

CIÊNCIAS BIOLÓGICAS

NATÁLIA FERMINO CALZADO

**Aspectos evolutivos e transcricionais da
família gênica da catalase em genomas
vegetais**



Rio Claro
2016

NATÁLIA FERMINO CALZADO

ASPECTOS EVOLUTIVOS E TRANSCRICIONAIS DA FAMÍLIA
GÊNICA DA CATALASE EM GENOMAS VEGETAIS

Orientador: Douglas Silva Domingues

Trabalho de Conclusão de Curso apresentado ao
Instituto de Biociências da Universidade
Estadual Paulista “Júlio de Mesquita Filho” -
Câmpus de Rio Claro, para obtenção do grau de
Bacharela em Ciências Biológicas

Rio Claro

2016

581.15 Calzado, Natália Fermino
C171a Aspectos evolutivos e transcripcionais da família gênica da
catalase em genomas vegetais / Natália Fermino Calzado. -
Rio Claro, 2016
30 f. : il., figs., gráfs., tabs.

Trabalho de conclusão de curso (bacharelado - Ciências
biológicas) - Universidade Estadual Paulista, Instituto de
Biociências de Rio Claro

Orientador: Douglas Silva Domingues

1. Genética vegetal. 2. Evolução molecular. 3. Sistema
antioxidante. I. Título.

AGRADECIMENTOS

Ao Departamento de Botânica e seus funcionários pelo espaço e ajuda.

A orientação do Prof. Douglas Silva Domingues e a colaboração fundamental do colega de estágio Daniel Dovigo, sem a ajuda e o apoio de ambos não seria possível realizar esse trabalho.

A todos da turma CBI2010 por serem mais do que colegas de classe e sim uma família de verdade. Em especial aos amigos Júlia, Luiza, Moreno, Marina e Isabela por todas as conversas, as risadas, o carinho imensurável e por transformarem Rio Claro em um novo lar e dividirem comigo os melhores (e os piores também) momentos desses anos. Sem vocês essa graduação seria muito mais difícil e guardo um carinho enorme por todos. Também agradeço aos amigos Kauã e Maria Fernanda por estarem sempre presentes me apoiando, incentivando e tornando meu cotidiano muito mais leve e divertido.

Aos professores que encontrei durante a graduação e não me deixaram desanimar com o curso. Um agradecimento especial ao Prof. Dr. Mario Sergio Palma por me orientar nos primeiros anos de graduação e me ensinar e apoiar tanto.

Ao namorado, Danilo, por ser tão presente e me ajudar em todos os momentos. Sem o carinho e companheirismo constante não seria possível concluir esse trabalho. Obrigada por me ouvir em todas as crises de desespero, por também dividir tantos momentos felizes e por ser o melhor amigo que eu poderia ter.

Agradeço também a minha família por sempre confiar e acreditar em mim e por sempre estar por perto.

Obrigada.

RESUMO

A fotossíntese e a respiração aeróbica são importantes mecanismos de obtenção de energia nos seres vivos. Ambos os processos passam pela oxidação celular, que pode ser intensificada em situações de estresse, gerando uma superprodução de espécies reativas a oxigênio (EROs), dos quais um exemplo é o peróxido de hidrogênio (H_2O_2). As células possuem vários sistemas para impedir danos causados por EROs, entre os quais o sistema enzimático antioxidante. Uma das enzimas envolvidas neste sistema é a catalase. Essa enzima atua na degradação de H_2O_2 e é codificada por uma família gênica que em *Arabidopsis thaliana* é composta por três genes. O objetivo do presente trabalho foi realizar a anotação e análise filogenética da família gênica catalase para 14 espécies vegetais, bem como inferências transcricionais para espécies selecionadas com base em dados da literatura. Apenas quatro espécies possuíram família gênica composta por 3 genes codificantes para catalase. Análises evolutivas indicaram que as catalases em dicotiledôneas são mais conservadas que em monocotiledôneas, que existem grupos filogenéticos de catalase ausentes na espécie-modelo *A. thaliana* que os genes desta família gênica passam por seleção purificadora, mas com padrão de substituição dissimilar entre códons. Inferências transcricionais em espécies selecionadas indicam que cada membro da família gênica tem um perfil transcricional tecido-específico e diferencialmente regulado por estresses. Dessa forma, o presente trabalho traz elementos que evidenciam a importância de análises evolutivas conjuntamente a análises bioquímicas para a compreensão de respostas moleculares do metabolismo celular.

Palavras chave: Catalase, Evolução molecular, Sistema antioxidante

ABSTRACT

The photosynthesis and aerobic respiration are important energy getting mechanisms in living beings. Both processes pass through cell oxidation, which can be enhanced under stress, causing an overproduction of reactive oxygen species (ROS), of which an example is hydrogen peroxide (H₂O₂). Cells have multiple systems to prevent damage caused by ROS, among which the antioxidant enzyme system. One of the enzymes involved in this system is catalase. This enzyme acts on H₂O₂ degradation and is encoded by a gene family that in *Arabidopsis thaliana* is composed of three genes. The aim of this study was to annotate and analysis phylogenetic the catalase gene family in 14 plant species, as well as transcriptional regulations for species selected based on literature data. Only four gene family owned species had 3 genes codificantes to catalase. Evolutionary analyses indicated that the catalase in dicots are more conserved than in monocots, there are missing phylogenetic groups of catalase in model species *A. thaliana* genes of this gene family go through purifying selection, but with substitution pattern dissimilar between codons. Inferences Transcriptional in selected species indicate that each member of the gene family has a tissue specific transcriptional profile they are differentially regulated by stress. Thus, this work brings elements that emphasize the importance of evolutionary analyses together with biochemical analysis for understanding cellular metabolism molecular responses.

Keywords: Catalase, molecular evolution, antioxidant system

SUMÁRIO

1	INTRODUÇÃO.....	6
2	OBJETIVOS.....	10
3	MATERIAIS E MÉTODOS.....	11
	3.1 Espécies selecionadas para análise	11
	3.2 Anotação da família catalase em genomas vegetais	11
4	RESULTADO E DISCUSSÃO.....	14
5	CONSIDERAÇÕES FINAIS	25
6	REFERÊNCIAS	26

1 INTRODUÇÃO

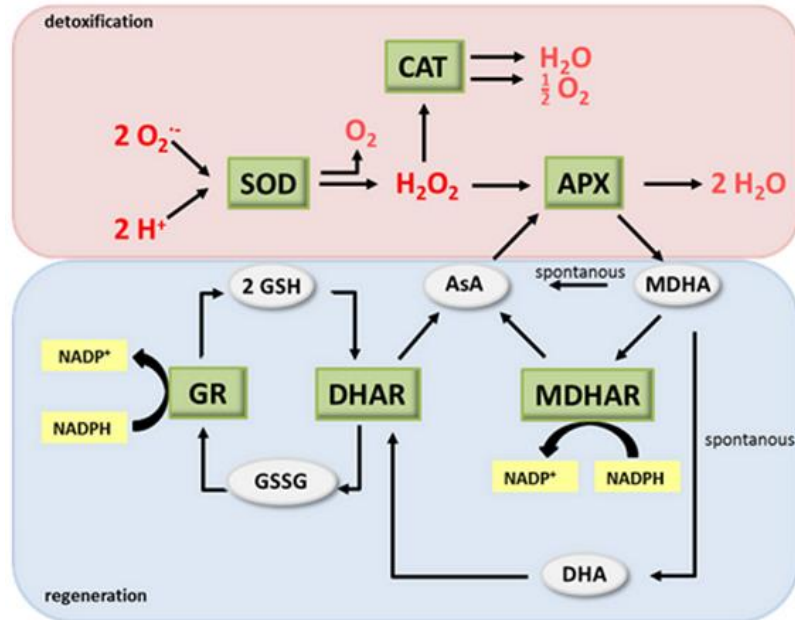
Nas plantas, a geração de energia está relacionada aos processos de respiração e fotossíntese. Ambos os processos tratam, grosso modo, da transferência de elétrons. Em condições favoráveis de crescimento vegetal, a respiração e a fotossíntese ajudam a manter o equilíbrio entre produção e consumo de espécies reativas de oxigênio (ERO) (Suzuki et al., 2012). O termo EROs inclui qualquer derivado molecular de oxigênio, se referindo principalmente a radicais livres como superóxido (O_2^-), a hidroxila (OH) e também ao peróxido de hidrogênio (H_2O_2) (Mhamdi et al., 2010).

