UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
INSTITUTO DE BIOCIÊNCIAS - RIO CLARO

CIÊNCIAS BIOLÓGICAS

DANIEL RODRIGO DOVIGO

A FAMÍLIA GÊNICA DA ASCORBATO PEROXIDASE EM GENOMAS DE *COFFEA* SPP: GENÔMICA COMPARATIVA E ASPECTOS EVOLUTIVOS

Rio Claro
2018

DANIEL RODRIGO DOVIGO

A FAMÍLIA GÊNICA DA ASCORBATO PEROXIDASE EM GENOMAS DE *COFFEA* SPP: GENÔMICA COMPARATIVA E ASPECTOS EVOLUTIVOS

Orientador: Dr. 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 Licenciado em Ciências Biológicas.

Rio Claro
2018

D743f

Dovigo, Daniel Rodrigo
A família gênica da ascorbato peroxidase em genomas de Coffea spp : Genômica comparativa e aspectos evolutivos / Daniel Rodrigo Dovigo. -- Rio Claro, 2018
31 p. + 1 CD-ROM

Trabalho de conclusão de curso (Licenciatura - Ciências Biológicas) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Rio Claro
Orientador: Douglas Silva Domingues

1. Bioinformática. 2. Genômica Comparativa. 3. Metabolismo Antioxidante. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

AGRADECIMENTOS

Provavelmente essa é a parte mais difícil de ser escrita em nesse trabalho.

Se eu esquecer alguém... espero que possa me perdoar depois.

Em primeiro lugar, agradeço minha família consanguínea. Mãe, Pai e Tato... vocês construíram a base do homem que eu me tornei até o momento. Palavras nunca serão o bastante para agradecer todo esse amor. Antes das eleições eu tinha agradecido mais gente que eu já me arrependi. Mantenho minha avó Cida e meu vô Lando que estão presentes em saudade – meu coração é de vocês para sempre – e minha prima Gabrielly. Juntos nós somos a luta, gatinha!

Em segundo lugar o agradecimento à família das muitas irmãs e irmãos que a vida colocou no meu caminho: Juca, Gabi Naomi, as duas Tatas, Thales, Giulia, Rubiño, Mayara, Zor, Julia (Pãozinho), Marinoca, Carol, Prin, Casquinho, Denise, Henrique, Fino, Magui, Xu, Bo, Lisso, Ju Vilalba, Dé, Flor, Aline, Chassi e Cabô. Por mais que a distância nos separe fisicamente, sei que só cheguei onde estou por ter o amor e o suporte de todos vocês. Vocês, todos vocês, são a fonte da minha coragem e garra.

Agradeço também meu orientador, Douglas (Doug), que é uma grande inspiração (por mais que ele talvez não saiba disso). Também a todos os colegas do GeTransPlant, em especial a Natalia (Gulis) que tem me aguentado e apoiado firmemente desde o primeiro dia que embarcamos nessa jornada.

Um beijo enorme para Ligia e Helena, que moram comigo atualmente na Vila do Chaves, vocês têm sido a inesperada cerejinha do bolo neste ano difícil.

Para a minha turma, CBN14, com quem divido histórias e momentos, um abraço especial. Um carinho especial também para os outros colegas do curso e da universidade, de outras turmas, que me ajudaram a me formar. Só quem fez *desordem* na graduação sabe o valor do colega que se lembra de você que não é da turma dele.

Enfim, para todos os citados

muitíssimo obrigado,

de todo o coração.

Daniel.

1 Coríntios 13

1. Ainda que eu falasse as línguas dos homens e dos anjos, e não tivesse amor, seria como o metal que soa ou como o sino que tine.

2. E ainda que tivesse o dom de profecia, e conhecesse todos os mistérios e toda a ciência, e ainda que tivesse toda a fé, de maneira tal que transportasse os montes, e não tivesse amor, nada seria.

3. E ainda que distribuísse toda a minha fortuna para sustento dos pobres, e ainda que entregasse o meu corpo para ser queimado, e não tivesse amor, nada disso me aproveitaria.

4. O amor é sofredor, é benigno; o amor não é invejoso; o amor não trata com leviandade, não se ensoberbece.

5. Não se porta com indecência, não busca os seus interesses, não se irrita, não suspeita mal;

6. Não folga com a injustiça, mas folga com a verdade;

7. Tudo sofre, tudo crê, tudo espera, tudo suporta.

8. O amor nunca falha; mas havendo profecias, serão aniquiladas; havendo línguas, cessarão; havendo ciência, desaparecerá;

9. Porque, em parte, conhecemos, e em parte profetizamos;

10. Mas, quando vier o que é perfeito, então o que o é em parte será aniquilado.

11. Quando eu era menino, falava como menino, sentia como menino, discorria como menino, mas, logo que cheguei a ser homem, acabei com as coisas de menino.

12. Porque agora vemos por espelho em enigma, mas então veremos face a face; agora conheço em parte, mas então conhecerei como também sou conhecido.

13. Agora, pois, permanecem a fé, a esperança e o amor, estes três, mas o maior destes é o amor.

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 eles o sistema enzimático antioxidante.

Umas das enzimas presentes neste sistema são as Ascorbato Peroxidases (*APx*). As *APx* são enzimas que catalisam a decomposição H_2O_2 em água, utilizando o ascorbato como doador de elétrons. Elas são enzimas codificadas por uma família multigênica composta por 7 a 9 genes em plantas. A despeito do conhecimento da regulação da atividade enzimática e transcricional de *APx* em espécies do gênero *Coffea*, pouco se conhece sobre questões evolutivas da família gênica neste gênero.

Com isso, o objetivo principal do presente trabalho foi identificar e caracterizar evolutivamente a família gênica codificadora de *APx* no genoma de três espécies do gênero *Coffea*: *C. canephora, C. eugenioides* e *C. arabica*. Com a utilização de genes de *Arabidopsis thaliana* como ponto de partida para a análise de genes em *Coffea*, foram identificados seis genes codificantes de *APx* em *C. eugenioides* e sete genes em *C. canephora*, espécies diploides de *Coffea*. Em *C. arabica* foram identificados 13 genes, corroborando a da origem alotetraploide de *C. arabica* a partir da fusão de *C. canephora* e *C. eugenioides*. Foi possível ainda identificar a localização celular das enzimas codificadas, bem como padrões filogéticos para classificação de isoformas. A pressão evolutiva pelo qual passam os genes também foi analisada, constatando-se a seleção purificadora em todas as sequências. Dados públicos de expressão em *C. canephora* foram obtidos e observou-se a maior expressão do gene Cc_01G11800, representante das *APx,* com maior expressão em estames. Para as isoformas *APx-R*, representada pelo gene Cc_08g00640, e *APx-L*, pelo gene Cc_10g06720, a maior atividade foi constatada nas folhas. Os dados de expressão demonstram que a regulação molecular de cada gene codificante é distinta, resultando em um padrão complexo de expressão das *APx*.

PALAVRAS-CHAVE: bioinformática, genômica comparativa, metabolismo antioxidante.

SUMÁRIO

1. INTRODUÇÃO 6
2. OBJETIVOS 9
3. MATERIAL E MÉTODOS 9
   3.1. Genomas selecionados para análise 9
   3.2. Anotação da família *APx* em genomas vegetais 10
   3.3. Perfil transcricional da família *APx* em *C. canephora* com dados públicos de RNA-seq 12
4. RESULTADOS E DISCUSSÕES 13
   4.1. Número de cópias da família gênica e análise filogenética 13
   4.2. Perfis de seleção entre parálogos 21
   4.3. Perfil transcricional da família APx com base em dados públicos de RNA-seq 24
5. CONCLUSÃO 28
6. REFERÊNCIAS 29

1. INTRODUÇÃO

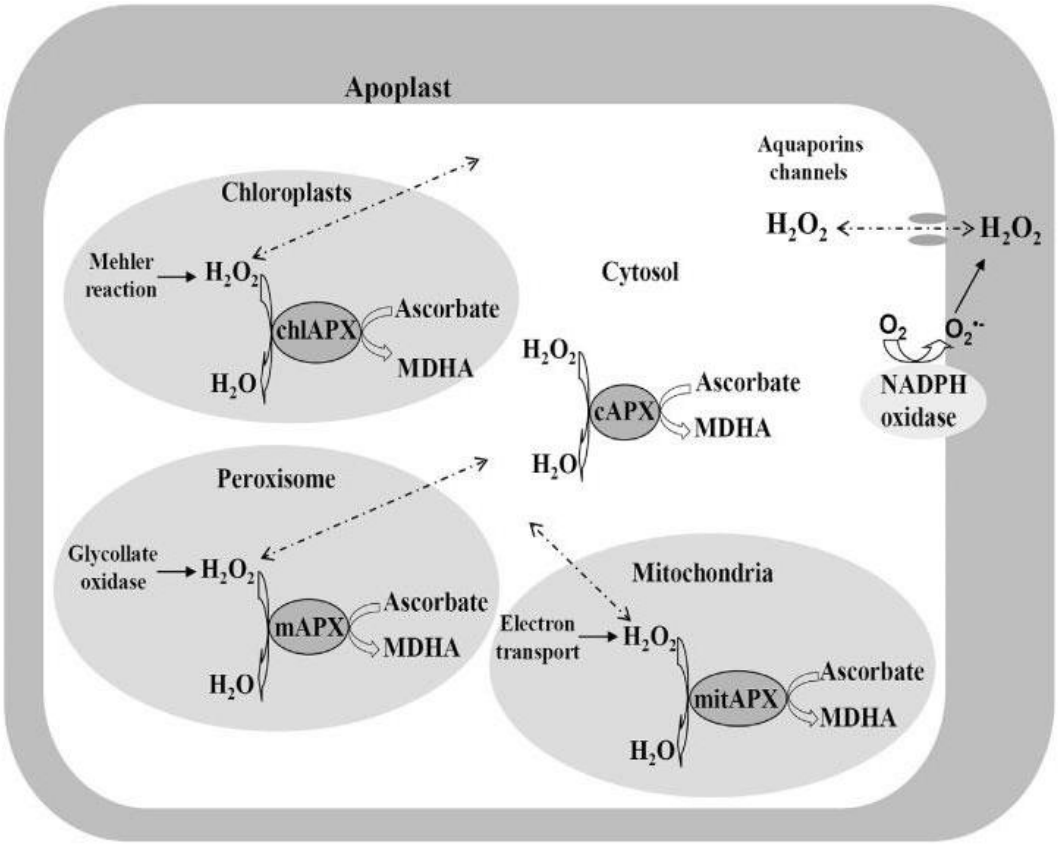
A respiração aeróbica foi uma evolução na obtenção de energia nos seres vivos, inclusive nas plantas. No entanto, durante os processos de fotossíntese e respiração, ampliaram-se os eventos de oxidação e redução de moléculas e o fluxo de elétrons no interior das células levando à geração de subprodutos decorrentes da redução parcial do oxigênio molecular - chamados de espécies reativas de oxigênio (EROs). Em plantas, espécies reativas de oxigênio também são produzidas durante a fotorrespiração. Esta redução parcial ocorre porque a conversão de meia molécula de O_2 à duas moléculas de água (H_2O) requer a transferência de quatro elétrons, entretanto, o oxigênio molecular tem preferência por vias univalentes de redução, o que leva à produção de moléculas parcialmente reduzidas (SCANDALIOS, 2002).

