI et al., 2010).

A catalase (EC 1.11.1.6) é encontrada predominantemente nos peroxissomos. Outras localizações subcelulares (cloroplastos, mitocôndrias e citosol) ainda são um tema controverso, necessitando de mais estudos para confirmar essas informações. A função da catalase na célula é a detoxificação do  $\text{H}_2\text{O}_2$ , efetuando uma reação típica em que 2 moléculas de  $\text{H}_2\text{O}_2$  sofrem dismutação em água e  $\text{O}_2$ , conforme reação mostrada abaixo (SCANDALIOS et al., 1997; MHAMDI et al., 2010).



Neste mecanismo de reação, o  $\text{O}_2$  advém de apenas uma molécula de peróxido (a segunda a ser oxidada), enquanto o primeiro peróxido é catalisado para formar a primeira molécula de água, enquanto o oxigênio restante permanece ligado duplamente ao íon de ferro e que será utilizado, posteriormente, para a produção da segunda molécula de água. No decorrer da reação, são formados intermediários de ferro que estão ligados a aminoácidos na estrutura da proteína que medeiam a ligação com o  $\text{H}_2\text{O}_2$  e a depender do momento da reação estão como  $\text{Fe}^{\text{III}}$  ou  $\text{Fe}^{\text{IV}}$ . O primeiro  $\text{H}_2\text{O}_2$  é oxidado a  $\text{H}_2\text{O}$ , gerando  $\text{Fe}^{\text{IV}}$ , ligado duplamente ao oxigênio restante do  $\text{H}_2\text{O}_2$  que entrou (composto I). O próximo  $\text{H}_2\text{O}_2$  se ligará ao composto I, havendo a formação de  $\text{O}_2$  e da segunda molécula de  $\text{H}_2\text{O}$  (ZAMOCKY; FURTMÜLLER; OBINGER, 2008; MHAMDI et al., 2010). Esta reação pode ser inibida através da ligação da enzima com o cianeto (ANJUM et al., 2016).

Este é um mecanismo típico das catalases monofuncionais – homotetrâmeros com um único grupo heme por subunidade - encontradas na maior parte dos organismos, desde procariontes, fungos animais e plantas. As outras duas famílias gênicas incluem as catalases bifuncionais (catalases-peroxidases com grupo heme) e as catalases associadas ao manganês, ausentes de grupo heme. Esta classificação se baseia em características bioquímicas e físicas das enzimas (ZAMOCKY; FURTMÜLLER; OBINGER, 2008; WANG et al., 2017).

Em plantas, a catalase já tem sido alvo de vários estudos, com trabalhos enfatizando a busca por uma compreensão da função fisiológica dessa enzima e os diferentes estímulos que contribuem para a expressão gênica, regulação pós-transcricional e atividade enzimática. Porém, apesar do grande enfoque dado a essa

enzima, devido a ser altamente expressa e parte do sistema de defesa antioxidante, pouco ainda se conhece acerca da função fisiológica das diferentes isoformas, sua distribuição subcelular, os mecanismos que regulam essa distribuição e a verdadeira significância fisiológica dessas enzimas em plantas (MHAMDI; NOCTOR; BAKER, 2012).

Em espécies de plantas angiospermas estudadas, a catalase é codificada por uma pequena família multigênica, variando de 3 genes em *Arabidopsis thaliana* (FRUGOLI et al., 1996), milho (GUAN et al., 1991; REDINBAUGH; WADSWORTH; SCANDALIOS, 1988) e tabaco (HAVIR; BRISSON; ZELITCH, 1996) e 2 genes em cevada (SKADSEN; SCHULZE-LEFERT; HERBST, 1995) e mamona (SUZUKI et al., 1994). Além disso, os genes possuem elevada identidade entre si e entre genes de outras espécies vegetais (FRUGOLI et al., 1998; LEE; AN, 2005; MHAMDI; NOCTOR; BAKER, 2012), mostrando haver uma conservação na família multigênica.

A nomenclatura dos genes da catalase é inconsistente entre espécies. O sistema mais usual, refere-se ao adotado para as espécies de milho e *Arabidopsis*, nomeados como genes *cat1*, *cat2* e *cat3*. Um outro sistema usado se baseia nos genes da catalase de Tabaco (WILLEKENS et al., 1994b). Neste sistema, os genes são integrados em classes, de acordo com os tecidos em que são mais expressos.

A classe I se refere à catalase que tem um papel primordial na metabolização do H<sub>2</sub>O<sub>2</sub> durante a fotorrespiração em folhas. Na classe II, a função da catalase parece estar associada ao sistema vascular. Também relata-se ter função para a proteção contra estresses ambientais. A classe III está relacionada a flores e sementes. Esta classe também assume papel na degradação de ácidos graxos em glioxissomos (JOO; LEE; SONG, 2014). Um terceiro sistema de nomenclatura foi introduzido a partir da anotação de genes da catalase em arroz, usando-se das letras A, B e C, para os três genes (MORI et al., 1992; MORITA et al., 1994).

Esses sistemas de nomenclaturas usuais acabam tornando a integração de dados de diferentes artigos científicos uma tarefa de difícil execução, já que não há o conhecimento de quais genes são ortólogos entre si, ou seja, o gene *cat1* em *Arabidopsis* não corresponde filogeneticamente ao mesmo gene *cat1* em milho, assim como no restante da família multigênica. Sobretudo, como os genes da catalase apresentam um perfil de expressão espaço-temporal e sensível as diferentes condições ambientais, isso tudo dificulta na decisão de inserir os genes na nomenclatura definida por Willekens e colaboradores (1994b) (DU et al., 2008).

Essas desconformidades a respeito da classificação da catalase, somada a alta identidade entre membros gênicos, dificulta ou inviabiliza inferir a relação evolutiva/funcional de ortólogos entre espécies. Ademais, ainda há um desconhecimento acerca da distribuição e diversificação dessa família gênica durante a evolução de plantas angiospermas. Os poucos trabalhos que obtêm árvores filogenéticas utilizam poucas espécies na análise, não chegando a ter resultados conclusivos nesse tema (GUAN; SCANDALIOS, 1996; FRUGOLI et al., 1996; FRUGOLI et al., 1998).

Devido a catalase se mostrar uma proteína fácil de se trabalhar, em decorrência da facilidade de execução dos ensaios e estar presente em praticamente todos os organismos, existe uma ampla diversidade de trabalhos com esta enzima. Em biologia molecular, está enzima vem sendo bastante explorada com vistas a se determinar o perfil de expressão dos seus genes em resposta a diversos estresses (CHELIKANI; FITA; LOEWEN, 2004).

Todavia, apesar da abundância de dados publicados na literatura acerca da expressão da catalase frente a estresses, ainda é difícil se determinar um perfil claro de expressão. Ademais, informações filogenéticas relacionadas a expressão dos genes da catalase podem fornecer informações relevantes para um maior entendimento da função fisiológica da catalase de plantas. Dessa forma, é de grande importância o entendimento das relações filogenéticas entre os genes da catalase de um grande número de espécies no intuito de se compreender como eles se distribuem entre as ordens e famílias, tentando associar a filogenia a informações de expressão com espécie (s) em que há vastos estudos de expressão gênica.

#### **2.4 Duplicação gênica e a origem de famílias multigênicas**

A duplicação é uma forma de mutação onde uma região genômica é replicada e, em alguns casos, ela pode ser introduzida em um local fisicamente separado. Este é um processo fundamental para a diversificação e adaptação biológica. Em plantas, onde há uma abundância de genes duplicados, estima-se que o processo de duplicação completa do genoma (*Whole-Genome Duplication – WGD*) tenha ocorrido a cerca de 200 milhões de anos atrás, durante a evolução das Angiospermas. Ademais, como essas duplicações estão presentes em abundância nos genomas de

plantas, isso as torna uma fonte potencial para estudo, propiciando um ambiente rico para a descoberta de novas funções (PANCHY; LEHTI-SHIU; SHIU, 2016).

É o processo de duplicação de genes e o acúmulo de mutações em suas respectivas cópias, levando a divergência entre elas, que origina as famílias multigênicas. Essas duplicações podem ser originárias de eventos tais como: *crossing over*, retrotransposição, transposição duplicada de DNA e poliploidização (WGD). Por conseguinte, a característica marcante dos genomas de angiospermas é a presença de genes duplicados (MAGADUM et al., 2013). Somente na planta modelo *A. thaliana* foi constatada a ocorrência de três eventos de duplicação do genoma (referidos como  $\alpha$ ,  $\beta$  e  $\gamma$ ) (GUO, et al., 2013; PANCHY; LEHTI-SHIU; SHIU, 2016).

No geral, na maioria dos genes de plantas anotados, constata-se a presença de uma segunda cópia. Normalmente, não é vantajoso para uma espécie possuir duas cópias de um mesmo gene, logo essa segunda cópia apresenta redundância funcional em relação ao gene inicial (subfuncionalização), entretanto, ela também pode ter outros dois diferentes destinos. Uma das cópias pode se tornar um pseudogene, uma cópia de um gene ancestral funcional, sendo uma sequência não-codificante sem função. Além disso, uma das cópias pode continuar desempenhando a sua função original, enquanto a outra acumula mutações e, portanto, pode passar a exercer uma função diferente (neofuncionalização) (PANCHY; LEHTI-SHIU; SHIU, 2016).

Hoje, com a disponibilidade de genomas completos sequenciados, é possível estudar relações evolutivas entre genes de diferentes genomas, em um sistema de genes homólogos que inclui genes ortólogos e parálogos. Os primeiros são genes homólogos, derivados de um mesmo ancestral que após um evento de especiação divergiram. Já os parálogos, são genes homólogos que sofreram duplicação, sendo um evento restrito a uma linhagem evolutiva.

De modo geral, genes relacionados a regulação transcricional, transdução de sinais e resposta a estresses tendem a ter parálogos. Genes que codificam enzimas que atuam na proteção da célula contra o estresse oxidativo, tais como os da catalase são encontrados como parálogos em algumas espécies, dentre as quais podemos citar: *Populus tremula*, *Cucurbita Pepo* e *Raphanus sativus* (KWON; LEE; AN, 2007). Além disso, nas espécies de monocotiledôneas *Oryza sativa* e *Zea mays* os três genes da catalase são ortólogos entre si (GUAN; SCANDALIOS, 1996; HIGO; HIGO, 1996).

## 2.5 Estudos com a catalase na espécie *Zea mays*

A catalase tem sido mais extensivamente estudada na espécie *Zea mays*, desde a década de 90 por Scandalios e colaboradores. O milho possui 3 genes da catalase (*cat1*, *cat2* e *cat3*) que codificam três isoformas distintas (CAT1, CAT2 e CAT3) (SCANDALIOS et al., 1997), sendo esse número encontrado na maior parte das espécies estudadas até o momento.

Como já demonstrado por Scandalios, Acevedo e Ruzca (2000) a catalase no milho é expressa temporal e espacialmente durante a germinação. As três isoformas exibem um perfil complexo de desenvolvimento através do ciclo de vida do milho, o que dificulta o estudo e elucidação de um padrão de expressão geral dessa família multigênica durante o desenvolvimento. Além de tudo, há fortes evidências que esses genes podem ser regulados pós-transcricionalmente (SCANDALIOS et al., 1997; GUAN; SCANDALIOS, 1998).

O gene da CAT1 em milho tem sido expresso durante os estágios iniciais da germinação da semente. Além do mais, esse gene tem sido o único encontrado até agora em endosperma leitoso de sementes e no pólen após a antese. O gene *cat1* também é expresso na aleurona em estágios tardios de sementes e depois de 9 – 10 dias após a polinização (DAP) pode ser encontrado no escutelo. Além disso, o gene *cat1*, também pode ser encontrado em raízes de plantas de milho. O *cat1* também responde a condições de estresse, tais como salinidade, desidratação e altas temperaturas. Ademais, prévios trabalhos têm destacado que esse gene é modulado por ABA (SCANDALIOS; GUAN; POLIDOROS, 1997; GUAN; SCANDALIOS, 1998; SCANDALIOS; ACEVEDO; RUZCA, 2000; MENEZES-BENAVENTE et al., 2004).

O gene *cat2* pode ser encontrado na aleurona durante estágios mais tardios do desenvolvimento da semente. No escutelo em 10 DAP, a enzima CAT2 é a única presente. Assim como o *cat1*, o *cat2* responde diferencialmente a ABA e igualmente ao primeiro, essa resposta depende também do estágio de desenvolvimento (SCANDALIOS; GUAN; POLIDOROS, 1997). Esse gene também é expresso em folhas após exposição a luz (POLIDOROS; SCANDALIOS, 1997). Ademais, Guan e Scandalios (1995) demonstraram que o *cat2* responde positivamente a uma ampla variedade de doses de ácido salicílico em estágios embrionários.

O gene *cat3* é detectado no pericarpo (tecido materno) durante o desenvolvimento do óvulo e durante os estágios iniciais do desenvolvimento da

semente, decaindo logo em seguida. A CAT3 é a isoenzima predominante em zimogramas do coleóptilo e em raízes. O gene *cat3* também é o único expresso no caule (SCANDALIOS; GUAN; POLIDOROS, 1997). Além disso, esse gene é transcricionalmente regulado pelo ritmo circadiano, tendo sua máxima expressão no final do fotoperíodo, aproximadamente 10 horas de exposição a luz, com o mínimo de transcritos acumulados tardiamente no final do período da noite (POLIDOROS; SCANDALIOS, 1997). Os transcritos de *cat3* também são induzidos em embriões de milho sob tratamento com a toxina cercosporina purificada, havendo aumento da expressão, sendo dose-dependente (WILLIAMSON; SCANDALIOS, 1992).

As comparações realizadas entre os genes da catalase do milho mostram que as proteínas deduzidas são mais similares a catalases de outras espécies do que entre elas, não havendo elevada conservação na sequência proteica. Em uma análise filogenética feita por Guan e Scandalios (1996) usando 16 espécies, as catalases de plantas podem ser divididas em três grupos. O primeiro grupo inclui ambas catalases de eudicotiledôneas e monocotiledôneas (milho, cevada, algodão, *N. plumbaginifolia*, arroz e soja), o segundo grupo contendo apenas *N. plumbaginifolia* e tomate, e o terceiro grupo inclui catalases específicas de monocotiledôneas, compreendendo CAT3 do milho, CAT2 em cevada e CAT-A de arroz, suportado por um valor de *bootstrap* de 100%.

Entretanto, em decorrência do número limitado de espécies usadas no estudo, não se pode conhecer com profundidade como ocorreu a distribuição e diversificação dessa família gênica durante a evolução de plantas angiospermas. Ademais, como as atuais classificações não podem ser aplicadas a todos os genes da catalase, em decorrência principalmente da escassa quantidade de informações de expressão para a maioria das espécies (dificultando a inserção na classificação proposta por Willekens e colaboradores (1994b), sendo a mais generalista), se faz necessário a análise em um grande número de espécies, abrangendo monocotiledôneas e eudicotiledôneas, de forma a se sugerir uma classificação mais abrangente e adequada para essa família multigênica. Logo, isso ajudaria na compreensão dos trabalhos já existentes com a catalase, interligando os dados de forma a se ter resultados mais conclusivos.

Aliado a essa proposta de classificação, a análise de expressão *in silico* com uma espécie que já tem sido alvo de estudos mais gerais com a catalase, nos permite avançar para se realizar uma conclusão com uma base de dados mais consistente.

Além do mais, diferente de espécies modelo como *Arabidopsis*, o milho possui características de interesse econômico de relevância, como alto consumo de grãos em determinadas populações, uso com fins alimentícios para animais e produção de etanol (RANUM; PEÑA-ROSAS; GARCIA-CASAL, 2014).

Desse modo, a disponibilidade do genoma do milho e outras pesquisas genéticas (marcadores moleculares, mapas genéticos e dados transcriptômicos, etc) nos permite focar em estudos funcionais de genes com o objetivo final de gerar conhecimento útil, o qual pode ser aplicado para obtenção de características de relevância econômica, já que a catalase é uma enzima que responde a estresses ambientais.

### 3 OBJETIVOS

#### 3.1 Objetivo Geral

Analisar a distribuição filogenética dos genes da catalase em angiospermas e a funcionalidade (expressão *in silico*) em milho (*Zea mays*) em diferentes tecidos e condições de estresse.

#### 3.2 Objetivos Específicos

- Anotar manualmente os genes da família mutigênica da catalase em genomas de angiospermas disponíveis;
- Mostrar a distribuição filogenética da catalase;
- Sugerir um sistema de classificação para a catalase em plantas angiospermas baseado em análise filogenética e estrutura dos genes;
- Realizar análise de expressão gênica, *in silico*, utilizando bioprojetos de RNA-Seq para a espécie *Zea mays* em diferentes tecidos e condições de estresse.



## 4 MATERIAL E MÉTODOS

### 4.1 Identificação e anotação manual das sequências da catalase

As sequências das catalases foram identificadas através da ferramenta BLAST (ALTSCHUL et al., 1997), realizando-se buscas nos seguintes bancos de dados do NCBI (<https://www.ncbi.nlm.nih.gov/>): Nucleotide Collection (NR), Reference mRNA Sequences (REFSEQ\_RNA), Reference Genomic Sequences (REFSEQ\_GENOMIC), Expressed Sequence Tags (EST), Sequence Read Archive (SRA), Whole-Genomes Shotgun Reads (WGS), Transcriptome Shotgun Assembly (TSA) e Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>). Inicialmente, os genes foram identificados e selecionados de cada genoma utilizando uma sequência de proteína da catalase da *A. thaliana* como referência, através do tBLASTn.

Após selecionados, os genes foram manualmente anotados (determinação da posição da 5'-UTR, códon de iniciação, éxons, códon de parada e 3'-UTR) utilizando como referência sequências de cDNAs da catalase depositados no refseq\_rna (NCBI) com o auxílio da ferramenta BLASTn. Após a dedução dos cDNAs, estes foram traduzidos em sequências de aminoácidos usando o ExPASy (<https://web.expasy.org/translate/>). As proteínas deduzidas foram validadas a partir de sequências de catalase disponíveis no *GenBank* usando a ferramenta BLASTp.

As sequências de cDNA parciais encontradas foram complementadas através de montagens em contigs através do CAP3 (HUANG; MADAN, 1999), usando dados transcriptômicos disponíveis no SRA (*Sequence Read Archive*) e/ou do banco de ESTs. Feito isso, as mesmas foram validadas através do BLASTx. A estratégia de anotação empregada foi a mesma proposta por Costa et al., (2014).

### 4.2 Análise filogenética

Em vista da elevada conservação entre os aminoácidos observada entre sequências de proteína da catalase da mesma ou de diferentes espécies, a análise filogenética foi conduzida a partir das sequências de cDNA, tendo em vista o menor grau de identidade entre essas sequências, o que permite obter resultados mais robustos.

Na construção da árvore, as sequências de cDNA das algas *Chlamydomonas reinhardtii* (Chlamydomonadales) e *Volvox carteri* (Volvocales) foram utilizadas como grupo externo. O alinhamento entre as sequências de cDNA das catalases foi realizado através do programa MEGA7 (TAMURA et al., 2001) usando como ferramenta de alinhamento o Muscle, através do método de agrupamento UPGMA. (EDGAR, 2004). As árvores foram construídas por meio do software MrBayes 3.2.2 (HUELSENBECK; RONQUIST, 2001), utilizando para a procura das topologias o método de Monte Carlo via Cadeia de Markov (MCMC). A partir da utilização da metodologia Bayesiana, os valores de confiança são dados em probabilidade posterior, números entre 0 e 1. O valor de probabilidade posterior de um clado representa uma inferência a respeito da probabilidade daquele clado representar uma história evolutiva que mais se aproxima da história real. Valores mais próximos a 1 são os mais relevantes (BRAUN; JUNQUEIRA; VERLI, 2014).

O modelo de substituição nucleotídica foi escolhido e usado como informação *a priori* para a construção das árvores filogenéticas. A escolha foi baseada na metodologia de máxima verossimilhança, seguida do critério mínimo de informação teórica (*Minimum Theoretical Information Criterion* - AIC) implementado pelo software MrModeltest 2.3. O melhor modelo de evolução escolhido para as sequências de cDNAs das catalases foi o GTR+I+G (TAVARÉ, 1986). Foi usado na análise 1 milhão de gerações, sendo feito um descarte de 25% como *burnin*. As demais topologias foram utilizadas para obter a árvore consenso de acordo com as topologias de maior probabilidade posterior através da metodologia Bayesiana. Todos os passos necessários para a realização da análise Bayesiana foram realizadas através do servidor online CIPRES Science Gateway V. 3.3 (<https://www.phylo.org/>). Por fim, as árvores filogenéticas foram visualizadas e editadas no software FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

### 4.3 Estrutura éxon-ínton dos genes da catalase

A disposição dos éxons e íntrons de todos os genes das catalases das espécies angiospermas anotados foi realizada com o uso da ferramenta online *Gene Structure Display Server* (GSDS 2.0) (HU et al., 2015). Esta configuração é obtida através do alinhamento do cDNA com a sua respectiva sequência gênica. Conforme a classificação, os genes que codificam cada um dos membros da família multigênica

da catalase foram dispostos e separadas em grupo de acordo com a isoforma que codificam (CAT1, CAT2, CAT3 ou CAT4) para a identificação do número de éxons e íntrons de cada membro. Além disso, buscou-se verificar as mudanças ocorridas nessa disposição no intuito de se conhecer o padrão obtido por perda ou ganho de éxons e/ou íntrons durante a evolução dessa família multigênica.

#### 4.4 Análise das regiões promotoras dos genes da catalase do milho

A presença de possíveis elementos *cis* regulatórios responsivos à estresse [ABRE (ACGTG), MYB (CTAACCA/CNGTTR/WAACCA/TAAGT), WRKY (TGAC), MYC (CANNTG)], na região *upstream* de 1000 pares de base do códon de início da tradução foi analisada por meio de dados disponíveis nos seguintes servidores: New Place (<https://integbio.jp/dbcatalog/en/record/nbdc00168>) (HIGO et al., 1999), PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>) (LESCOT et al., 2002) e AtPAN (<http://atpan.itps.ncku.edu.tw/>) (CHEN et al., 2012).