As organelas com alta atividade de transferência de elétrons, como cloroplasto e mitocôndria, são as principais fontes de EROs nas células (Mittler et al., 2004). As espécies reativas de oxigênio atuam na regulação de vários processos biológicos das células, como crescimento celular, morte celular, sinalização hormonal e respostas a estresse. Ou seja, apesar de gerar subprodutos tóxicos para a célula, as EROs também exercem um papel importante para o ciclo celular (Mittler et al., 2004). A produção excessiva de EROs pode ser desencadeada por fatores de origem biótica ou abiótica, como a ação de patógenos, deficiência hídrica, salinidade e extremos de temperatura. Para restabelecer a homeostase, as células contam com diversos mecanismos de defesa.

Os níveis de EROs são controlados por vias enzimáticas e não enzimáticas. As vias enzimáticas são denominadas sistema antioxidante e possuem como principais enzimas de detoxificação de EROs a glutatona peroxidase (GPX), ascorbato peroxidase (APX) e a catalase (CAT) (Sofet et al., 2015; Figura 1). A conversão do radical superóxido (O_2^-) a O_2 e H_2O_2 (peróxido de hidrogênio) ocorre pelo sistema superóxido dismutase (SOD). A catalase e ascorbato peroxidase são responsáveis por desintegrar o peróxido de hidrogênio (H_2O_2) (Gross et al., 2013). O peróxido de hidrogênio é a EROs mais estável, e variações em seus níveis podem funcionar como sinalizadores em diversos processos, como resposta a ação de patógenos, expressão gênica e morte celular (Vuosku et al., 2015). O equilíbrio entre SOD e enzimas do sistema antioxidante é fundamental para restaurar a homeostase celular normalizando os níveis de O_2^- e H_2O_2 (Mittler et al., 2004).

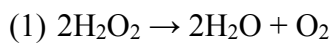
A figura 1 exemplifica como ocorre a degradação de EROs pelo sistema antioxidante enzimático.

Figura 1: Sistema antioxidante em plantas. Legenda: AsA, ácido ascórbico; DHA, desidroascorbato; SOD, superóxido dismutase; CAT, catalase; APX, ascorbato peroxidase; MDHA, monodehidroascorbato; MDHAR, MDHA redutase; GR, glutathiona redutase; GSH glutathiona; GSSG, glutathiona dissulfeto



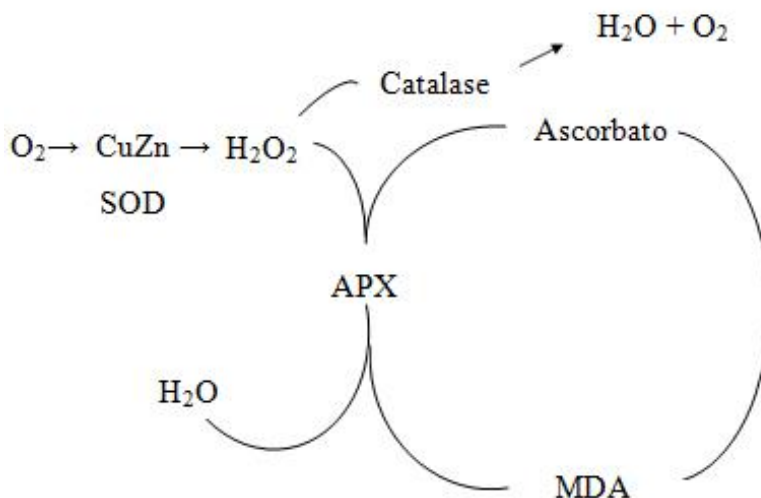
(retirado de Gross et al., 2013).

A catalase é uma proteína tetramérica monofuncional, com alta afinidade por H_2O_2 presente na maioria dos organismos eucariontes. Sua função é decompor peróxido de hidrogênio (H_2O_2) em água e oxigênio (1) (Hu et al., 2016). A atividade bioquímica e transcricional da catalase pode ser encontrada em diferentes tecidos vegetais e exercendo funções variadas, mas sempre relacionadas ao estresse oxidativo das células.



Os peroxissomos são organelas que têm alta produção de EROs, e principalmente de H_2O_2 formada diretamente pela glicolato oxidase durante a fotorespiração (Mhamdi et al., 2010); em função disso, a catalase é altamente presente nos peroxissomos. Nessa organela a catalase age em conjunto com a SOD, o ascorbato e o monodehidroascorbato (MDA), como exemplificado na figura 2.

Figura 2: Ação da catalase e outras enzimas antioxidantes promovendo a decomposição de peróxido de hidrogênio em água e oxigênio



(retirado de Mittler et al., 2004).

A enzima catalase de classe 1 está relacionada ao metabolismo de H_2O_2 , a de classe 2 é associada ao sistema de transporte das plantas e a catalase classe 3 é encontrada principalmente em flores e frutos com baixos níveis em folhas (Vuosku et al, 2015)

O nível transcricional de genes que codificam a catalase é induzido por situações de estresse, como o estresse salínico, hídrico, temperaturas altas ou muito baixas, ação de patógenos ou por meio do tratamento com ácido absísico ou H_2O_2 (Mhamdi et al., 2012). Em dicotiledôneas, acredita-se que a catalase é codificada por uma família gênica composta por três genes (CAT1, CAT2 e CAT3) (Du et al., 2008).

Com a crescente disponibilização de dados públicos de genomas e transcriptomas vegetais ocorrida nos últimos anos, ganham relevância estudos relativos à compreensão da distribuição genômica e respostas transcricionais de genes codificantes de atividades-chave no metabolismo vegetal, antes parcialmente caracterizados. O relacionamento entre informações bioquímicas e genômicas nos permite avaliar se é possível atribuir a uma determinada cópia genômica de catalase a principal contribuição de atividade química da enzima em um determinado tecido. Adicionalmente, pode-se avaliar se a seleção natural atuou diferencialmente na manutenção ou diversificação destes genes ao longo da escala evolutiva.

Estão disponíveis na literatura dados sobre a atividade bioquímica da catalase em resposta a estresse para diversas espécies vegetais que atualmente dispõe de recursos

genômicos de larga escala. Isso ocorre, por exemplo, em espécies de café (Marraccini et al., 2012), tabaco (Clark et al., 2000), pêssego (Hernández et al., 2004; Ahmand et al., 2011), algodão (Chaudhary et al., 2009; Wang et al., 2012), banana (Figuroa et al., 2012; D'Hont et al., 2012) e outras gramíneas. Estes dados agora podem ser melhor compreendidos em termos moleculares por estratégias de análise comparativa de genomas.

2 OBJETIVOS

O objetivo geral do trabalho foi caracterizar aspectos genômicos, evolutivos e transcricionais da família gênica catalase em genomas de angiospermas.

Especificamente pretendeu-se:

- Identificar genes codificantes da enzima catalase em genomas de plantas angiospermas;
- Avaliar qual perfil de seleção ocorre entre os membros da família gênica catalase;
- Avaliar se cada membro da família gênica possui um padrão transcricional específico em espécies selecionadas;
- Identificar qual gene é o principal contribuinte para a atividade transcricional de catalase em espécies selecionadas.

3 MATERIAIS E MÉTODOS

3.1 Espécies selecionadas para análise

Selecionamos para análise as seguintes espécies de eudicotiledôneas: *Coffea canephora* (cafeeiro Robusta ou Conilon), *Nicotiana tabacum* (tabaco), *Prunus persica* (pêssego), *Amborella trichopoda*, *Capsella rubella*, *Citrus sinensis*, *Citrullus lanatus* (melancia), *Fragaria vesca* (morango), *Gossypium raimondii* (algodão), *Musa acuminata* (banana), *Brachypodium distachyon*, *Sorghum bicolor* (sorgo) e *Setaria italica* (Tabela 1). Estas espécies foram selecionadas por apresentarem genoma completo sequenciado, bem como dados transcricionais ou de quantificação da atividade de catalase. Para *C. canephora*, os dados utilizados foram obtidos em www.coffee-genome.org. Para tabaco, os dados foram obtidos em <https://solgenomics.net/tools/blast/>. Os dados genômicos das demais espécies foram obtidos na **plataforma** PLAZA (Proost et al., 2015).

3.2 Anotação da família catalase em genomas vegetais

Os procedimentos gerais de anotação da família gênica catalase foram baseados em artigos recentes que utilizam informações genômicas com os mesmos objetivos para diferentes famílias gênicas (Almeida et al., 2011; Lira et al., 2014; Tian et al., 2015; Xie et al., 2016).