A produção de EROs durante o metabolismo celular é inevitável, mesmo em condições ótimas de crescimento. Entretanto, a produção destas moléculas é aumentada quando o organismo é exposto a adversidades bióticas ou abióticas como estresse hídrico ou salino, altas e baixas temperaturas, ataque de patógenos ou quando submetidos à tratamento com ácido abscísico ou com peróxido de hidrogênio, como revisados em Suzuki et al. (2012). Para evitar a superprodução de EROs no interior das células, as plantas contam com mecanismos que buscam diminuir concentração destas moléculas em situações de estresse (MITTLER, 2002). Um destes mecanismos é o sistema antioxidante, que apresenta componentes não-enzimáticos e enzimáticos que buscam limitar os níveis intracelulares de EROs, amenizando o dano oxidativo às estruturas celulares. Dentre os componentes enzimáticos destacam-se as enzimas superóxido dismutase, catalase, glutationa peroxidase e a ascorbato peroxidase (*APx*), foco de estudo no presente trabalho (CAVERZAN et al., 2012).

As ascorbato peroxidases são enzimas que contém heme como grupo prostético – isto é, um grupo contendo Fe+ ligado à enzima auxiliando em sua atividade biológica – e que catalisam a conversão de peróxido de hidrogênio a água, por meio da oxidação do substrato ascorbato. O processo de detoxificação do H_2O_2 catalisado pela enzima ascorbato peroxidase é acompanhada por uma série de reações que, em conjunto, formam uma das vias antioxidantes mais importantes em plantas - o ciclo do ascorbato-glutationa. Neste ciclo, o ascorbato e a glutationa são utilizados como fonte redutora para a detoxificação do H_2O_2 e os compostos oxidados são recuperados por várias reações com gasto de energia. Nas plantas, o ciclo do ascorbato-glutationa é particularmente importante nos cloroplastos, grandes produtores de

H_2O_2, mas opera também no citosol, nos peroxissomos e nas mitocôndrias, conforme ilustrado na Figura 1.

Figura 1. Ação e localização celular do ciclo da ascorbato-glutationa.



Fonte: Foyer e Noctor (2011)

Estas enzimas são codificadas por famílias multigênicas que costumam apresentar de 7 a 9 membros em plantas dicotiledôneas diplóides. As isoformas da família, classificadas filogeneticamente em *APx*, *APx-R* e *APx-L*, estão divididas de acordo com a sua localização subcelular e resíduos catalíticos. Em *APx-R* algumas substituições na sequência proteica, como a de triptofano por fenilalanina na região central da proteína e a ausência de uma arginina altamente conservada, importante na formação de sítios ativos da família *APx* e do sítio heme, conferem a classificação ao subgrupo. Estes genes não costumam apresentar eventos de duplicação por interagirem com o genoma das organelas onde se localizam e causarem alterações. Por isso, acabam tendo sua síntese controlada nas células (DUNAND et al., 2011). Já na isoforma *APx-L* a diferenciação acontece pela ausência quase total de vários dos resíduos importantes para a formação dos sítios catalíticos e heme observado nas *APx*. Apesar das alterações, *APx-R* e *APx-L* apresentam regiões com afinidades à substratos diferentes de *APx*,

conferindo a estes subgrupos capacidade distinta de atuação. Sua ação já foi observada, por exemplo, no sistema antioxidante das sementes (LAZZAROTTO et al., 2015). Quanto à localização, podem-se encontrar isoformas da família solúveis no citosol e no estroma do cloroplasto, enquanto outras, ancoradas às membranas, podem ser encontradas nos peroxissomos e nos tilacóides (TEIXEIRA et al., 2004; LAZZAROTTO et al., 2015).

Atualmente, com a crescente disponibilização de dados públicos de genomas e transcriptomas vegetais, uma nova e relevante frente de estudos em Biologia Vegetal consiste na anotação e compreensão da distribuição genômica e respostas transcricionais de genes antes parcialmente caracterizados em plantas. Por exemplo, em eucalipto, foi demonstrado que a família gênica ascorbato peroxidase passou por um processo de duplicação em tandem (LI et al., 2015).

No presente trabalho, uma estratégia similar foi utilizada para responder esta pergunta no gênero *Coffea*, integrando informações passíveis de cruzamento de dados de diferentes plataformas. Neste caso, existem informações da regulação transcricional e/ou enzimática de *APx* em cafeeiro (VIEIRA et al., 2012; MARRACCINI et al., 2012), que poderão ser analisadas sob um outro prisma, comparando análises genômicas com dados bioquímicos e transcricionais. Por exemplo, foi observado por Marraccini et al. (2012) que há diferença na expressão de *APx* em folhas entre os membros da família multigênica, quando em situação de estresse hídrico, demonstrando que nem todos os genes são expressos simultaneamente em situações de alta produção de EROs. Foi também observado que há uma maior atividade enzimática de *APx* nos frutos, em estágios progressivos de amadurecimento, em função da altitude onde os cafeeiros estão plantados (SANTOS et al., 2008). Uma maior atividade de *APx* em folhas foi identificada quando em solo deficiente em magnésio e nitrogênio (REIS et al., 2015; DA SILVA et al., 2017), sob a exposição de temperaturas elevadas e ainda maior com a combinação da deficiência nutricional a este fator (DA SILVA et al., 2017) . Em folhas com plantas submetidas à maior exposição de radiação solar pode-se observar também aumento na atividade de *APx*, mensuradas nos meses de agosto, outubro e dezembro com maior atividade registrada no mês de outubro, o mês com maior incidência solar (CHAVES et al., 2007).

O gênero *Coffea* pertence à família Rubiaceae e possui 124 espécies (DAVIS et al., 2011). Duas delas são de maior importância comercial: *Coffea canephora*, também conhecida como café Robusta (ou Conilon) e *Coffea arabica*. O Brasil é o maior exportador mundial há mais de 150 anos e a bebida preparada a partir das espécies citadas é apreciada em escala

mundial, com consumo que atinge cerca de 12 kg por pessoa ao ano em países europeus (ICO, 2018). As duas espécies de *Coffea* spp. com valor agronômico possuem ploidia distinta: enquanto *C. canephora* é uma espécie diplóide (2n=2x=22), *C. arabica* é um alotetraplóide (2n=4x=44), resultado de uma provável fusão entre os genomas de *C. canephora* e *C. eugenioides* (CLARINDO & CARVALHO, 2008; VIDAL et al., 2010).

C. canephora já tem seu genoma sequenciado (DENOEUD et al., 2014). As espécies *C. eugenioides* e *C. arabica* recentemente passaram a ter seus genomas sequenciados (DE KOCHKO, 2015). Estes dados foram utilizados no presente trabalho, o que permite compreender o panorama genômico e transcricional da família gênica *APx*.

2. OBJETIVOS

O objetivo geral do projeto é compreender aspectos genômicos, evolutivos e transcricionais da família gênica ascorbato peroxidase, *APx*, em genomas de três espécies do gênero *Coffea*: *C. canephora*, *C. eugenioides* e *C. arabica*.

Especificamente, foram analisados:

- O número de cópias da família gênica *APx, APx-R* e *APx-L* nas espécies estudadas;
- Os perfis de seleção da família gênica *APx* entre seus parálogos;

O perfil transcricional da família *APx* em *C. canephora* com dados públicos de RNA-seq.

3. MATERIAL E MÉTODOS

3.1. Genomas selecionados para análise

Os dados genômicos de *C. arabica*, *C. eugenioides* e *C. canephora* são provenientes do *Arabica Coffee Genome Consortium* (ACGC, http://tinyurl.com/z56hcjg), do qual o orientador é membro. Este consórcio tem disponível versões preliminares (não publicadas) do genoma destas espécies. O *download* dos dados foi feito do repositório FTP ftp.solgenomics.net, que

possui acesso restrito a membros do ACGC. Para *C. canephora*, também foram utilizados dados da plataforma genômica PLAZA 4.0 (VAN BEL et al., 2017).

Para estudos comparativos, foram também obtidos da plataforma PLAZA 4.0 dados genômicos das eudicotiledôneas *Arabidopsis thaliana, Coffea canephora, Eucalyptus grandis e Citrullus lanatus*, e das monocotiledôneas *Brachypodium distachyon*, *Musa acuminata* e *Setaria italica.*

3.2. Anotação da família *APx* em genomas vegetais

Os procedimentos gerais de anotação desta família gênica são baseados em artigos similares que utilizam informações genômicas com os mesmos objetivos para anotação de famílias gênicas (p.ex. LI et al., 2015; TIAN et al., 2015).

Sequências proteicas da família da ascorbato peroxidase em *Arabidopsis thaliana* foram utilizadas como ponto de partida para análise dos demais genomas vegetais. Inicialmente elas foram obtidas no banco de dados PeroxiBase (FAWAL et al., 2012) e também identificadas no banco PLAZA 4.0. A ocorrência dos domínios conservados da família *APx* nos genes de *A. thaliana* foi confirmada por meio de comparação contra a base de dados InterproScan (http://www.ebi.ac.uk/Tools/InterProScan/). Todas *APx* de *Arabidopsis* contém o domínio InterPro código IPR002207 (Class I peroxidase). Elas foram então utilizadas como sondas eletrônicas (*queries*) para procuras no banco de dados PLAZA 4.0, utilizando o programa BLAST (CAMACHO et al., 2009), versão tblastn, no proteoma predito das espécies selecionadas.

No caso das espécies de *Coffea* do consórcio ACGC, as procuras por BLAST foram realizadas utilizando o programa MyBlast (http://eln.iis.sinica.edu.tw/myblast/?q=home), contra as predições de sequências codificantes e proteínas, com realização de novas predições, quando necessário.