A busca por elementos *cis* se concentrou nas famílias citadas acima, amplamente conhecidas por respostas a estresses, contudo demais elementos encontrados em grande quantidade (famílias GATABOX, IBOX e DOF) e outras famílias de elementos *cis*, também responsivas a estresses porém, menos conhecidas (PREATPRODH, LTRECOREATCOR15, DRECRTCOREAT, CBFHV, SBOXATRBCS, GT1GMSCAM4, ACGTATERD1, GCCCORE, SEBFCONSSTPR10A, LTRECOREATCOR15, ACGTATERD1) também foram consideradas durante a análise.

#### 4.5 Análise da expressão *in silico* utilizando dados de RNA-Seq para *Zea mays*

No intuito de avançar na compreensão da função de cada membro, a expressão gênica foi avaliada através de dados de RNA-seq disponíveis no *GenBank* de diferentes condições de desenvolvimento e estresse, usando a espécie *Zea mays* como modelo de estudo. As condições experimentais de cada bioprojeto explorado estão detalhadas na tabela 1.

Para as análises de expressão *in silico*, foram selecionados experimentos depositados no banco de dados SRA (*Sequence Read Archive*) do *GenBank*. No geral,

foram selecionados experimentos de transcriptômica com triplicatas biológicas da espécie *Zea mays* em desenvolvimento e em condições de estresse abiótico e biótico. Os arquivos com os dados dos experimentos foram baixados diretamente do SRA (<https://www.ncbi.nlm.nih.gov/sra>). Em seguida, esses dados (bibliotecas de cDNA) foram submetidos a diversos softwares computacionais aplicados em análise de dados de RNA-Seq.

Inicialmente, as bibliotecas de cDNA foram analisadas quanto a qualidade das *reads*. Essa etapa foi realizada através do programa FastQC v.0.11.7, que verifica a qualidade das *reads* (ANDREWS, 2010). Em seguida, a ferramenta Trimmomatic v.0.36 (<http://www.usadellab.org/cms/?page=trimmomatic>) (BOLGER et al., 2014) foi aplicada, quando necessário, com o intuito de se remover as sequências dos adaptadores utilizados na construção da biblioteca de cDNA, assim como também visando a remoção das *reads* de qualidade inferior, com um índice de qualidade *Phred* (Q) < 20, e das *reads* de tamanho muito reduzido (< 50 pb a depender do tamanho da *read*).

Após a avaliação da qualidade das *reads*, aquelas que apresentaram qualidade alta/média foram usados para a etapa de mapeamento com os genes de interesse, utilizando os parâmetros padrões das ferramentas TopHat v. 2.1.1 (KIM et al., 2013) e Bowtie v. 2.2.9 (LANGMEAD; SALZBERG, 2012). Para se obter a contagem bruta do número de *reads* que mapearam contra os genes de interesse foi empregada a ferramenta HTSeq v. 0.9.1 ([https://htseq.readthedocs.io/en/release\\_0.9.1/](https://htseq.readthedocs.io/en/release_0.9.1/)) (ANDERS et al., 2015) utilizando-se como base a anotação de referência (GTF).

Os valores de contagem bruta dos genes foram normalizados de acordo com o tamanho da biblioteca utilizando o método de normalização contagem por milhão (CPM) presente na *cpm function* do pacote *edgeR* (ROBINSON et al., 2010). Apenas transcritos com CPM  $\geq 2$  em pelo menos três das bibliotecas foram selecionados para análise de expressão.

Para verificação do comportamento das replicatas, foi conduzida uma análise de componentes principais (ACP) baseado no arquivo de contagem após normalização e filtragem. Para os dados de expressão em tecidos e/ou em desenvolvimento, foi usado o método de normalização a partir da função *counts*, a qual normaliza os dados a partir dos *size factors*, conforme descrito por Anders e Huber (2010).

A ACP, a normalização e os genes diferencialmente expressos (GDEs) foram determinados através do programa R studio v. 3.4.1 (RCORE TEAM), utilizando o pacote DESeq2 v. 1.30.0 (<http://bioconductor.org/packages/release/bioc/html/DESeq2.html>) (LOVE; HUBER; ANDERS, 2014). Foram considerados como diferencialmente expressos apenas os genes que apresentaram valor de  $p$  ( $p$  value) ajustado  $< 0,05$  (BENJAMINI; HOCHBERG, 1995) e  $|\log_2 \text{fold-change}| (\geq 1)$ .

Tabela 1 - Experimentos de RNA-Seq utilizados para se obter o perfil de expressão dos genes *cat1*, *cat2* e *cat3* em *Zea mays*.

Acesso do Bioprojeto	Cultivar	Tecido (s)	Estágio da planta	Tratamentos	Detalhamento experimental	Replicatas	Referência
PRJNA291919	B73	Raiz primária	V4	Controle e estresse osmótico	Mudas com um comprimento de raiz primária de 2 a 4 cm foram transferidas para rolos de papel embebidos em soluções de polietilenoglicol (PEG8000) com um potencial de água de 0,8 MPa para déficit hídrico, sendo o controle feito com água destilada. Após 6 h de incubação em solução aerada de PEG ou água destilada, as raízes primárias foram dissecadas em quatro tecidos distintos. Primeiramente, os 2 mm apicais da raiz foram coletados, contendo capuz de raiz e zona meristemática. Segundo, a zona proximal adjacente à ponta da raiz até a parte da raiz onde os pêlos radiculares se tornaram visíveis foi amostrada, o que corresponde à zona de alongamento. Finalmente, a zona de diferenciação distal, da zona de raiz do cabelo ao coleorriza, foi mecanicamente separada em córtex e estela.	Três	-
PRJNA436207	Yayu889 (tolerante a doença)	Folhas	R3/R4	Controle e infectado	As plantas foram infectadas com o fungo <i>Cercospora zea-maydis</i> , compreendendo os estágios 1, 2 e 3, correspondentes a 89, 91 e 93 dias após o plantio.	Três	-

Fonte: elaborado pela autora.

Tabela 1 - Continuação.

Acesso do Bioprojeto	Cultivar	Tecido (s)	Estágio da planta	Tratamentos	Detalhamento experimental	Replicatas	Referência
PRJNA80075	B73	Óvulos, saco embrionário e pólen	V3 – R1	-	Plantas foram cultivadas em casa de vegetação (16 dias). Em seguida óvulos maduros e sacos embrionários foram isolados de estilo-estigmas de aprox. 10 cm. Os óvulos foram removidos com pinça e colocados em placas de Petri para digestão da parede celular. Os sacos embrionários foram removidos dos óvulos usando agulhas de dissecação e colocados em microtubos com solução tampão. Em seguida, as amostras foram submetidas a refrigeração com nitrogênio líquido e após isso, foi realizada a extração do RNA.	Três	-
PRJNA217053	B73	Vários	V8	-	Diversos tecidos foram coletados de plantas em estágio V8, crescidas em casa de vegetação.	Três	Walley et al. (2016)

**Fonte:** elaborado pela autora.

## 4.6 Análise estatística

A análise dos dados de expressão obtidos a partir dos experimentos de RNA-Seq foi realizada usando o software GraphPad Prism 5.0. Os resultados foram expressos como a média +/- desvio padrão das três replicatas biológicas. As diferenças nas variáveis foram analisadas por *one-way* ANOVA, seguido pelo teste de Bonferroni. As diferenças foram consideradas significante em  $p < 0,05$ .

## 5 RESULTADOS

### 5.1 Identificação, análise filogenética e classificação da catalase

Dentre os genomas de angiospermas disponíveis nos bancos de dados do NCBI (<https://www.ncbi.nlm.nih.gov/>) e Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>), foram selecionados e anotados genes da catalase em 110 genomas, compreendendo 26 ordens, as quais englobaram tanto monocotiledôneas, quanto eudicotiledôneas. No total, 288 sequências gênicas foram identificadas, com o número de genes variando entre as espécies de um a sete genes.

De maneira geral, nas ordens de monocotiledôneas foi encontrado um número menor de genes, com três genes sendo o número mais prevalente. Nas eudicotiledôneas, esse número também se manteve, contudo, uma maior variação entre ordens foi encontrada (Figura 1).

De forma genérica, nas sequências proteicas deduzidas a partir dos cDNAs anotados, foi encontrada uma elevada identidade, tanto entre sequências de uma mesma espécie, quanto entre sequências de espécies diferentes (entre 70% a 99%). Também se constatou uma conservação no tamanho da sequência de 492 aminoácidos, havendo poucas variações, sendo estas encontradas em algumas espécies da ordem Poales, como em *Ocimum tenuiflorum*, *Zea mays*, *Panicum hallii*, *Dichanthelium oligosanthes*, *Setaria italica*, *Sorghum bicolor*, *Saccharum hybrid cultivar*, *Triticum urartu* e *Aegilops tauschii*. Em algumas dessas espécies o comprimento das sequências variou de 492 a 496 aminoácidos.



**Figura 1** - Distribuição dos genes das subfamílias da catalase em angiospermas. A filogenia de angiospermas apresentada está de acordo com as normas estabelecidas pelo *Angiosperm Phylogeny Group* (APG IV, 2016).

Angiospermas		Ordens	Subfamílias da catalase
Angiospermas basais		Amborellales	CAT1a1, CAT1a2
Monocotiledôneas		Zingiberales	CAT3a, CAT3b, CAT1a, CAT1b
		Poales	CAT2, CAT3, CAT1a1, CAT1a2
		Arecales	CAT3, CAT1, CAT1a, CAT1a1, CAT1a2
		Asparagales	CAT3, CAT1a, CAT1b
		Dioscoreales	CAT3, CAT1
		Alismatales	CAT1b
Eudicotiledôneas basais		Proteales	CAT1a1, CAT1a2
Eudicotiledôneas	Rosidaeas	Cucurbitales	CAT4a, CAT4b, CAT1a, CAT1a1, CAT1a2
		Fagales	CAT4a, CAT4b1, CAT4b2, CAT4b, CAT4c
		Rosales	CAT4a, CAT4a1, CAT4a2, CAT1, CAT1b, CAT4
		Fabales	CAT4a, CAT4b, CAT4a1, CAT4a2, CAT4b1, CAT4b2, CAT1
		Oxalidales	CAT4
		Malpighiales	CAT1, CAT4a1, CAT4a2, CAT4a, CAT4b, CAT4, CAT4b1, CAT4b2
		Malvales	CAT4, CAT4a1, CAT4a2, CAT1, CAT4a
		Brassicales	CAT1, CAT4a1, CAT4a2, CAT1a, CAT1b, CAT1c, CAT1a1, CAT1a2, CAT1b1, CAT1b2, CAT1c
		Sapindales	CAT4a, CAT4b
		Myrtales	CAT4a, CAT4b, CAT4c
		Vitales	CAT1, CAT4
		Caryophyllales	CAT1a, CAT1b, CAT1b1, CAT1b2, CAT1
		Asteridaeas	Ericales
	Gentianales		CAT1, CAT4a1, CAT4a2, CAT4
	Lamiales		CAT4, CAT4a, CAT4b
	Solanales		CAT1, CAT1a1, CAT1a2, CAT4, CAT4a, CAT4b, CAT4a1, CAT4a2
	Asterales		CAT4a, CAT4b, CAT4c, CAT4a1, CAT4a2, CAT4b1, CAT4b2
	Apiales		CAT4a1, CAT4a2, CAT4, CAT4b1, CAT4b2

Fonte: dados da pesquisa

A partir da árvore filogenética construída com base na metodologia Bayesiana, usando o método de agrupamento UPGMA, observa-se a necessidade de se propor uma nova classificação. Isso se dá devido, principalmente a classificações anteriores propostas para tabaco, arroz e *Arabidopsis* não poderem ser estendidas para as demais espécies, pois a disposição dos genes dessas espécies não permite uma nomeação com base na filogenia, já que as classificações baseadas na filogenia dos genes são uma das classificações que mais fornecem informações relevantes.

A partir dessa análise, quatro subfamílias de catalase em angiospermas foram nomeadas de acordo com as sequências de catalase de *Zea mays*. Para deixar a nomenclatura mais clara e objetiva algumas sequências tiveram a sua nomenclatura modificada, contudo como poucas sequências eram nomeadas no NCBI essas alterações foram mínimas. Além disso, para as espécies mais estudadas a nomenclatura proposta é disposta juntamente com a nomenclatura anteriormente definida (Figura 2).

Figura 2 - Classificação proposta para espécies com classificação previamente definida no NCBI e/ou na literatura.

Espécie	Classificação (NCBI/artigo)			Referência	Classificação proposta		
	CAT1	CAT2			CAT2	CAT1	
<i>Hordeum vulgare</i>	CAT1	CAT2		Skadsen, Schulze-lefert e Herbst (1995)	CAT2	CAT1	
<i>Musa acuminata</i>	MaCAT2			Figuroa-Yáñez et al. (2012)	CAT3a		
<i>Oryza sativa</i>	CATA	CATB	CATC	Mori et al. 1992 e Morita et al. 1994	CAT3	CAT1	CAT2
<i>Saccharum hybrid cultivar</i>	CAT1	CAT2		Su et al. (2014)	CAT3	CAT1	
<i>Zostera marina</i>	ZMCAT1			Zang et al. (2018)	ZmCAT1		
<i>Arabidopsis thaliana</i>	CAT1	CAT2	CAT3	Frugoli et al. (1996)	CAT1c	CAT1a	CAT1b
<i>Capsicum annuum</i>	CAT2	CAT1	CAT3	Lee et al. (2005)	CAT4a	CAT4b	CAT1
<i>Nicotiana tabacum</i>	CAT3	CAT1		Havir, Brisson e Zelitch (1996)	CAT4a	CAT1	CAT4b
<i>Prunus persica</i>	CAT3	CAT1	CAT2	Bagnoli et al. (2004)	CAT4a1	CAT4a2	
<i>Ricinus communis</i>	CAT1	CAT2	CAT1	SUZUKI et al. (1994)	CAT4a	CAT4b	CAT1
<i>Solanum lycopersicum</i>	CAT3	CAT1	CAT2	Kabir e Wang (2011)	CAT4a	CAT4b	CAT1

Fonte: dados da pesquisa

Inicialmente, o primeiro clado, que denota maior proximidade com o grupo externo, integra a subfamília CAT2, conforme indicado na árvore filogenética em verde. Esta subfamília é encontrada apenas em monocotiledôneas e é exclusiva de

espécies da ordem Poales. Tal clado é suportado por um valor de probabilidade posterior de 1. Este clado agrupa somente 4% das sequências do total de genes identificados. O segundo clado integra a subfamília CAT3 (em amarelo na árvore), tal família gênica é encontrada apenas em monocotiledôneas (Figura 3). Esse clado agrupa todas as ordens de monocotiledôneas analisadas nesse trabalho, sendo elas Poales, Zingiberales, Dioscoreales, Asparagales, Arecales e Alismatales, possui 8% das sequências do total de genes identificados e é suportado por um valor de probabilidade posterior de 1. O terceiro clado agrupa a subfamília da CAT1 (em vermelho na árvore). Essa subfamília é compartilhada por ambas as classes (monocotiledôneas e eudicotiledôneas) e se trata de um clado amplo, pois, aproximadamente 39% das sequências do total analisado compõem esse clado.

Por último, a subfamília CAT4 (em azul na árvore), compondo o subclado quatro, apresenta uma notável proximidade com o clado 3 (subfamília da CAT1), no entanto, é composto somente por espécies de eudicotiledôneas (Figura 3). Aproximadamente 49% das sequências do total, compõem esse clado, suportado por um valor de probabilidade posterior relativamente baixo de 0,5, o que sugere uma relação ambígua. Todas as informações referentes ao número de genes de cada espécie e sua classificação, assim como os acrônimos usados para nomeação na árvore filogenética e na estrutura éxon-íntron estão dispostas no ANEXO A.

No que concerne as relações filogenéticas entre os grupos na árvore, as monocotiledôneas apresentam as sequências da catalase (cDNA) com uma divergência de identidade entre os membros gênicos de uma mesma espécie, como é o caso da espécie *Zea mays*. Poucas duplicações recentes estão presentes nas espécies de monocotiledôneas estudadas aqui, com exceções as Zingiberales, onde foi constatado duplicações da CAT1 e da CAT3, as quais foram nomeadas como CAT1a/CAT1b e CAT3a/CAT3b, respectivamente. Contudo, algumas ordens de eudicotiledôneas ainda compartilham a CAT1, que é o caso das ordens Brassicales, Solanales, Caryophyllales, Arecales, Dioscoreales, Asparagales, Proteales, Amborellales, Malpighiales, Malvales, Vitales, Gentianales e Ericales. Dessa forma, constata-se que o surgimento da CAT1 ocorreu antes do grande evento de separação de monocotiledôneas e eudicotiledôneas. Em todas as demais eudicotiledôneas analisadas observou-se uma maior predominância da CAT4.

Em decorrência do padrão gerado a partir da reconstrução filogenética, selecionou-se o milho como espécie-chave para se estudar a expressão da catalase,

tanto por as três subfamílias estarem presentes na espécie, quanto pela existência de uma ampla base de dados transcriptômicos. Além disso, o milho possui uma literatura relativamente ampla com a catalase e aliado à isso, essa espécie possui grande importância econômica, sendo que os estudos de expressão com a catalase podem ser úteis para futuros trabalhos objetivando o desenvolvimento de marcadores associados a estresses.

Figura 3 - Árvore filogenética com as seqüências de cDNAs da catalase, obtidos a partir da anotação em 110 genomas de plantas angiospermas.

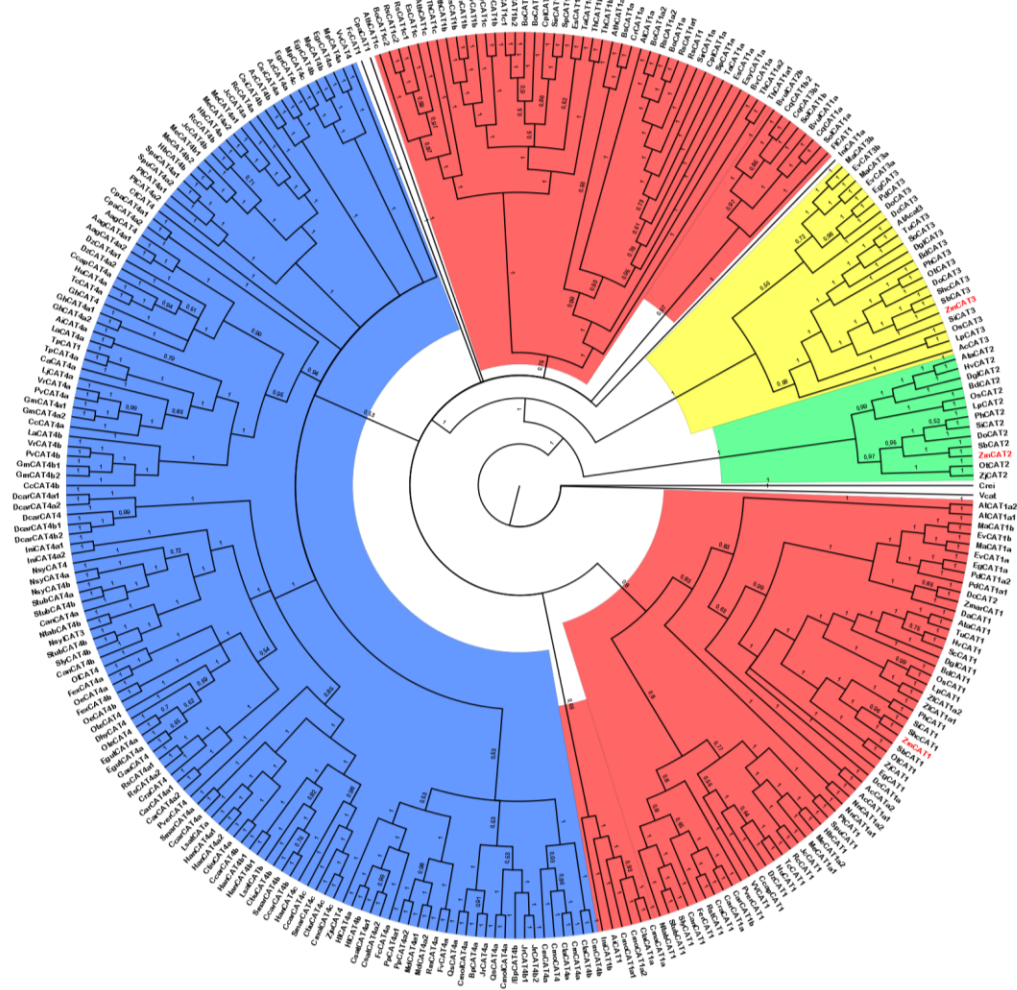


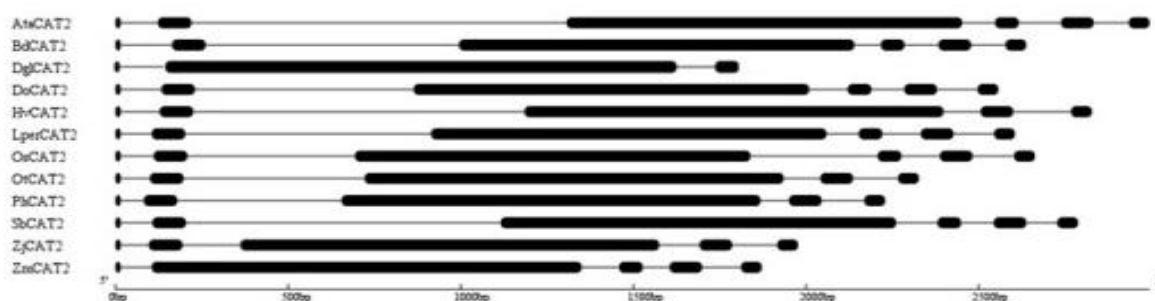
Figura 3: dados da pesquisa.