Foi utilizado como ponto de partida as sequências de nucleotídeos e aminoácidos os genes codificantes para catalase da espécie-modelo *Arabidopsis thaliana*, obtidas no banco de dados Peroxibase, específico de peroxidases (Fawal et al., 2013). Todas as sequências protéicas da família gênica catalase em *Arabidopsis thaliana* possuem o domínio proteico código IPR011614 na base de dados Interpro (<https://www.ebi.ac.uk/interpro/>).

Com base nestas informações, as sequências proteicas de *A. thaliana* foram utilizadas como sondas eletrônicas (*queries*) para procuras por TBLASTN (Gertz et al., 2006) contra sequências codificantes dos genomas selecionados. Sequências que apresentaram um *score*

acima de 200 e *e-value* abaixo de e^{-50} foram selecionadas para análise manual, com *download* da sequência do quadro aberto de leitura e proteica. Adicionalmente, para as espécies com sequência genômica disponível na plataforma PLAZA, foi observado se os genes obtidos nesta procura continham o domínio IPR011614. Para as demais espécies, esta procura foi feita manualmente no banco de dados Interpro (<https://www.ebi.ac.uk/interpro/>). Foram utilizadas para análises subsequentes apenas genes cujo do quadro aberto de leitura apresenta de 80 a 120% do tamanho dos genes de *A. thaliana*. Esta medida visa eliminar da análise genes com erros de predição ou pseudogenes.

As sequências proteicas selecionadas foram utilizadas para análise filogenética. Elas foram alinhadas utilizando-se o programa MUSCLE (Edgar, 2004) implementado no programa MEGA 7 (Kumar et al., 2016). Este alinhamento foi utilizado para fazer teste de modelo de seleção a ser utilizado nas análises filogenéticas, no próprio programa MEGA 7. Por meio desta análise, foi definido que seria utilizada a matriz de substituição de Le e Gascuel (2008) (LG), com 5 faixas de distribuição (distribuição gama) de taxas de substituição no alinhamento. A árvore filogenética dos genes analisados foi desenvolvida pelo método de máxima verossimilhança, com bootstrap de 1000 réplicas e não foram utilizadas para inferências filogenéticas posições ausentes em mais de 20% das sequências analisadas.

Para cada espécie analisada, foram estimadas taxas de substituição sinônimas (dS) e não-sinônimas (dN) nas sequências codificantes no software MEGA 7. Foi utilizado o método de Nei-Gojobori (distância p), com cálculo de erro-padrão. Para avaliar a pressão seletiva que atua nessas sequências foi calculada a razão dN/dS (também chamada de ω), onde $dN/dS = 1$, $dN/dS < 1$ e $dN/dS > 1$ representam evolução neutra, seleção purificadora e seleção positiva Darwiniana, respectivamente.

Para se comparar modelos evolutivos que poderiam explicar o padrão evolutivo da família gênica em plantas, três modelos de evolução foram testados no programa CODEML da ferramenta PAML (Yang, 2007), por meio da interface gráfica PAMLX (Xu et al. 2013). Para determinação das sequências de nucleotídeos a serem utilizadas nesta análise, primeiramente as sequências de aminoácidos de catalase foram alinhadas na ferramenta MUSCLE. O padrão de alinhamento de aminoácidos foi utilizado para a determinação do alinhamento de códon, por meio da ferramenta PAL2NAL (Suyama et al., 2006). Posições que continham *gaps* foram removidas.

Foram testados no programa CODEML os modelos evolutivos M0, M1a e M2a. O modelo M0 parte do princípio que todos os códon tem o mesmo nível de dN e dS; o modelo

M1a propõe duas classes de códons, nos quais alguns códons tem evolução neutra ($dN/dS = 1$) e outros seleção purificadora; o modelo M2a indica três categorias de evolução de códons: seleção purificadora, evolução neutra e seleção positiva. Para os três modelos, foi obtido o valor de dN/dS . Para determinar qual dos 3 modelos melhor explica o padrão evolutivo de catalases, foi calculadaa razão de verossimilhança (LTR) dos 3 modelos. Com este valor, é calculado o dobro da diferença de subtração entre os modelos (2-delta log), para as três possibilidades (M0 *versus* M1a; M0 *versus* M2a e M1 *versus* M2a) Estes valores são comparados com a distribuição de χ^2 a 95% de significância, conforme proposto por Posada (2003).

Para as inferências transcricionais dos genes de catalase de *C. canephora*, *N. tabacum* e *B. distachyon* foram utilizados dados de RT-PCR quantitativo em tempo real (qRT-PCR) obtidos da literatura (Golemiec et al., 2014; Kang et al., 2016; Marraccini et al., 2012). Os *primers* utilizados nas reações de qPCR ou as sequências de referência utilizadas para desenho dos *primers* foram comparados por BLASTN com os genes anotados neste trabalho.

Dados de expressão de transcritos por RNA-seq em *C. canephora* e *S. bicolor* e *C. canephora* foram obtidos em bancos de dados públicos. Os dados de *C. canephora* foram obtidos do *site* Coffee Genome Hub (<http://coffee-genome.org/>) e de sorgo, do *site* Morokoshi Sorghum transcriptome (<http://sorghum.riken.jp/morokoshi/Home.html>).

4 RESULTADOS E DISCUSSÃO

Três sequências de proteína catalase de *Arabidopsis thaliana*, foram utilizadas como *query* no banco de dados Plaza para encontrar genes parálogos de catalase em treze espécies vegetais.

Nessa análise obtivemos um total de 44 sequências, sendo 2 genes de *A. trichopoda*, 3 genes de *C. rubella*, 2 genes de *C. sinensis*, 2 genes de *C. lanatus*, 1 gene de *F. vesca*, 4 genes de *G. raimondii*, 2 genes de *P. persica*, 10 genes de *N. tabacum*, 4 genes de *C. canephora*, 2 genes de *S. bicolor*, 3 genes de *S. italica*, 3 genes de *B. distachyon* e 3 genes de *M. acuminata*. A tabela 1 resume a informação geral dos genes identificados

Tabela 1: Família gênica de catalase para cada espécie selecionada. Legenda: gene ID – identificação do gene na base de dados; CDS – tamanho em nucleotídeos do quadro aberto de leitura; Proteína – tamanho em aminoácidos da proteína obtida; PI – ponto isoelétrico; MW – massa molecular; Chr – cromossomo onde se localiza o gene.

Espécie	Gene ID	CDS	Proteína	P.I.	M.W.	Chr
<i>Arabidopsis thaliana</i>	AT1G20630	1479	492	7.4411	56762.07	Chr 01
	AT1G20620	1479	492	7.6937	56695.34	Chr 01
	AT4G35090	1479	492	7.1209	56931.29	Chr 04
<i>Amborella trichopoda</i>	ATR_00001G00070	1472	493	7.2447	56987.28	Chr 01
	ATR_00001G00080	1472	493	7.2512	57049.81	Chr 01
<i>Capsella rubella</i>	CRU_001G19270	1479	492	7.7234	56677.31	Chr 01
	CRU_001G19280	1479	492	7.4414	56736.21	Chr 01
	CRU_007G04900	1479	492	7.2450	56894.34	Chr 07
<i>Citrus sinensis</i>	CS00002G06660	1479	492	7.3756	56909.50	Chr 02
	CS00002G06650			7.1074	55377.61	Chr 02
<i>Citrullus lanatus</i>	CL11G12650	1479	492	7.0097	57038.13	Chr 11
	CL11G12640	1479	492	7.7925	57570.84	Chr 11
<i>Fragaria vesca</i>	FV5G40900	1479	492	7.0820	56939.03	Chr 05

