

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE BIOCÊNCIAS
DEPARTAMENTO DE BIOQUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA**

RAFFAEL AZEVEDO DE CARVALHO OLIVEIRA

**ANÁLISE EVOLUTIVA DO SISTEMA ANTIOXIDANTE GÊNICO DE
*Arabidopsis thaliana***

NATAL, 2017

RAFFAEL AZEVEDO DE CARVALHO OLIVEIRA

**ANÁLISE EVOLUTIVA DO SISTEMA ANTIOXIDANTE GÊNICO DE
*Arabidopsis thaliana***

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica da Universidade Federal do Rio Grande do Norte como requisito parcial para obtenção do título de mestre em Bioquímica.

Orientador: Prof. Dr. Rodrigo Juliani Siqueira Dalmolin.

NATAL, 2017

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial do Centro de Biociências - CB

Oliveira, Raffael Azevedo de Carvalho.

Análise Evolutiva do Systems Antioxidante Gênico de Arabidopsis thaliana / Raffael Azevedo de Carvalho Oliveira. - Natal, 2017.

147 f.: il.

Dissertação (Mestrado) - Universidade Federal do Rio Grande do Norte. Centro de Biociências. Programa de Pós-Graduação em Bioquímica.

Orientador: Prof. Dr. Rodrigo Juliani Siqueira Dalmolin.

1. Biologia de sistemas - Dissertação. 2. Evolução - Dissertação. 3. Defesa antioxidante - Dissertação. I. Dalmolin, Rodrigo Juliani Siqueira. II. Universidade Federal do Rio Grande do Norte. III. Título.

RN/UF/BSE-CB

CDU 573

RAFFAEL AZEVEDO DE CARVALHO OLIVEIRA

Análise Evolutiva do Sistema Antioxidante de *Arabidopsis thaliana*

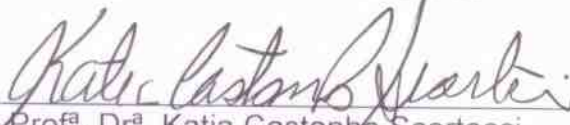
Dissertação apresentada ao Programa de Pós-graduação em Bioquímica, da Universidade Federal do Rio Grande do Norte, como requisito parcial para obtenção do título de Mestre em Bioquímica.

Aprovado em: 17/03/2017.

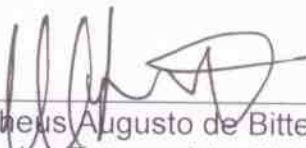
BANCA EXAMINADORA



Prof. Dr. Rodrigo Juliani Siqueira Dalmolin
Orientador/Presidente - UFRN



Prof.ª Dr.ª Katia Castanho Scortecci
Examinadora Interna ao Programa - UFRN



Prof. Dr. Matheus Augusto de Bittencourt Pasquali
Examinador Externo à Instituição - UFCG

Dedico esta obra ao meu avô, **Amaro Trajano de Oliveira** (*in memoriam*),
um curioso por natureza.

AGRADECIMENTOS

Primeiramente, agradeço a Deus pelas oportunidades dadas a mim;

Agradeço a meus familiares, pelo apoio emocional, financeiro e institucional. Especialmente a **Amaro Trajano, Ana de Carvalho, Amarylis, Atamir e Lucilene**, sem contar com a minha “família torta”: **Mariane e Camila**. Obrigado pelo apoio!

Agradeço imensamente a meu orientador e grande amigo, **Rodrigo Dalmolin**, pela confiança e persistência. Obrigado pelas lições, pelas cobranças e pela paciência. Esses 2 anos me fizeram crescer muito e espero continuar crescendo mais e mais.

Aos meus colegas do BioME: **André, Vandeclécio, Diego Coelho, Iara, Karla, Paulo, Diego Gomes, Ricard, Leônidas, Viviane, Laíse, Danilo Martins, Clóvis Reis, Raulzito Fernandes, Danilo Imparato, Abraão** e todos mais da casa mais do barulho da UFRN! blulululu

Aos meus amigos fora do meio acadêmico: **Casagrande, Hao, Tony, Arthur** e tantos outros que fizeram parte dessa jornada, mesmo que indiretamente! Seus lindos! <3

Aos membros da Igreja Batista de Ponta Negra, que vibravam com minhas vitórias, preocupavam-se com minhas preocupações e oraram tanto por mim. Sou eternamente grato, principalmente ao **pastor Allan** e sua esposa, **Vanessa Pires**. Meu eterno muito obrigado!

A minha melhor amiga, minha companheira, cúmplice, ajudante, ombro amigo, lugar onde posso repousar, eterna namorada e esposa: **Vanessa Maria de Andrade Pedroso Azevedo**. Obrigado por estar nessa comigo e já se prontificar pra próxima. Te amo!