As regiões com alinhamento de *score* acima de 200 foram analisadas individualmente. Dentre os genes selecionados, os que apresentaram algum tipo de erro de anotação foram submetidos à análises nas ferramentas FGNESH+ e/ou FGENESH (disponível em

www.softberry.com), realizando nova predição de genes nas regiões cromossômicas selecionadas por BLAST. Foi utilizado como referencial que as sequências codificantes e/ou de proteína deveriam ter entre 90 e 110% do tamanho das sequências identificadas em *Arabidopsis*. A ocorrência dos domínios conservados da família *APx* foi confirmada por meio de comparação contra a base de dados InterproScan (http://www.ebi.ac.uk/Tools/InterProScan/) e todas as sequências analisadas neste trabalho contém o domínio IPR002207. Foram ainda identificados manualmente os resíduos proteicos característicos de *APx*, *APx-R* e *APx-L* (LAZZAROTTO et al., 2015).

Também foram verificados o peso molecular (MW), o ponto isoelétrico (pI) e a localização celular das proteínas identificadas. O MW e o pI foram determinados por meio do uso da ferramenta Compute pI/MW (http://web.expasy.org/cgi-bin/compute_pi/pi_tool) e a possível localização celular foi verificada utilizando os programas TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/), Plant-mPloc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) e CELLO (http://cello.life.nctu.edu.tw/). Uma anotação manual da localização celular também foi realizada levando em consideração os resultados obtidos nos programas supracitados. As *APx* indicadas como cloroplastidiais tiveram sua localização confirmada com o programa ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/).

As sequências de aminoácidos de *APx* obtidas foram alinhadas utilizando-se o programa MUSCLE (EDGAR, 2004), divididas por isotipo de *APx* (*APx, APX-R* e *APx-L*) e localização celular. Foi realizado um teste para identificar o melhor modelo de substituição de aminoácidos no programa MEGA X (KUMAR et al., 2018). As árvores filogenéticas também foram geradas neste programa utilizando-se a metodologia de máxima verossimilhança, *bootstrap* de 1000 réplicas e o modelo de substituição sugerido pelo programa.

Foram também avaliadas taxas de substituição sinônimas (dS) e não-sinônimas (dN) nas sequências codificantes de *APx*, utilizando-se a mesma estratégia de Jourda et al. (2014) em banana e Tian et al. (2015) em *Populus*. O alinhamento de proteínas foi convertido em alinhamento de nucleotídeos no programa PAL2NAL (http://www.bork.embl.de/pal2nal/).

Para verificar a pressão seletiva atuante nessas sequências foi calculada a razão dN/dS no programa MEGA X, no qual dN/dS = 1, dN/dS < 1 e dN/dS > 1 representam evolução neutra, seleção purificadora e seleção positiva Darwiniana, respectivamente. A determinação dos sítios

que sofrem seleção foi realizada com o conjunto de testes estatísticos disponível na plataforma Selecton (http://selecton.tau.ac.il/), com o modelo *mechanistic empirical combination* (MEC).

3.3. Perfil transcricional da família *APx* em *C. canephora* com dados públicos de RNA-seq

Os dados públicos de sequenciamento de RNA (RNA-seq) de diversos órgãos vegetativos e reprodutivos de *C. canephora* foram utilizados para inferências de expressão dos genes *APx*. Eles foram obtidos do site Coffee Genome Hub (http://coffee-genome.org), na unidade de *Reads per kilobase of transcript per million mapped reads* (RPKM). RPKM é uma unidade de medida utilizada para quantificar a expressão de genes em dados de RNA-seq, baseada na quantidade de *reads* encontradas no sequenciamento divido por 1.000.000 – número de escala utilizado – e pelo tamanho do gene analisado (https://www.rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/).

4. RESULTADOS E DISCUSSÕES

4.1. Número de cópias da família gênica e análise filogenética

Foi identificado um total de 77 sequências de APx além das 8 encontradas em *Arabidopsis thaliana*: 13 em *Coffea arabica*, 6 em *Coffea eugenioides*, 7 em *Coffea canephora* (em ambos os bancos de dados – PLAZA e ACGC), 9 em *Brachypodium distachyon*, 7 em *Citrulus lanatus*, 8 em *Eucalyptus grandis*, 10 em *Musa acuminata* e 10 em *Setaria italica*.

Deste total, foram identificadas 33 sequências parálogas de *APx* em 3 espécies de *Coffea:* 7 *de C. canephora* (tanto no banco de dados PLAZA como ACGC), 6 de *C. eugenioides* e 13 em *C. arabica.* As principais características dos genes identificados estão detalhadas na tabela 1.

As sequências de proteínas *APx* identificadas variam entre 227 e 487 aminoácidos.

Tabela 1. *Relação de genes APx identificados nas espécies estudadas. Genes marcados com * foram preditos com o programa Fgenesh+ e os marcados com ** foram preditos no programa Fgenesh.* ***Legenda:*** *Gene ID – Identificação dos genes nos bancos de dados; CDS – tamanho em nucleotídeos do quadro aberto de leitura; Prot. (Proteína) – tamanho em aminoácidos da proteína obtida; Isoforma – Classificação do gene na família Ascorbato Peroxidase; pI/MW – ponto isoelétrico/massa molecular; Localização Celular – Localização celular estimada pelas plataformas TargetP, Plant-mPloc, CELLO e a anotação final inferida pelo autor baseado nos resultados das plataformas três plataformas; Clorop. – Cloroplasto; Perox. – Peroxissomo; Citop. – Citoplasma; Mito. – Mitocôndria; Extracel. – Extracelular.*

					Localização Celular			
Gene ID	**CDS**	**Prot.**	**Isoforma**	**pI/MW (kD)**	**TargetP**	**Plant-mPloc**	**CELLO**	**Anotação Final**
				Arabidopsis thaliana				
AT3G09640	756	251	APX	5.87 / 28006.04	-	Perox.	Citop.	Perox.
AT1G07890	753	250	APX	5.72 / 27561.22	-	Perox.	Citop.	Perox.
AT4G09010	1191	396	APX-L	7.67 / 43476.17	-	Clorop.	Clorop.	Clorop.
AT4G35000	864	287	APX	6.47 / 31571.86	-	Perox.	Citop.	Perox.
AT4G08390	1119	372	APX	8.31 / 40407.32	Clorop.	Clorop./Mito.	Clorop.	Clorop.
AT1G77490	1281	426	APX	6.81 / 46092.30	Clorop.	Clorop./Mito.	Clorop.	Clorop.
AT4G32320	990	329	APX-R	8.99 / 36239.74	Clorop.	Citop.	Clorop.	Clorop.
AT4G35970	840	279	APX	8.80 / 30895.31	-	Perox.	Citop./Nuclear	Perox.
				Brachypodium distachyon				
Bradi1g16510	771	256	APX	5.28 / 27733.45	-	Perox.	Citop.	Perox.
Bradi5g03640	768	255	APX	9.15 / 27784.66	-	Perox.	Mito.	Mito.

Bradi5g20670	1095	364	APX-L	8.98 / 39083.35	Clorop.	Clorop.	Mito.	Clorop.
Bradi4g41180	945	314	APX	6.57 / 34434.82	Mito.	Clorop./Mito.	Mito./Citop.	Mito.
Bradi3g40330	900	299	APX-R	9.05 / 32123.73	Clorop.	Citop.	Clorop.	Clorop.
Bradi5g10490	1038	345	APX	8.77 / 37465.19	Clorop.	Clorop./Mito.	Clorop./Mito.	Clorop.
Bradi3g42340	870	289	APX	7.70 / 31542.93	Mito.	Perox.	Citop./Clorop.	Mito.
Bradi1g65820	753	250	APX	5.71 / 27496.21	-	Perox.	Citop.	Perox.
Bradi3g45700	1323	440	APX	5.72 / 47461.26	Clorop.	Clorop./Mito.	Clorop.	Clorop.
				Citrullus lanatus				
Cla015833.g	861	286	APX	6.67 / 31559.95	-	Perox.	Citop.	Perox.
Cla022327.g	999	332	APX-R	6.97 / 36227.39	Clorop.	Citop.	Clorop.	Clorop.
Cla013927.g	1344	447	APX	7.63 / 48721.25	Clorop.	Clorop./Mito.	Clorop.	Clorop.
Cla023026.g	1050	349	APX-L	6.06 / 38057.14	Clorop.	Clorop.	Extracel./Clorop.	Clorop.
Cla013254.g	753	250	APX	5.60 / 27274.98	-	Perox.	Citop.	Perox.
Cla008291.g*	774	258	APX	5.24 / 28390.96	-	Perox.	Citop.	Perox.
Cla014301.g	891	296	APX	8.30 / 32914.62	-	Perox.	Citop.	Perox.
				Eucalyptus grandis				
Eucgr.C01740	1110	369	APX-R	8.44 / 39616.06	Clorop.	Citop.	Clorop.	Clorop.
Eucgr.F00373	1071	356	APX-L	6.50 / 38364.43	Clorop.	Clorop.	Clorop.	Clorop.
Eucgr.F04344	1341	446	APX	8.71 / 48225.91	Clorop.	Clorop./Mito.	Clorop.	Clorop.
Eucgr.A01180	750	249	APX	6.07 / 27459.21	-	Perox.	Citop.	Perox.
Eucgr.B02456	750	249	APX	5.29 / 27271.91	-	Perox.	Citop.	Perox.
Eucgr.J02463	800	267	APX	6.19 / 29456.58	Mito.	Perox.	Citop.	Perox.
Eucgr.I01408	864	287	APX	6.67 / 31522.91	-	Perox.	Citop./Clorop.	Perox.
Eucgr.J00965	753	250	APX	6.08 / 27334.21	-	Perox.	Citop.	Perox.
				Musa acuminata				
MAC05G1291	864	287	APX	6.40 / 31693.08	-	Perox.	Citop.	Perox.
MAC08G1647	1122	373	APX	6.72 / 40641.69	Clorop.	Clorop./Mito.	Clorop./Mito.	Clorop.
MAC08G2470	807	268	APX-R	8.88 / 29127.65	Mito.	Citop.	Mito./Citop.	Mito.
MAC03G1031	867	288	APX	6.76 / 31738.08	-	Perox.	Citop.	Perox.
MAC03G0759	1056	351	APX-L	8.78 / 38173.53	Mito.	Clorop.	Mito./Clorop.	Mito.
MAC06G0853	750	249	APX	5.13 / 27307.73	-	Perox.	Citop.	Perox.
MAC05G0325	750	249	APX	5.31 / 27374.02	-	Perox.	Citop.	Perox.
MAC10G1518	1320	439	APX	8.04 / 48557.44	Clorop.	Clorop./Mito.	Clorop.	Clorop.
MAC06G2184	930	309	APX	6.51 / 33503.51	Clorop.	Clorop./Mito.	Mito./Citop.	Clorop.
MAC03G0014	750	249	APX	5.30 / 27286.90	-	Perox.	Citop.	Perox.
				Setaria italica				
Seita.9G444200	753	250	APX	5.69 / 27329.02	-	Perox.	Citop.	Perox.
Seita.7G225300	1062	353	APX-L	8.17 / 38125.02	Clorop.	Clorop.	Mito./Clorop.	Clorop.
Seita.7G023900	879	292	APX	6.67 / 31981.42	-	Perox.	Citop.	Perox.
Seita.1G192100	1437	478	APX	5.18 / 51198.41	Clorop.	Clorop./Mito.	Clorop.	Clorop.