<i>Gossypium</i>	GR01G19870	1170	389	7.2076	45165.23	Chr 01
<i>raimondii</i>	GR03G00280	1479	492	7.4530	56927.19	Chr 03
	GR09G18760	1479	492	7.4134	56961.31	Chr 09
	GR02G11590	1479	492	7.1729	56817.20	Chr 02
<i>Prunus persica</i>	PPE_005G01200	1479	492	7.4524	57071.38	Chr 05
	PPE_005G01190	1479	492	7.4076	56982.34	Chr 05
<i>Nicotiana</i>	mRNA_184252	1479	492	7.4180	56947.48	
<i>tabacum</i>	mRNA_184196	1479	492	7.4022	56947.48	
	mRNA_178552	1479	492	7.2630	56802.13	
	mRNA_96510	1479	492	7.2254	56917.30	
	mRNA_66566	1479	492	7.3408	56848.13	
	mRNA_56818	1479	492	7.1947	56824.02	
	mRNA_178553	1479	492	7.1492	54962.08	
	mRNA_66565	1479	492	6.9577	54869.80	
	mRNA_141689	1479	492	7.0848	56824.27	
	mRNA_24920	1479	492	7.0848	56824.27	
<i>Coffea</i>	Cc07_G11720	1410	470	7.1026	54178.38	Chr 07
<i>canephora</i>	Cc07_G11710	1479	493	7.1471	56832.02	Chr 07
	Cc10_G00570	1479	493	7.2579	56791.12	Chr 10
	Cc10_G00580	1479	493	7.2719	57003.42	Chr 10
<i>Sorghum</i>	SB10G030840	1479	492	7.1068	56841.33	Chr 01
<i>bicolor</i>	SB04G001130	1479	492	7.2018	53868.09	Chr 04
<i>Setaria italica</i>	SI001G12120	1479	492	7.0093	56626.93	Chr 01
	SI004G29580	1479	492	7.3130	56696.20	Chr 04
	SI009G57040	1491	492	7.6472	56836.10	Chr 09
<i>Brachypodium</i>	BD1G76330			7.2909	56746.93	Chr 01
<i>distachyon</i>	BD3G01470	1479	492	7.0615	57134.63	Chr 03
	BD1G29800	1479	492	7.1683	56585.11	Chr 01
<i>Musa</i>	MA03G14660	1479	492	6.8862	57103.44	Chr 03
<i>acuminata</i>	MA05G16310	1389	462	7.7398	53751.79	Chr 05
	MA08G13790	1479	492	7.1879	57237.50	Chr 08

Os genes obtidos possuem proteínas com sequências polipeptídicas entre 442-492 resíduos de aminoácidos, ponto isoelétrico entre pH 6,8 e 7,7 e o peso molecular variou entre 45,1 e 57,2 kDa.

Estudos na espécie modelo *A. thaliana* (Frugoli et al., 1996) indicavam que a família gênica catalase era composta por três genes. A mesma informação era encontrada em trabalhos referentes as espécies como, por exemplo, *Nicotiana tabacum* (Golemiec et al., 2014). Entretanto, as análises feitas com genomas completos sequenciados indicam que a família gênica catalase pode apresentar família gênica composta por dois genes, como é o caso das espécies: *A. trichopoda*, *C. sinensis*, *P. persica*, *F. vesca*, *S. bicolor*, mas também pode apresentar família gênica composta por mais de três genes, como é o caso de *N. tabacum*, *G. raimondii* e *C. canephora*. Entre as espécies selecionadas para o presente trabalho, apenas *C. rubella*, *M. acuminata*, *B. distachyon* e *S. italica* apresentaram família gênica composta por três genes de catalase, assim como em *A. thaliana*.

As análises de substituição sinônimas e não sinônimas de cada espécie (Tabela 2), mostram *B. distachyon* como a espécie com a maior taxa de substituições. No entanto, todas as espécies analisadas tiveram dN/dS menor que 1, indicando seleção purificadora entre os genes de catalase.

Tabela 2: Taxas de substituição sinônimas e não sinônimas entre as espécies com 3 ou mais genes.

Especie	dS	dN	(dN/dS)
<i>Brachypodium distachyon</i>	1,622 ± 0,309	0,208 ± 0,013	0,128
<i>Setaria italica</i>	0,495 ± 0,581	0,177 ± 0,013	0,357
<i>Musa acuminata</i>	1,497 ± 0,370	0,106 ± 0,010	0,070
<i>Arabidopsis thaliana</i>	1,251 ± 0,135	0,117 ± 0,010	0,093
<i>Coffea canephora</i>	1,291 ± 0,339	0,079 ± 0,007	0,061
<i>Capsella rubela</i>	1,283 ± 0,135	0,120 ± 0,010	0,093
<i>Gossypium raimondii</i>	1,283 ± 0,143	0,120 ± 0,010	0,093
<i>Nicotiana tabacum</i>	1,087 ± 0,126	0,077 ± 0,006	0,070
Todas	1,317 ± 0,083	0,086 ± 0,024	0,065
Dicotiledôneas	1,345 ± 0,085	0,112 ± 0,008	0,083
Monocotiledôneas	1,014 ± 0,102	0,141 ± 0,010	0,139

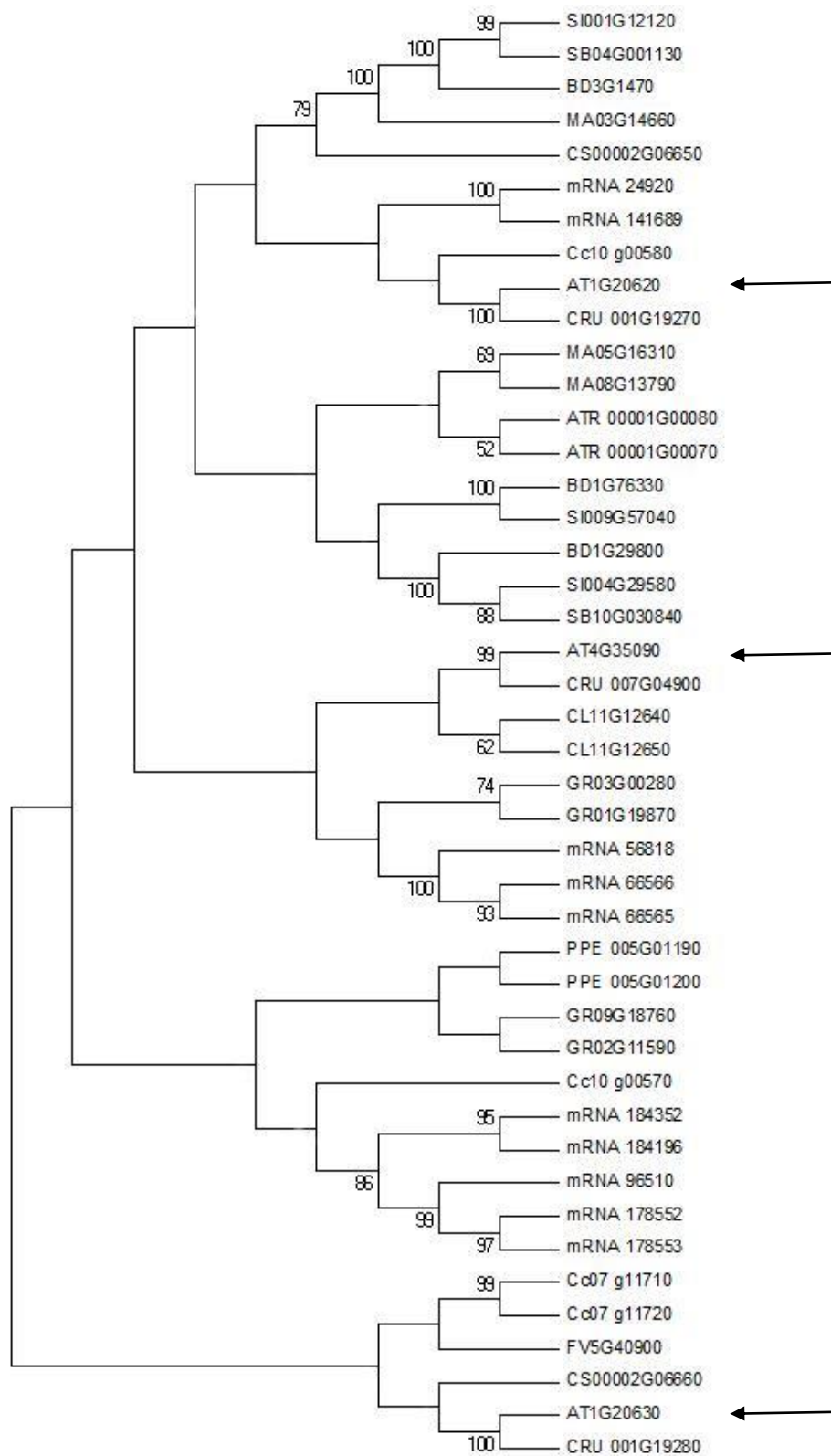
O teste de razão de verossimilhança (Tabela 3) mostrou que entre os modelos testados, M1 foi aceito em relação a M0, indicando que há no mínimo dois padrões distintos de evolução de códons nas sequências analisadas. O modelo M2 teve resultado não-significativo contra os outros dois modelos testados, sendo assim descartado das análises.(Tabela 3).