## 5.2 Caracterização estrutural dos membros gênicos/proteínas da catalase de Angiospermas

A estrutura de éxons e íntrons foi determinada e comparada entre os genes da catalase de diferentes espécies com a mesma classificação. No geral o que se observa é um padrão de variação no número de íntrons, fazendo com que os membros gênicos tivessem entre 3 e 10 éxons. Como, no geral, as catalases anotadas apresentaram um tamanho com poucas variações no número de aminoácidos, a ORF (*Open Reading Frame*) de praticamente todos os genes teve um tamanho de 1479 pb. De alguns genes não foi possível prever a estrutura, pois estes foram obtidos através de montagem do cDNA completo através de bancos de transcriptômica, não sendo feita a anotação.

No gene da CAT2 (clado 1), específico de espécies de monocotiledôneas, sendo encontrado em espécies da ordem Poales, observou-se uma variação do número de três a seis éxons. No total, 50% das sequências apresentaram seis éxons em sua estrutura e 40% apresentaram cinco éxons (o milho pertence a este grupo), como um resultado da deleção do íntron 3, exceto em *Zea mays* em que o íntron deletado foi o de número 2. Interessantemente, nesse grupo, *cat2* de *Dactylis glomerata* exibiu um padrão incomum, com apenas três éxons devido a deleção dos íntrons 2, 3 e 4. Com relação à proteína, apenas o gene *Zmcat2* apresentou variação no tamanho, com 493 aminoácidos, com um resíduo adicional de glutamato na posição 282. A estrutura com os éxons e íntrons dos genes *cat2* está representada na figura 4.

Figura 4 - Estrutura éxon-íntron dos genes pertencentes a subfamília CAT2. Os éxons são representados por caixas pretas e os íntrons por linhas finas.

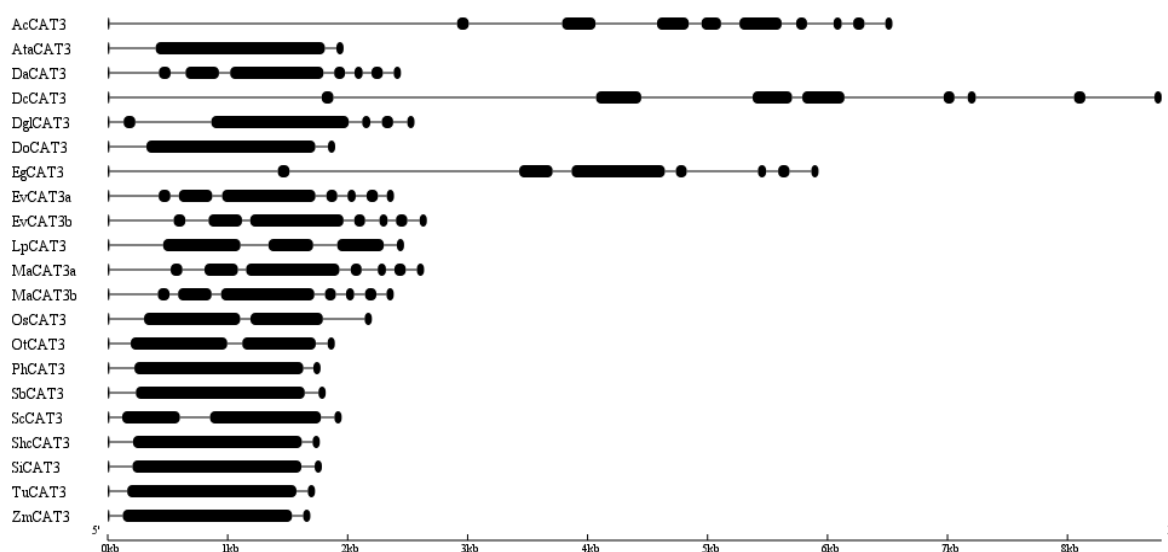


Fonte: elaborada pela autora.

No gene da CAT3 (clado 2), também específico de monocotiledôneas, notou-se uma variação ainda maior na estrutura de éxons e íntrons. Nesse gene, observa-se

uma variação de três a dez éxons. No geral, 40% das sequências possuem três éxons (o milho pertence a este grupo) e 36% possuem 8 éxons. Os 24% restantes possuem de quatro a nove éxons. Neste gene já se observa um padrão bem mais diversificado de distribuição de éxons/introns, além do tamanho da sequência da proteína ser acompanhado por variações no número de aminoácidos. Essas variações ocorrem na posição 406 e 410 da sequência, com deleções e inserções de aminoácidos, sendo caracterizada como uma região muito variável (Figura 5).

Figura 5 - Estrutura éxon-íntron dos genes pertencentes a subfamília CAT3.

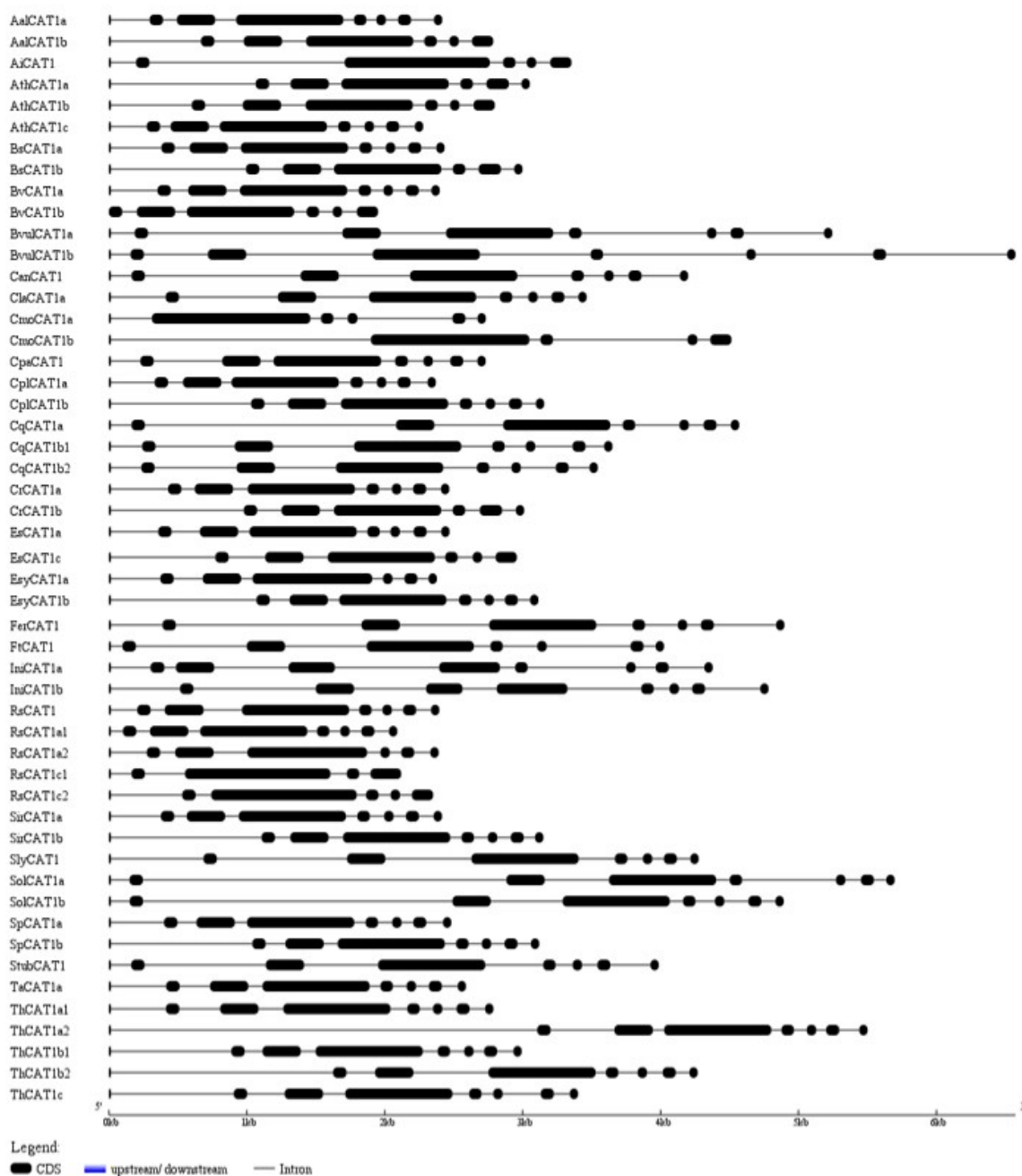


Fonte: elaborado pela autora.

No gene *cat1* (clado 3), compartilhado por monocotiledôneas e eudicotiledôneas também se observa variação no número de éxons/introns. Nesse gene, percebe-se uma variação de quatro a nove éxons. Entretanto, observa-se um padrão comum de oito éxons, sendo que 80% das sequências possuem esse número de éxons, (o milho pertence a este grupo) 9% possuem sete e os 11% restantes variam de nove à quatro éxons. Pouca variação é encontrada no tamanho da sequência proteica. Apenas *Zlcat1a1* e *BoCAT1a2* possuem variação, sendo 494 e 491 aminoácidos, respectivamente. Neste último gene constatou-se a presença de um códon de parada bem no centro da sequência, o que levanta o questionamento de que possa se tratar de um pseudogene. A estrutura éxon-íntron está representada na figura 6.

Por último, o gene *cat4* possui uma estrutura pouco variável. No geral, 80% das sequências possuem 8 éxons, com os 20% restantes variando de nove a seis éxons. Poucos polimorfismos são encontrados quanto ao tamanho da proteína.

Figura 6 - Estrutura éxon-intron dos genes pertencentes a subfamília CAT1.



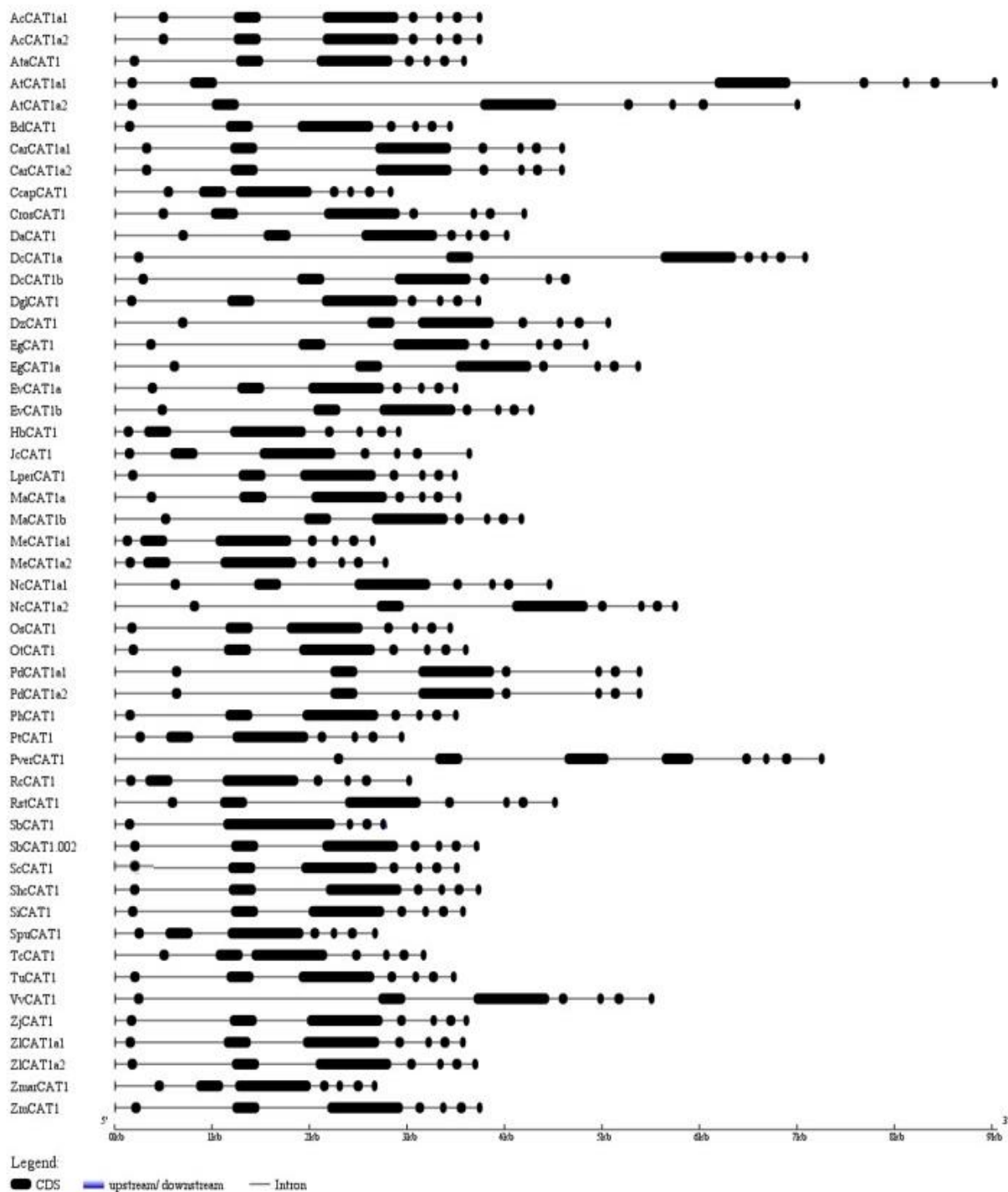
Fonte: elaborado pela autora.

Analisando a estrutura dos genes que compõem ambas as subfamílias da catalase, percebe-se que os eventos de deleção de íntrons foram preponderantes no processo evolutivo desses genes. Como os genes para a *cat1* e *cat4* são resultado de processos de duplicação mais recentes, o número de éxons/íntrons em sua estrutura



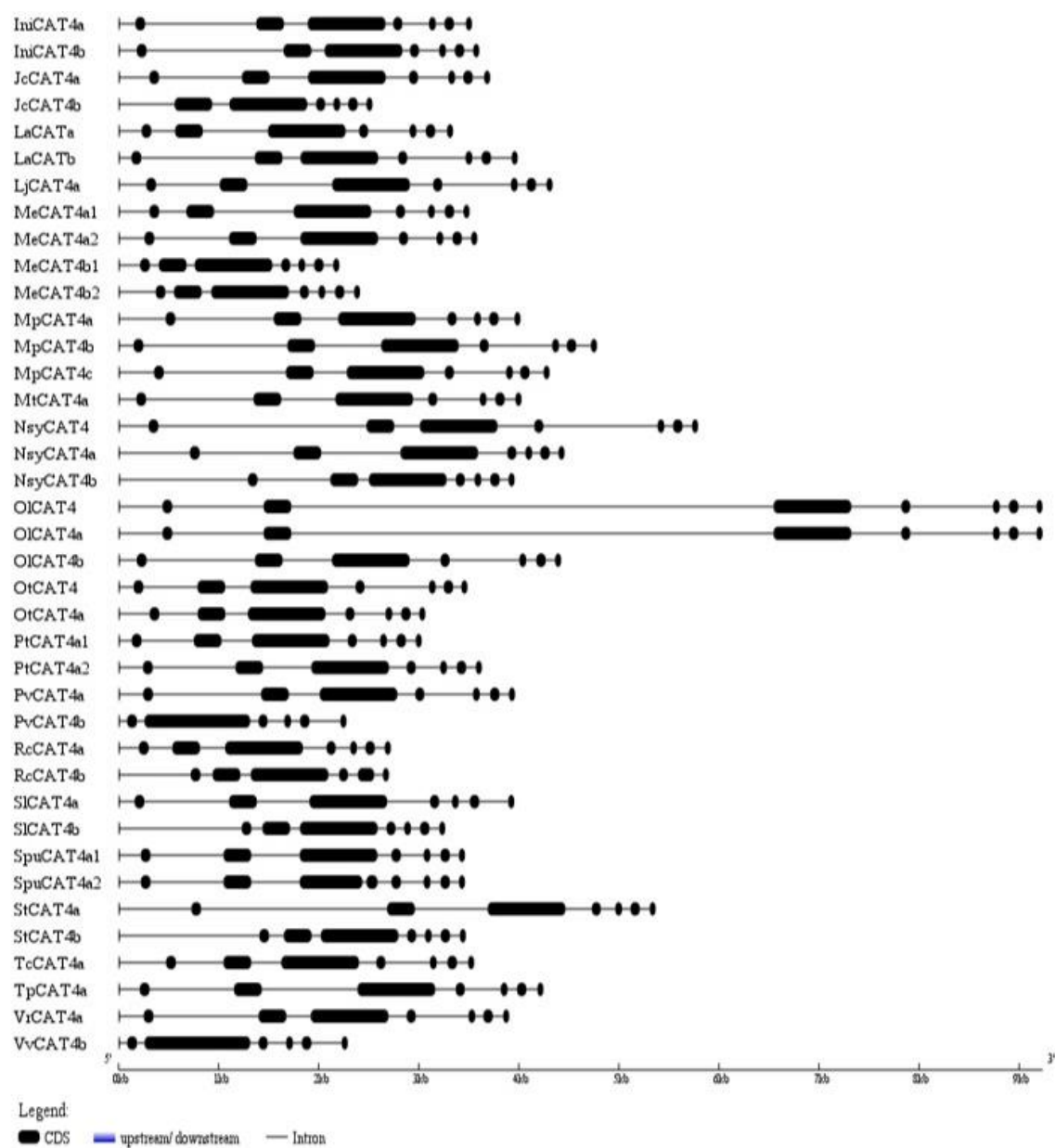
não teve muitas alterações, diferentemente dos genes da *cat2* e *cat3*, com alterações em sua estrutura bem mais evidentes.

Figura 6 - Continuação



Fonte: elaborado pela autora.

Figura 7 - Estrutura éxon-íntron dos genes pertencentes a subfamília CAT4.



Fonte: elaborado pela autora.

Figura 7 - Continuação.

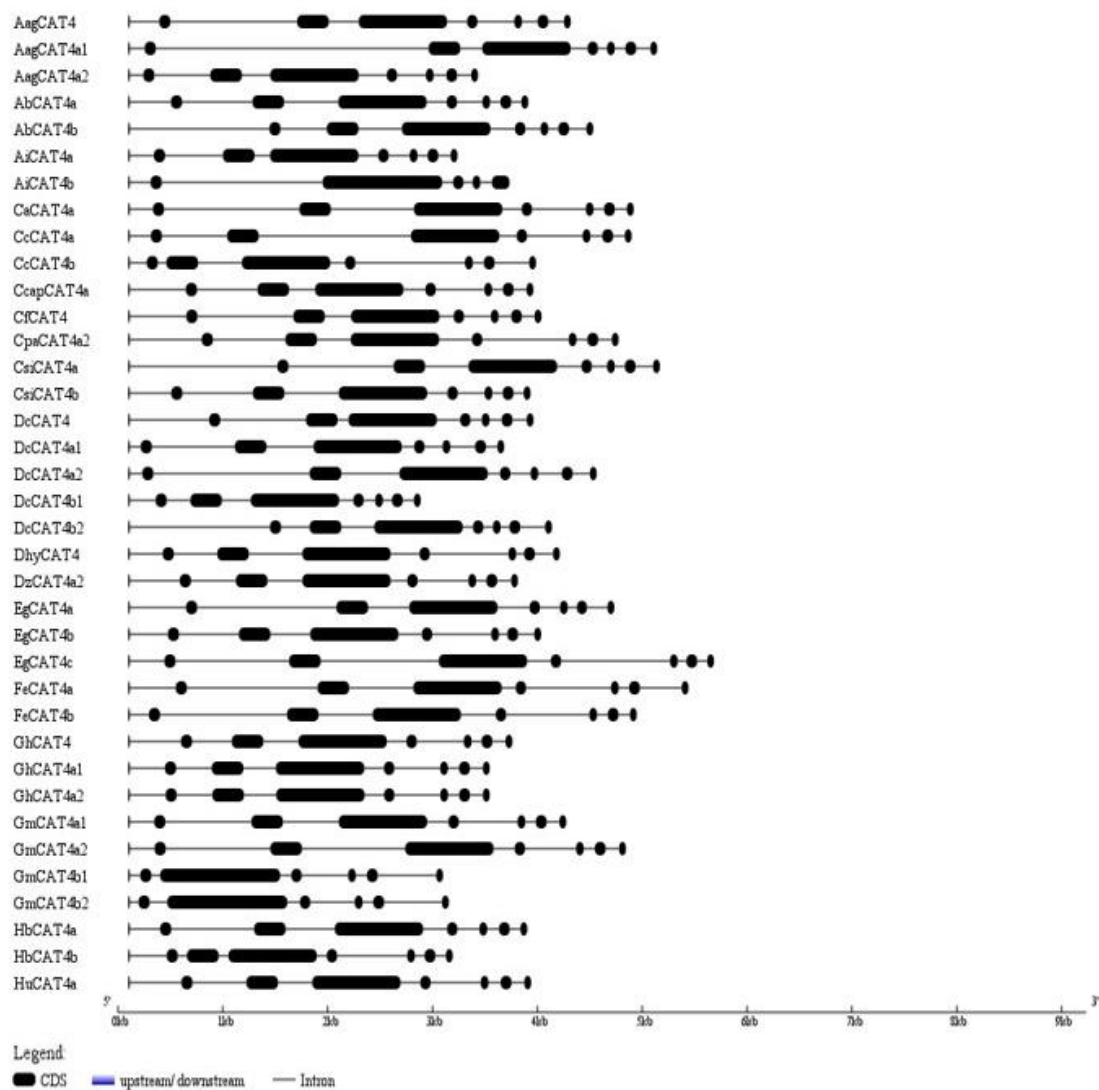
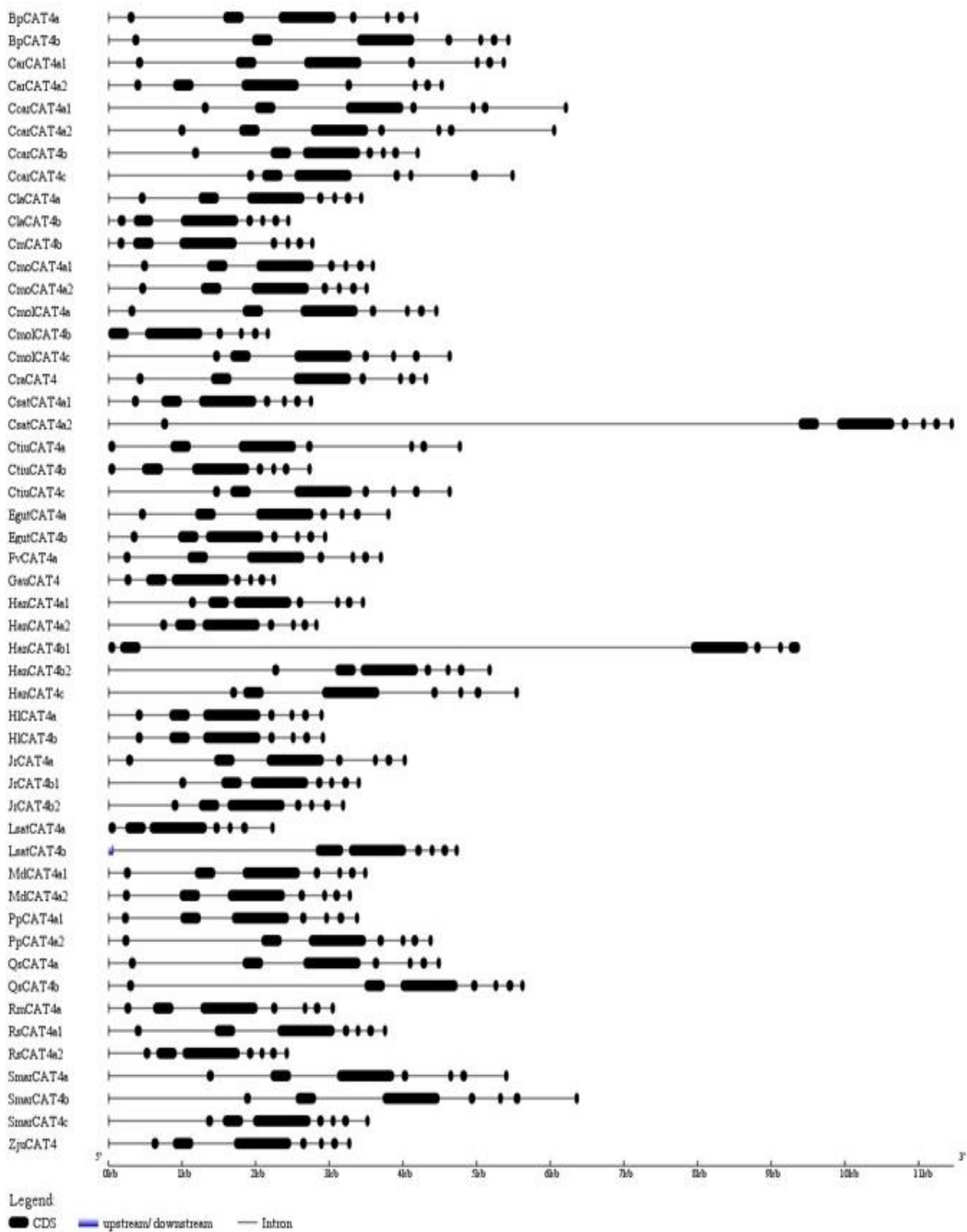