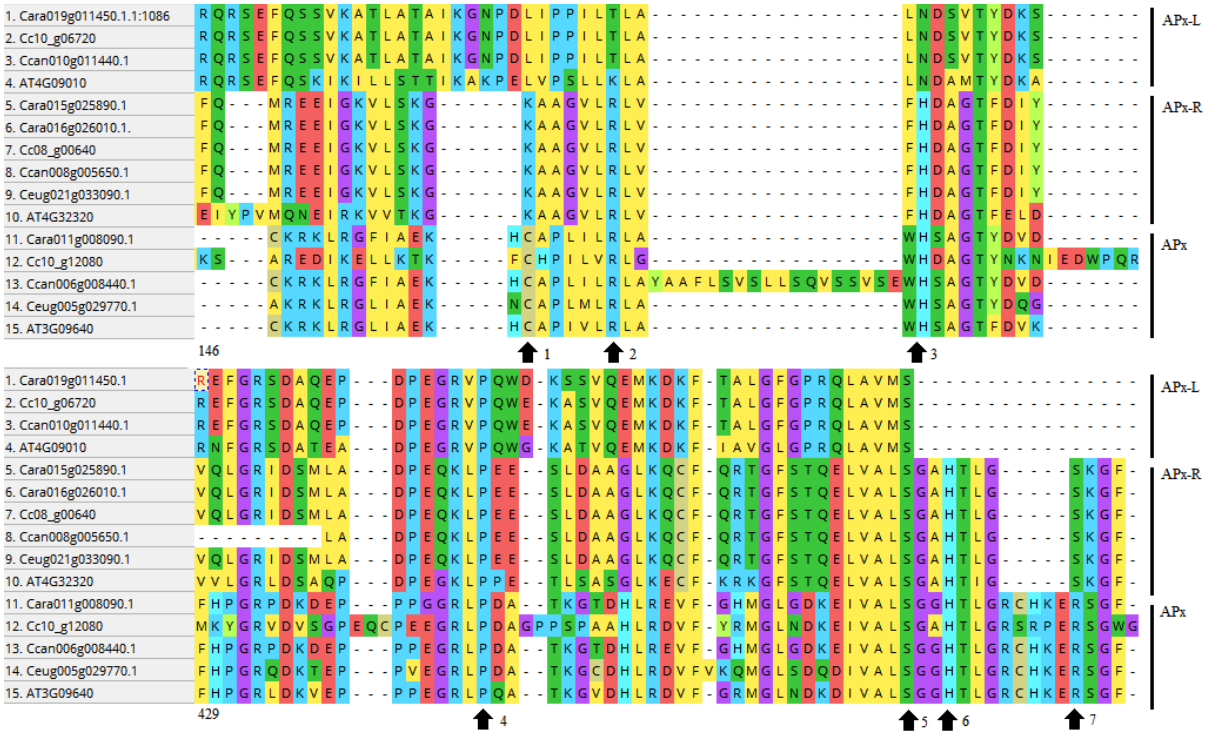
Seita.7G102200	1029	342	APX	8.79 / 36981.69	Clorop.	Clorop./Mito.	Mito.	Clorop.
Seita.2G444000	753	250	APX	5.18 / 27160.69	-	Perox.	Citop.	Perox.
Seita.7G283600	936	311	APX	7.13 / 33942.50	Mito.	Clorop./Mito.	Mito.	Mito.
Seita.3G268800**	942	313	APX	10.11 / 33965.66	Clorop.	Clorop.	Mito./Nuclear	Clorop.
Seita.6G210000	1005	334	APX-R	5.92 / 36045.74	Clorop.	Citop.	Clorop.	Clorop.
Seita.6G239800	867	288	APX	7.02 / 31419.81	-	Perox.	Citop.	Perox.
				Coffea canephora (PLAZA)				
Cc10_g06720	969	322	APX-L	6.85 / 35016.69	Clorop.	Clorop.	Clorop.	Clorop.
Cc10_g12080	1464	487	APX	6.81 / 52867.95	Clorop.	Clorop./Mito.	Clorop.	Clorop.
Cc07_g11020	843	280	APX	7.06 / 31153.48	-	Perox.	Citop.	Perox.
Cc02_g16260	870	289	APX	7.10 / 31845.29	-	Perox.	Citop.	Perox.
Cc08_g00640	996	331	APX-R	8.56 / 36146.63	Clorop.	Citop.	Clorop.	Clorop.
Cc06_g03490	732	243	APX	5.70 / 26961.57	-	Perox.	Citop.	Perox.
Cc01_g11800	753	250	APX	5.53 / 27306.95	-	Perox.	Citop.	Perox.
				Coffea canephora (ACGC)				
Ccan006g008440.1	786	261	APX	5.60 / 28830.67	-	Citop.	Citop.	Perox.
Ccan001g024320.1	864	287	APX	5.49 / 31717.17	-	Citop.	Citop.	Perox.
Ccan007g015520.1	837	279	APX	6.67 / 31025.31	-	Perox.	Citop.	Perox.
Ccan002g021350.1	870	289	APX	7.10 / 31845.29	-	Perox.	Citop.	Perox.
Ccan010g011440.1	1098	366	APX-L	8.92 / 40175.75	-	Clorop.	Extracel./Clorop.	Clorop.
Ccan010g020540.1**	1206	401	APX	5.18 / 43879.44	-	Citop.	Citop.	Perox.
Ccan008g005650.1	870	289	APX-R	8.88 / 31883.64	Clorop.	Citop.	Clorop.	Clorop.
				Coffea eugenioides				
Ceug005g029770.1	807	268	APX	5.83 / 29110.01	-	Perox.	Citop.	Perox.
Ceug046g016440.1	843	280	APX	7.06 / 31123.45	-	Perox.	Citop.	Perox.
Ceug044g023620.1	861	286	APX	8.60 / 31544.92	Mito.	Perox.	Mito./Citop.	Mito.
Ceug083g021580.1	1314	437	APX	7.66 / 47646.12	Clorop.	Clorop./Mito.	Clorop.	Clorop.
Ceug083g012540.1	1170	389	APX-L	8.64 / 42693.41	-	Clorop.	Clorop./Extracel.	Clorop.
Ceug021g033090.1	1014	337	APX-R	8.25 / 36868.45	Clorop.	Citop.	Clorop.	Clorop.
				Coffea arabica				
Cara011g008090.1	732	244	APX	5.71 / 27044.70	-	Perox.	Citop.	Perox.
Cara002g020870.1	879	293	APX	5.78 / 32021.27	-	Perox.	Citop./Clorop.	Perox.
Cara012g008170.1	681	227	APX	5.53 / 25239.67	-	Citop.	Citop.	Perox.
Cara001g019190.1	831	277	APX	5.32 / 30412.61	-	Perox.	Citop.	Perox.
Cara013g014130.1	837	279	APX	6.41 / 30986.28	-	Perox.	Citop.	Perox.
Cara003g021250.1	867	289	APX	7.10 / 31845.29	-	Perox.	Citop.	Perox.
Cara004g021420.1	867	289	APX	7.10 / 31887.34	-	Perox.	Citop./Clorop.	Perox.
Cara014g015810.1	978	326	APX	9.21 / 36750.28	-	Perox.	Citop./Mito.	Perox.
Cara019g018980.1*	1251	417	APX	7.10 / 45354.45	Clorop.	Clorop./Mito.	Clorop.	Clorop.
Cara019g011450.1	1086	362	APX-L	8.65 / 39450.89	-	Clorop.	Mito.	Clorop.
Cara020g019680.1**	1143	381	APX	5.63 / 42137.63	-	Clorop./Mito.	Citop.	Clorop.

Cara015g025890.1	1047	349	APX-R	7.62 / 38175.06	-	Citop.	Clorop.	Clorop.
Cara016g026010.1	1101	367	APX-R	7.67 / 39897.73	-	Citop.	Clorop.	Clorop.

O número de sequências por espécie variou entre 6 e 13. *Coffea arabica*, uma espécie alotetraploide derivado da provável fusão entre *C. eugenioides* e *C. canephora* (CLARINDO & CARVALHO, 2008; VIDAL et al., 2010), apresentou um total equivalente à soma do número encontrado em suas espécies de origem – 6 e 7 respectivamente.

Pela observação direta do alinhamento de sequências de aminoácidos de *APx* foi possível identificar os resíduos de aminoácidos típicos de *APx-R* e *APx_L* (Lazarotto et al.,2015). Em todas as espécies foi possível identificar as isoformas *APx, APx-R* e *APx-L*. Na figura 2 é destacada a identificação destas isoformas em espécies de *Coffea*.