Tabela 3: Teste de modelos de evolução entre genes codificantes de catalase.

Grupo	Modelo	lnL	2Δl (M1-M0)	Parâmetros	ρ0-ρ1
Dicotiledôneas	M0	-1264.0368214	33,8640188	ω0= 0,11125	100000 0
	M1	-1247.4362046		ω0=0,07673 ω1= 100000 ρ0= 0,86854 ρ1= 0,13146	
Monocotiledôneas	M0	-7820.8120102	163,5865724	ω0= 0,09896	100000 0
	M1	-7739.018724		ω0=0,06226 ω1= 100000 ρ0= 0,93107 ρ1= 0,06893	
Todas	M0	-1622.6743245	47,937964	ω0= 0,09896	100000 0
	M1	-1598.7174263		ω0= 0,07521 ω1= 100000 ρ0= 0,90076 ρ1= 0,09924	

A análise filogenética da proteína catalase (Figura 3) evidencia que cada gene de catalase da espécie-modelo *A. thaliana* encontra-se em um clado distinto. Foi também observado que entre as espécies selecionadas, existem dois grupos de catalase que não estão presentes em *A. thaliana*.

Figura 3: Árvore filogenética da família gênica da proteína catalase em 14 espécies vegetais. Os valores representados nos clados indicam a porcentagem do *bootstrap* obtido.



Os dados obtidos para número efetivo de códons (Tabela 4) e substituições sinônimas e não-sinônimas (Tabela 3) mostram que catalases em dicotiledôneas são mais conservadas que em monocotiledôneas, porém utilizam um maior número de códons para a sua expressão. Observa-se também que as catalases de plantas dicotiledôneas sofreram maior seleção purificadora.

Tabela 4: Número efetivo de códons e número de sequência de proteína para cada espécie

Espécie	Número de códons	Número de sequências
<i>Brachypodium</i>	42,84	03
<i>Distachyum</i>		
<i>Setaria itálica</i>	38,84	03
<i>Musa acuminata</i>	51,03	03
<i>Sorghum bicolor</i>	46,41	02
<i>Arabidopsis thaliana</i>	49,65	03
<i>Coffea canephora</i>	51,76	04
<i>Capsella rubela</i>	48,90	03
<i>Gossypium raimondii</i>	52,69	04
<i>Nicotiana tabacum</i>	49,98	10
<i>Fragaria vesca</i>	51,49	01
<i>Prunus persica</i>	50,04	02
<i>Citrus sinensis</i>	52,15	02
<i>Citrullus lanatus</i>	51,18	02
<i>Amborella trichopoda</i>	52,94	02
Todas	52,87	44
Dicotiledôneas	51,18	33
Monocotiledôneas	45,11	11

As inferências transcricionais de membros da família catalase em *N. tabacum* e *B. distachyon* foram realizadas através de re-análise de dados de qRT-PCR (Kang et al., 2016; Golemic et al., 2014). Em *N. tabacum*, cada primer alinhou-se com dois genes para cada de catalase, evidenciando o caráter tetraploide do genoma analisado. Os genes mRNA_66566; mRNA_66565 e mRNA_56818 se alinham melhor com o primer de CAT1, os genes

mRNA_24920 e mRNA_141689 se alinham melhor com CAT2 e os genes mRNA_184352 e mRNA_184196 se alinham melhor com CAT3. Em todos os casos, a identidade foi maior ou igual a 96%.

Para *B. distachyon* foi possível identificar que a catalase analisada no artigo de Kag et al., (2016) corresponde ao gene BD1G76330. Neste trabalho foi observado que a atividade da catalase diminui após a imposição de estresse hídrico.

Figura5: Resultado do alinhamento do *primer* BdCAT e gene de *Brachypodium distachyon* BD1G76330

BD1G76330
Sequence ID: Query_108513 Length: 1479 Number of Matches: 3

Range 1: 907 to 926 [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
40.1 bits(20)	2e-08	20/20(100%)	0/20(0%)	Plus/Plus

```

Query 1   GTCGGGAGGCTGGTGCTGAA 20
          |||
Sbjct 907 GTCGGGAGGCTGGTGCTGAA 926
  
```

Os resultados de análises transcricionais por RNA-seq de *Sorghum bicolor* e *Coffea canephora* estão apresentados nas figuras 5 e 6.

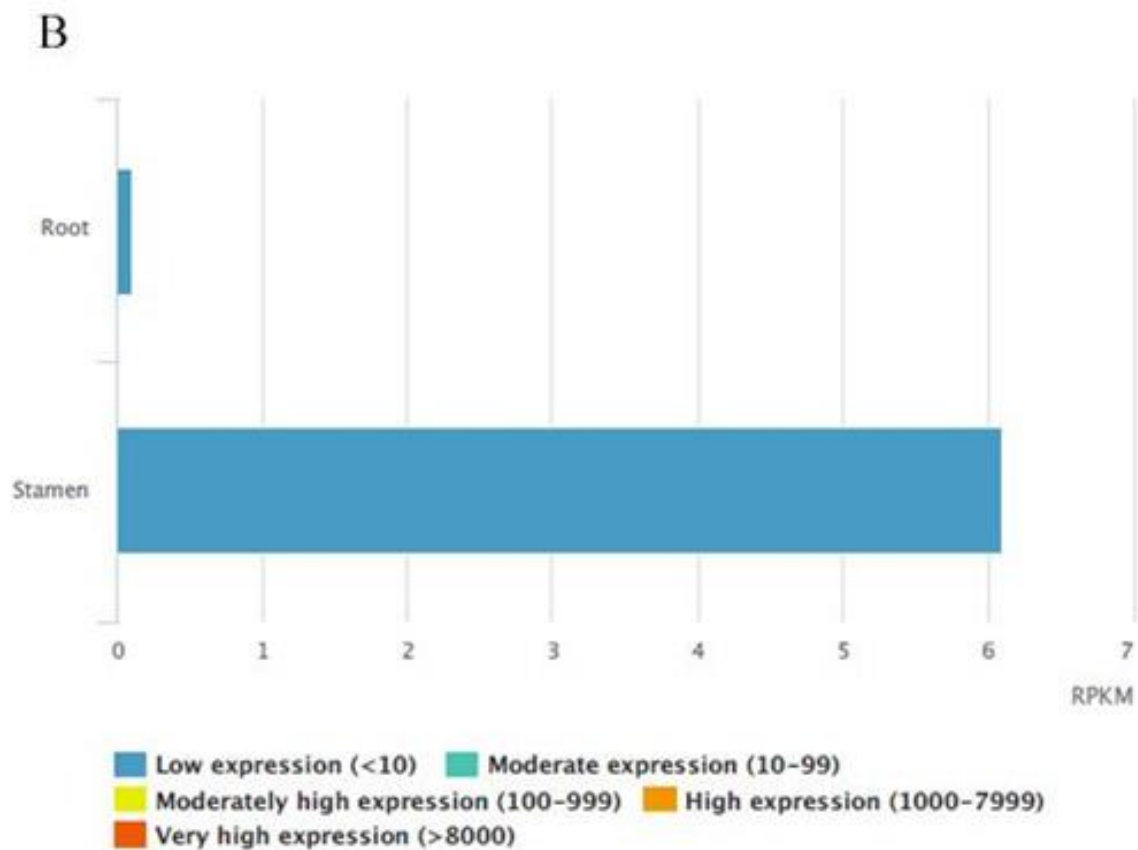
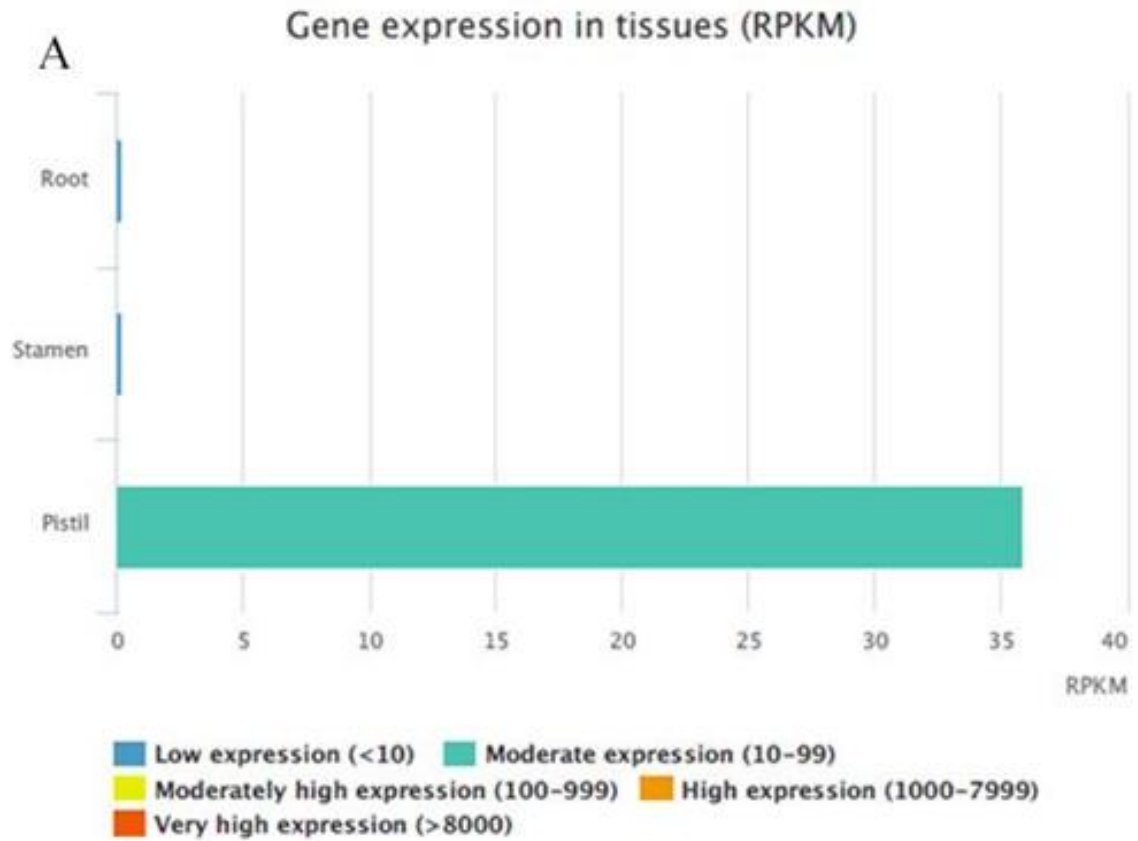
As análises transcricionais de catalase em *C. canephora* e *S. bicolor* mostram que cada cópia genômica de catalase possui padrão distinto de intensidade de expressão e especificidade tecidual.