A todas as outras pessoas que contribuíram: **OBRIGADO!!**

Do or do not. There is no try.
Yoda

RESUMO

As espécies reativas de oxigênio (EROs) são conhecidos subprodutos do metabolismo aeróbio e apesar de ter função sinalizadora, sua produção exacerbada pode levar a dano celular. As plantas desenvolveram um sistema antioxidante complexo, formado por componentes enzimáticos e não-enzimáticos, para se proteger contra a superprodução de EROs. Porém, a evolução desse sistema antioxidante ainda não está clara. Aqui o objetivo do trabalho foi descrever uma rede gênica que modela o sistema antioxidante de *Arabidopsis thaliana* e também investigar sua origem evolutiva. Primeiramente foram obtidos os genes a partir de bancos de dados públicos como *Gene Ontology*, que tivessem uma relação direta com a remoção de EROs dos sistemas, bem como alguns substratos/produtos. Então, foi modelada uma rede de interação proteína-proteína para *A. thaliana*. Posteriormente, utilizando informação de ortologia de 238 espécies anotadas no *STRING*, pudemos inferir a raiz evolutiva de cada um dos genes a fim de reconstruir a história evolutiva da rede gênica antioxidante de *A. thaliana*. Na rede construída, dois grupos interconectados foram encontrados, sendo um formado por proteínas SOD, tiol-redox e peroxidases; outro formado unicamente por peroxidases classe III. Cada um dos grupos provavelmente teve sua origem em momentos diferentes da escala evolutiva, porém as peroxidases classe III se mostraram os componentes mais recentes da rede. De acordo com nossos resultados, esse grupo de genes continua em expansão ao longo da evolução das plantas.

Palavras-chave: evolução; antioxidante vegetal; peroxidase classe III; biologia de sistemas

ABSTRACT

Reactive oxygen species (ROS) are byproducts of aerobic metabolism and may cause oxidative damage to biomolecules. Plants have a complex antioxidant system, involving enzymatic and non-enzymatic compounds, to protect against ROS exacerbated production. The evolutionary origin of enzymatic defense in plants is not totally clear. Here we describe an antioxidant gene network for *A. thaliana* and investigate the evolutionary origin of this network. We gathered from public repositories 208 *A. thaliana* genes directly involved with ROS detoxification and proposed an *A. thaliana* antioxidant gene network. Using orthology information of 238 Eukaryotes from STRINGdb we have inferred the evolutionary root of each gene to reconstruct the evolutionary history of *A. thaliana* antioxidant gene network. We found two interconnected clusters: one formed by SOD-related, Thiol-redox, and peroxidases; and other formed entirely by class III peroxidases. Each cluster emerged in different periods of evolution and class III peroxidases are the most recent components of the network. According to our results, class III peroxidases are in expansion throughout plant evolution.

Keywords: evolution; plant antioxidant; class III peroxidase; systems biology

LISTA DE FIGURAS

FIGURA SUPLEMENTAR 1: Árvore de espécies construída com o <i>geneplast</i>	66
FIGURA SUPLEMENTAR 2 A 71: Inferência evolutiva dos grupos de ortólogos utilizados, obtidos de <i>A. thaliana</i>	67
FIGURA SUPLEMENTAR 72: <i>Boxplots</i> mostrando a distribuição dos ortólogos de Peroxidase classe III COGs de <i>A. thaliana</i> e <i>O. sativa</i> como base.....	139

LISTA DE TABELAS

TABELA SUPLEMENTAR 1: Descrição dos genes utilizados na análise.....	140
TABELA SUPLEMENTAR 2: Descrição dos metabólitos utilizados na análise.....	145
TABELA SUPLEMENTAR 3: Abundância de genes tiol-redox, SOD, peroxidases classe III e não-classe III em grupos de ortólogos de <i>A. thaliana</i>	146
TABELA SUPLEMENTAR 4: Abundância de genes peroxidase classe III em grupos de ortólogos de <i>O. Sativa</i>	148
TABELA SUPLEMENTAR 5: Distribuição dos 208 genes em 70 grupos de ortólogos.....	149

LISTA DE ABREVIATURAS

O₂ – Oxigênio molecular

ATP – Adenosina Trifosfato

O₂^{•-} - Superóxido

H₂O₂ – Peróxido de hidrogênio

OH[•] - Radical Hidroxila

EROs – Espécies Reativas de Oxigênio

GSH – Glutathiona Reduzida

AsA – Ascorbato Reduzido

CAT – Catalase

SOD – Superóxido dismutase

APX – Ascorbato Peroxidase

PRX - Peroxirredoxina

GPX – Glutathiona Peroxidase

WGD – *Whole Genome Duplication* (Duplicação Completa do Genoma)

MF – *Molecular Function* (Função Molecular)

CC – *Cellular Component* (Componente Celular)

BP – *Biological Process* (Processo Biológico)

TAIR – *The Arabidopsis Information Resource* (Repositório de Informações sobre Arabidopsis)

COG – *Cluster of Orthologous Group* (Grupo de Ortólogos)

LCA – *Last Common Ancestor* (Último Ancestral Comum)

NOGs – *Non-Supervised Orthologous Groups* (Grupo de Ortólogos Não-Supervisionado)

GO – *Gene Ontology*

PPI – *Protein-Protein Interaction* (Interação Proteína-Proteína)