Figura 2. *Alinhamento de APx (em aminoácidos) de espécies de Coffea e de A. thaliana, com representantes de todas as isoformas de APx. Em 1 e 7 são indicados sítios de ligação com ascorbato, formados por cisteína e arginina respectivamente, presentes apenas em APx. Em 2, 3, 5 e 6 são indicados motivos relacionados à atividade catalítica das APx; em 2 a ausência de arginina nas APx-L; em 3 a substituição de triptofano em APx por fenilalanina em APx-R; em 5 uma região conservada de serinas e em 6 uma região de argininas conservadas apenas em APx. Em 4 um resíduo de prolina, responsável pela ligação das APx com anéis aromáticos, presentes em todas as isoformas. As observações são baseadas em Lazarotto et al. (2015).*

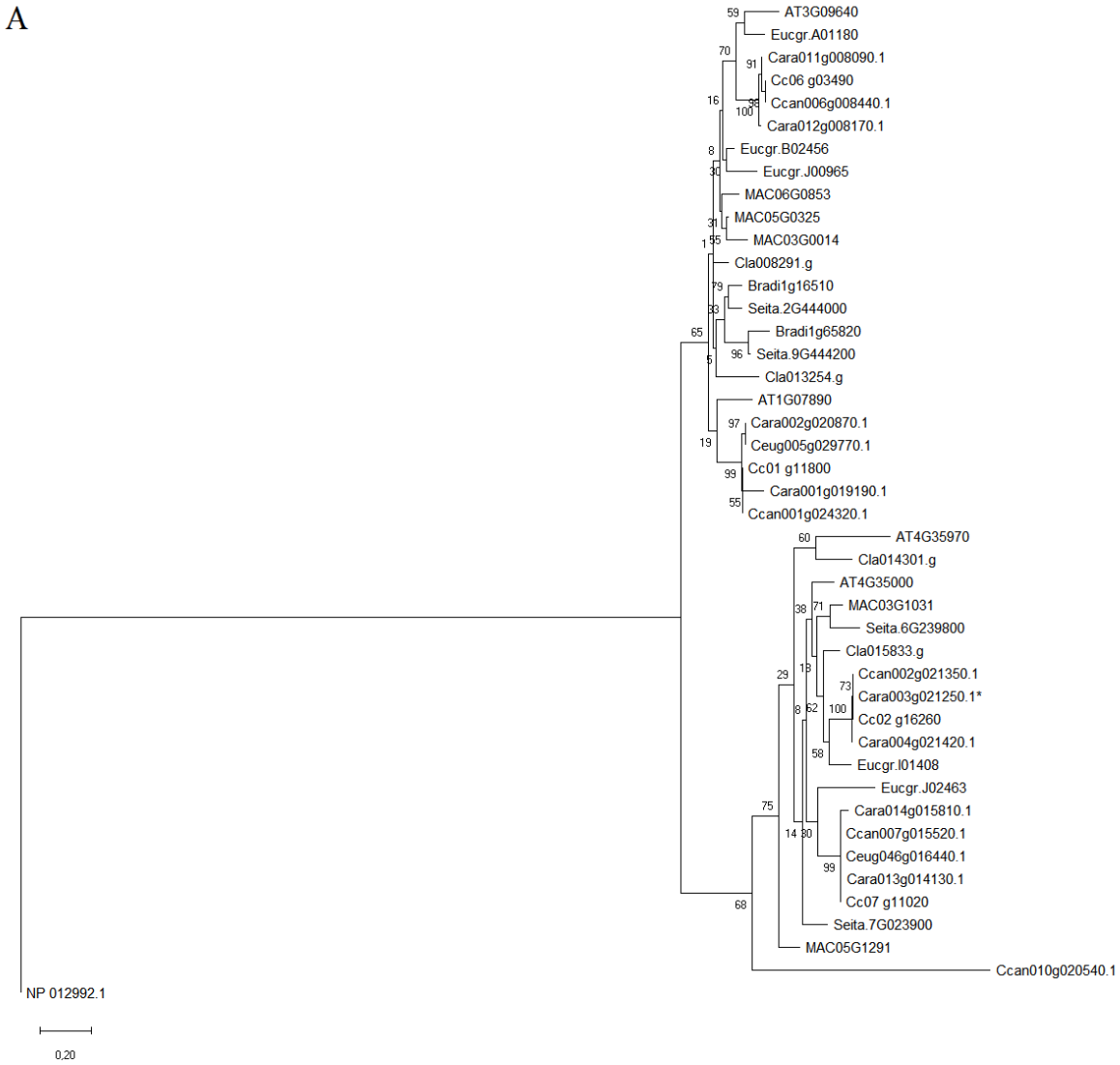


Fonte: elaborado pelo autor

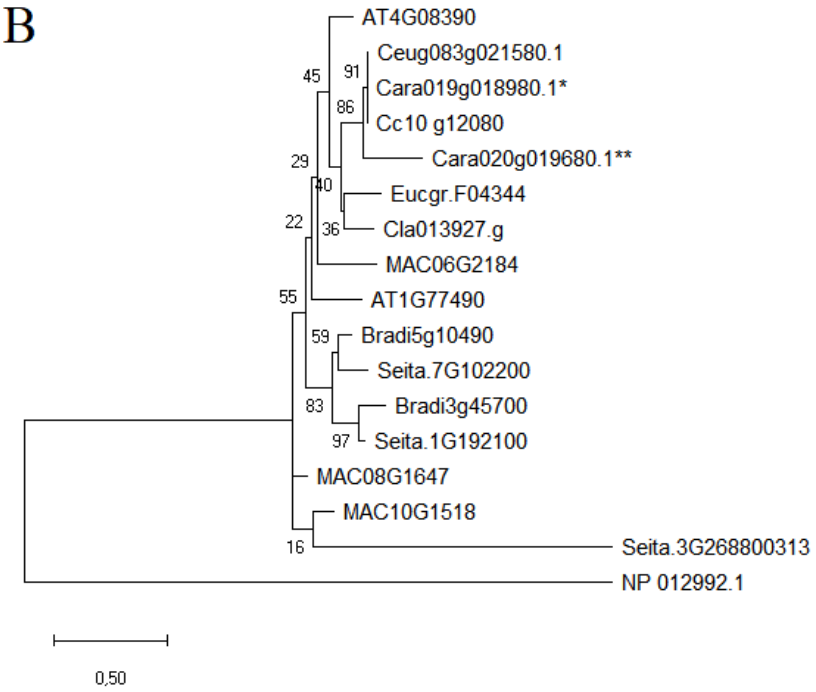
As sequências apresentadas na tabela 1 foram utilizadas na construção árvores filogenéticas. Problemas de baixa confiabilidade estatística impossibilitaram a análise conjunta de todas as 85 sequências. Por isso adotamos a estratégia de construir árvores de acordo com as localizações celulares e as isoformas da família gênica. Foi feita uma análise filogenética de sequências de peroxissomos e citoplasma (Figura 3A), uma análise filogenética de sequências de cloroplastos (Figura 3B), uma de sequências com localização celular mitocondrial (Figura 3C) e uma análise de sequências das isoformas *APx-R* e *APx-L* (Figura 3D).

***Figura 3**. Árvores filogenéticas construídas a partir das sequências proteicas dos genes APx. Foi utilizado o método de Máxima Verossimilhança, bootstrap de 1000 replicatas. Foram utilizados apenas sítios da matriz com ao menos 95% das posições sem gaps. Em A: Sequências de Peroxissomos e Citoplasma no modelo LG (LE e GASCUEL. 2008). Em B: Sequências de Cloroplastos no modelo JTT (JONES et al., 1992). Em C: Sequências de Mitocôndrias no modelo WAG (WHELAN e GOLDMAN, 2001). Em D: Sequências de APx-R e APx-L no modelo JTT. Em todas as 4 árvores foi utilizada uma peroxidase de Saccharomyces cereviseae como grupo externo (acesso GenBank NP 012992.1), na qual a árvore foi enraizada.*