Figura 7 - Continuação.



### 5.3 Possíveis elementos *cis* responsivos a estresses nas regiões promotoras dos genes da Catalase de *Zea mays*

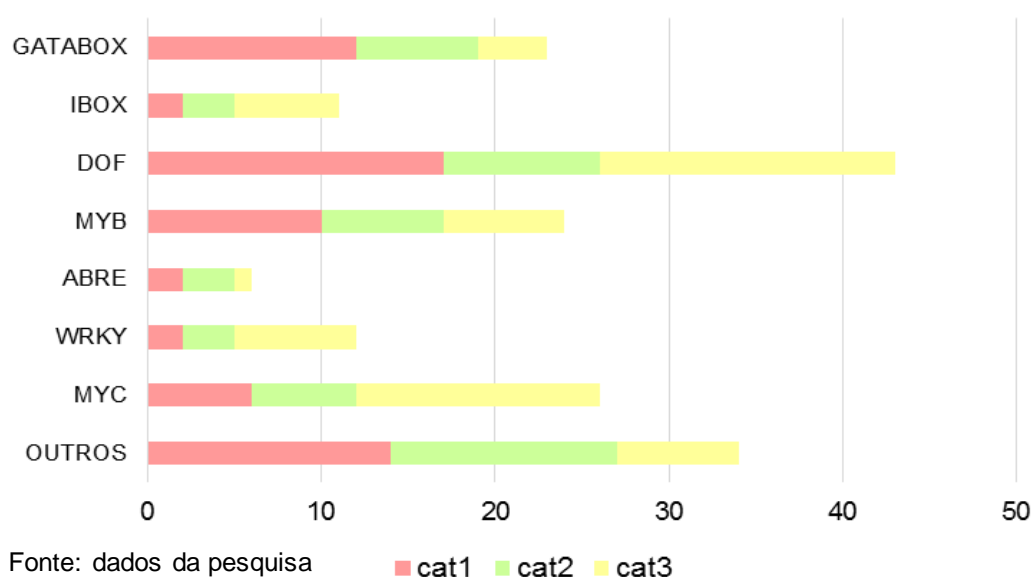
As análises das regiões promotoras dos genes *cat1*, *cat2* e *cat3* do milho, conforme mostrado na figura 8, revelaram a presença de elementos *cis*, sendo divididos em três grandes grupos: o primeiro grupo, composto pelos elementos relacionados a estresses, tais como as famílias ABRE (ACGTG), WRKY (TGAC), MYB (AACGG/YAACKG/CNGTTR) e MYC (CANNTG), presentes em maior abundância na região promotora de todos os genes. Já o segundo grupo é composto pelos elementos relacionados a genes responsivos a luz, como as famílias de elementos *cis* GATABOX (GATA), IBOX (GATAAG/GATAA) e DOF (AAAG), ambos também muito abundantes, e o terceiro grupo composto por outros elementos que foram encontrados em um número pequeno (variando de um a quatro), estando relacionados a respostas a estresses, conforme representado na figura 8.

Todos os três genes apresentaram elementos *cis* que se distribuíram nos três grupos. Uma maior diversidade de elementos *cis* foi encontrada no promotor do gene da CAT1 (mostrado em vermelho na figura 8) e uma maior quantidade no promotor de gene da CAT3 (em amarelo).

No gene da CAT2 (mostrado em verde), um total de 19 elementos *cis* foram identificados, sendo em sua maior parte, os elementos *cis* pertencentes à família MYB, com uma maior diversidade de outros elementos *cis* responsivos à patógenos, pertencentes ao grupo três. No gene da CAT3, o qual apresentou um maior número de elementos *cis* na sua região promotora, foi identificado um total de 29 elementos, os quais, em sua grande maioria, pertencente à família MYC. Por último, no gene da CAT1, foi encontrado um total de 20 elementos *cis*, sendo em sua maioria da família MYB. Nesse gene uma maior diversidade de outros elementos *cis* responsivos à estresses abióticos foi encontrada, incluídos no grupo três, em torno de sete elementos, aparecendo por volta de duas vezes cada.

Os elementos *cis* relacionados a respostas a luz foram abundantes, com uma maior quantidade de elementos *cis* pertencentes a família DOF, em todos os genes da catalase.

Figura 8 - Número de elementos *cis* regulatórios presentes na região promotora dos genes *cat1*, *cat2* e *cat3* de *Zea mays*.



### 5.3 Expressão *in silico* dos genes da catalase em *Zea mays*

A expressão dos genes da catalase *in silico* foi avaliada a partir de dados transcriptômicos da espécie *Zea mays* em condições de estresse osmótico, ataque por patógenos e em diferentes tecidos durante o desenvolvimento. Em todos os experimentos avaliados, a análise dos componentes principais revelou uniformidade entre as replicatas.

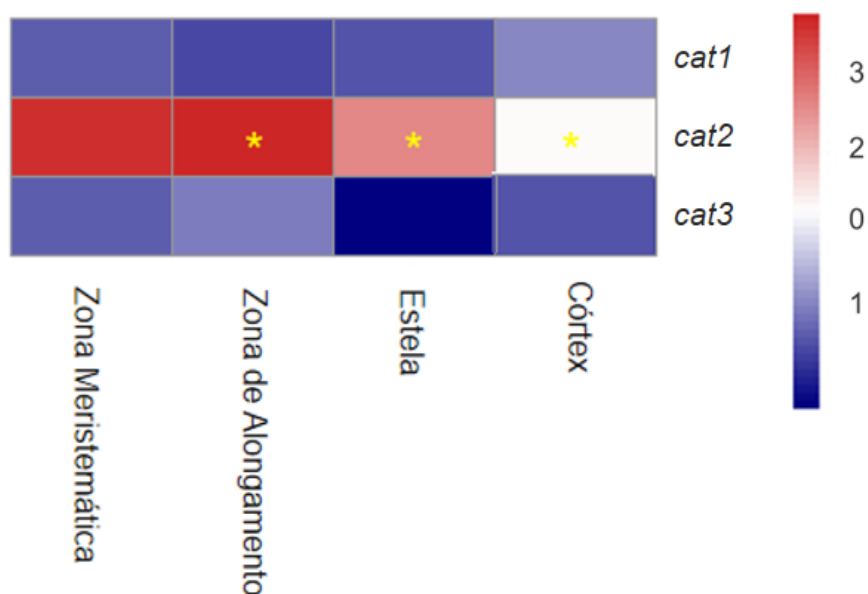
No geral, os resultados mostram que todos os genes foram expressos, no entanto, o perfil de expressão foi variável e dependente do tecido e do estágio de desenvolvimento da planta, além do tratamento ao qual a planta é exposta.

#### 5.3.1 Expressão *in silico* dos genes da catalase em plantas de milho submetidas a estresse osmótico

Os dados de RNA-Seq para esse experimento foram obtidos a partir da extração e sequenciamento do RNA de plantas de milho da cultivar B73 submetidas a estresse osmótico, avaliando-se diferentes partes da raiz dessas plantas, o experimento foi composto por três replicatas biológicas. O estágio avaliado foi o V4, onde o ponto de crescimento das plantas ainda se encontra abaixo da superfície do solo, apresentando um caule pequeno e ainda não completamente desenvolvido. Neste estágio, devido

ao crescimento acontecer a partir da raiz, condições de temperaturas extremas (alta e baixa), assim como pouco suprimento de água, como avaliado no estudo, podem afetar negativamente o crescimento da planta, culminando em um aumento do tempo entre um estágio e outro, alongando o ciclo da cultura e acarretando prejuízos à produção.

Figura 9 - Perfil de expressão dos genes *cat1*, *cat2* e *cat3* em diferentes partes da raiz primária obtidas de plantas de milho submetidas à estresse osmótico. A escala de cor representa os valores de Log2fold-change obtidos após normalização dos dados. O asterisco (\*) representa um *p-value* > 0,05.



Fonte: dados da pesquisa

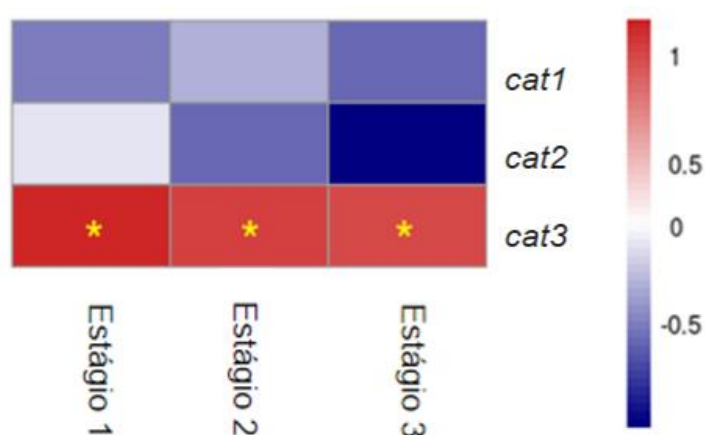
Nesse experimento, o gene *cat2* foi regulado positivamente na zona de alongamento, zona meristemática e na estela (Figura 9). Contudo, apenas na zona de alongamento e na estela este gene apresentou um valor de *p* abaixo de 0,05, sendo um resultado consistente. Os outros genes *cat3* e *cat1*, foram regulados negativamente em relação ao controle, não apresentando valores de *p* considerados confiáveis.

### 5.3.2 Expressão *in silico* dos genes da catalase em plantas de milho com a doença da mancha cinzenta

As plantas de milho pertencentes a cultivar Yayu889, resistente a GLS (do inglês *Grey leaf spot*), foram submetidas a infecção pelo fungo *Cercospora zea-maydis*. O

experimento foi composto por três replicatas biológicas. Os experimentos foram conduzidos com as plantas infectadas nos estágios 1, 2 e 3. O gene da CAT3 foi regulado positivamente e apresentou diferenças no perfil de expressão em todos os estágios avaliados, sendo mais expresso no estágio 3, já o gene *cat1* foi pouco expresso em todos os estágios e o gene *cat3* foi regulado negativamente em relação ao controle, em ambos os estágios avaliados (Figura 10).

Figura 10 - Perfil de expressão dos genes *cat1*, *cat2* e *cat3* de *Zea mays* sob ataque pelo fungo *Cercospora zea-maydis*. A escala de cor representa os valores de Log2fold-change obtidos após normalização dos dados. O asterisco (\*) representa um *p-value* < 0,05.



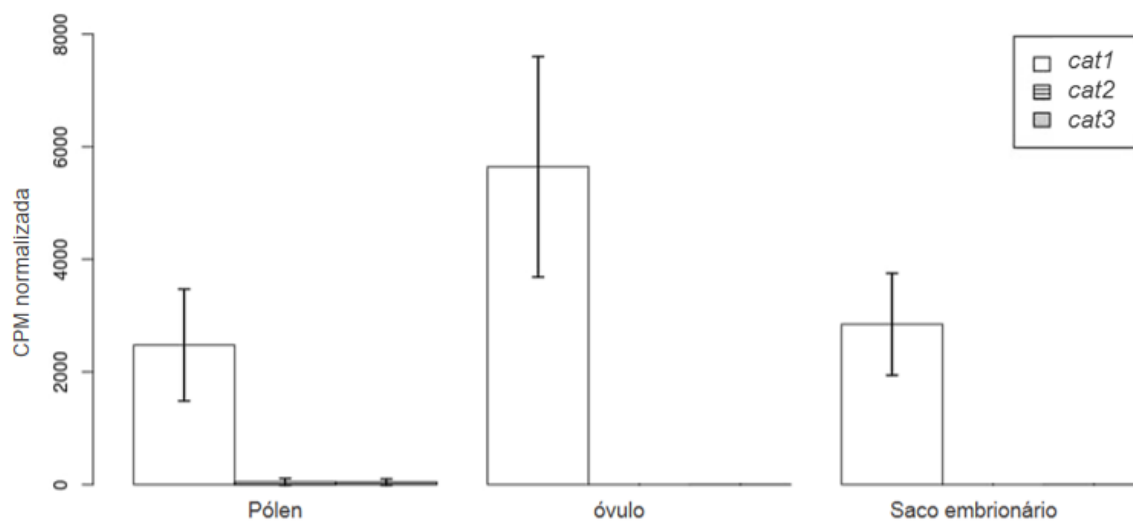
Fonte: dados da pesquisa

### 5.3.3 Expressão *in silico* dos genes da catalase em plantas de milho em diferentes tecidos reprodutivos

Nesse experimento plantas pertencentes a cultivar B73 foram usadas para o isolamento de RNA dos seguintes tecidos reprodutivos: óvulo e saco embrionário, isolados de estilo-estigmas (cabelos de milho) de aproximadamente 10 cm, e pólen. Após as análises, apenas o gene *cat1* foi predominantemente expresso em todos os tecidos estudados, contudo não houve diferença significativa entre os tecidos ( $p < 0,05$ ). Os demais genes apresentaram baixas contagens nos tecidos avaliados, não sendo observado expressão (Figura 11).



Figura 11 - Perfil de expressão dos genes *cat1*, *cat2* e *cat3* de *Zea mays* em diferentes tecidos reprodutivos. A análise estatística (*one-way* ANOVA seguido pelo teste de Bonferroni) foi aplicada para cada gene separadamente.

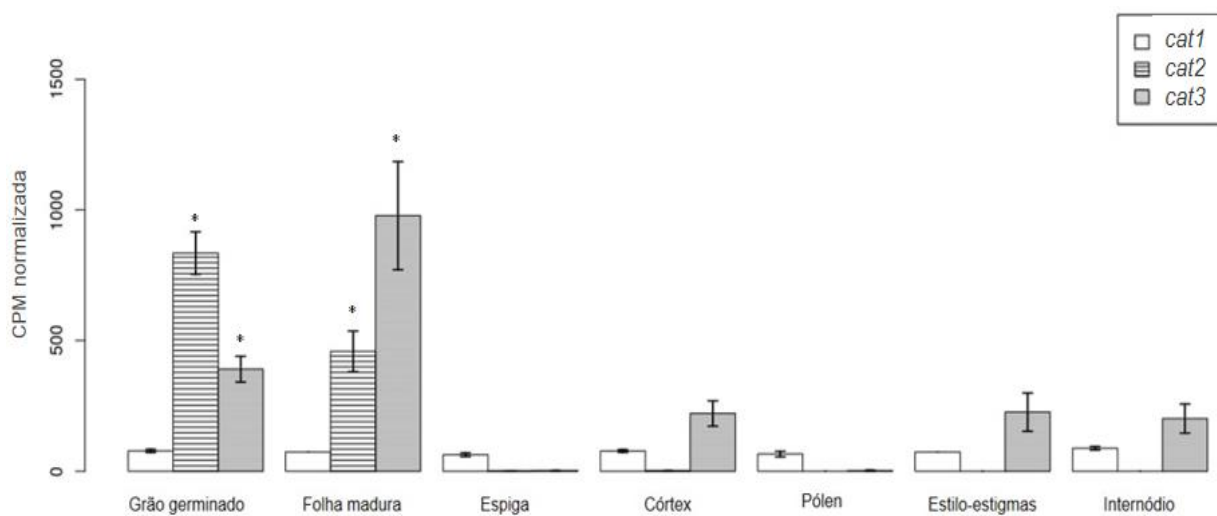


Fonte: dados da pesquisa

#### 5.3.4 Expressão *in silico* dos genes da catalase em plantas de milho em diferentes tecidos

Para este experimento, o RNA foi extraído dos seguintes tecidos: Internódio, pólen, estilo-estigmas (cabelos de milho), córtex, espigas, folha madura e embriões (20 DAP – dias após a polinização). Na folha madura e no grão germinado os genes da *cat2* e *cat3* foram mais expressos, sendo estatisticamente diferentes a 0,05% de probabilidade em ambos os tecidos. Em relação aos outros tecidos analisados, espiga, pólen, córtex, estilo-estigmas e internódio foi detectado um valor de contagem baixo, não mostrando assim, um nível de expressão relevante.

Figura 12 - Perfil de expressão dos genes *cat1*, *cat2* e *cat3* de *Zea mays* em diferentes tecidos. A análise estatística (*one-way* ANOVA seguido pelo teste de Bonferroni) foi aplicada para cada gene separadamente. As diferenças significativas em  $p < 0,05$  são indicadas por asterisco (\*).



Fonte: dados da pesquisa

## 6 DISCUSSÃO

### 6.1 Distribuição das subfamílias da catalase em angiospermas

No presente estudo, os genes da catalase foram identificados em espécies de angiospermas. Esses genes foram caracterizados quanto a estrutura éxon-ínton e em seguida, a expressão *in silico* com os genes *cat2*, *cat3* e *cat1* da espécie *Zea mays* foi avaliada sob diferentes condições de estresse e em diferentes tecidos durante o desenvolvimento.

A catalase se mostra como uma enzima bastante conservada no decorrer da evolução de plantas angiospermas, como destaca Guan e Scandalios (1996) no seu estudo analisando 16 sequências de catalase, dentre monocotiledôneas e eudicotiledôneas. Na maior parte das espécies aqui estudadas, três genes foram identificados. Nas monocotiledôneas, um padrão de três genes vem sendo encontrado em trabalhos prévios com arroz e milho, sendo essas as espécies mais estudadas (MORI et al., 1992; MORITA et al., 1994; SCANDALIOS; GUAN; POLIDOROS, 1997; JOO; LEE; SONG, 2014; VIGHI et al., 2016).

Contudo, apesar do padrão de três genes ser o mais encontrado nas espécies estudadas até o momento, em algumas espécies foram identificados um número maior de genes, que é o caso das espécies da ordem zingiberales (*Musa acuminata* e *Ensete ventricosum*, com quatro genes), Apiales (*Daucus carota*, com seis genes), Cucurbitales (*Cucurbita moschata*, com quatro genes), Fabales (*Glycine max*, com quatro genes), Malpighiales (*Manihot esculenta*, com seis genes), Brassicales (*Brassica oleraceae* e *Raphanus sativus*, com sete genes e cinco, respectivamente), Gentianales (*Coffea arabica*, com quatro genes), Solanales (*Ipomoea nil*, com quatro genes) e Asterales (*Helianthus annuus*, com cinco genes) (Figura 3 e ANEXO A).

Nas Zingiberales, os quatro genes da catalase foram nomeados de *cat3a/cat3b* e *cat1a/cat1b* (Figura 1) sendo assim, considerados dois grupos de genes parálogos. Essa duplicação parece recente na história evolutiva dessa ordem, tendo em vista que os parálogos guardam uma identidade entre eles de cerca de 90%. Nesta ordem, a CAT2 não está presente, assim como em algumas espécies de Poales, como *Triticum urartu*, *Saccharum hybrid cultivar* e *Secale cereale* (Figura 3). Trabalhos na literatura também apontam apenas esses dois genes nessas duas primeiras espécies citadas (SARUYAMA; MATSUMURA, 1999; SU et al., 2014; LIU et al., 2015). Parece que a

perda da subfamília CAT2 ocorreu primeiro em Poales e se manteve ao longo evolução das comelinídeas.