Mn – Manganês

Cu-Zn – Cobre-Zinco

Fe – Ferro

SUMÁRIO

1. INTRODUÇÃO.....	14
1.1. Espécies reativas de oxigênio e mecanismo antioxidante.....	14
1.1.1. Peroxidases classe III.....	18
1.2. Evolução dos Sistemas Antioxidantes em Plantas.....	22
1.3. Biologia de Sistemas.....	26
2. MATERIAIS E MÉTODOS.....	31
2.1. Escolha dos genes de interesse de <i>A. thaliana</i>	31
2.2. Construção do modelo de interação.....	32
2.3. Análise de inferência evolutiva.....	33
2.4. Análise de abundância de peroxidases classe III.....	34
3. RESULTADOS.....	36
3.1. Rede gênica antioxidante de <i>A. thaliana</i>	36
3.2. Origem evolutiva da rede gênica antioxidante de <i>A. thaliana</i>	40
3.3. Distribuição dos genes ortólogos de peroxidase classe III em monocotiledôneas e eudicotiledôneas.....	43
4. DISCUSSÃO.....	45
5. CONCLUSÃO.....	54
6. BIBLIOGRAFIA.....	55
7. ANEXOS.....	64

1. INTRODUÇÃO

1.1. Espécies reativas de oxigênio e mecanismo antioxidante

O oxigênio (O_2) fomenta diversas formas de vida na Terra há milhões de anos. Está presente nas rotas bioquímicas dos seres vivos aeróbios principalmente como acceptor final nas cadeias de transporte de elétrons. Nessas cadeias, a transferência de elétrons para o O_2 confere um gradiente de prótons. Este gradiente possibilita a geração de adenosina trifosfato (ATP), que funcionam como moeda energética universal em diversos organismos (DE GARA *et al.*, 2010). Apesar de ser um mantenedor da vida na Terra para diferentes organismos, como os vegetais, mamíferos e tantos outros, o oxigênio tanto é benéfico como maléfico. Sua estrutura química permite que seja convertido do seu estado basal (oxigênio tripleto) para formas mais reativas, como superóxido, oxigênio singleto, e radical hidroxila, por exemplo (Figura 1) (APEL; HIRT, 2004).

A primeira espécie reativa a ser produzida a partir do oxigênio é o radical superóxido ($O_2^{\cdot-}$), que tem sua origem principalmente nas cadeias respiratórias da cloroplastos e mitocôndrias, bem como em reações de defesa contra patógenos (LOW; MERIDA, 1996). A redução do $O_2^{\cdot-}$ leva a formação de hidroxiperóxidos, como o peróxido de hidrogênio (H_2O_2). Essa espécie reativa pode atravessar membranas e danificar estruturas celulares (NANDA *et al.*, 2010) Por fim, a redução do H_2O_2 por íons metálicos na reação de Fenton leva a formação do radical hidroxila OH^{\cdot} .

Devido a sua alta reatividade, essas EROs tendem a reagir com biomoléculas (peroxidação lipídica, carbonilação proteica, degradação de ácidos nucleicos, etc) de diferentes meios. Em ácidos nucleicos, a interação com EROs pode levar a oxidações

nas bases nitrogenadas, causando a mudança de bases; outras alterações que desestabilizam o material genético são: a oxidação da desoxirribose, quebra da dupla-fita (no caso do DNA) e remoção de nucleotídeos. As alterações em proteínas também são muito significativas. Interações dessas espécies reativas com aminoácidos levam a uma reação em cadeia: alterações em aminoácidos podem causar diferença em carga e/ou polaridade. Consequentemente, a proteína tem um dobramento ineficiente e, por conseguinte, sofre alterações em sua funcionalidade. Em lipídeos, as EROs podem romper as cadeias de hidrocarbonetos. A desestruturação dos lipídeos, principalmente os de membrana, levam ao aumento na fluidez da membrana, consequentemente, aumentando sua permeabilidade. Em todos os casos apresentados anteriormente, a interação de EROs com biomoléculas podem levar a situações incompatíveis com a vida (SHARMA *et al.*, 2012). Por fim, a redução do H_2O_2 por íons metálicos na reação de Fenton leva a formação do radical hidroxila OH^\bullet .

Devido a sua alta reatividade, essas EROs tendem a reagir com biomoléculas (peroxidação lipídica, carbonilação proteica, degradação de ácidos nucleicos, etc) de diferentes meios. Em ácidos nucleicos, a interação com EROs pode levar a oxidações nas bases nitrogenadas, causando a mudança de bases; outras alterações que desestabilizam o material genético são: a oxidação da desoxirribose, quebra da dupla-fita (no caso do DNA) e remoção de nucleotídeos. As alterações em proteínas também são muito significativas. Interações dessas espécies reativas com aminoácidos levam a uma reação em cadeia: alterações em aminoácidos podem causar diferença em carga e/ou polaridade. Consequentemente, a proteína tem um dobramento ineficiente e, por conseguinte, sofre alterações em sua funcionalidade. Em lipídeos, as EROs podem

romper as cadeias de hidrocarbonetos. A desestruturação dos lipídeos, principalmente os de membrana, levam ao aumento na fluidez da membrana, conseqüentemente, aumentando sua permeabilidade. Em todos os casos apresentados anteriormente, a interação de EROs com biomoléculas podem levar a situações incompatíveis com a vida (APEL; HIRT, 2004)