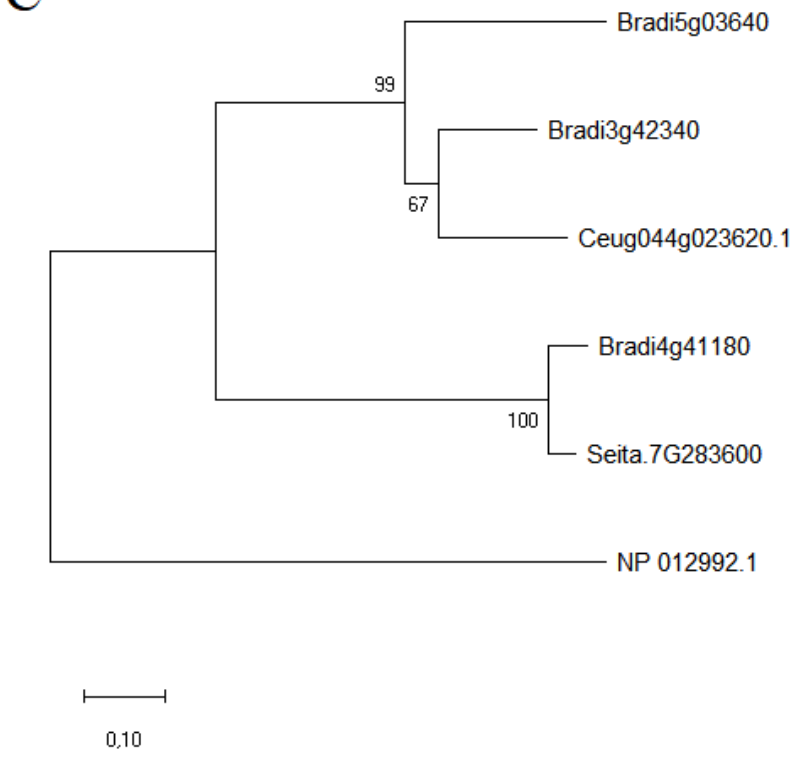
A

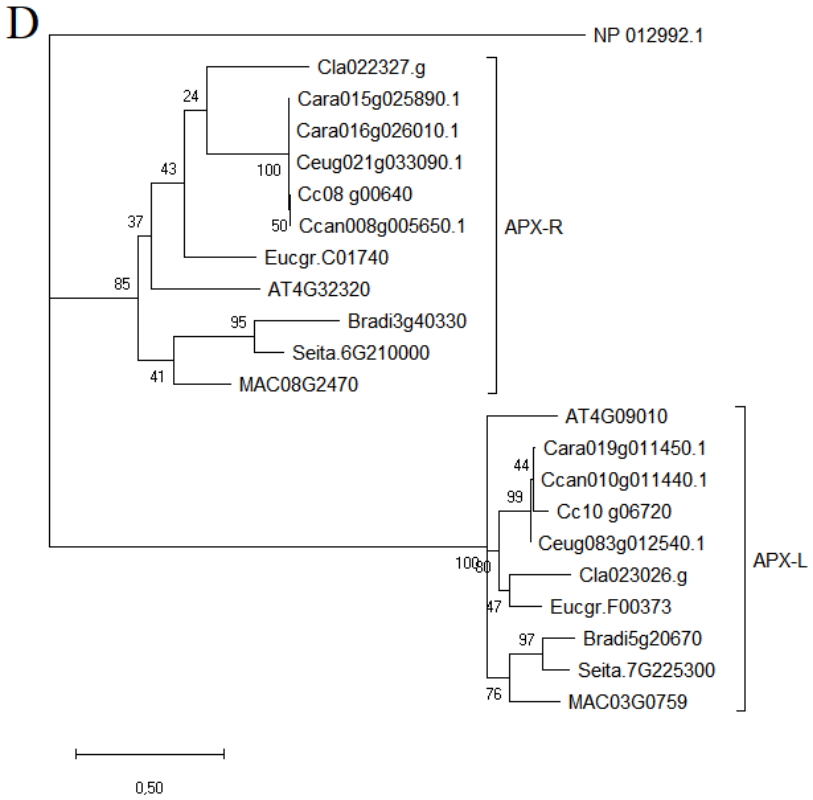

AT3G09640
Eucgr.A01180
Cara011g008090.1
Cc06 g03490
Ccan006g008440.1
Cara012g008170.1
Eucgr.B02456
Eucgr.J00965
MAC06G0853
MAC05G0325
MAC03G0014
Cla008291.g
Bradi1g16510
Seita.2G444000
Bradi1g65820
Seita.9G444200
Cla013254.g
AT1G07890
Cara002g020870.1
Ceug005g029770.1
Cc01 g11800
Cara001g019190.1
Ccan001g024320.1
AT4G35970
Cla014301.g
AT4G35000
MAC03G1031
Seita.6G239800
Cla015833.g
Ccan002g021350.1
Cara003g021250.1*
Cc02 g16260
Cara004g021420.1
Eucgr.I01408
Eucgr.J02463
Cara014g015810.1
Ccan007g015520.1
Ceug046g016440.1
Cara013g014130.1
Cc07 g11020
Seita.7G023900
MAC05G1291
Ccan010g020540.1
NP 012992.1
0,20

B


AT4G08390
Ceug083g021580.1
Cara019g018980.1*
Cc10 g12080
Cara020g019680.1**
Eucgr.F04344
Cla013927.g
MAC06G2184
AT1G77490
Bradi5g10490
Seita.7G102200
Bradi3g45700
Seita.1G192100
MAC08G1647
MAC10G1518
Seita.3G268800313
NP 012992.1
45
91
86
29
40
22
36
55
59
83
97
16
0,50

C


Bradi5g03640
Bradi3g42340
Ceug044g023620.1
Bradi4g41180
Seita.7G283600
NP 012992.1
99
67
100
0,10



Na árvore construída com sequências de peroxissomos e citoplasma é possível identificar dois clados principais. As sequências de *A. thaliana, E. grandis* e *B. distachyon* utilizadas no presente trabalho foram também utilizadas em Ozyigit et al. (2016), permitindo a comparação de resultados. Naquele trabalho, o clado contendo AT3G09640 e AT1G07890 agrupa sequências menores. Já o clado que contém as sequências AT4G35000 e AT4G35970 agrupa as maiores. Com isso, pudemos realizar caracterização genômica comparativa das *APx* a partir do agrupamento das sequencias das três espécies e sua localização nos dois grandes clados (Figura 3A). A árvore das sequencias cloroplastidiais apresenta dados similares aos obtidos por Ozyigit et al. (2016); as sequências AT4G08390 e AT1G77490 foram descritas como longas e no presente trabalho também aparecem agrupadas (Figura 3B).

Segundo Ozyigit et al. (2016), as sequências AT4G09010 e AT4G32320 foram indicadas como enzimas *APx* cloroplastidiais de tamanhos médio e curto em comparação com outras sequências. No presente trabalho, tais sequências foram classificadas como isoformas *APx-L* e *APx-R*, respectivamente, conforme descrito por Lazzarotto et al. (2015). Foi possível identificar as isoformas *APx-L* e *APx-R* no gênero *Coffea,* com uma cópia nas espécies diploides e *Coffea* e duas cópias de *APx-R* em *C. arabica.*

É possível observar que sequências de *E. grandis* são próximas filogeneticamente das sequências de *Coffea*, um indicativo da proximidade na evolução da família *APx* entre tais espécies.

Não foi possível distinguir em nenhuma das árvores um padrão entre espécies de monocotiledôneas e eudicotiledôneas, o que corrobora a teoria que o ancestral de angiospermas já possuía mais de um grupo de *APx* (Lazarotto et al., 2015).

4.2. Perfis de seleção entre parálogos

Verificamos que os genes da família gênica sofrem pressão purificadora, isto é, dN/dS<1 (Tabela 2). Os maiores níveis de pressão purificadora são observados nos genes cloroplastidiais (Tabela 2). No entanto as maiores taxas de substituição foram observadas em *APx* de localização mitocondrial.

Tabela 2. *Taxas de substituição sinônimas (DS) e não-sinônimas (DN) dos genes estudados e as características observadas nos resultados do programa Selecton.*

	DS	DN	DN/DS	Selecton
Peroxissomo	1,91 ± 0,06	0,24 ± 0,02	0,1257	Maioria purificador
Mitocondria	1,95 ± 0,17	0,37 ± 0,03	0,1897	Maioria purificador
Cloroplasto	1,67 ± 0,08	0,14 ± 0,01	0,0838	Início positivo, fim purificador
APx-L	1,67 ± 0,12	0,22 ± 0,02	0,1317	Início positivo, fim purificador
Apx-R	1,32 ± 0,12	0,23 ± 0,02	0,1742	Início positivo, fim purificador

A análise do perfil de seleção individual de cada um dos aminoácidos também foi realizada, uma vez que em diferentes regiões de uma proteína podem sofrer pressões evolutivas distintas. Foi constatado na maior parte das sequências uma tendência de pressão purificadora. A pressão de seleção positiva aparece apenas nas regiões iniciais das sequências de Cloroplasto, *APx-R* e *APx-L* (Fig. 4C, 4D e 4E) indo até aproximadamente a posição 95. É nessa região onde localizam-se os peptídeos sinais, regiões responsáveis por destinar as proteínas para sua localização (cloroplastidial nas três sequências analisadas) e que podem explicar o perfil de seleção positiva.

Figura 4. *Perfil de seleção individual dos aminoácidos utilizando sequências de A. thaliana (ou com maior score frente aos genes da espécie) como referência. Em A: Sequências de Peroxissomo com AT1G07890 como referência. Em B: Sequências de Mitocôndria com Seita.7G283600 como referência. Em C: Sequências de Cloroplasto com AT1G77490 como referências. Em D: Sequências de APx-R com AT4G32320 como referências. Em E: Sequências de APx-L com AT4G09010 como referência.*



C

```
1               11              21              31              41
MSVSLSAASH LLCSSTRVSL SPAVTSSSSS PVVALSSSTS PHSLGSVASS
51              61              71              81              91
SLFPHSSFVL QKKHPINGTS TRMISPKCAA SDAAQLISAK EDIKVLLRTK
101             111             121             131             141
FCHPILVRLG WHDAGTYNKN IEEWPLRGGA NGSLRFEAEL KHAANAGLLN
151             161             171             181             191
ALKLIQPLKD KYPNISYADL FQLASATAIE EAGGPDIPMK YGRVDVVAPE
201             211             221             231             241
QCPEEGRLPD AGPPSPADHL RDVFYRMGLD DKEIVALSGA HTLGRARPDR
251             261             271             281             291
SGWGKPETKY TKTGPGEAGG QSWTVKWLKF DNSYFKDIKE KRDDDLLVLP
301             311             321             331             341
TDAALFEDPS FKNYAEKYAE DVAAFFKDYA EAHAKLSNLG AKFDPPEGIV
351             361             371             381             391
IENVPEKFVA AKYSTGKKEL SDSMKKKIRA EYEAIGGSPD KPLPTNYFLN
401             411             421
IIIAIGVLVL LSTLFGGNNN SDFSGF
```

D

```
1               11              21              31              41
MTTTTASLVK TFLFRCDSFS SFKFKCKFES PAKTRLLSPA TEKHVVRSSR
51              61              71              81              91
AWRIRCLSDD PGSSHVFVAS RRKMVVLLST VQLLSHMLPQ NGNAAEIYPV
101             111             121             131             141
MQNEIRKVVT KGKAAGVLRL VFHDAGTFEL DDHSGGINGS IAYELERPEN
151             161             171             181             191
IGLKKSLKVL AKAKVKVDEI QPVSWADMIS VAGSEAVSIC GGPTIPVVLG
201             211             221             231             241
RLDSAQPDPE GKLPPETLSA SGLKECFKRK GFSTQELVAL SGAHTIGSKG
251             261             271             281             291
FGDPTVFDNA YYKILLEKPW TSTSKMTSMV GLPSDHALVQ DDECLRWVKR
301             311             321
YAEDQDKFFE DFTNAYIKLV NSGAKWNML
```

E

```
1            11           21           31           41
MDPYDWEIWI TENPQKCILH KTKTHSPFSS DNSFSFLLSS LSHHKPEMGG
51           61           71           81           91
VSFLSTVPSF TNTTNHQHLT TLSSSSHRSA VIRCSKIEPQ VSGESLAFHR
101          111          121          131          141
RDVLKLAGTA VGMELIGNGF INNVGDAKAA DLNQRRQRSE FQSKIKILLS
151          161          171          181          191
TTIKAKPELV PSLLKLALND AMTYDKATKS GGANGSIRFS SELSRAENEG
201          211          221          231          241
LSDGLSLIEE VKKEIDSISK GGPISYADII QLAGQSAVKF TYLASAIRKC
251          261          271          281          291
GGNEEKGNLL YTAYGSAGQW GLFDRNFGRS DATEADPEGR VPQWGKATVQ
301          311          321          331          341
EMKDKFIAVG LGPRQLAVMS AFLGPDQAAT EQLLATDPQV APWVQKYQRS
351          361          371          381          391
RETVSQTDYE VDLITAFTKL SCLGQQINFE AYTYPVERIN LSKLKL
```

Fonte: Selecton (http://selecton.tau.ac.il/)