Acredita-se que nas espécies da ordem Poales citadas acima, esse gene foi perdido no decorrer da evolução. Investigações realizadas quanto a um papel fisiológico para o gene *cat2*, em espécies com os três genes (milho e arroz), mostram que esse gene é responsivo a várias condições em que os genes *cat1* e *cat3* também foram expressos. O que se observa nos trabalhos analisados com milho e arroz, onde os genes *cat1* e *cat3* formam um grupo de ortólogos com o trigo (identidade acima de 95%) a ausência do gene *cat2*, parece ser compensada por esses outros dois genes. Ou seja, parece que a presença de três genes para a catalase não é um fator de primordial importância que dita um sistema de remoção de H<sub>2</sub>O<sub>2</sub> eficiente.

Além disso, trabalhos em *Arabidopsis thaliana* já mostram um mecanismo de compensação entre dois genes da catalase. Du et al. (2008) analisando linhas mutantes de *Arabidopsis* (plantas deficientes em catalase), mostram que quando o gene *cat2* (a nomenclatura entre as catalases de *Arabidopsis* e milho não são correspondentes) está ausente, os genes *cat1* e *cat3* conseguem compensar, melhorando os seus perfis de expressão. Ademais, Hu et al. (2010), destacam haver uma redundância funcional entre genes de *Arabidopsis* (*cat1* e *cat3*), portanto, esses genes podem desempenhar papéis similares.

Comumente as enzimas que compõem o sistema antioxidante são codificadas por famílias multigênicas. Essas enzimas possuem uma regulação complexa, além disso, geralmente, seus substratos desempenham um papel como molécula sinalizadora atuando em diversas vias. A compreensão da sua regulação coordenada parece algo extremamente complexo. Além disso, pelo próprio papel desempenhado pela catalase na célula e por fazer parte de um sistema antioxidante de limpeza, entende-se que a sua expressão pode ser desencadeada por uma ampla variedade de fatores (ZIMMERMANN et al., 2006).

Outro fator a se considerar a esse respeito é a própria localização subcelular da catalase, o que contribui para que cada uma das isoformas desempenhe papéis similares durante a exposição das plantas a estressores externos. Em seu trabalho Hu et al. (2010), destacaram a existência de redundância funcional entre dois genes da catalase em *Arabidopsis*. Contudo, a especialização funcional dessas isoformas codificadas não pode ser descartada, já que os estudos sobre caracterização e

expressão gênica com a catalase em plantas ainda são poucos, trazendo limitação acerca da discussão de tais temas.

Ainda, tratando-se de Poales, a espécie *Hordeum vulgare* possui apenas as subfamílias CAT2 e CAT1, tendo perdido a CAT3. Similarmente, na espécie *Zoysia japonica* esse gene também não foi encontrado. Em um estudo com os genes da catalase em *H. vulgare*, Skadsen, Schulze-Lefert e Herbst (1995), detectaram apenas dois genes que são correspondentes aos identificados neste estudo. Além disso, em algumas Poales encontra-se apenas 1 tipo de catalase, como é o caso de *Zizania latifolia*, com os genes da subfamília CAT1, nomeados *zcat1a1* e *zcat1a2* com 96% de identidade. Em outras monocotiledôneas, tais como, *Nelumbo nucifera* (*nncat1a1* e *nncat1a2*, com 91% de identidade) e *Ananas comosus* (*accat1a1* e *accat1a2*, com 99% de identidade) observou-se esse mesmo padrão (Figura 3). Tais observações podem ser reflexo de um processo evolutivo dinâmico desta família multigênica, onde nem sempre relações de grupos de ortólogos foram formados.

Nas espécies de eudicotiledôneas aqui estudadas, também foi identificado, em sua maioria, um padrão geral de três genes (Figura 1). Na literatura, esse modelo também é relatado em trabalhos com *Solanum lycopersicum* (KABIR; WANG, 2011), *Arabidopsis thaliana* (FRUGOLI et al., 1996; Du et al., 2008; ALAM; GHOSH, 2018), *Capsicum annuum* (LEE; AN, 2005), *Cucumis sativus* L. (HU et al., 2016), *Nicotiana tabaccum* (HAVIR; BRISSON; ZELITCH, 1996), *Nicotiana plumbaginifolia* (WILLEKENS et al., 1995a) e *Cucurbita* sp (ESAKA et al., 1997).

Nas espécies de eudicotiledôneas estudadas percebeu-se uma maior predominância de genes que podem ter sido originados de processos de duplicações gênicas recentes, tendo em vista a identidade elevada observada entre os membros gênicos (Figuras 3). Um caso bem típico pode ser observado na espécie *Glycine max*, onde foram identificados 4 genes, os quais estão dispostos em dois grupos de genes parálogos (Figura 3). Em geral, na ordem Fabales, esse foi o padrão observado nas plantas analisadas. Em *G. max*, esses genes estão duplicados, provavelmente em decorrência de eventos de duplicações do genoma que ocorreu há cerca de 59 e 13 milhões de anos atrás. Nesse sentido, aproximadamente 75% dos genes da soja estão presentes com múltiplas cópias (SCHMUTZ et al., 2010). Os genes da catalase encontrados em *G. max* foram renomeados como *gmcat4a1*, *gmcat4a2*, *gmcat4b1* e *gmcat4b2* (Figura 3).

Em *Daucus carota*, também se observa um número de genes acima do normalmente encontrado e assim como em *G. max*, parecem ser oriundos de duplicações recentes. Os seis genes identificados mostram elevada identidade entre eles (Figura 3). O que se sabe é que o genoma de *Daucus carota*, assim como de todas as eudicotiledôneas sofreu processos de duplicações, contudo Iorizzo et al. (2016) identificaram dois eventos de duplicação completa específicos da espécie, Dc- $\alpha$  e Dc- $\beta$ . Durante esse evento evolutivo, os segmentos genômicos contendo as sequências gênicas da catalase podem ter sido duplicadas e gerado o padrão de genes encontrado nesta espécie.

Assim como em *D. carota*, a espécie *Brassica oleracea* (couve) apresentou um número de genes para a catalase bastante fora do padrão, com um total de sete genes (Figura 3). No genoma desta espécie Liu e colaboradores (2014), constataram diversos rearranjos cromossômicos e perdas assimétricas de genes em blocos genômicos duplicados. Esses autores ainda identificaram um largo número de gene parálogos. Esses eventos genômicos ocorridos nessas espécies podem ser uma provável explicação para o número de genes da catalase encontrados.

Quanto a outras ordens estudadas no presente trabalho, tais como Rosales, Fabales, Sapindales, Vitales e Ericales, um total de dois genes foi identificado (Figura 1). Dentre essas ordens, somente um estudo com *Prunus persica* (pêssego), pertencente a ordem Rosales, foi encontrado na literatura. Nesse estudo, os autores também relataram a presença de apenas dois genes (BAGNOLI et al., 2004).

De forma geral, parece haver uma certa divergência quanto ao número de genes encontrados nas espécies estudadas, principalmente em eudicotiledôneas (Figura 1). Vale salientar que poucos trabalhos na literatura com a catalase abordaram um número tão grande e diverso de espécies. Conforme foi observado existem variações intra-específicas e também inter-específicas.

Conforme apontam Flagel e Wendel (2009), o número de membros gênicos de uma família multigênica sofre flutuações frequentes entre táxons, sendo que a ampla diversidade genética encontrada em plantas terrestres não poderia ser sustentada por um conjunto igualmente diverso de novos genes. Além disso, como foi bastante observado em várias espécies estudadas, diversos genes compartilham uma identidade muito elevada entre eles. Isso pode apontar um provável funcionamento em conjunto desses genes, caracterizando o processo de subfuncionalização (PANCHY; LEHTI-SHIU; SHIU, 2016).

## 6.2 Catalase: sugestão de uma nova classificação em angiospermas

Muito comumente encontrado na literatura, diversos trabalhos tentam classificar genes e/ou proteínas, de forma tanto a propor um sistema de classificação usual, quanto de forma a uniformizar a nomenclatura que geralmente se encontra confusa.

Com a catalase, já existem algumas nomenclaturas usuais (MORI et al., 1992; MORITA et al., 1994; WILLEKENS et al., 1994b). Contudo, de uma forma geral, em uma boa parte dos trabalhos publicados, a nomeação dos genes é realizada a partir de alinhamentos com sequências de catalase (gene ou proteína) com nomenclatura definida, como é o caso de *Arabidopsis* e milho. Contudo o uso dessa metodologia para nomear os genes/proteínas da catalase de outras espécies, tornou a compreensão dos trabalhos realizados uma tarefa de difícil execução, pois há ausência de correspondência entre os genes de diferentes espécies.

O que se pretende aqui é gerar uma classificação baseada em identidade de sequência de forma a refletir as relações filogenéticas entre elas. Trabalhos com essa temática já são bastante úteis para auxiliar na execução de novos estudos, facilitando a comunicação no meio científico, livre de contradições e fornecendo informações valiosas para o campo da genômica comparativa (COSTA et al., 2014; LIU et al., 2015; FINET et al., 2016).

A classificação proposta nesse trabalho foi feita com base na nomenclatura encontrada na espécie *Zea mays*, em decorrência da disposição dos cDNAs dessa espécie na árvore (Figura 3). Em um estudo realizado por Frugoli e colaboradores (1998), analisando as relações filogenéticas entre sequências gênicas da catalase (39 sequências), esses autores relataram a existência de três subclados tomando a espécie *Z. mays* como referência, algo que não se observa quando se leva em considerações espécies modelos, como *A. thaliana*. Ademais, vários trabalhos também relatam a disposição das catalases do milho em árvores filogenéticas em clados distintos, conforme encontrado no presente trabalho, apontando os genes da catalase dessa espécie como sendo úteis para gerar uma classificação (GUAN; SCANDALIOS 1996; FRUGOLI et al., 1996; KLOTZ; KLASSEN; LOEWEN, 1997; BAGNOLI et al., 2004).

Os agrupamentos na árvore gerada, conforme mostrado na figura 3, são similares aos agrupamentos formados no trabalho de Guan e Scandalios (1996).

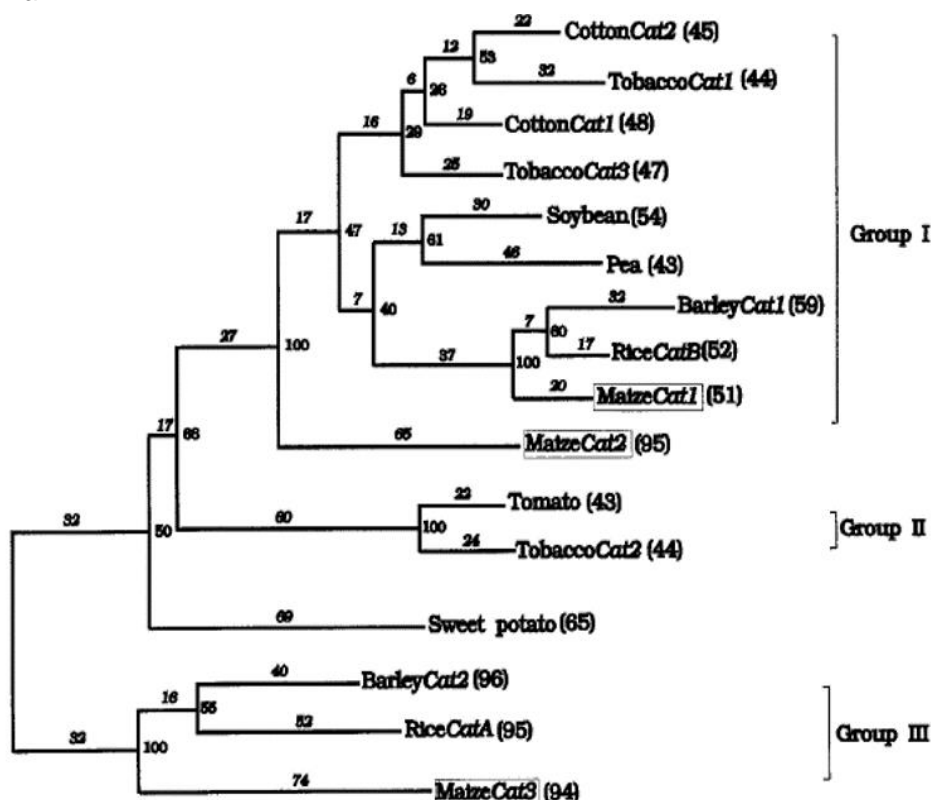
Esses autores também usaram como partida as catalases de *Zea mays* para nomeação das subfamílias da catalase identificadas nas espécies estudadas. Assim como discutido por esses autores, eles defendem a existência, com base nos agrupamentos formados, de quatro subfamílias da catalase, conforme também foi constatado no presente trabalho (Figura 3).

Segundo Guan e Scandalios (1996), conforme mostrado na figura 13, adaptada dos autores, existe uma subfamília da catalase compartilhada por monocotiledôneas e eudicotiledôneas (Grupo 1), outra subfamília encontrada somente em eudicotiledôneas (Grupo 2) e outra subfamília encontrada somente em monocotiledôneas (Grupo 3). Esses autores também relatam a existência de uma outra subfamília da catalase encontrada apenas em *Z. mays*, fracamente relacionada com o primeiro grupo. Os autores declaram que a posição desta última catalase não se mostra clara através da reconstrução filogenética. Uma provável explicação é o baixo número de sequências usadas na análise, não contando com um número suficiente e diverso de espécies para se poder encontrar sequências com elevada identidade com a CAT2.

A partir da constatação, através de alinhamentos realizados, que as sequências usadas pelos autores acima são as mesmas usadas no presente estudo, observa-se que os agrupamentos gerados pelos autores são, de certa forma, equivalentes aos obtidos no presente estudo.



Figura 13 - Árvore filogenética baseada em seqüências de aminoácidos gerada a partir do método da parcimônia.



Fonte: Adaptada de Guan e Scandalios 1996.

Observando os grupos formados no trabalho dos autores acima, adotando a nomenclatura aqui proposta, teríamos a seguinte distribuição: CAT1 (grupo I), CAT2 (grupo isolado composto por apenas a CAT2 de milho, sendo que a sua posição não é clara, em decorrência do pouco número de seqüências analisadas), CAT3 (grupo III) e CAT4, (formado por uma seqüência de *Ipomoea batatas* (em inglês, sweet potato) e mais duas seqüências, uma de tomate e a outra de tabaco, identificado no presente estudo com a subfamília da CAT4). O que corrobora com a existência das quatro subfamílias da catalase evidenciadas no presente trabalho e fortalece a classificação proposta.

Contudo, ainda sobre o trabalho de Guan e Scandalios (1996), cabe ressaltar que a análise filogenética usada conta com apenas 16 seqüências e não utiliza uma árvore enraizada (ausência de *outgroup*), de forma a se estabelecer uma relação com uma seqüência de um organismo mais antigo e se determinar a forma mais primitiva da catalase em plantas. Apesar dessas duas desvantagens do estudo, ainda assim, percebe-se uma relação de concordância com os dados aqui mostrados.

Klotz, Klassen e Loewen (1997), alcançaram resultados similares aos dos autores acima e aos dos resultados do presente trabalho. Todavia, esses autores relatam a CAT2 de *Zea mays* como um quebra-cabeça, pois esta sequência sempre está situada isoladamente em um clado, assim como relatado em outros estudos. Contudo, em análises dessa sequência gênica, eles concluem que a CAT2 está mais proximamente relacionada com o grupo ao qual a CAT3 faz parte, como foi encontrado no presente estudo. Essa catalase se apresenta, em vários estudos, isolada de outras sequências, ou seja, nas árvores filogenéticas ela não agrupa com outras sequências. Isso se dá, pelo fato de que, nesses estudos, outras sequências de CAT2 não são avaliadas, em decorrência da não caracterização dessas sequências no ano de publicação dos artigos.

Dessa forma, além da nomeação dos genes tomando como base a nomenclatura do milho, essa mesma espécie, possui uma maior quantidade de dados na literatura que podem vir a auxiliar numa compreensão mais profunda de uma função para os genes da catalase. Dessa forma, o milho veio a ter um papel chave para se definir uma classificação para a catalase.

Cabe salientar aqui que algumas correspondências foram encontradas entre a classificação aqui proposta e a de Willekens e colaboradores (1994b). As catalases de *Oryza sativa* e *Zea mays*, se encaixam em ambas as classificações. Novas investigações devem ser feitas para verificar se a exemplo das espécies acima mencionadas, as monocotiledôneas no geral compartilham essa característica. Todavia, em eudicotiledôneas o mesmo não foi observado.

### **6.3 Estrutura éxon-ínton dos genes da catalase: um padrão evolutivo**

A estrutura de éxons e íntrons também foi avaliada com vistas a se buscar evidências para uma compreensão da evolução molecular dos genes da catalase. Além de se realizar uma análise filogenética com as sequências, a estrutura de éxons e íntrons fornece informações que podem ser úteis para se traçar um perfil evolutivo desses genes durante a evolução de plantas angiospermas.

Uma maior divergência no número de éxons foi encontrada em monocotiledôneas, principalmente nos genes que codificam para a CAT2 e CAT3, específicas desta classe (Figuras 4 e 5). Este resultado também foi encontrado por Iwamoto et al. (1998), comparando a estrutura de genes em doze sequências

genômicas de seis espécies vegetais, onde os genes para a catalase em milho, arroz e trigo foram diferentes. Estes autores também relataram uma conservação no número de éxons nas eudicotiledôneas avaliadas (*Ricinus communis*, *Arabidopsis thaliana*, *Glycine max* e *Solanum tuberosum*) de oito éxons, também observado neste trabalho, onde 80% das sequências possuem 8 éxons, pertencendo a subfamília CAT4 (Figura 7).

No presente trabalho, também foi evidenciado a conservação de certos éxons nas sequências identificadas no geral. O primeiro e o último éxon são muito conservados no tamanho. O primeiro éxon com 15 nucleotídeos (codificando uma sequência de cinco aminoácidos) e o último éxon com sessenta nucleotídeos (codificando uma sequência de vinte aminoácidos). Todavia, a sequência desses éxons confere uma alta variabilidade entre as sequências de aminoácidos que codificam. No estudo de Iwamoto et al. (1998), esses autores também verificaram essa conservação no tamanho da sequência peptídica codificada por tais éxons. Além disso, neste último éxon, mais próximo a posição 3'UTR, estão as sequências conhecidas como PTS (*Peroxisomal Targeting Signal*), sequências de tripeptídeos, na maioria das vezes, posicionadas na região C-terminal, que direcionam a proteína para o peroxissomo e que são altamente variáveis entre espécies (MHAMDI et al., 2010).

Os resultados encontrados por Guan e Scandalios (1996) e Iwamoto et al. (1998), suportam a hipótese de que depois da divergência evolucionária de monocotiledôneas e eudicotiledôneas, consecutivas duplicações de genes, seguida pela perda diferencial de íntrons, ocorrendo diferencialmente em cada classe, conduziu a formação de três genes (ou até quatro genes em eudicotiledôneas).

Em um estudo realizado por Koralewski e Krutovsky (2011) o número de éxons, assim como o número de genes, aumenta de acordo com a complexidade da espécie. Ademais, esses autores relatam que, geralmente éxons mais curtos, o que foi bastante observado nas espécies de eudicotiledôneas, aumentam o potencial para a *Splicing* alternativo, mecanismo de grande importância que gera múltiplos produtos gênicos a partir de um único gene, incrementando a diversidade de proteínas num organismo.

No decorrer da evolução, os genes que codificam para as subfamílias da catalase específicas de monocotiledôneas (CAT2 e CAT3) (Figuras 4 e 5), mostram uma estrutura gênica composta por poucos éxons (perda de íntrons e éxons maiores), com tamanhos variados. Conforme relata Guan e Scandalios (1996), em

monocotiledôneas, há em geral, uma tendência durante a evolução de uma redução no número de íntrons, como foi relatado nos genes *cat2* e *cat3*, específicos dessa classe. O gene da CAT2, com maior proximidade com o grupo externo, denotando ser uma forma mais primitiva da catalase, possui poucos éxons. Todavia, observa-se um terceiro éxon com um comprimento em torno de 1000 nucleotídeos, comum a todas as sequências (Figura 4). Esses genes também não apresentaram elevada identidade com outras sequências, tendo identidade alta apenas entre eles.

O gene que codifica para a CAT3 já possui uma maior diversidade no número de éxons. (Figura 5). Em alguns genes desse clado, o segundo éxon ainda mantém um número de nucleotídeos em torno de 1000 pb, Essa característica parece ser intrínseca de espécies da ordem Poales, pois praticamente todos os genes da CAT3 nas espécies dessa ordem apresentam essa característica, somente algumas espécies, tais como *Leersia perrieri*, *Oryza sativa*, *Oropetium thomaeum* e *Secale cereale*, esse éxon parece ter sofrido uma fragmentação (inserção de íntrons). Esse éxon também é conservado nos genes da CAT2 (encontrada somente em Poales) (Figura 4). Em quaisquer outros membros dessas duas subfamílias, assim como das demais subfamílias (CAT1 e CAT4), esta característica não foi observada (Figuras 6 e 7).

No gene da CAT1, conservação no número de éxons com um número maior, são as características mais marcantes (Figura 6). Nessas sequências, no segundo e no quarto éxon, observa-se alta conservação, com 97 e 777 nucleotídeos, respectivamente. Esta última observação também é válida para o gene da CAT4 (Figura 7).