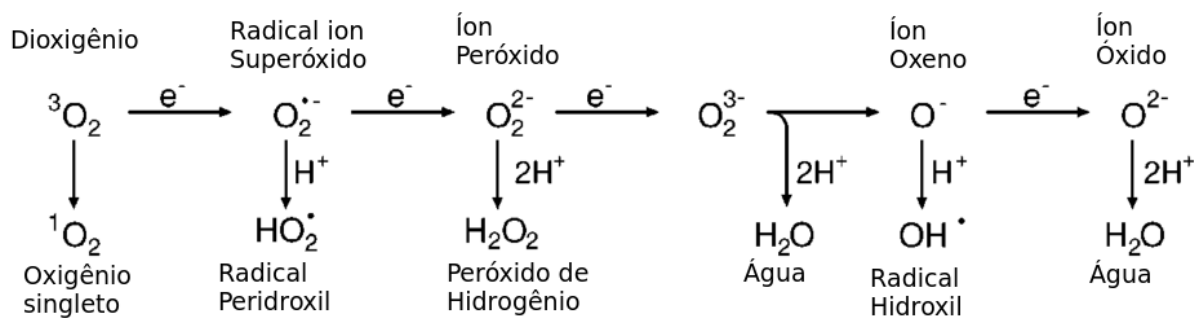


Figura 1: Esquema de geração de EROs a partir do estado basal do O_2 . Adaptado de Apel & Hirt, 2004.

Para controlar a superprodução de EROs, diferentes organismos desenvolveram estratégias de lidar com uma provável alta carga destas substâncias em seus sistemas, uns mais do que os outros. Com as plantas não é diferente. Sua condição sésil (*i.e.*, imóvel) leva a uma constante exposição a fatores abióticos diversos, que variam conforme o bioma em que determinados organismos estão inseridos, tais como salinidade, seca, frio, calor, radiação, poluentes, pesticidas, alta luminosidade, etc., além da infecção por outros organismos (vírus, bactérias, fungos, etc) (GILL; TUTEJA, 2010). Estas formas de estresse abiótico levam a diversas condições internas. No geral,

as principais formas de produção de EROs são pela diminuição da fixação de CO₂, aumento na atividade fotossintética, gerando mais EROs pelos cloroplastos e desbalanço da absorção e uso da luz para fotossíntese, pela inibição do ciclo Calvin-Benson (SHARMA *et al.*, 2012). Os mecanismos de combate ao desbalanço redox vem sendo estudado por diversos pesquisadores ao longo do tempo e é definido como sistema antioxidante (HALLIWELL, 2006). Esse sistema, em plantas, possui componentes não-enzimáticos e enzimáticos. O primeiro consiste em elementos como glutathiona reduzida (GSH), ascorbato (AsA), tocoferóis, compostos fenólicos, carotenóides, flavonóides, entre outros (AHMAD *et al.*, 2010). Estes compostos são oxidados para reduzir os potenciais danos causados pelas EROs. Além disso, os carotenóides também podem aumentar a resistência a estresse oxidativo em presença de alta luminosidade (APEL; HIRT, 2004). Além destes compostos, existem também as enzimas antioxidantes, que catalisam a transformação de espécies mais reativas do oxigênio em formas menos ativas. São elas a catalase (CAT), superóxido dismutases (SOD), ascorbato peroxidases (APX), peroxirredoxinas (PRX), glutathiona peroxidase (GPX), entre proteínas que não apresentam função catalítica (MITTLER *et al.*, 2004). As CAT são responsáveis pela conversão do H₂O₂ em água; as SOD convertem o O₂^{•-} em H₂O₂; as APX, GPX e PRX também realizam a conversão do H₂O₂ em água, porém se utilizam de ascorbato reduzido, glutathiona reduzida e um resíduo de cisteína conservado, respectivamente, para tal. Apesar da divisão didática em enzimática e não enzimática, os componentes que compõem a defesa antioxidante trabalham conjuntamente. A parte não-enzimática muitas vezes serve como fonte primordial de elétrons, como a GSH, para ser oxidados. Esse trabalho em conjunto pode ser

observada em reações como o ciclo glutationa-ascorbato, que é caracterizado por um sistema de tamponamento redox (FOYER; NOCTOR, 2011).