Em *C. canephora* observa-se dois genes (Cc07_G11720 e Cc07_11710) com intensidade de expressão baixa, mas com alta especificidade tecidual, ativos apenas em tecidos reprodutivos (pistilo e estame, respectivamente). Em contrapartida, os outros dois genes (Cc10_G00570 e Cc10_G00580) possuem maior intensidade de expressão e encontram-se ativos em diversos órgãos vegetativos

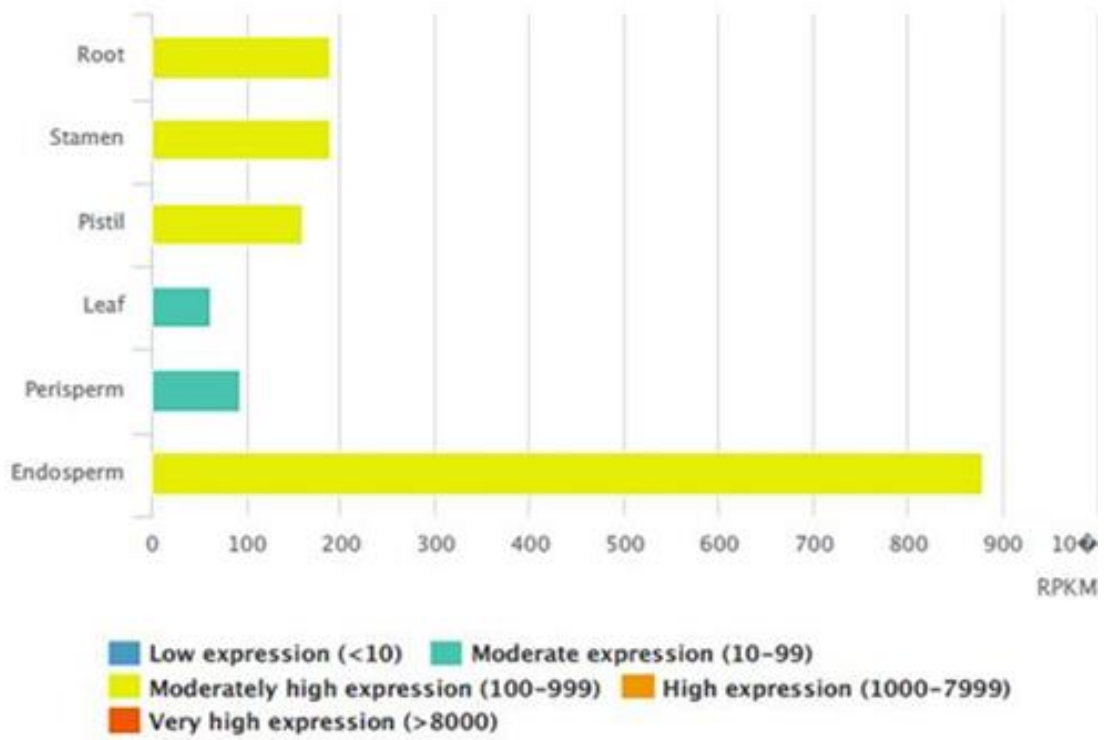
Com isso, podemos observar que diferentes cópias de catalase possuem funções distintas em tecidos vegetais. O gene Cc07_G11720 corresponde a catalase 1 e o gene Cc10_G00570 corresponde a catalase 2 do artigo de Marraccini et al. (2012). Naquele artigo, onde foi avaliada a atividade transcricional destes genes em resposta ao estresse hídrico, observou-se que o gene Cc07_G11720 foi reprimido pelo estresse hídrico enquanto Cc10_G00570 foi induzido pelo mesmo estresse. Dessa forma, o presente trabalho consegue demonstrar que genes de catalase com expressão específica em tecidos reprodutivos são reprimidos pelo estresse, enquanto o padrão oposto ocorre com o gene ativo em tecidos vegetativos.

Em *S. bicolor* foi também possível observar um padrão de tecido-especificidade para cada cópia gênica. Enquanto o gene SB04G001130 tem maior intensidade de expressão em folhas, o gene SB10G030840 tem maior intensidade de expressão em sementes.

Figura 4: Atividade transcricional de genes de catalase em *Coffea canephora*. A) gene Cc07_G11720 B) gene Cc07_11710 C) gene Cc10_G00570 D) Cc10_G00580