Tem sido bastante considerado que a evolução de genes em eucariotos tem ocorrido a partir de perda e ganho de íntrons (ROGOZIN et al., 2005; ZHANG et al., 2014). Geralmente, a perda de íntrons é um fenômeno mais frequente em curtas distâncias evolucionárias, enquanto inserções de íntrons acompanham transições evolucionárias maiores, como é o caso do evento de separação de monocotiledôneas e eudicotiledôneas. Isso reforça o nosso achado da maioria dos genes da catalase em eudicotiledôneas apresentarem um maior número de íntrons do que em monocotiledôneas.

As análises feitas suportam a hipótese que a estrutura éxon-íntron da família multigênica da catalase é muito dinâmica e que a evolução deste gene pode ter envolvido numerosas perdas e/ou ganho de íntrons. Além do mais, ao analisarmos o

perfil éxon-íntron dos genes de acordo com a sua história evolutiva, percebe-se além do que já foi mencionado acima, que o padrão de inserções de íntrons ocorreu com a introdução de íntrons no terceiro éxon da CAT2, gerando os perfis observados em algumas sequências da CAT3 e mais tipicamente nas sequências da CAT1 e CAT4. Esse fenômeno, onde os íntrons são inseridos dinamicamente, geralmente no centro de um éxon, foi demonstrado por Wang e Stein (2013) analisando a estrutura éxon-íntron de organismos eucariotos de diferentes espécies, incluindo as espécies vegetais *Oryza sativa*, *Arabidopsis thaliana*, *Sorghum bicolor* e *Zea mays*.

Esses autores também relatam que esse processo de inserção é proporcional ao tamanho do éxon. Como esse terceiro éxon no gene *cat2* apresenta um tamanho de cerca de 1000 nucleotídeos (Figura 4), sendo o maior éxon caracterizado nas sequências de catalase, os perfis das próximas catalases na escala evolutiva, demonstram esse éxon interrompido por várias sequências de íntrons. Este processo de inserção do íntron é equivalente a fragmentação do éxon (*splitting*).

#### **6.4 História evolutiva da Família multigênica da catalase em Angiospermas**

As catalases presentes em plantas, no geral, são as catalases monofuncionais ou comumente conhecidas como catalases típicas, encontradas também em alguns procariotos, fungos e em animais. Evolutivamente, essa catalase parece ter surgido em procariotos, inicialmente, e se mantido ainda em cianobactérias, contudo, têm sido relatadas como incomuns nesses últimos organismos (BERNROITNER et al., 2009). Todavia, na espécie de cianobactéria filamentosa *Nostoc punctiforme* tem sido encontrado uma sequência de uma catalase monofuncional, contudo seu papel fisiológico ainda é desconhecido (PASSARDI et al., 2007).

Nas algas, algumas sequências de catalases monofuncionais já têm sido caracterizadas, como é o caso da alga verde *Chlamydomonas reinhardtii* (KATO et al., 1997) e da alga vermelha marinha *Pyropia yezoensis* (LI et al., 2012). Ambas contam com apenas um gene, contudo em *C. reinhardtii* esse gene parece codificar para mais de uma isoforma. Nessa espécie, somente uma sequência gênica foi identificada, assim também como em *Volvox carteri*, onde ambas as espécies de algas compunham o grupo externo (Figura 3).

Na espécie de pinheiro *Pinus taeda* (gimnosperma) várias isoformas foram identificadas em gel não-desnaturante, (MULLEN; GIFFORD, 1993), contudo nenhuma sequência gênica foi caracterizada até o momento.

Em decorrência dos poucos estudos voltados para a caracterização de genes das catalases em outros organismos, é difícil se estimar um perfil evolutivo começando de um provável ancestral até as plantas atuais. Contudo, a partir das evidências dispostas nesses estudos, é provável que o surgimento das três subfamílias de catalases, mais comumente encontrada em plantas, tenha ocorrido depois da divergência de algas e plantas. Um provável ancestral teria apenas um gene dessa catalase monofuncional e através de subseqüentes duplicações, no decorrer da evolução, deu-se o surgimento das três (agora com a redefinição para quatro) subfamílias em plantas. Esse perfil de evolução da catalase também é defendido nos trabalhos de Von Ossowski, Hausner e Loewen (1993) e Klotz, Klassen e Loewen (1997).

É bem provável que as algas venham a ter apenas um gene da catalase típica por também comportar outros tipos de catalases, como a KcatG (catalase bifuncional) e as MnCATs (Manganês-catalases, não possuem grupo heme), diferente das plantas que possuem somente a família das catalases monofuncionais.

Em conjunto, as relações filogenéticas e a arquitetura dos genes, fornecem pistas para a determinação de uma história evolutiva desta enzima em plantas angiospermas. A CAT2 pode ser considerada uma forma mais primitiva da catalase, encontrada em monocotiledôneas, restrita a Poales. Essa catalase, sofreu duplicação, originando a CAT3, compartilhada por monocotiledôneas, abrangendo bem mais ordens. A CAT3 então, duplicou, originando a CAT1, antes da divergência entre monocotiledôneas e eudicotiledôneas. Em seguida, a CAT1 presente em eudicotiledôneas, sofreu duplicação, originando a CAT4, uma sequência da catalase mais recente na evolução e ainda com uma forte relação com a CAT1, tendo sido um evento de duplicação recente na evolução. Após essas respectivas duplicações, houve uma provável especialização adaptativa das catalases.

## 6.5 Papel fisiológico dos membros gênicos da catalase

A catalase desempenha um importante papel no sistema de defesa antioxidante por degradar  $H_2O_2$  em  $H_2O$  e  $O_2$ . Diversos estudos têm sido realizados até o momento, principalmente nas espécies *Zea mays*, *Oryza sativa*, *Arabidopsis thaliana* e *Nicotiana tabacum*, investigando o papel fisiológico da catalase. Todavia, o presente trabalho vem acrescentar a esse campo de estudos, informações de expressão *in silico* dos genes da catalase a partir da análise de dados de RNA-Seq da espécie *Zea mays*.

Sob estresse osmótico, a análise dos dados revelou o gene da CAT2 sendo regulado positivamente nas raízes (estela e a zona de alongamento) (Figura 9). A estela está relacionada a funções no transporte de água, nutrientes e fotossintatos. Já a zona de alongamento na raiz, localizada acima da zona meristemática ou de proliferação, é onde a divisão celular cessa e há um rápido crescimento celular por alongamento.

Na raiz, esse gene ainda não tinha sido detectado. Todavia, os trabalhos que analisam a expressão de genes da catalase nesse órgão ainda são escassos. Contudo, a raiz é um órgão chave no estudo de expressão gênica em condições de seca (estresse osmótico), em decorrência das características morfológicas e anatômicas que permitem a aquisição de água e nutrientes do solo. Em grande parte dos estudos transcriptômicos realizados em raiz, os genes relacionados a enzimas antioxidantes, tais como APX, GR, SOD e CAT, geralmente são detectados (JANIÁK; KWAŚNIEWSKI; SZAREJKO, 2015). Além disso, APX e CAT são as principais peroxidases responsáveis pela manutenção dos níveis de  $H_2O_2$  adequados nas raízes (CUNHA et al., 2016).

Ademais, esse gene, em plantas de milho, foi detectado em resposta a aplicações exógenas de ABA em embriões quando em estágios tardios da embriogênese (GUAN; SCANDALIOS, 1998; POLIDOROS; MYLONA; SCANDALIOS, 2001). Na literatura, esse gene presente em linhas transgênicas de tabaco, tem tornado as plantas resistentes à estresses oxidativos (POLIDOROS; MYLONA; SCANDALIOS, 2001).

Em arroz, o seu ortólogo OsCATC, também tem sido expresso em plantas transgênicas de tabaco, tornando-as resistentes ao estresse osmótico. Também tem sido relatado a sua expressão durante desidratação e salinidade, estando relacionado a fotorrespiração (IWAMOTO; HIGO; HIGO, 2000; YAMANE et al., 2010; ALAM;

GHOSH, 2018). Na espécie *Saccharum officinarum* (cana de açúcar), quando transformada com a proteína recombinante do gene *cat2*, esse gene revelou ser *up-regulated* em diferentes estresses, como estresse oxidativo (H<sub>2</sub>O<sub>2</sub>), metal pesado (CuCl<sub>2</sub>), hiper-osmótico (PEG e sal) e hormônios vegetais (AS, ABA e MeJA) (SU et al., 2014).

Este gene também foi expresso em tecidos foliares (folha madura), como já foi identificado anteriormente em milho e em arroz (SCANDALIOS et al., 1984; IWAMOTO; HIGO; HIGO, 2000) e em grãos germinados (Figura 12). Em milho e em arroz, este gene foi destacado por apresentar expressão em sementes germinadas. Em *Hordeum vulgare* (cevada) ele foi encontrado em sementes em maturação (SKADSEN; SCHULZE-LEFERT; HERBST, 1995; GUAN; SCANDALIOS, 2000; IWAMOTO; HIGO; HIGO, 2000).

A catalase é uma enzima geralmente encontrada em sementes, já que ela participa ativamente do ciclo do glioxissomo, muito ativo em sementes. Neste ciclo, os ácidos graxos armazenados são convertidos em carboidratos para posterior utilização e manutenção do embrião (BAILLY et al., 2004). A CAT2 se encontra predominante nesses órgãos e, além disso, responde a vários estresses ambientais. Dessa forma, em espécies da família Poaceae (ela foi encontrada apenas em algumas espécies dessa família), que estão amplamente dispersas no globo terrestre, ela poderia atuar no sistema de limpeza antioxidante da semente, contribuindo para a homeostase, mesmo em condições adversas, já que essas espécies são encontradas nos mais variados ambientes, sendo exposta a variados estresses que poderiam afetar drasticamente a sobrevivência dessas plantas.

Em ataque por fungo, o gene *cat3* teve expressão diferencial, sendo regulado positivamente (Figura 10). Esse gene também foi muito expresso, em um estudo realizado com a toxina cercosporina purificada desse mesmo fungo. Os níveis de mRNA aumentaram linearmente junto com os níveis da enzima CAT3 (WILLIAMSON; SCANDALIOS, 1992). Além disso, esses autores relataram que os níveis de mRNA aumentaram de acordo com a dose de cercosporina. O ortólogo do gene *cat3* em *O. sativa* mostrou regulação positiva em resposta aos patógenos *Meloidogyne graminicola* e *Nilaparvata lugens* (ALAM; GHOSH, 2017).

Ainda nesse gene, quatorze elementos da família MYC foram detectados (Figura 08), sendo descritos em *Arabidopsis* por desempenhar um papel na sinalização via ácido jasmônico (POZO et al., 2008). Contudo, estudos sobre a catalase relacionada



a defesa contra patógenos ainda são escassos e de uma forma geral, o mecanismo pelo qual isso acontece ainda é desconhecido.

O gene da CAT3 foi encontrado expresso em folhas (Figura 12), como anteriormente destacado por Guan e Scandalios (1998). Em estudos realizados com esse gene em milho, ele tem sido modulado por diferentes concentrações de ABA (GUAN; SCANDALIOS, 1998). Em arroz, ele tem mostrado expressão em genótipos tolerantes a salinidade e expressão em condições de frio. Em trigo, este gene é expresso em condições de estresse osmótico (LUNA et al., 2004), em banana, expresso em resposta a baixas temperaturas e dano físico (FIGUEROA-YÁÑEZ et al., 2012) e em *Zostera marina*, esse gene foi regulado positivamente sob baixas temperaturas (<10°C) (ZANG et et., 2018).

Além disso, na análise da região promotora do gene *cat3*, foi observado o maior número de elementos *cis* responsivos a estresses, cerca de 29 elementos, no geral, com uma predominância da família MYC e WRKY (Figura 8), sendo este último envolvido em diversas respostas associadas à estresses abióticos/bióticos (PHUKAN; JEENA; SHUKLA, 2016). Ademais, ambos os genes da CAT3 em milho, arroz e em trigo, têm sido modulados por ritmo circadiano (LUNA et al., 2004; REDINBAUGH; SABRE; SCANDALIOS, 1990; IWAMOTO; HIGO; HIGO, 2000). Mais estudos são necessários para se confirmar se essa é uma característica intrínseca das catalases da subfamília CAT3.

Em relação ao gene da CAT1, foi detectado expressão apenas em tecidos reprodutivos (Figura 11), apresentando expressão baixa nos demais tecidos analisados e respondendo negativamente aos estresses avaliados. Contudo, na literatura este gene tem sido expresso em uma variada gama de tecidos de milho, tais como: endosperma leitoso, aleurona, pericarpo, sementes em desenvolvimento, raízes, meristema e escutelo em plantas jovens (SCANDALIOS et al., 1984). Segundo Menezes-Benavente et al. (2004), este gene foi induzido por sal e de acordo com Guan e Scandalios (1998, 2000) este gene é modulado por ABA, estresse osmótico e desidratação.

Em arroz, o gene *cat1* se mostra expresso em sementes e raízes, bem como fortemente induzido por ABA e sob estresse osmótico, salinidade e frio (JOO; LEE; SONG, 2014; VIGHY et al., 2016). Ademais, nesse gene apenas os elementos *cis* da família MYB, foram encontrados em maior abundância, cerca de 10 elementos *cis*

(Figura 8). Esta família geralmente está associada a respostas a estresses abióticos, tais como seca, desidratação e à baixas temperaturas (YOU et al., 2015).

Conforme destaca Guan e Scandalios (1996), a CAT1 funcionaria como uma catalase básica, com funções voltadas para a detoxificação do H<sub>2</sub>O<sub>2</sub> nas células em diferentes tecidos em algumas condições de estresse, desempenhando uma função geral. Já a CAT2 e a CAT3 são altamente reguladas e mostram especificidade tecidual, atuando para funções adicionais e diversificadas, respondendo a uma ampla gama de estímulos. Contudo, a partir dos estudos encontrados na literatura até hoje, o perfil geral da catalase é mais complexo do que se pensava. Ambas as isoformas apresentam expressão em vários tecidos e respondem a estresses.

É provável o surgimento dos três genes a partir de eventos de duplicação gênica, tendo ambos se mantido por pressões ambientais, durante evolução aeróbica. Observa-se uma diversificação funcional nas isoformas que pode estar atrelada a mudanças nos elementos regulatórios, que podem ser resultado de uma seleção natural do ambiente. Contudo, a manutenção de genes duplicados que respondem a estresses, principalmente que estejam relacionados a metabolismo oxidativo, pode fornecer diversas vantagens para a planta.

Para os genes da subfamília CAT4, presentes na maioria das espécies avaliadas, cerca de 88% das sequências identificadas, percebe-se também uma grande diversificação da função (Figuras 1 e 3). Em plantas como *Solanum lycopersicum*, os genes que codificam para CAT4a e CAT4b já mostram papéis funcionais diversos. Segundo Kabir e Wang (2011), o gene *cat4a* está relacionado a papéis de detoxificação do peróxido sob vários estresses e ainda atuando na homeostase em tecidos foliares, enquanto que o gene *cat4b* está associado a fotorrespiração e em adição a isso, este gene também responde positivamente a condição de salinidade.

Na espécie *Ricinus communis*, ambos os genes *cat4a* e *cat4b* são expressos diferencialmente. O gene *cat4a* é expresso abundantemente no endosperma e no cotilédone, enquanto o *cat4b* é expresso em diversos órgãos (cotilédones, endosperma, folhas, raiz e hipocótilos) (SUZUKI et al., 1994).

Os genes citados acima (*cat4a* e *cat4b* em *Solanum lycopersicum* e em *Ricinus communis*) compartilham identidade entre eles acima de 80% e apesar de estarem incluídos dentro de uma mesma subfamília, apresentam diferentes funcionalidades. Essas diferenças na resposta de genes da catalase, pertencentes a uma mesma

subfamília (no caso, CAT4), também foram observadas nas espécies *Prunus persica* e *Capsicum annuum* (BAGNOLI et al., 2004; LEE; AN, 2005).

Ademais, é difícil se inferir uma relação entre os agrupamentos formados e as funções dos membros gênicos da subfamília da catalase, conforme já relatado por Frugoli et al. (1996, 1998). Muitas lacunas ainda existem acerca de vários assuntos relativos a esta família multigênica, tais como localização celular, regulação e a própria função, o que limita a extrapolação dos dados e a especificar características de cada membro, nos levando a tirar conclusões mais gerais. Todavia, o presente trabalho acrescenta informações relevantes a essa temática, principalmente no que concerne à evolução e classificação dos genes, servindo de base para se conduzir novos estudos com essa família multigênica e avançar mais para se chegar a especificação de um papel fisiológico para cada membro gênico.

## 7 CONCLUSÃO

- O número de genes que codificam para a catalase em angiospermas varia de um a sete genes, a depender da espécie e da ordem a qual ela pertence;

- A estrutura éxon-íntron dos genes varia a depender da subfamília a qual ele pertence. No geral, os genes da catalase apresentam a seguinte estrutura éxon-íntron predominante: *cat1* (8 éxons e 7 íntrons), *cat2* (6 éxons e 5 íntrons; 5 éxons e 4 íntrons), *cat3* (3 éxons e 2 íntrons; 8 éxons e 7 íntrons) e *cat4* (8 éxons e 7 íntrons). A evolução da estrutura dos genes foi acompanhada de eventos de deleção de íntrons;

- No decorrer da evolução a catalase sofreu uma dramática diversificação da função, atuando na detoxificação do H<sub>2</sub>O<sub>2</sub> em plantas sob condições de estresse. Os genes *cat2* e *cat3* são expressos em tecidos foliares e sementes e respondem positivamente a condições de estresse, já o gene *cat1* é expresso, predominantemente, em tecidos reprodutivos. Todos os três genes estudados possuem uma expressão dependente do tecido e do estágio de desenvolvimento da planta;

- Foi possível se estabelecer um sistema de classificação para os genes da catalase baseado na filogenia dos genes, englobando diversas espécies de angiospermas;

- Os resultados obtidos nesse trabalho são úteis para elucidar a distribuição evolutiva dos genes da catalase em angiospermas. Essas informações podem ser usadas para o estabelecimento de um sistema de classificação amplo, buscando uma inter-relação entre filogenia, estrutura e função gênica.

## REFERÊNCIAS

- ALAM, N. B.; GHOSH, A. Comprehensive analysis and transcript profiling of *Arabidopsis thaliana* and *Oryza sativa* catalase gene family suggests their specific roles in development and stress responses. **Plant Physiology and Biochemistry**, [S.l.], v. 123, p.54-64, fev. 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/29223068>. Acesso em: 07 out. 2018.
- ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research** v. 25, p. 3389-3402, 1997.
- ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biology**, [S.l.], v. 11, n. 10, p.1-12, 2010. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-10-r106>. Acesso em: 10 nov. 2018.
- ANDERS, S.; PYL, P.T.; HUBER, W. HTSeq: a Python framework to work with high throughput sequencing data. **Bioinformatics**, v. 31, p. 166-169, 2015.
- ANDREWS, S. FastQC: a quality control tool for high throughput sequence data. 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> Acesso em: 10 nov. 2018.
- ANJUM, N. A.; SHARMA, P.; GILL, S. S.; HASANUZZAMAN, M.; KHAN, E. A.; KACHHAP, K; MOHAMED, A. A; THANGAVEL, P.; DEVI, G. D.; VASUDHEVAN, P.; SOFO, A.; KHAN, N. A.; MISRA, A. N.; LUKATKIN, A. S.; SINGH, H. P.; PEREIRA, E.; TUTEJA, N. Catalase and ascorbate peroxidase - representative H<sub>2</sub>O<sub>2</sub>-detoxifying heme enzymes in plants. **Environmental Science and Pollution Research**, [S.l.], v. 23, n. 19, p.19002-19029, 23 ago. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27549233>. Acesso em: 16 out. 2018.
- ATKINSON. N. J.; URWIN. P. E. The interaction of plant biotic and abiotic stresses: from genes to the field. **Journal of Experimental Botany**, [S.l.], v. 63, n. 10, p.3523-3543, 30 mar. 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/22467407>. Acesso em: 15 ago. 2018.
- BAGNOLI, F.; DANTI, S.; MAGHERINI, V.; COZZA, R.; INNOCENTI, A. M.; RACCHI, M. L. Molecular cloning, characterization and expression of two catalase genes from peach. **Functional Plant Biology**, [S.l.], v. 31, n. 4, p.349-357, 2004. Disponível em: <http://www.publish.csiro.au/FP/FP03203>. Acesso em: 11 out. 2018.
- BAILLY, C.; LEYMARIE, J.; LEHNER, A.; ROUSSEAU, S.; CÔME, D. CORBINEAU, F. Catalase activity and expression in developing sunflower seeds as related to drying. **Journal of Experimental Botany**, [S.l.], v. 55, n. 396, p.475-483, 22 jan. 2004. Disponível em: <https://academic.oup.com/jxb/article/55/396/475/489098>. Acesso em: 08 nov. 2018.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289–300, 1995.

BERNROITNER, M.; ZAMOCKY, M.; FURTMULLER, P. G.; PESCHEK, G. A.; OBINGER, C. Occurrence, phylogenetic structure, and function of catalases and peroxidases in cyanobacteria. **Journal of Experimental Botany**, [S.l.], v. 60, n. 2, p.423-440, 6 jan. 2009. Disponível em: <https://academic.oup.com/jxb/article/60/2/423/632000>. Acesso em: 24 out. 2018.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, [S.l.], v. 30, n. 15, p.2114-2120, 1 abr. 2014. Disponível em: <https://academic.oup.com/bioinformatics/article/30/15/2114/2390096>. Acesso em: 10 nov. 2018.

BRAUN, R. L.; JUNQUEIRA, D. M.; VERLI, H. Filogenia Molecular. In: VERLI, H. **Bioinformática: da Biologia a Flexibilidade Molecular**. Porto Alegre: [S.l.], 2014. Cap. 5. p. 90-124. Disponível em: <https://www.ufrgs.br/bioinfo/ebook/>. Acesso em: 1 set. 2018.

CHELIKANI, P.; FITA, I.; LOEWEN, P. C. Diversity of structures and properties among catalases. **Cellular and Molecular Life Sciences (cmls)**, [S.l.], v. 61, n. 2, p.192-208, 1 jan. 2004. Disponível em: <https://link.springer.com/article/10.1007%2Fs00018-003-3206-5>. Acesso em: 15 out. 2018.