Apesar de ser danosas em qualquer organismo (quando em concentração elevada), as EROs também possuem o seu lado benéfico. Em plantas, EROs como o H_2O_2 , em produção controlada, pode ser utilizado como molécula sinalizadora. Pesquisas recentes mostram que esta molécula tem importantes papéis na germinação das sementes (BARBA-ESPÍN *et al.*, 2011), fechamento e abertura de estômatos (GE *et al.*, 2015), produção hormonal (BARBA-ESPIN *et al.*, 2010), entre outras funções. Por isso, a manutenção de uma produção basal de EROs deve existir para que haja a progressão no desenvolvimento da planta, através dos processos de respiração oxidativa (MITTLER *et al.*, 2004). Esse balanço entre a produção de EROs e sua remoção em plantas necessita de uma regulação temporal e espacial, a fim de evitar danos a biomoléculas e permitir que essas moléculas cumpram seu papel de sinalizadores (BAILEY-SERRES; MITTLER, 2006; DIETZ; MITTLER; NOCTOR, 2016).

1.1.1. Peroxidases classe III

Além de apresentar um arsenal diversificado de enzimas que mantém o equilíbrio redox celular, as plantas possuem peroxidases específicas. As peroxidases de classe III, também chamadas de peroxidases clássicas de plantas, são glicoproteínas, pertencentes a um subtipo de heme-peroxidases presente em todas as plantas terrestres, sendo postuladas como um dos grandes contribuintes para a colonização da terra pelas plantas (WENG; CHAPPLE, 2010). Essas enzimas apresentam homologia com outras peroxidases, as de classe I e II. A primeira,

distribuída principalmente entre os procariotos (não se estendendo aos animais) e a segunda, exclusivamente sintetizada por fungos (COSIO; DUNAND, 2009). Sua estrutura bioquímica é baseada na estrutura da peroxidase clássica 1C de *Armoracia rusticana*, a horseradish peroxidase 1C e consiste em um grupo heme central (protohemina IX ou heme b), predominância de α -hélices; quatro pontes dissulfeto e uma ponte salina, enterrada na porção hidrofóbica da estrutura proteica, que ajudam a manter a estabilidade durante o dobramento da mesma (COSIO; DUNAND, 2009; WELINDER *et al.*, 2002).

Arabidopsis thaliana, a planta modelo, possui 73 isoformas de peroxidases classe III em seu genoma. Esse número pode variar de acordo com a espécie: em *Oryza sativa*, são 138 isoformas; *Zea mays* são 107 e *Populus trichocarpa*, 93, respectivamente (ALMAGRO *et al.*, 2009; REN *et al.*, 2014; WANG *et al.*, 2015). Seu mecanismo de ação difere de outras peroxidases, como as APX, que agem no ambiente intracelular removendo peróxido apenas: as peroxidase classe III, em determinadas etapas, podem também produzir H_2O_2 (WELINDER *et al.*, 2002). As peroxidases classe III são secretadas para o ambiente extracelular via retículo endoplasmático, e por isso e devido a sua grande variedade de isoformas, possuem funções diversas, que acompanham a planta desde a germinação até a senescência. Por exemplo, Wang e colaboradores estudaram a expressão gênica de peroxidases classe III em *Zea mays* e encontraram que algumas isoformas seriam expressas no período pós-germinativo da planta, o que poderia indicar a produção de H_2O_2 , auxiliando na germinação (PASSARDI, F. *et al.*, 2005; WANG *et al.*, 2015). Também supõe-se que essa função de controlar a concentração de H_2O_2 por meio da

produção/remoção, pode estar relacionada com alongamento celular em tecidos de raiz, folhas e caule (DUNAND *et al.*, 2003). As peroxidases classe III também têm relação com a morte celular programada (PASSARDI, F. *et al.*, 2005). Além, disso diversos trabalhos apontam para outras funções que auxiliam na manutenção da homeostase. Na figura 2 podemos ver diversas dessas funções, como: catabolismo de auxina; deposição de suberina e lignina, que auxiliam no enrijecimento da parede celular; sensores fisiológicos de estresses bióticos e abióticos; desenvolvimento do fruto, além de defesa contra patógenos e insetos herbívoros (ALMAGRO *et al.*, 2009; COSIO; DUNAND, 2009; ZAMOCKY; FURTMÜLLER; OBINGER, 2008). Apesar de contar com uma função de remoção do peróxido, as peroxidases classe III podem também produzir essa ERO a partir da oxidação da auxina, sendo importante no alongamento das células do caule e raízes (RODRÍGUEZ; GRUNBERG; TALEISNIK, 2002; WANG *et al.*, 2015). Além destas, outras funções estão descritas para essa classe proteica na figura 2 (COSIO; DUNAND, 2009).