C



D

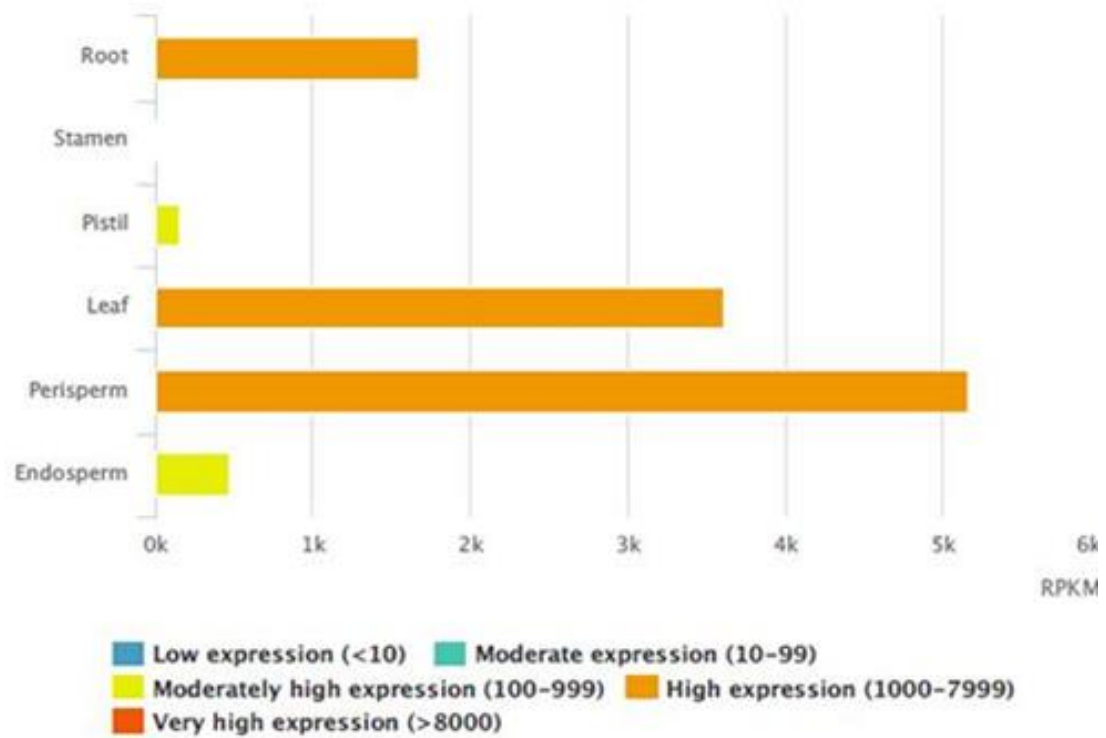


Figura 5: Resultado de BLASTN para A) *primer* CcCAT1 e gene de *Coffea canephora* Cc07_g11720 B) *primer* CcCAT2 e gene de *Coffea canephora* Cc10_g00580

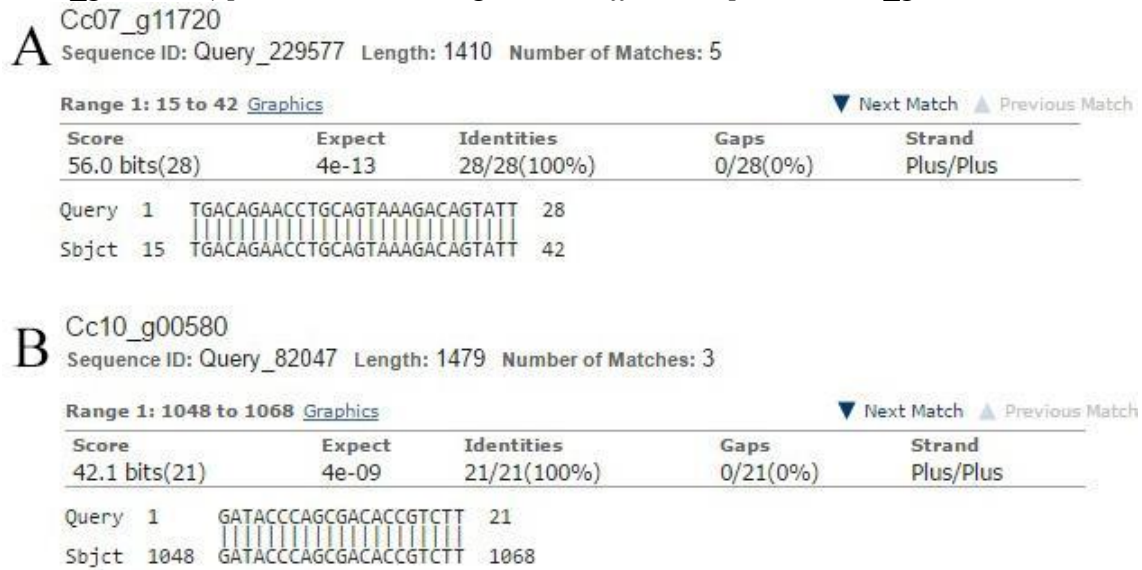
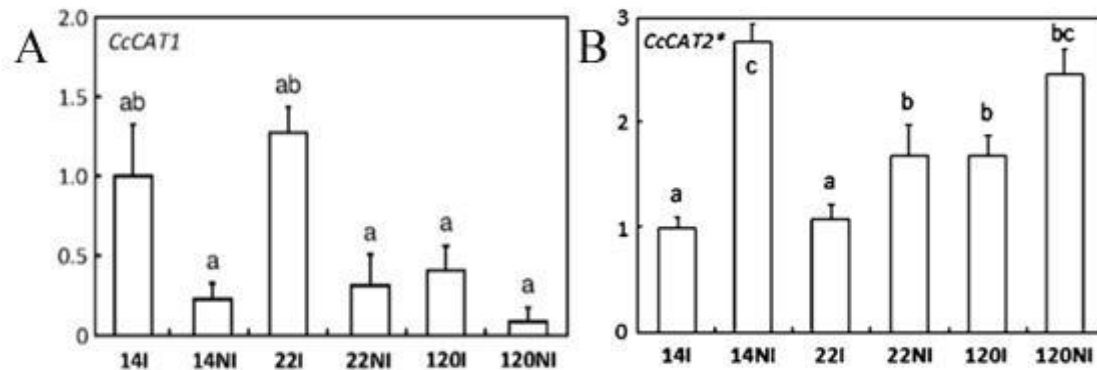
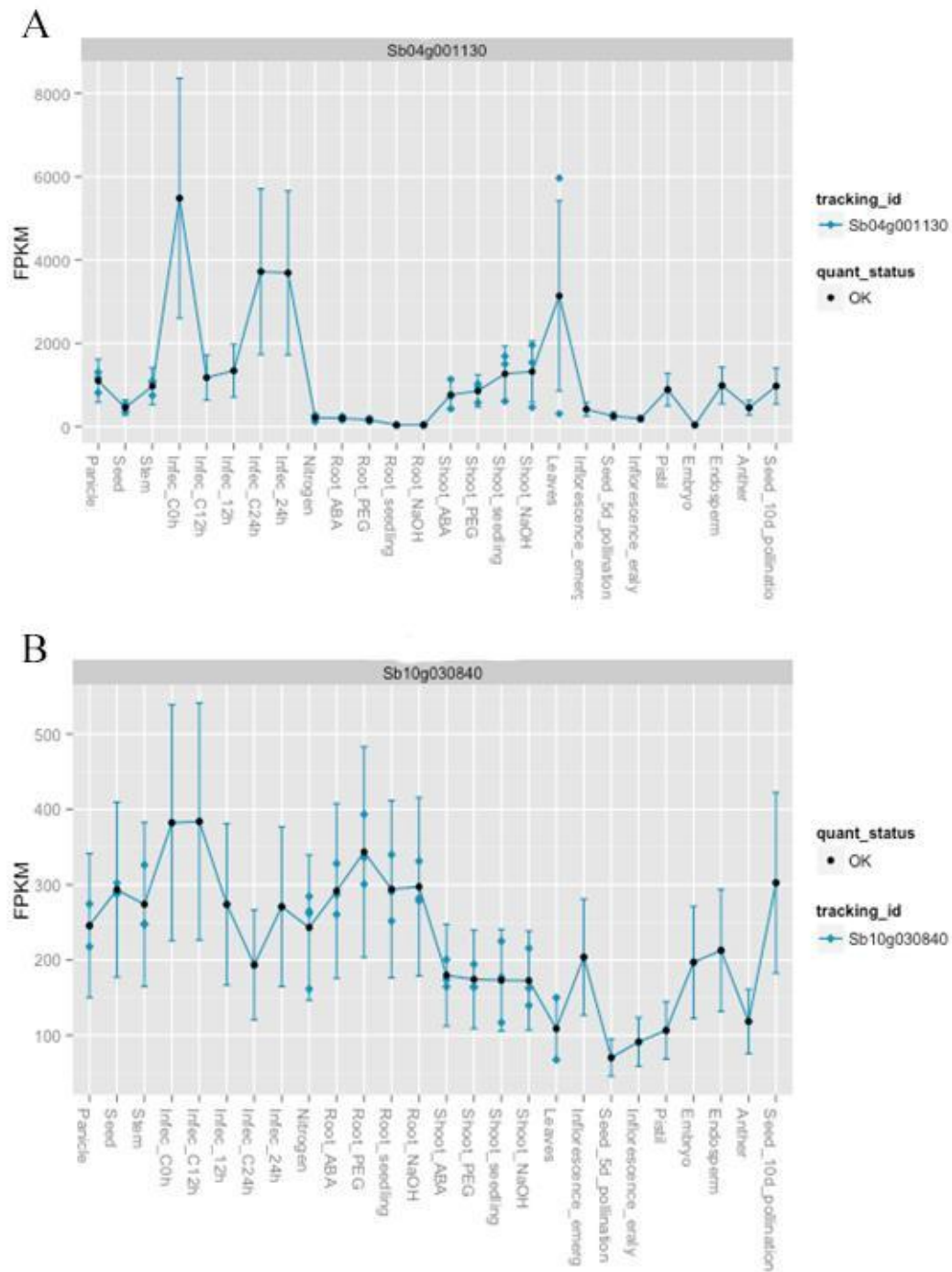


Figura 6: Atividade transcricional de CcCAT1 e CcCAT2. Em estresse hídrico a CcCAT2 (B) é mais expressa do que a CcCAT1 (A)



(retirado de Marriccini et al., 2006)

Figura 7: Atividade transcricional de *Sorghum bicolor*. A) gene SB04G001130 B) gene SB10G030840



5 CONSIDERAÇÕES FINAIS

A família gênica da catalase nem sempre é composta por três genes, podendo ser formada por mais que três genes (como observado em 3 espécies) ou menos que três genes (como observado em 5 espécies). Análises filogenéticas da família gênica catalase apresentam dois grupos de catalases que não são encontrados na espécie-modelo *Arabidopsis thaliana*.