CHEN, Y. A., WEN, Y. C., CHANG, W. C. AtPAN: an integrated system for reconstructing transcriptional regulatory networks in *Arabidopsis thaliana*. **BMC genomics**, v. 13, p. 85, 2012.

CHOUDHURY, F. K.; RIVERO, R. M.; BLUMWALD, E.; MITTLER, R. Reactive oxygen species, abiotic stress and stress combination. **The Plant Journal**, [S.l.], v. 90, n. 5, p.856-867, 1 nov. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27801967>. Acesso em: 16 ago. 2018.

COSTA, J. H.; MCDONALD, A. E.; ARNHOLDT-SCHMITT, B.; MELO, D. F. A classification scheme for alternative oxidases reveals the taxonomic distribution and evolutionary history of the enzyme in angiosperms. **Mitochondrion**, [S.l.], v. 19, p.172-183, nov. 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/24751423>. Acesso em: 27 ago. 2018.

CUNHA, J. R.; LIMA NETO, M. C.; CARVALHO, F. E. L.; MARTINS, M. O.; JARDIM-MESSEDER, D.; MARGIS-PINHEIRO, M.; SILVEIRA, J. A. G. Salinity and osmotic stress trigger different antioxidant responses related to cytosolic ascorbate peroxidase knockdown in rice roots. **Environmental and Experimental Botany**, [S.l.], v. 131, p.58-67, nov. 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0098847216301344>. Acesso em: 10 fev. 2019.

DU. Y.; WANG. P.; CHEN. J.; SONG. C. Comprehensive Functional Analysis of the Catalase Gene Family in *Arabidopsis thaliana*. **Journal of Integrative Plant Biology**, [S.l.], v. 10, n. 50, p.1318-1326, jun. 2008. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19017119>. Acesso em: 01 ago. 2018.

EDGAR. R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. **Bmc Bioinformatics**, [S.l.], v. 5, n. 1, p.1-19, 2004. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/15318951>. Acesso em: 16 maio 2018.

ESAKA, M.; YAMADA, N.; KITABAYASHI, M.; SETOGUCHI, Y.; TSUGEKI, R.; KONDO, M.; NISHIMURA, M. CDNA cloning and differential gene expression of three catalases in pumpkin. **Plant Molecular Biology**, [S.l.], v. 33, n. 1, p.141-155, 1997. Disponível em: <https://link.springer.com/article/10.1023/A:1005742916292>. Acesso em: 07 out. 2018.

FIGUEROA-YÁÑEZ, L.; CANO-SOSA, J.; CASTAÑO, E.; ARROYO-HERRERA, A.; CAAMAL-VELAZQUEZ, J. H.; SANCHEZ-TEYER, F.; LÓPEZ-GÓMEZ, R.; SANTOS-BRIONES, C. L.; RODRÍGUEZ-ZAPATA, L. Phylogenetic relationships and expression in response to low temperature of a catalase gene in banana (*Musa acuminata* cv. "Grand Nain") fruit. **Plant Cell, Tissue and Organ Culture (pctoc)**, [S.l.], v. 109, n. 3, p.429-438, 7 jan. 2012. Disponível em: <https://link.springer.com/article/10.1007/s11240-011-0107-4>. Acesso em: 10 out. 2018.

FINET, C.; FLOYD, S. K.; CONWAY, S. J.; ZHONG, B.; SCUTT, C. P.; BOWMAN, J. L. Evolution of the YABBY gene family in seed plants. **Evolution & Development**, [S.l.], v. 18, n. 2, p.116-126, 13 jan. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26763689>. Acesso em: 12 out. 2018.

FOYER. C. H.; RUBAN. A. V.; NOCTOR. G. Viewing oxidative stress through the lens of oxidative signalling rather than damage. **Biochemical Journal**, [S.l.], v. 474, n. 6, p.877-883, 7 mar. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28270560>. Acesso em: 10 fev. 2019.

FRUGOLI, J. A.; ZHONG, H. H.; NUCCIO M. L.; MCCOURT, P.; MCPEEK, M. A.; THOMAS, T. L.; MCCLUNG, C. R. Catalase is encoded by a multigene family in *Arabidopsis thaliana* (L.) Heynh. **Plant Physiology**, [S.l.], v. 1, n. 112, p.327-336, set. 1996. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/8819328>. Acesso em: 08 out. 2018.

FRUGOLI, J. A.; MCPEEK, M. A.; THOMAS, T. L.; MCCLUNG, C. R. Intron loss and gain during evolution of the catalase gene family in angiosperms. **Genetics**, [S.l.], v. 1, n. 149, p.355-365, jan. 1998. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1460146/>. Acesso em: 01 ago. 2018.

GUAN, L. RUZSA, S.; SKADSEN, R. W.; SCANDALIOS, J. G. Comparison of the Cat2 Complementary DNA Sequences of a Normal Catalase Activity Line (W64A) and a High Catalase Activity Line (R6-67) of Maize. **Plant Physiology**, [S.l.], v. 96, n.

4, p.1379-1381, 1 ago. 1991. Disponível em:

<http://www.plantphysiol.org/content/96/4/1379.long>. Acesso em: 01 ago. 2018.

GUAN, L.; SCANDALIOS, J. G. Catalase transcript accumulation in response to dehydration and osmotic stress in leaves of maize viviparous mutants. **Redox Report**, [S.l.], v. 5, n. 6, p.377-383, dez. 2000. Disponível em:

<https://www.ncbi.nlm.nih.gov/pubmed/11140749>. Acesso em: 05 set. 2018.

GUAN, L.; SCANDALIOS, J. G. Developmentally related responses of maize catalase genes to salicylic acid. **Proceedings of The National Academy of Sciences**, [S.l.], v. 92, n. 13, p.5930-5934, 20 jun. 1995. Disponível em:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC41615/>. Acesso em: 02 out. 2018.

GUAN, L.; SCANDALIOS, J. G. Effects of the plant growth regulator abscisic acid and high osmoticum on the developmental expression of the maize catalase genes. **Physiologia Plantarum**, [S.l.], v. 104, n. 3, p.413-422, nov. 1998. Disponível em:

<https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1399-3054.1998.1040317.x>. Acesso em: 05 set. 2018.

GUAN, L.; SCANDALIOS, J. G. Molecular evolution of maize catalases and their relationship to other eukaryotic and prokaryotic catalases. **Journal of Molecular Evolution**, [S.l.], v. 42, n. 5, p.570-579, maio 1996. Disponível em:

<https://link.springer.com/article/10.1007/BF02352287>. Acesso em: 05 set. 2018.

GUO, H.; LEE, T. H.; WANG, X.; PATERSON, A. H. Function Relaxation Followed by Diversifying Selection after Whole-Genome Duplication in Flowering Plants. **Plant Physiology**, [S.l.], v. 162, n. 2, p.769-778, 11 abr. 2013. Disponível em:

<https://www.ncbi.nlm.nih.gov/pubmed/23580595>. Acesso em: 11 set. 2018.

HALFORD, N. G.; CURTIS, T. Y.; CHEN, Z.; HUANG, J. Effects of abiotic stress and crop management on cereal grain composition. **Journal of Experimental Botany**, [S.l.], v. 66, n. 5, p.1145-1156, 26 nov. 2014. Disponível em:

<https://www.ncbi.nlm.nih.gov/pubmed/25428997>. Acesso em: 10 ago. 2018.

HAVIR, E. A.; BRISSON, L. F.; ZELITCH, I. Distribution of catalase isoforms in *Nicotiana tabacum*. **Phytochemistry**, [S.l.], v. 41, n. 3, p.699-702, fev. 1996.

Disponível em: <https://www.sciencedirect.com/science/article/pii/0031942295006745>. Acesso em: 06 set. 2018.

HIGO, K.; UGAWA, Y.; IWAMOTO, M.; KORENAGA, T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. **Nucleic acids research**, v. 27, p. 297-300, 1999.

HIGO, K.; HIGO, H. Cloning and characterization of the rice CatA catalase gene, a homologue of the maize Cat3 gene. **Plant Molecular Biology**, [S.l.], v. 30, n. 3, p.505-521, fev. 1996. Disponível em:

<https://link.springer.com/article/10.1007/BF00049328>. Acesso em: 26 set. 2018.

HU, B.; JIN, J.; GUO, A. Y.; ZHANG, H.; LUO, J.; GAO, G. GSDS 2.0: an upgraded gene feature visualization server. **Bioinformatics**, v. 31, p. 1296-1297, 2014.



HU, L.; YANG, Y.; JIANG, L.; LIU, S. The catalase gene family in cucumber: The catalase gene family in cucumber: genome-wide identification and organization. **Genetics and Molecular Biology**, [S.l.], v. 39, n. 3, p.408-415, 25 jul. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5004828/>. Acesso em: 07 out. 2018.

HU, Y. LIU, S.; YUAN, H.; LI, J.; YAN, D.; ZHANG, J.; LU, Y. Functional comparison of catalase genes in the elimination of photorespiratory H<sub>2</sub>O<sub>2</sub> using promoter- and 3'-untranslated region exchange experiments in the *Arabidopsis cat2* photorespiratory mutant. **Plant, Cell & Environment**, [S.l.], v. 33, n. 10, p.1656-1670, 7 set. 2010.

HUANG, X.; MADAN, A. CAP3: a DNA sequence assembly program. **Genome Research**. 9, 868–877, 1999.

HUELSENBECK, J. P.; RONQUIST, F. MRBAYES: Bayesian inference of phylogenetic trees. **Bioinformatics**, v. 17, n. 8, p. 754-755, 2001.

INUPAKUTIKA, M.; SENGUPTA, S.; DEVIREDDY, A. R.; AZAD, R. K.; MITTLER, R. The evolution of reactive oxygen species metabolism. **Journal of Experimental Botany**, [S.l.], v. 67, n. 21, p.5933-5943, 14 out. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27742750>. Acesso em: 01 ago. 2018.

IORIZZO, M.; ELLISON, S.; SENALIK, D.; ZENG, P.; SATAPOOMIN, P.; HUANG, J.; BOWMAN, M.; IOVENE, M.; SANSEVERINO, W.; CAVAGNARO, P.; YILDIZ, M.; MACKO-PODGÓRNI, A.; MORANSKA, E.; GRZEBELUS, E.; GRZEBELUS, D.; ASHRAFI, H.; ZHENG, Z.; CHENG, S.; SPOONER, D.; DEYNZE, A. V.; SIMON, P. A. high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. **Nature Genetics**, [S.l.], v. 48, n. 6, p.657-666, 9 maio 2016. Disponível em: <https://www.nature.com/articles/ng.3565>. Acesso em: 08 out. 2018.

ISLAM, W.; NOMAN, A.; QASIM, M.; WANG, L. Plant Responses to Pathogen Attack: Small RNAs in Focus. **International Journal of Molecular Sciences**, [S.l.], v. 19, n. 2, p.1-23, 8 fev. 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/29419801>. Acesso em: 15 ago. 2018.

IWAMOTO, M.; MAEKAWA, M.; SAITO, A.; HIGO, H.; HIGO, K. Evolutionary relationship of plant catalase genes inferred from exon-intron structures: isozyme divergence after the separation of monocots and dicots. **Tad Theoretical and Applied Genetics**, [S.l.], v. 97, n. 1-2, p.9-19, 29 jul. 1998. Disponível em: <https://link.springer.com/article/10.1007/s001220050861>. Acesso em: 04 out. 2018.

JANIAK, A.; KWAŚNIEWSKI, M.; SZAREJKO, I. Gene expression regulation in roots under drought. **Journal of Experimental Botany**, [S.l.], v. 67, n. 4, p.1003-1014, 11 dez. 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26663562>. Acesso em: 02 out. 2018.

JOO, J.; LEE, Y. H.; SONG, S. I. Rice CatA, CatB, and CatC are involved in environmental stress response, root growth, and photorespiration,

respectively. **Journal of Plant Biology**, [S.I.], v. 57, n. 6, p.375-382, 25 nov. 2014. Disponível em: <https://link.springer.com/article/10.1007/s12374-014-0383-8>. Acesso em: 08 out. 2018.

JWA, N.; HWANG, B. K. Convergent Evolution of Pathogen Effectors toward Reactive Oxygen Species Signaling Networks in Plants. **Frontiers in Plant Science**, [S.I.], v. 8, p.1-12, 29 set. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5627460/>. Acesso em: 20 ago. 2018.

KABIR, M. H.; WANG, M. H. Functional studies on two catalase genes from tomato (*Solanum lycopersicum* L.). **The Journal of Horticultural Science and Biotechnology**, [S.I.], v. 86, n. 1, p.84-90, jan. 2011. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/14620316.2011.11512730>. Acesso em: 07 out. 2018.

KATO, J. Characterization of catalase from green algae *Chlamydomonas reinhardtii*. **Journal of Plant Physiology**, [S.I.], v. 151, n. 3, p.262-268, jan. 1997. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0176161797802519>. Acesso em: 23 out. 2018.

KIM, D; PERTEA, G.; TRAPNELL, C.; PIMENTEL, H.; KELLEY, R.; SALZBERG, S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome Biology**, v. 14, n. 4, 2013.

KLOTZ, M. G.; KLASSEN, G. R.; LOEWEN, P. C. Phylogenetic relationships among prokaryotic and eukaryotic catalases. **Molecular Biology and Evolution**, [S.I.], v. 14, n. 9, p.951-958, 1 set. 1997. Disponível em: <https://academic.oup.com/mbe/article/14/9/951/1014674>. Acesso em: 24 out. 2018.

KORALEWSKI, T. E.; KRUTOVSKY, K. V. Evolution of Exon-Intron Structure and Alternative Splicing. **Plos One**, [S.I.], v. 6, n. 3, p.1-10, 25 mar. 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3064661/>. Acesso em: 18 out. 2018.

KWON, S. I.; LEE, H.; AN, C. S. Differential Expression of Three Catalase Genes in the Small Radish (*Rhaphanus sativus* L. var. *sativus*). **Molecules and Cells**, [S.I.], v. 1, n. 24, p.37-44, ago. 2007. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/17846497>. Acesso em: 14 out. 2018.

LANDI, S.; HAUSMAN, J.; GUERRIERO, G.; ESPOSITO, S. Poaceae vs. Abiotic Stress: Focus on Drought and Salt Stress. Recent Insights and Perspectives. **Frontiers In Plant Science**, [S.I.], v. 8, p.1-9, 11 jul. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28744298>. Acesso em: 09 out. 2018.

LANGMEAD, B; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357-359, 2012.

LEE, S. H.; AN, C. S. Differential Expression of Three Catalase Genes in Hot Pepper (*Capsicum annuum* L.). **Molecular and Cells**, [S.I.], v. 20, n. 2, p.247-255, 2005.

Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/16267400>. Acesso em: 23 ago. 2018.

LESCOT, M., DÉHAIS, P., THUIS, G., MARCHAL, K., MOREAU, Y., VAN DE PEER, Y., ROUZÉ, P., ROMBAUTS, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. **Nucleic acids research**, v. 30, p. 325-327, 2002.

LI, XIAN C.; XING, Y. Z.; JIANG, X.; QIAO, J.; TAN, H. L.; TIAN, Y.; ZHOU, B. Identification and characterization of the catalase gene *bcCAT* from the red alga *Pyropia yezoensis* (bangiales, rhodophyta). **Journal of Phycology**, [S.l.], v. 48, n. 3, p.664-669, 10 maio 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27011082>. Acesso em: 23 out. 2018.

LIU, Q.; WANG, Z.; XU, X.; ZHANG, H.; LI, C. Genome-Wide Analysis of C2H2 Zinc-Finger Family Transcription Factors and Their Responses to Abiotic Stresses in Poplar (*Populus trichocarpa*). **Plos One**, [S.l.], v. 10, n. 8, p.1-25, 3 ago. 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26237514>. Acesso em: 12 jan. 2018.

LIU, S.; LIU, Y.; YANG, X.; TONG, C. [...] The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. **Nature Communications**, [S.l.], v. 5, n. 1, p.1-11, 23 maio 2014. Disponível em: <https://www.nature.com/articles/ncomms4930>. Acesso em: 12 out. 2018.

LIU, Y.; LIUA, Y.; YAOA, Y.; HUA, X.; XINGA, S.; XUA, L. Cloning and allelic variation of two novel catalase genes (*SoCAT-1* and *SsCAT-1*) in *Saccharum officinarum* L. and *Saccharum spontaneum* L. **Biotechnology & Biotechnological Equipment**, [S.l.], v. 29, n. 3, p.431-440, 6 mar. 2015. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/13102818.2015.1018839>. Acesso em: 04 out. 2018.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome biology**, v. 15, n. 12, p. 550, 2014.

MAGADUM, S.; BANERJEE, U.; MURUGAN, P.; GANGAPUR, D.; RAVIKESAVAN, R. Gene duplication as a major force in evolution. **Journal of Genetics**, [S.l.], v. 01, n. 92, p.1-7, abr. 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/23640422>. Acesso em: 10 set. 2018.

MARIANI, L.; FERRANTE, A. Agronomic Management for Enhancing Plant Tolerance to Abiotic Stresses - Drought, Salinity, Hypoxia, and Lodging. **Horticulturae**, [S.l.], v. 3, n. 4, p.1-18, 1 dez. 2017. Disponível em: <http://www.mdpi.com/2311-7524/3/4/52>. Acesso em: 06 ago. 2018.

MEENA, K. K.; SORTY, A. M.; BITLA, U. M.; CHOUDHARY, K.; GUPTA, P.; PAREEK, A.; SINGH, D.; PRABHA, R.; SAHU, P. K.; GUPTA, V. K.; SINGH H. B.; KRISHANANI, K. K.; MINHAS, P. S. Abiotic Stress Responses and Microbe-Mediated Mitigation in Plants. **Frontiers in Plant Science**, [S.l.], v. 8, p.1-25, 9 fev.

2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28232845>. Acesso em: 08 ago. 2018.

MENEZES-BENAVENTE, L.; KERNODLE S. P.; MARGIS-PINHEIRO, M.; SCANDALIOS, J. G. Salt-induced antioxidant metabolism defenses in maize (*Zea mays L.*) seedlings. **Redox Report**, [S.l.], v. 9, n. 1, p.29-36, fev. 2004. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/15035825>. Acesso em: 07 nov. 2018.

MHAMDI, A.; NOCTOR, G.; BAKER, A. Plant catalases: Peroxisomal redox guardians. **Archives of Biochemistry and biophysics**, [S.l.], v. 525, n. 2, p.181-194, set. 2012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0003986112001476?via%3Dihub>. Acesso em: 10 dez. 2018.

MHAMDI, A.; QUEVAL, G.; CHAOUCH, S.; VANDERAUWERA, S.; BREUSEGEM, F. V.; NOCTOR, G. Catalase function in plants: a focus on *Arabidopsis* mutants as stress-mimic models. **Journal of Experimental Botany**, [S.l.], v. 61, n. 15, p. 4197-4220, set. 2010.

MORI, H.; HIGO, K.; HIGO, H.; MINOBE, Y.; MATSUI, H.; CHIBA, S. Nucleotide and derived amino acid sequence of a catalase cDNA isolated from rice immature seeds. **Plant Molecular Biology**, [S.l.], v. 18, p.973-976, 1992.

MORITA, S.; TASAKA, M.; FUJISAWA, H.; USHIMARU, T.; TSUJI, H. A cDNA Clone Encoding a Rice Catalase Isozyme. **Plant Physiology**, [S.l.], v. 105, n. 3, p.1015-1016, 1 jul. 1994. Disponível em: <http://www.plantphysiol.org/content/105/3/1015.long>. Acesso em: 11 set. 2018.

MULLEN, R. T.; GIFFORD, D. J. Purification and Characterization of Catalase from Loblolly Pine (*Pinus taeda L.*) Megagametophytes. **Plant Physiology**, [S.l.], v. 103, n. 2, p.477-483, 1 out. 1993. Disponível em: <http://www.plantphysiol.org/content/103/2/477.long>. Acesso em: 24 out. 2018.

PANCHY, N.; LEHTI-SHIU, M. D.; SHIU, S. Evolution of gene duplication in plants. **Plant Physiology**, [S.l.], p.1-23, 10 jun. 2016. Disponível em: <http://www.plantphysiol.org/content/171/4/2294>. Acesso em: 10 set. 2018.

PANDEY, P.; IRULAPPAN, V.; BAGAVATHIANNAN, M. V.; SENTHIL-KUMAR, M. Impact of Combined Abiotic and Biotic Stresses on Plant Growth and Avenues for Crop Improvement by Exploiting Physio-morphological Traits. **Frontiers in Plant Science**, [S.l.], v. 8, p.1-15, 18 abr. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28458674>. Acesso em: 05 ago. 2018.

PASSARDI, F.; ZAMOCKY, M.; FAVET, J.; JAKOPITSCH, C.; PENEL, C.; OBINGER, C.; DUNAND, C. Phylogenetic distribution of catalase-peroxidases: Are there patches of order in chaos?. **Gene**, [S.l.], v. 397, n. 1-2, p.101-113, ago. 2007. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0378111907001928?via%3Dihub>. Acesso em: 24 out. 2018.

PHUKAN, U. J.; JEENA, G. S.; SHUKLA, R. K. WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants. **Frontiers in Plant Science**, [S.l.], v. 7, p.1-14, 3 jun. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4891567/pdf/fpls-07-00760.pdf>. Acesso em: 08 out. 2018.

POLIDOROS, A. N.; MYLONA, P. V.; SCANDALIOS, J. G. Transgenic tobacco plants expressing the maize Cat2 gene have altered catalase levels that affect plant-pathogen interactions and resistance to oxidative stress. **Transgenic Research**, [S.l.], v. 6, n. 10, p.555-69, dez. 2001. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/11817543>. Acesso em: 26 set. 2018.