4.3. Perfil transcricional da família APx com base em dados públicos de RNA-seq

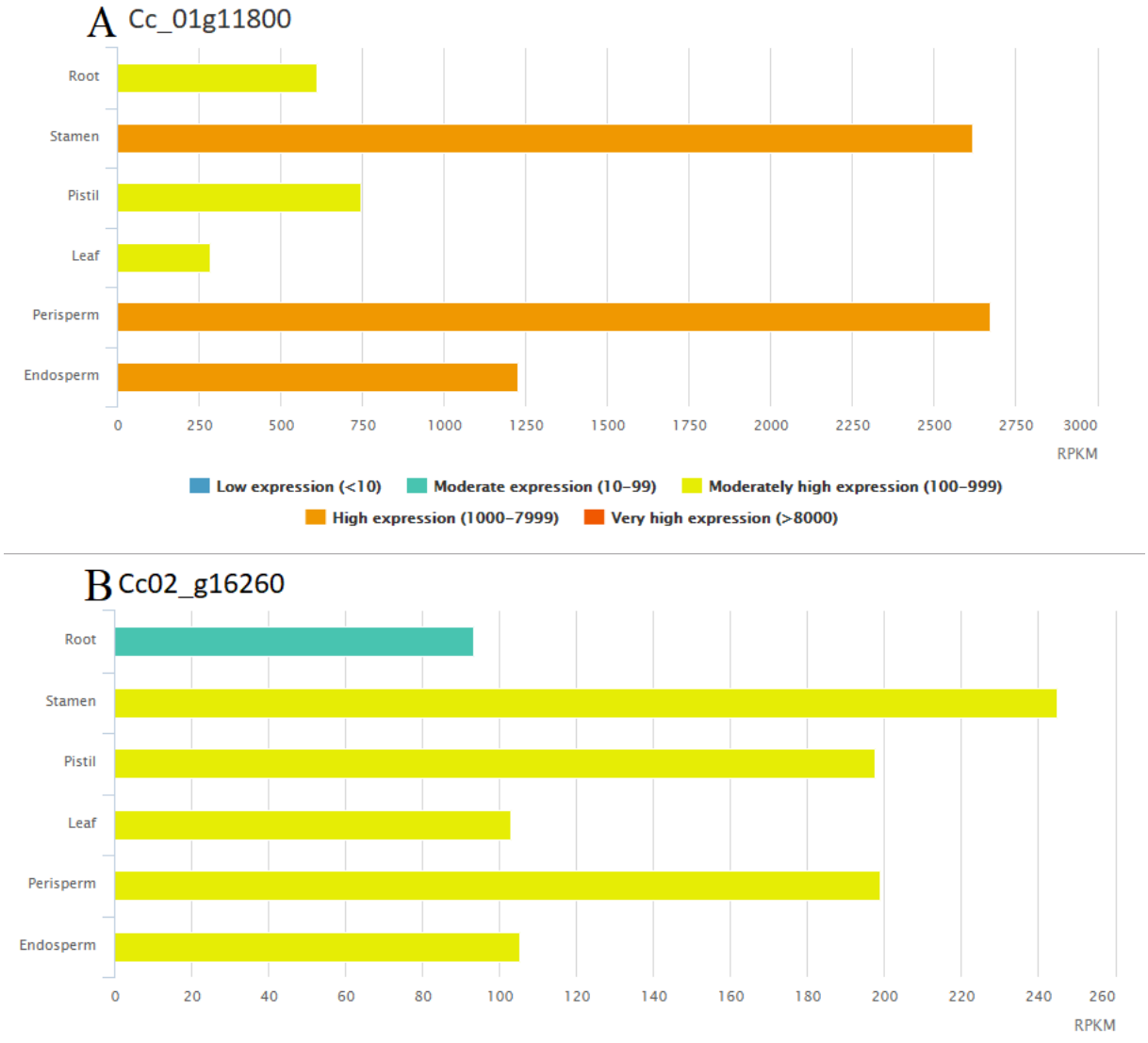
Com base nos dados públicos de expressão dos genes de *C. canephora* notou-se que o gene Cc_01G11800 é o mais expresso nesta espécie (Figura 5). Tal gene é responsável pela síntese de uma proteína de *APx* em peroxissomos e sua expressão é mais alta em estames, no perisperma e em folhas. Em folhas também ocorre uma alta expressão do gene Cc_02G16260.

As isoformas *APx-R* (Cc_08g00640) e *APx-L* (Cc_10g06720) aparecem mais expressas nas folhas. A expressão de *APx*, de forma geral, é maior nas folhas e nos estames. Sendo ambos são órgãos de bastante atividade metabólica em plantas, é compreensível uma expressão elevada de enzimas relacionadas ao sistema antioxidante.

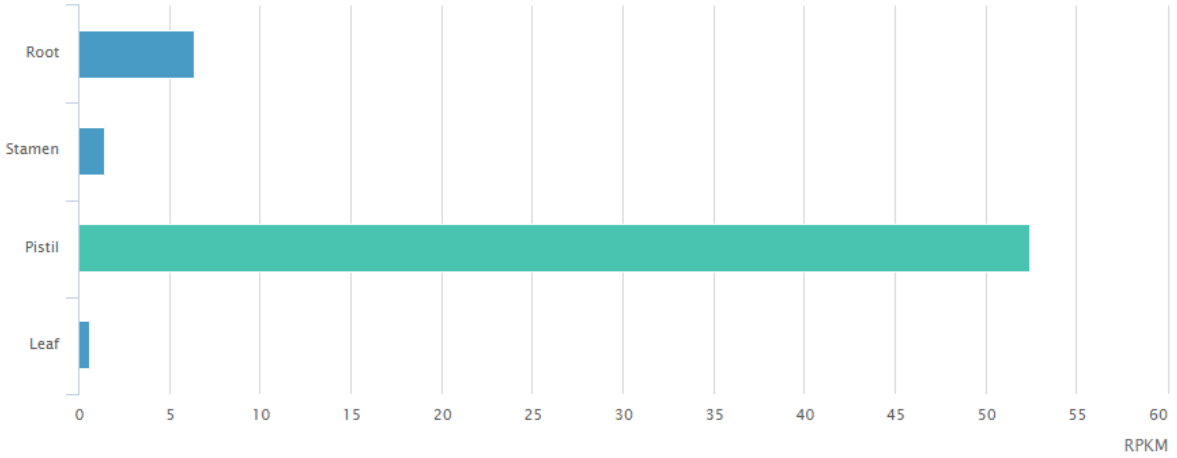
O gene com maior expressão aqui analisado, Cc_01G11800, aparece reunido no mesmo clado (Figura 3A) que AT1G07890, que teve sua expressão analisada por Ozyigit et al. (2016). Este gene, por sua vez, apresenta maior expressão em raízes enquanto o gene de *C. canephora* possui maior expressão em estame. Além desta diferença, ambos apresentam expressão

moderadamente alta no endosperma. Os dados de perisperma não foram avaliados em *A. thaliana*, impossibilitando a comparação.

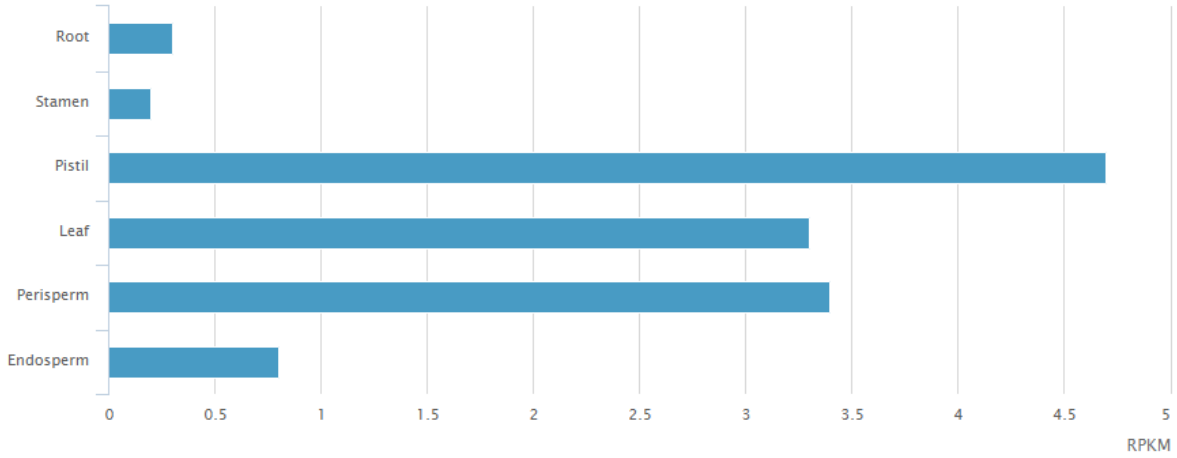
Figura 5. Expressão (em RPKM) dos genes da família gênica *APx* em *C. canephora* nos órgãos vegetais. As figuras de A a G mostram individualmente a expressão para cada um dos genes.