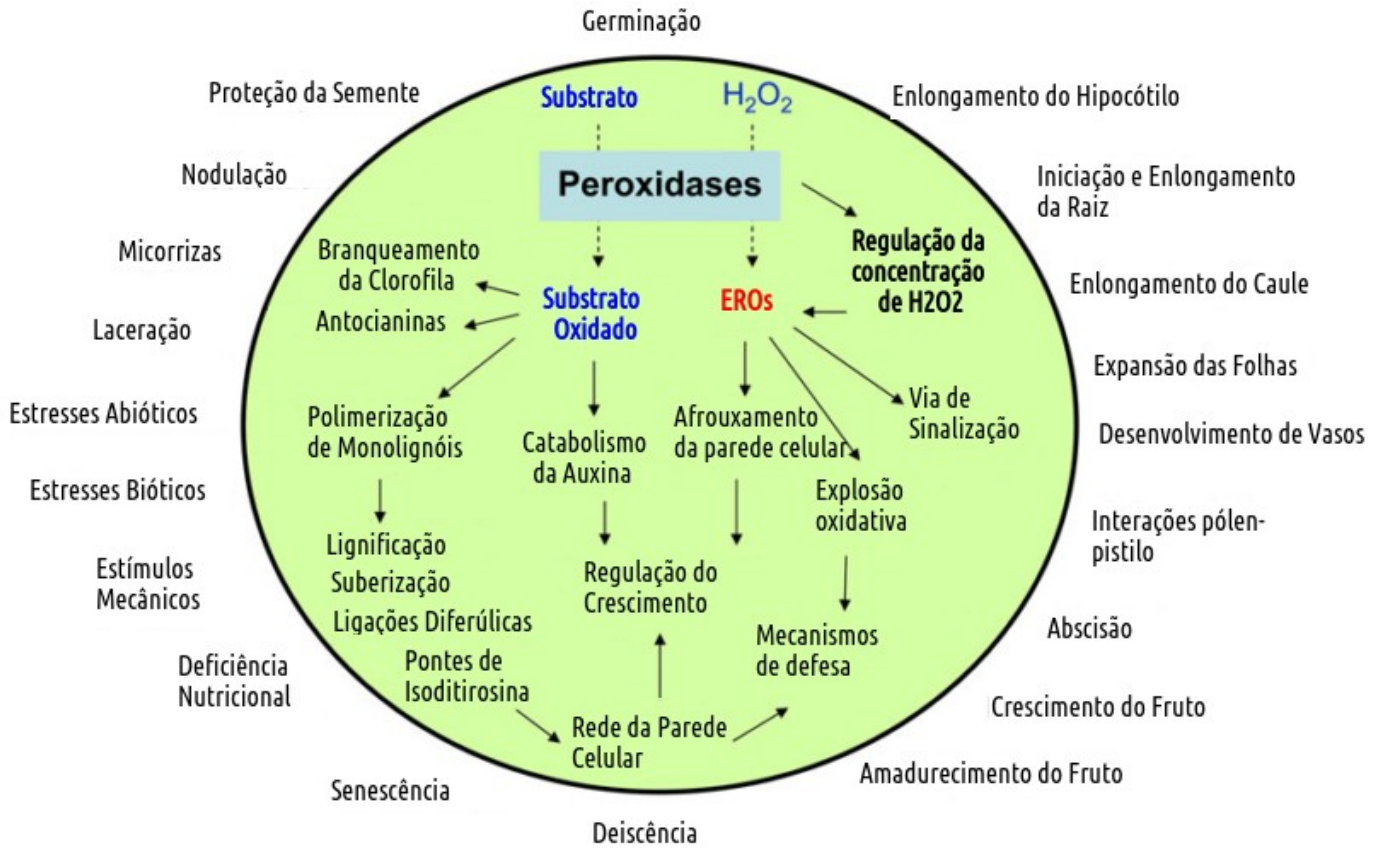


Figura 2: Esquema indicativo das diversas funções exercidas pelas peroxidases classe III. Adaptado de (COSIO; DUNAND, 2009). As peroxidases classe III não metabolizam apenas o H_2O_2 mas também outros peróxidos orgânicos. Estes peróxidos orgânicos oxidados são utilizados em processos fisiológicos da própria planta, enquanto que (principalmente) o H_2O_2 leva a vias de sinalização.

1.2. Evolução dos Sistemas Antioxidantes em Plantas

A atmosfera primitiva era composta principalmente de hélio, dióxido de carbono e hidrogênio, provenientes das explosões solares. Colisões com grandes objetos no espaço permitiram que compostos voláteis, como a água, fossem incorporados à atmosfera terrestre. Posteriormente, o oxigênio da água presente na Terra primitiva foi sendo fotodissociado e passou a constituir a atmosfera (BENZIE, 2000). Com a evolução dos organismos fermentadores, que produziam oxigênio como produto de suas reações, a concentração desse gás foi aumentando na atmosfera, até aproximadamente 2 bilhões de anos atrás, quando se ocorreu o grande evento de oxidação (ANBAR, 2008; BENZIE, 2000; INUPAKUTIKA *et al.*, 2016). Essa mudança na constituição da atmosfera provavelmente foi um dos motivos que levou a evolução de processos biológicos como as cadeias respiratórias e também a fotossíntese. As plantas presentes atualmente em todo o planeta existem desse modo devido a longos processos evolutivos. A adaptação para utilização do oxigênio como acceptor de elétrons é precedida pelo desenvolvimento de um sistema pronto para remover os subprodutos da oxidação deste gás (MILLER, 2012). Sua rápida conversão em $O_2^{\cdot-}$ induziu a formação de estratégias para prevenção de danos por EROs. Inicialmente foram as barreiras físicas e químicas. As barreiras físicas provavelmente foram as primeiras a surgir e eram (e continuam sendo) primordiais na compartimentalização de biomoléculas (empacotamento do DNA em histonas, formação de membranas, etc); em plantas, também haviam barreiras químicas, os pigmentos celulares, como os carotenóides e antocianinas, por exemplo. Estes compostos são amplamente conhecidos por seus papéis antioxidantes e por seus papéis como pigmentos