As análises evolutivas mostram que a família gênica da catalase sofre pressão seletiva purificadora, sendo que essa pressão é maior em monocotiledôneas.

A catalase possui diferentes padrões de intensidade de expressão e de especificidade tecidual para cada membro de sua famíliagênica. Por meio de análises comparativas entre resultados de qRT-PCR e sequenciamento de RNA, foi possível observar que existem padrões específicos de expressão em resposta a estresses e de tecido-especificidade entre os membros da família gênica. Estes resultados reforçam a necessidade de análises genômicas comparativas amplas para avaliação da função biológica da atividade bioquímica de uma enzima.

REFERÊNCIAS

- AHMAD, Riaz et al, Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection, **BMC genomics**, v, 12, n, 1, p, 1, 2011.
- ALMEIDA, Juliana et al. Genetic dissection of vitamin E biosynthesis in tomato. **Journal of experimental botany**, v. 62, n. 11, p. 3781-3798, 2011.
- CHAUDHARY, Bhupendra et al, Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (*Gossypium*), **Bmc Genomics**, v, 10, n, 1, p, 1, 2009
- CLARK, Daniel et al, Nitric oxide inhibition of tobacco catalase and ascorbate peroxidase, **Molecular Plant-Microbe Interactions**, v, 13, n, 12, p, 1380-1384, 2000.
- D'HONT, Angélique et al, The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants, **Nature**, v, 488, n, 7410, p, 213-217, 2012,,
- DU, Yan-Yan et al, Comprehensive functional analysis of the catalase gene family in *Arabidopsis thaliana*, **Journal of Integrative Plant Biology**, v, 50, n, 10, p, 1318-1326, 2008.
- EDGAR, Robert C, MUSCLE: multiple sequence alignment with high accuracy and high throughput, **Nucleic acids research**, v, 32, n, 5, p, 1792-1797, 2004.
- FAWAL, Nizar et al, PeroxiBase: a database for large-scale evolutionary analysis of peroxidases, *Nucleic acids research*, v, 41, n, D1, p, D441-D444, 2013.
- FIGUEROA-YÁÑEZ, Luis et al, 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)**, v, 109, n, 3, p, 429-438, 2012.
- FRUGOLI, Julia A. et al. Catalase is encoded by a multigene family in *Arabidopsis thaliana* (L.) Heynh. **Plant Physiology**, v. 112, n. 1, p. 327-336, 1996.
- GERTZ, E. Michael et al. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC biology*, v. 4, n. 1, p. 1, 2006.
- GOLEMIEC, Elżbieta et al. A dissection of the effects of ethylene, H₂O₂ and high irradiance on antioxidants and several genes associated with stress and senescence in tobacco leaves. *Journal of plant physiology*, v. 171, n. 3, p. 269-275, 2014.
- GROß, Felicitas; DURNER, Jörg; GAUPELS, Frank, Nitric oxide, antioxidants and prooxidants in plant defence responses, *Front, PlantSci*, v, 4, n, 419, p, 10,3389, 2013.
- HERNÁNDEZ, José Antonio et al, Oxidative stress induced by long-term plum pox virus infection in peach (*Prunus persica*), **Physiologia Plantarum**, v, 122, n, 4, p, 486-495, 2004.

- HU, Lifang et al. The catalase gene family in cucumber: genome-wide identification and organization. **Genetics and molecular biology**, n. AHEAD, p. 0-0, 2016.
- JOURDA, Cyril et al, Expansion of banana (*Musa acuminata*) gene families involved in ethylene biosynthesis and signalling after lineage-specific whole-genome duplications, *New Phytologist*, v, 202, n, 3, p, 986-1000, 2014.
- KANG, Hanhan et al. Overexpression of wheat ubiquitin gene, Ta-Ub2, improves abiotic stress tolerance of *Brachypodium distachyon*. *Plant Science*, v. 248, p. 102-115, 2016.
- KELLER, Oliver et al, A novel hybrid gene prediction method employing protein multiples equence alignments, **Bioinformatics**, v, 27, n, 6, p, 757-763, 2011.
- KORNOBIS, Etienne et al, TRUFA: a user-friendly web server for de novo RNA-seq analysis using cluster computing, **Evolutionary bioinformatics online**, v, 11, p, 97, 2015.
- KUMAR, Sudhir; STECHER, Glen; TAMURA, Koichiro, MEGA7: Molecular Evolutionary Genetics Analysis version 7,0 for bigger datasets, **Molecular Biology and Evolution**, p, msw054, 2016.
- LE, Si Quang; GASCUEL, Olivier. An improved general amino acid replacement matrix. *Molecular biology and*
- Posada D. 2003. Selecting models of evolution. In *The Phylogenetic Handbook*. Vandamme AM, Salemi M (eds). Cambridge University press. pp. 256-282.
- LIRA, Bruno Silvestre et al. Plant degreening: evolution and expression of tomato (*Solanum lycopersicum*) dephytylation enzymes. **Gene**, v. 546, n. 2, p. 359-366, 2014.
- MAJOROS, William H.; PERTEA, Mihaela; SALZBERG, Steven L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. **Bioinformatics**, v. 20, n. 16, p. 2878-2879, 2004.
- MHAMDI, Amna; NOCTOR, Graham; BAKER, Alison, Plant catalases: peroxisomal redox guardians, **Archives of Biochemistry and Biophysics**, v, 525, n, 2, p, 181-194, 2012.
- MHAMDI, Amna et al, Catalase function in plants: a focus on Arabidopsis mutants as stress-mimic models, **Journal of Experimental Botany**, p, erq282, 2010.
- MITTLER, Ron, et al, "Reactive oxygen gene network of plants," *Trends in plant science* 9,10 (2004): 490-498
- SCANDALIOS, John G, et al, Oxidative stress responses—what have genome-scale studies taught us, **Genome Biol**, v, 3, n, 7, p, 1019,1-1019,6, 2002.
- SUZUKI, Nobuhiro et al, ROS and redox signalling in the response of plants to abiotic stress, **Plant, Cell & Environment**, v, 35, n, 2, p, 259-270, 2012.

- SUYAMA, Mikita; TORRENTS, David; BORK, Peer. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. **Nucleic acids research**, v. 34, n. suppl 2, p. W609-W612, 2006.
- TIAN, Fengxia et al, Genome-wide identification, classification, and expression analysis of 14-3-3 gene family in Populus, PloS one, v, 10, n, 4, p, e0123225, 2015.
- VUOSKU J, Sutela S, Kestila J, Jokela A, Sarjala T and Haggman H (2015) Expression of catalase and retinoblastoma-related protein genes associates with cell death processes in Scots pine zygotic embryogenesis, BMC Plant Biol 15:e88
- WANG, Kunbo et al, The draft genome of a diploid cotton *Gossypium raimondii*, **Nature genetics**, v, 44, n, 10, p, 1098-1103, 2012,
- XIE, Lulu et al. Phylogeny and Expression Analyses Reveal Important Roles for Plant PKS III Family during the Conquest of Land by Plants and Angiosperm Diversification. *Frontiers in Plant Science*, v. 7, 2016.
- XU, Bo; YANG, Ziheng. PAMLX: a graphical user interface for PAML. **Molecular biology and evolution**, v. 30, n. 12, p. 2723-2724, 2013.
- ZHANG, Gengyun et al, Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential, **Nature biotechnology**, v, 30, n, 6, p, 549-554, 2012,

Douglas Silva Domingues

Natália Fermino Calzado