POLIDOROS, A. N.; SCANDALIOS, J. G. Response of the Maize Catalases to Light. **Free Radical Biology and Medicine**, [S.l.], v. 23, n. 3, p.497-504, 1997. Disponível em: <https://www.sciencedirect.com/science/article/pii/S089158499700110X?via%3Dihub>. Acesso em: 05 set. 2018.

POZO, M. J.; ENT, S. V. D.; VAN LOON, L. C.; PIETERSE, C. M. J. Transcription factor MYC2 is involved in priming for enhanced defense during rhizobacteria-induced systemic resistance in *Arabidopsis thaliana*. **New Phytologist**, [S.l.], v. 180, n. 2, p.511-523, out. 2008. Disponível em: <https://nph.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1469-8137.2008.02578.x>. Acesso em: 08 out. 2018.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>

RANUM, P.; PEÑA-ROSAS, J. P.; GARCIA-CASAL, M. N. Global maize production, utilization, and consumption. **Annals of the New York Academy of Sciences**, [S.l.], v. 1312, n. 1, p.105-112, 20 mar. 2014. Disponível em: <https://nyaspubs.onlinelibrary.wiley.com/doi/epdf/10.1111/nyas.12396>. Acesso em: 06 set. 2018.

REDINBAUGH, M. G.; WADSWORTH, G. J.; SCANDALIOS, J. G. Characterization of catalase transcripts and their differential expression in maize. **Biochimica et Biophysica Acta (bba) - Gene Structure And Expression**, [S.l.], v. 951, n. 1, p.104-116, nov. 1988. Disponível em: <https://www.sciencedirect.com/science/article/pii/0167478188900309?via%3Dihub>. Acesso em: 06 set. 2018.

REDINBAUGH, M. G.; SABRE, M.; SCANDALIOS, J. G. Expression of the maize Cat3 catalase gene is under the influence of a circadian rhythm. **Proceedings of the National Academy of Sciences**, [S.l.], v. 87, n. 17, p.6853-6857, 1990.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139-140, 2010.

ROGOZIN, I. B.; SVERDLOV, A. V.; BABENKO, V. N.; KOONIN, E. V. Analysis of evolution of exon-intron structure of eukaryotic genes. **Briefings In Bioinformatics**, [S.l.], v. 2, n. 6, p.118-134, jun. 2005. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/15975222>. Acesso em: 22 out. 2018.

SARUYAMA, H.; MATSUMURA, T. Cloning and Characterization of a cDNA Encoding Catalase in Wheat. **DNA Sequence**, [S.l.], v. 10, n. 1, p.31-35, jan. 1999. Disponível em: <https://www.tandfonline.com/doi/abs/10.3109/10425179909033933>. Acesso em: 04 out. 2018.

SCANDALIOS, J. G.; ACEVEDO, A.; RUZSA, S. Catalase gene expression in response to chronic high temperature stress in maize. **Plant Science**, [S.l.], v. 156, n. 1, p.103-110, jul. 2000. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0168945200002351?via%3Dihub>. Acesso em: 05 set. 2018.

SCANDALIOS, J. G.; GUAN, L. M.; POLIDOROS, A. N. Catalases in plants: gene structure, properties, regulation, and expression. In: Scandalios JG. (ed) **Oxidative Stress and the Molecular Biology of Antioxidant Defenses**. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1997; 343–406. Disponível em: <https://cshmonographs.org/index.php/monographs/article/view/4093>. Acesso em: 04 set. 2018.

SCANDALIOS, J. G.; TSAFTARIS, A. S.; CHANDLEE, J. M.; SKADSEN, R. W. Expression of the developmentally regulated catalase (cat) genes in maize. **Developmental Genetics**, [S.l.], v. 4, n. 4, p.281-293, 1983. Disponível em: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/dvg.1020040406>. Acesso em: 10 out. 2018.

SCHMUTZ, J.; CANNON, S. B.; SCHLUETER, J. MA, J.; MITROS, T.; NELSON, W.; HYTEN, D. L.; SONG, Q.; THELEN, J. J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G. D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M. K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG, X.; SHINOZAKI, K.; NGUYEN, H. T.; WING, R. A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R. C.; JACKSON, S. A. Genome sequence of the palaeopolyploid soybean. **Nature**, [S.l.], v. 463, n. 7278, p.178-183, jan. 2010. Disponível em: <https://www.nature.com/articles/nature08670>. Acesso em: 08 out. 2018.

SEWELAM, N.; KAZAN, K.; SCHENK, P. M. Global Plant Stress Sinalina: Reactive Oxygen Species at the Cross-Road. **Frontiers in Plant Science**, [S.l.], v. 7, n. 187, p.1-21, 23 fev. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4763064/>. Acesso em: 15 ago. 2018.

SHARMA, P.; JHA, A. B.; DUBEY, R. S.; PESSARAKLI, M. Reactive Oxygen Species, Oxidative Damage, and Antioxidative Defense Mechanism in Plants under Stressful Conditions. **Journal of Botany**, [S.l.], v. 2012, p.1-26, 2012. Disponível em: <https://www.hindawi.com/journals/jb/2012/217037/>. Acesso em: 10 ago. 2018.

SKADSEN, R. W.; SCHULZE-LEFERT, P.; HERBST, J. M. Molecular cloning, characterization and expression analysis of two catalase isozyme genes in barley. **Plant Molecular Biology**, [S.l.], v. 05, n. 29, p.1005-1014, dez. 1995. Disponível em: <https://link.springer.com/article/10.1007/BF00014973>. Acesso em: 06 set. 2018.

SU, Y.; GUO, J.; LING, H.; CHEN, S.; WANG, S.; XU, L.; ALLAN, A. C.; QUE, Y. Isolation of a Novel Peroxisomal Catalase Gene from Sugarcane. Which Is Responsive to Biotic and Abiotic Stresses. **Plos One**, [S.l.], v. 9, n. 1, p.1-11, 2 jan. 2014. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084426>. Acesso em: 04 out. 2018.

SUZUKI, M.; ARIO, T.; HATTORI, T.; NAKAMURA, K.; ASAHI, T. Isolation and characterization of two tightly linked catalase genes from castor bean that are differentially regulated. **Plant Molecular Biology**, [S.l.], v. 03, n. 25, p.507-516, jun. 1995. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/8049373>. Acesso em: 06 set. 2018.

SUZUKI, N.; KATANO, K. Coordination Between ROS Regulatory Systems and Other Pathways Under Heat Stress and Pathogen Attack. **Frontiers in Plant Science**, [S.l.], v. 9, p.1-8, 16 abr. 2018. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00490/full>. Acesso em: 13 ago. 2018.

TAIZ, L.; ZEIGER, E.; MOLLER, I. M.; MURPHY, A. **Fisiologia e Desenvolvimento Vegetal**. 6. ed. Porto Alegre: Artmed, 2017.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular Biology and Evolution** v. 28, p. 2731-2739, 2011.

TAVARÉ, S. Some probabilistic and statistical problems in the analysis of DNA sequences, in *Some Mathematical Questions in Biology - DNA Sequence Analysis*, ed Miura R. M., editor. (Providence, RI: Amer Math Soc;), 57–86, 1986.

THE ANGIOSPERM PHYLOGENY GROUP. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. **Bot. Botanical Journal of the Linnean Society**. 186, 1–20, 2016.

TOHIDFAR, M.; KHOSRAVI, S. Transgenic crops with an improved resistance to biotic stresses. A review. **Biotechnology, Agronomy, Society and Environment**, [S.l.], v. 19, n. 1, p.62-70, dez. 2015. Disponível em: <https://popups.uliege.be/1780-4507/index.php?id=11844>. Acesso em: 18 ago. 2018.

TURKAN. I. Emerging roles for ROS and RNS – versatile molecules in plants. **Journal of Experimental Botany**, [S.l.], v. 68, n. 16, p.4413-4416, 20 jul.

2017. Disponível em: <https://academic.oup.com/jxb/article/68/16/4413/4080428>. Acesso em: 06 nov. 2019.

VIGHI, I. L.; BENITEZ, L. C.; AMARAL, M. N.; AULER, P. A.; MORAES, G. P.; RODRIGUES, G. S.; MAIA, L. C.; PINTO, L. S.; BRAGA, E. J. B. Changes in gene expression and catalase activity in *Oryza sativa* L. under abiotic stress. **Genetics and Molecular Research**, [S.l.], v. 15, n. 4, p.1-15, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27819727>. Acesso em: 13 abr. 2018.

VON OSSOWSKI, I.; HAUSNER, G.; LOEWEN, P. C. Molecular evolutionary analysis based on the amino acid sequence of catalase. **Journal of Molecular Evolution**, [S.l.], v. 37, p.71-76, jan. 1993. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/8360921>. Acesso em: 24 out. 2018.

WALLEY, J. W.; SARTOR, R. C.; SHEN, Z.; SCHMITZ, R. J.; WU, K. J.; URICH, M. A.; NERY, J. R.; SMITH, L. G.; SCHNABLE, J.C.; ECKER, J. R.; BRIGGS, S. P. Integration of omic networks in a developmental atlas of maize. **Science**, [S.l.], v. 353, n. 6301, p.814-818, 18 ago. 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27540173>. Acesso em: 09 nov. 2018.

WANG, L.; STEIN, L. D. Modeling the evolution dynamics of exon-intron structure with a general random fragmentation process. **Bmc Evolutionary Biology**, [S.l.], v. 13, n. 1, p.1-15, fev. 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/23448166>. Acesso em: 22 out. 2018.

WANG, L.; WU, X.; GAO, W.; ZHAO, M.; ZHANG, J.; HUANG, C. Differential Expression Patterns of *Pleurotus ostreatus* Catalase Genes during Developmental Stages and under Heat Stress. **Genes**, [S.l.], v. 8, n. 11, p.335-337, 21 nov. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/29160795>. Acesso em: 03 out. 2018.

WILLEKENS, H.; VILLARROEL, R.; VAN, M. M.; INZÉ, D.; VAN, C. W. Molecular identification of catalases from *Nicotiana plumbaginifolia* (L.). **Febs Letters**, [S.l.], v. 352, n. 1, p.79-83, 19 set. 1994a. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/7925949>. Acesso em: 07 out. 2018.

WILLEKENS, H.; LANGEBARTELS, C.; TIRÉ, C.; MONTAGU, M. V.; INZÉ, D.; Van, W. Camp Differential expression of catalase genes in *Nicotiana plumbaginifolia* (L.). **Proceedings of The National Academy of Sciences**, [S.l.], v. 91, n. 22, p.10450-10454, 25 out. 1994b.

WILLIAMSON, J. D.; SCANDALIOS, J. G. Differential response of maize catalases and superoxide dismutases to the photoactivated fungal toxin cercosporin. **The Plant Journal**, [S.l.], v. 2, n. 3, p.351-358, maio 1992. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/1284657>. Acesso em: 05 set. 2018.

YOU, J.; ZHANG, L.; SONG, B.; QI, X.; CHAN, Z. Systematic Analysis and Identification of Stress-Responsive Genes of the NAC Gene Family in *Brachypodium distachyon*. **Plos One**, [s.l.], v. 10, n. 3, p.1-20, 27 mar. 2015. Disponível em:



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376915/pdf/pone.0122027.pdf>. Acesso em: 08 out. 2018.

ZAMOCKY, M.; FURTMÜLLER, P. G.; OBINGER, C. Evolution of Catalases from Bacteria to Humans. **Antioxidants & Redox Signaling**, [S.l.], v. 10, n. 9, p.1527-1548, set. 2008. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/18498226>. Acesso em: 24 out. 2018.

ZANDALINAS, S.; BALFAGÓN, D.; ARBONA, V.; GÓMEZ-CADENAS, A. Modulation of Antioxidant Defense System Is Associated with Combined Drought and Heat Stress Tolerance in Citrus. **Frontiers in Plant Science**, [S.l.], v. 8, p.1-10, 7 jun. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28638395>. Acesso em: 06 nov. 2018.

ZANG, Y.; LIU, J.; TANG, X. X.; ZHOU, B. Description of a *Zostera marina* catalase gene involved in responses to temperature stress. **Peerj**, [S.l.], v. 6, p.1-20, 26 mar. 2018.

ZHANG, C.; GSCHWEND, A. R.; OUYANG, Y.; LONG, M. Evolution of Gene Structural Complexity: An Alternative-Splicing-Based Model Accounts for Intron-Containing Retrogenes. **Plant Physiology**, [S.l.], v. 165, n. 1, p.412-423, 11 fev. 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4012599/>. Acesso em: 09 jan. 2018

ZHU, JIAN-KANG. Abiotic Stress Signaling and Responses in Plants. **Cell**, [S.l.], v. 167, n. 2, p.313-324, out. 2016. Disponível em: [https://www.cell.com/cell/abstract/S0092-8674\(16\)31080-7](https://www.cell.com/cell/abstract/S0092-8674(16)31080-7). Acesso em 15 ago. 2018.

ZIMMERMANN, P.; HEINLEIN, C.; ORENDI, G.; ZENTGRAF, U. Senescence-specific regulation of catalases in *Arabidopsis thaliana* (L.) Heynh. **Plant, Cell and Environment**, [S.l.], v. 29, n. 6, p.1049-1060, jun. 2006. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/17080932>. Acesso em: 05 out. 2018.

**ANEXO A - DADOS DE TAXONOMIA, NÚMERO DE GENES DA CATALASE POR  
ESPÉCIE E DISTRIBUIÇÃO DE CLADOS COM BASE NA FILOGENIA.**

Ordem	Espécies	Acrônimo	Número de genes	Clado			
				CAT2	CAT3	CAT1	CAT4
Amborellales	<i>Amborella trichopoda</i>	At	2	0	0	2	0
Zingiberales	<i>Ensete ventricosum</i>	Ev	4	0	2	2	0
	<i>Musa acuminata</i>	Ma	4	0	2	2	0
Poales	<i>Zoysia japonica</i>	Zj	2	1	0	1	0
	<i>Oropetium thomaeum</i>	Ot	3	1	1	1	0
	<i>Zea mays</i>	Zm	3	1	1	1	0
	<i>Sorghum bicolor</i>	Sb	3	1	1	1	0
	<i>Dichanthelium digosanthes</i>	Do	2	1	1	0	0
	<i>Setaria italica</i>	Si	3	1	1	1	0
	<i>Panicum hallii</i>	Ph	3	1	1	1	0
	<i>Laersia perrieri</i>	Lp	3	1	1	1	0
	<i>Oryza sativa</i>	Os	3	1	1	1	0
	<i>Brachypodium distachyon</i>	Bd	3	1	1	1	0
	<i>Dactylis glomerata</i>	Dgl	3	1	1	1	0
	<i>Hordeum vulgare</i>	Hv	2	1	0	1	0
	<i>Aegilopsis tauschii</i>	Ata	3	1	1	1	0
	<i>Ananas comosus</i>	Ac	3	1	1	1	0
Asparagales	<i>Saccharum hybrid cultivar</i>	Shc	2	0	1	1	0
	<i>Secale cereale</i>	Sc	2	0	1	1	0
	<i>Triticum urartu</i>	Tu	2	0	1	1	0
Arecales	<i>Dendrobium catenatum</i>	Dc	2	0	1	1	0
	<i>Phoenix dactylifera</i>	Pd	2	0	1	1	0
Dioscoreales	<i>Elaeis guineensis</i>	Eg	3	0	1	2	0
Dioscoreales	<i>Dioscorea alata</i>	Da	2	0	1	1	0
Alismatales	<i>Zostera marina</i>	Zmar	1	0	0	1	0
Proteales	<i>Nelumbo nucifera</i>	Nn	2	0	0	2	0
Cucurbitales	<i>Cucurbita moschata</i>	Cmo	3	0	0	1	2
	<i>Citrullus lanatus</i>	Cla	3	0	0	1	2
	<i>Cucumis melo</i>	Cm	3	0	0	1	2
	<i>Quercus suber</i>	Qs	2	0	0	0	2
Fagales	<i>Castanea mollissima</i>	Cmol	3	0	0	0	3
	<i>Betula pendula</i>	Bp	2	0	0	0	2
	<i>Juglans regia</i>	Jr	3	0	0	0	3
	<i>Ficus carica</i>	Fc	2	0	0	1	1
	<i>Ziziphus jujuba</i>	Zju	1	0	0	0	1
	<i>Humulus lupulus</i>	Hl	2	0	0	0	2
Rosales	<i>Cannabis sativa</i>	Csat	2	0	0	0	2
	<i>Prunus persica</i>	Pper	2	0	0	0	2
	<i>Malus domestica</i>	Md	2	0	0	0	2
	<i>Rosa multiflora</i>	Rm	1	0	0	0	1
	<i>Fragaria vesca</i>	Fv	1	0	0	0	1
	<i>Arachis ipaensis</i>	Ai	2	0	0	1	1
	<i>Lupinus angustifolius</i>	La	2	0	0	0	2
	<i>Trifolium pratense</i>	Tpr	1	0	0	0	1
Fabales	<i>Medicago truncatula</i>	Mt	1	0	0	0	1
	<i>Cicer arietinum</i>	Ca	1	0	0	0	1
	<i>Lotus japonicus</i>	Lj	1	0	0	0	1
	<i>Vigna radiata</i>	Vr	2	0	0	0	2
	<i>Phaseolus vulgaris</i>	Pv	2	0	0	0	2
Oxalidales	<i>Glycine max</i>	Gm	4	0	0	0	4
	<i>Cajanus cajan</i>	Cc	2	0	0	0	2
	<i>Cephalotus follicularis</i>	Cf	1	0	0	0	1
	<i>Ricinus communis</i>	Rc	3	0	0	2	2
Malpighiales	<i>Jatropha curcas</i>	Jc	3	0	0	1	2
	<i>Manihot esculenta</i>	Me	6	0	0	2	4
	<i>Hevea brasiliensis</i>	Hb	3	0	0	1	2
	<i>Salix purpurea</i>	Spur	3	0	0	1	2
	<i>Populus trichocarpa</i>	Pt	2	0	0	1	1
	<i>Aquilaria agallocha</i>	Aag	3	0	0	0	3
	<i>Durio zibethinus</i>	Dz	3	0	0	1	2
Malvales	<i>Corchorus capsularis</i>	Ccap	2	0	0	1	1
	<i>Herrania umbratica</i>	Hu	2	0	0	2	2
	<i>Theobroma cacao</i>	Tc	2	0	0	2	2
	<i>Gossypium arboreum</i>	Gar	3	0	0	0	3

Ordem	Espécies	Acrônimo	Número de genes	Clado			
				CAT2	CAT3	CAT1	CAT4
Brassicales	<i>Tarenaya hassleriana</i>	Th	5	0	0	5	0
	<i>Barbarea vulgaris</i>	Bv	2	0	0	2	0
	<i>Euclidium syriacum</i>	Esy	3	0	0	3	0
	<i>Eutrema salsugineum</i>	Es	3	0	0	3	0
	<i>Thlaspi arvense</i>	Ta	2	0	0	2	0
	<i>Schrenkiella parvula</i>	Sp	2	0	0	2	0
	<i>Conringia planisiliqua</i>	Cpl	2	0	0	2	0
	<i>Sisymbrium irio</i>	Sir	2	0	0	2	0
	<i>Raphanus sativus</i>	Rsat	5	0	0	5	0
	<i>Brassica oleracea</i>	Bol	7	0	0	7	0
	<i>Arabis alpina</i>	Aal	2	0	0	2	0
	<i>Capsella rubella</i>	Cr	2	0	0	2	0
	<i>Boechera stricta</i>	Bs	2	0	0	2	0
	<i>Arabidopsis thaliana</i>	Ath	3	0	0	3	0
Sapindales	<i>Carica papaya</i>	Cpa	3	0	0	1	2
	<i>Azadirachta indica</i>	Ain	2	0	0	0	2
	<i>Citrus sinensis</i>	Csi	2	0	0	0	2
Myrtales	<i>Metrosideros polymorpha</i>	Mp	3	0	0	0	3
Vitales	<i>Eucalyptus grandis</i>	Egr	3	0	0	0	3
	<i>Vitis vinifera</i>	Vv	2	0	0	1	1
Caryophyllales	<i>Fagopyrum tataricum</i>	Ft	1	0	0	1	0
	<i>Spinacia oleracea</i>	Sol	2	0	0	2	0
	<i>Chenopodium quinoa</i>	Cq	3	0	0	3	0
Ericales	<i>Beta vulgaris</i>	Bvul	2	0	0	2	0
	<i>Primula veris</i>	Pver	2	0	0	1	1
Gentianales	<i>Rhazya stricta</i>	Rst	3	0	0	1	2
	<i>Catharanthus roseus</i>	Cra	2	0	0	1	1
Lamiales	<i>Coffea arabica</i>	Car	4	0	0	2	2
	<i>Olea europaea</i>	Oe	3	0	0	0	3
	<i>Fraxinus excelsior</i>	Fex	2	0	0	0	2
	<i>Ocimum tenuiflorum</i>	Ote	2	0	0	0	2
	<i>Dorcoceas hydrometricum</i>	Dhy	1	0	0	0	1
Solanales	<i>Erythranthe guttata</i>	Egut	2	0	0	0	2
	<i>Genilsea aurea</i>	Gau	1	0	0	0	1
	<i>Nicotiana sylvestris</i>	Nsyl	3	0	0	0	3
	<i>Ipomoea nil</i>	Ini	4	0	2	2	0
	<i>Nicotiana tabacum</i>	Ntab	3	0	0	1	2
	<i>Solanum tuberosum</i>	Stub	3	0	0	1	2
	<i>Solanum lycopersicum</i>	Sly	3	0	0	1	2
Asterales	<i>Capsicum annuum</i>	Can	3	0	0	1	2
	<i>Silybum marianum</i>	Smar	3	0	0	0	3
	<i>Cynara cardunculus</i>	Ccar	4	0	0	0	4
	<i>Lactuca sativa</i>	Lsat	2	0	0	0	2
Apiales	<i>Helianthus annuus</i>	Han	5	0	0	0	5
	<i>Carthamus tiuctorius</i>	Ctiu	3	0	0	0	3
	<i>Daucus carota</i>	Dcar	5	0	0	0	5

Fonte: dados da pesquisa.