(LARSON, 1988). Posteriormente, quando não mais efetivo, as primeiras enzimas começaram a surgir, como a SOD (BENZIE, 2000; INUPAKUTIKA *et al.*, 2016). A SOD provavelmente foi a primeira linha de defesa enzimática formada e após isso, outras enzimas surgiram para complementar o sistema antioxidante de modo geral. Com o tempo mais enzimas surgiram e passaram a complementar e melhorar a defesa fisiológica contra agente oxidantes. Essas enzimas possivelmente tiveram um papel crucial no desenvolvimento das plantas, que colonizaram a terra e passaram a desenvolver outras estratégias para sua consolidação neste ambiente (LEWIS; MCCOURT, 2004).

Dentre esses processos, a diversidade gênica em vegetais se deve principalmente aos eventos de duplicação completa do genoma (*whole genome duplication*, WGD). Esses eventos ocorreram em torno de 319 e 192 milhões de anos atrás, após a divergência das plantas com sementes e Angiospermas atuais (JIAO *et al.*, 2011). De acordo com a Figura 3, os genomas de Angiospermas (que surgiram após os dois grandes eventos de WGD), apresentam uma poliploidia (ou seja, mais cópias do seu genoma) muito vasta. Um dos fatores que remetem o sucesso na colonização da terra pode ser creditado aos eventos de WGD, visto que muitas das plantas verdes existentes apresentam uma diversidade metabólica e funcional graças aos processos de duplicação (LEWIS; MCCOURT, 2004). O resultado dessa duplicação pode ter gerado cópias de genes, responsáveis por produtos funcionais dos mais diversos, e esse processo é um dos maiores fornecedores de matéria-prima a ser trabalhada pela evolução (JIAO *et al.*, 2011). Porém, apesar de parecer uma fonte ideal para o aparecimento de novas proteínas, a WGD pode gerar efeitos desfavoráveis para a

planta, como reorganização genômica e também expressão gênica (TE BEEST *et al.*, 2012). Nos processos de WGD, assim como outros processos de duplicação gênica, podem levar a uma situação onde as novas cópias originadas podem gerar genes ortólogos. Estes são resultado da especiação, onde um gene ancestral comum dá origem a duas novas cópias em espécies diferentes (ortólogo); quando esse mesmo gene se mantém na mesma espécie, a segunda cópia pode ser classificada como um parálogo (GABALDÓN; KOONIN, 2013). A duplicação completa de genoma tem suas vantagens e desvantagens em plantas. Por causa desse processo, muitos genes puderam ter o que se chama de *mutation buffering* (do inglês, tamponamento de mutação), aumento na diversidade alélica e até mesmo neo-funcionalização e/ou sub-funcionalização. O *mutation buffering* se dá quando existe poliploidia em um organismo e as mutações em um determinado gene não afetam a funcionalidade dos seus produtos, ou seja, não há haploinsuficiência (SOLTIS; VISGER; SOLTIS, 2014). O aumento na diversidade alélica segue no mesmo sentido do *mutation buffering*. O interessante também são os processos de neo-funcionalização e/ou sub-funcionalização, em que essas novas cópias podem sofrer processos diferenciais e exercer atividades distintas da original, ou apenas realizar uma sub-função em uma mesma rota metabólica (PANCHY; LEHTI-SHIU; SHIU, 2016). As possíveis desvantagens se encontram na organização dos cromossomos e genes, que levam a um aumento no volume nuclear, que podem reorganizar todo o processo de regulação da expressão gênica. O aumento no número de cópias reflete na mitose, por exemplo, onde a planta demanda mais energia para que o processo seja feito com a mesma eficiência (COMAI, 2005; TE BEEST *et al.*, 2012). Apesar de bem estabelecido, ainda

não se sabe a origem de todos os genes oriundos de WGD, portanto, necessitando de cada vez mais pesquisas nessa área.