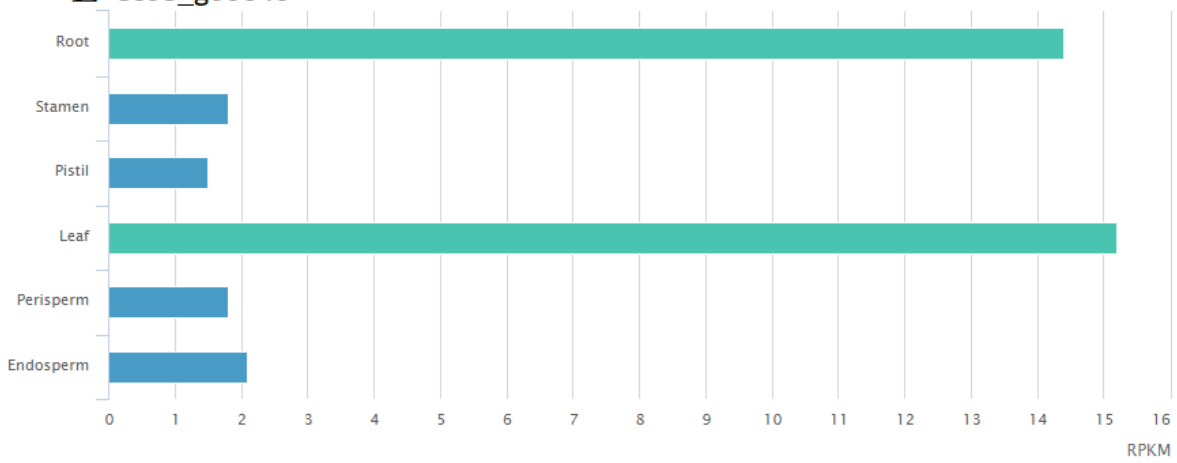
C Cc06_g03490



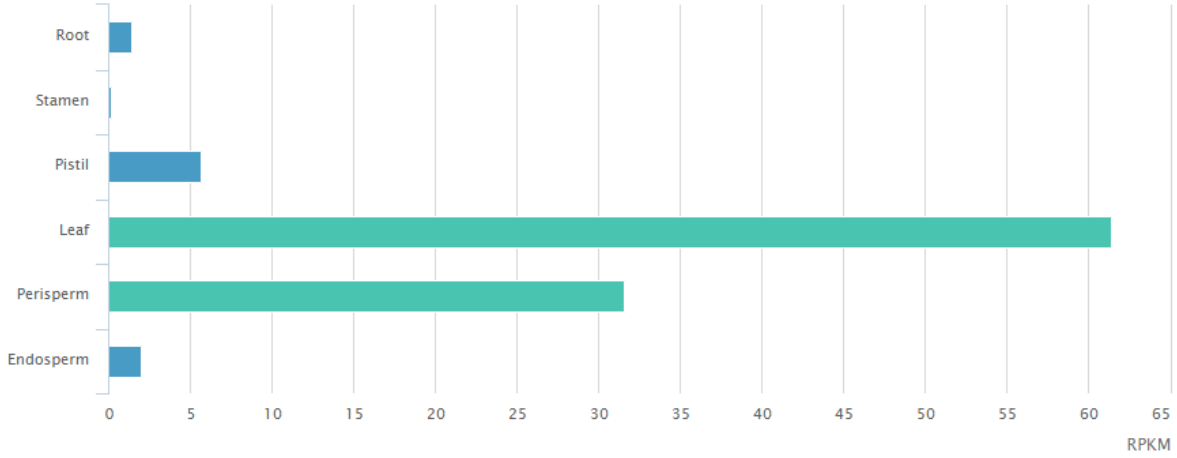
D Cc07_g11020



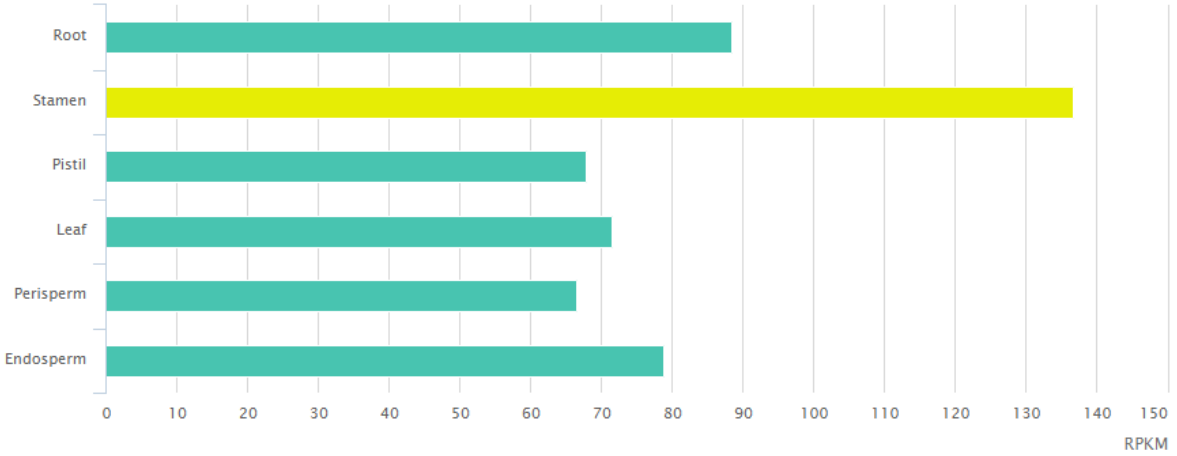
E Cc08_g00640



F Cc10_g06720



G Cc10_g12080



Fonte: Coffee Genome Hub (http://coffee-genome.org/)

5. CONCLUSÃO

Foi possível realizar a caracterização de genes de todas as isoformas de *APx* em 3 espécies do gênero *Coffea* e outras 6 espécies angiospermas, para efeito comparativo. Também foi possível identificar o número de genes para cada uma das três isoformas de *APx*. Pode-se verificar o tamanho das sequências de aminoácidos e de nucleotídeos, o ponto isoelétrico, o peso molecular e localização celular. Com as árvores filogenéticas construídas, foi possível observar os agrupamentos das isoformas de *APx* e a subdivisão dos clados em sequências de maior e menor tamanho, bem como incluir e agrupar sequências de café nunca antes relacionadas aos genes estudados. Foi possível também determinar as sequências de café provenientes de *C. eugenioides* e *C. canephora* que compõe o genoma da espécie hibrida *C. arabica*; também foi possível identificar dois genes relacionados à isoformas *APx-R* que não apresenta eventos de duplicação sugerindo que a presença dupla se dá pela diploidia da espécie.

Cálculos de dN/dS permitiram observar que a família gênica *APx* tem sofrido pressão purificadora, especialmente os genes cloroplastidiais. Foi possível ver que maioritariamente os aminoácidos das proteínas sofrem pressão purificadora, mas genes cloroplastidiais apresentem uma região de seleção positiva no início de suas sequências.

Com os dados de expressão observou-se os perfis transcricionais de *APx*, *APx-L* e *APx-R* em cafeeiros. Os genes de *APx* em café são mais expressos em comparação aos de *APx-R* e *APx-L*. Além disso também foi possível observar que o órgão em que se encontram mais expressos os genes de *APx-L* e *APx-R* são as folhas, enquanto os genes de *APx* são mais abundantes no estame e no pistilo.

Desta forma, o presente trabalho fornece as informações genômicas necessárias para maior subsídio a estudos em Bioquímica e Fisiologia do metabolismo em *Coffea*, especialmente quanto ao metabolismo antioxidante.

REFERÊNCIAS

CAMACHO, Christiam et al. BLAST+: architecture and applications. **BMC bioinformatics**, v. 10, n. 1, p. 421, 2009.

CAVERZAN, Andréia et al. Plant responses to stresses: role of ascorbate peroxidase in the antioxidant protection. **Genetics and molecular biology**, v. 35, n. 4, p. 1011-1019, 2012.

CHAVES, Agnaldo RM et al. Seasonal changes in photoprotective mechanisms of leaves from shaded and unshaded field-grown coffee (Coffea arabica L.) trees. **Trees**, v. 22, n. 3, p. 351, 2008.

CLARINDO, W. R.; CARVALHO, C. R. First Coffea arabica karyogram showing that this species is a true allotetraploid. **Plant Systematics and Evolution**, v. 274, n. 3-4, p. 237, 2008.

DA SILVA, Dayane Meireles et al. The effect of magnesium nutrition on the antioxidant response of coffee seedlings under heat stress. **Scientia Horticulturae**, v. 224, p. 115-125, 2017.

DAVIS, Aaron P. et al. Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. **Botanical Journal of the Linnean Society**, v. 167, n. 4, p. 357-377, 2011.

DE KOCHKO, Alexandre; CROUZILLAT, Dominique; ARABICA COFFEE GENOME CONSORTIUM. Aims and goals of the Arabica Coffee Genome Consortium (ACGC). **12th Solanaceae Conference**, Bordeaux. 2015.

DENOEUD, France et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n. 6201, p. 1181-1184, 2014.

DUNAND, Christophe et al. Ascorbate peroxidase-related (APx-R) is not a duplicable gene. **Plant signaling & behavior**, v. 6, n. 12, p. 1908-1913, 2011.

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, 2012.

FOYER, Christine H.; NOCTOR, Graham. Ascorbate and glutathione: the heart of the redox hub. **Plant Physiology**, v. 155, n. 1, p. 2-18, 2011.

ICO, International Cofee Organization. Historical Data on the Global Coffee Trade. **International Cofee Organization**. Londres, 2018. Disponível em: <http://www.ico.org/new_historical.asp?section=Statistics>. Acesso em: 12 de set. de 2018.

JONES, David T.; TAYLOR, William R.; THORNTON, Janet M. The rapid generation of mutation data matrices from protein sequences. **Bioinformatics**, v. 8, n. 3, p. 275-282, 1992.

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.

KUMAR, Sudhir et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. **Molecular biology and evolution**, v. 35, n. 6, p. 1547-1549, 2018.

LAZZAROTTO, Fernanda; ZOLET, Andreia Carina Turchetto; MARGIS-PINHEIRO, Márcia. Revisiting the Non-Animal Peroxidase Superfamily. **Trends in plant science**, 2015.

LE, Si Quang; GASCUEL, Olivier. An improved general amino acid replacement matrix. **Molecular biology and evolution**, v. 25, n. 7, p. 1307-1320, 2008.

LI, Qiang et al. Explosive tandem and segmental duplications of multigenic families in Eucalyptus grandis. **Genome biology and evolution**, v. 7, n. 4, p. 1068-1081, 2015.

MARRACCINI, Pierre et al. Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. **Journal of experimental botany**, p. ers103, 2012.

MITTLER, Ron. Oxidative stress, antioxidants and stress tolerance. **Trends in plant science**, v. 7, n. 9, p. 405-410, 2002.

OZYIGIT, Ibrahim I. et al. Identification and comparative analysis of H2O2-scavenging enzymes (ascorbate peroxidase and glutathione peroxidase) in selected plants employing bioinformatics approaches. **Frontiers in plant science**, v. 7, p. 301, 2016.

REIS, André R. et al. Antioxidant metabolism in coffee (Coffea arabica L.) plants in response to nitrogen supply. **Theoretical and Experimental Plant Physiology**, v. 27, n. 3-4, p. 203-213, 2015.

SANTOS, Meline Oliveira et al. Antioxidant System Differential Regulation is Involved in Coffee Ripening Time at Different Altitudes. **Tropical Plant Biology**, p. 1-10, 2018.

SCANDALIOS, John G. 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.

TEIXEIRA, Felipe Karam et al. Analysis of the molecular evolutionary history of the ascorbate peroxidase gene family: inferences from the rice genome. **Journal of Molecular Evolution**, v. 59, n. 6, p. 761-770, 2004.

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, 2015.

VAN BEL, Michiel et al. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. **Nucleic acids research**, v. 46, n. D1, p. D1190-D1196, 2017.

VIDAL, Ramon O. et al. A high-throughput data mining of single nucleotide polymorphisms in Coffea species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid Coffea arabica. **Plant Physiology**, v. 154, n. 3, p. 1053-1066, 2010.

VIEIRA, Natalia G. et al. Different Molecular Mechanisms Account for Drought Tolerance in *Coffea canephora* var. **Conilon. Tropical plant biology**, v. 6, n. 4, p. 181-190, 2013.

WHELAN, Simon; GOLDMAN, Nick. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. **Molecular biology and evolution**, v. 18, n. 5, p. 691-699, 2001.