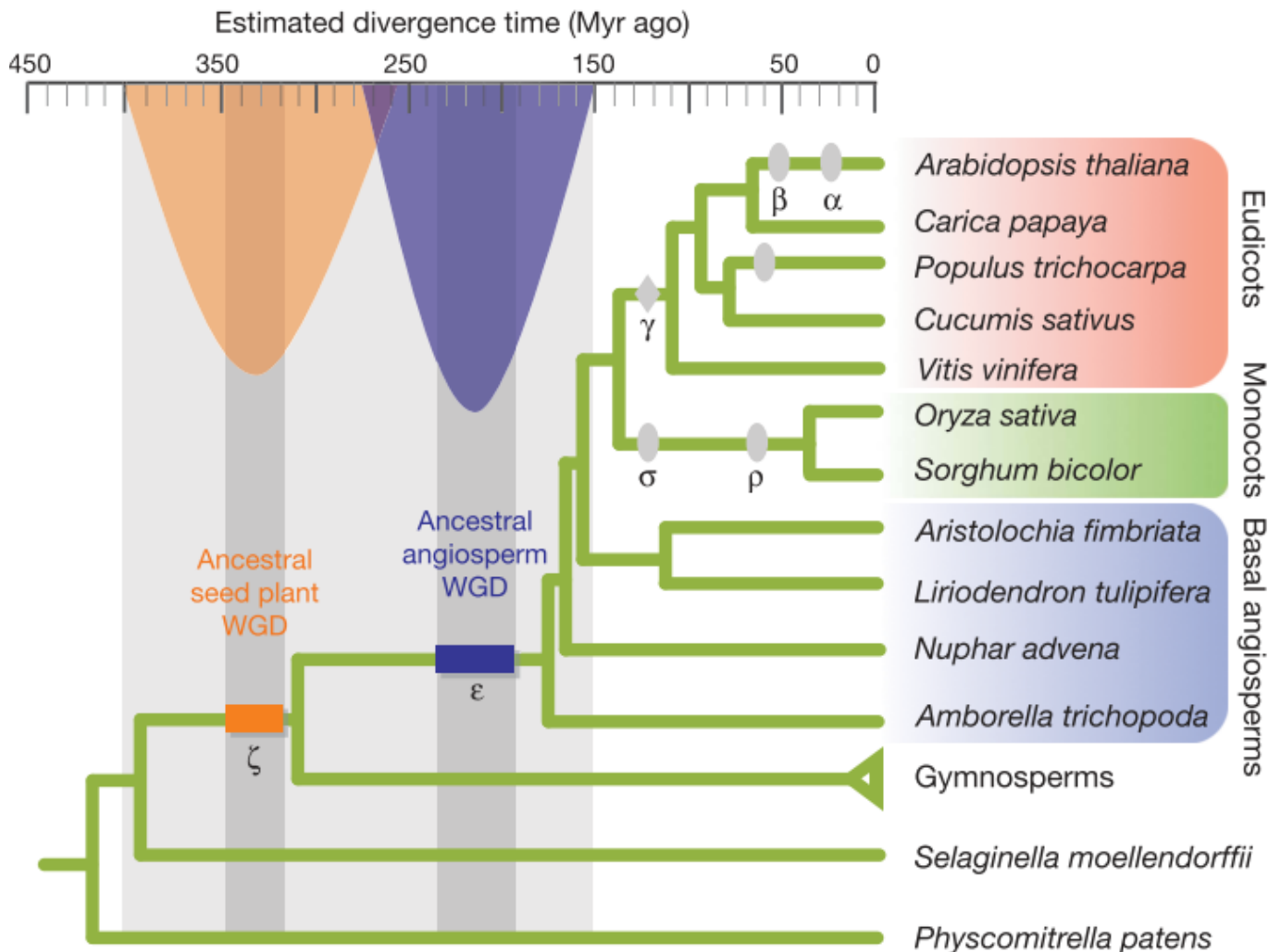


Figura 3: Filogenia esquemática do surgimento das plantas verdes (Viridiplantae). Em laranja e azul, vemos os dois principais pontos de WGD antes do aparecimento das Angiospermas e suas distribuições (parábolas). Elipses: duplicações aceitas e identificadas em genomas atualmente sequenciados. Losango: evento de triplicação provavelmente compartilhado pelas eudicotiledôneas. Retângulos: indicam regiões de confiança para distribuição das estimativas médias das idades de seus ancestrais. Retirado de (JIAO *et al.*, 2011)

1.3. Biologia de Sistemas

As proteínas que compoem a defesa antioxidante em plantas fazem parte de um sistema biológico, que nesse caso, trata-se do sistema antioxidante gênico de plantas. A definição de sistema é o conjunto de elementos que possuem relações entre si. Os sistemas biológicos, que são os alvos da biologia de sistemas, são tidos como sistemas complexos e são assim por um motivo simples: eles mostram a interconexão de sistemas simples. Essa interconexão revela propriedades que possivelmente não seriam perceptíveis ao analisar um sistema simples, o que possibilita a observação de características do conjunto, algo que seria perdido caso fosse analisado apenas um elemento. O estudo dos sistemas biológicos se dá pela biologia de sistemas, que é um campo da ciência que busca a compreensão dos problemas biológicos sob diferentes pontos de vista (VERLI, 2014). A óptica da visualização de dados biológicos pela biologia de sistemas pode ser observada em diferentes níveis hierárquicos e em diferentes escalas de tempo. (BARABASI; OLTVAI; BARABÁSI, 2004). Os sistemas biológicos podem ser estudados desde o nível mais abrangente, como ecologicamente, até escalas menores, como órgãos, tecidos, células e componentes ainda menores (KLIPP *et al.*, 2009). Essas abordagens permitem que se observe, por exemplo, uma condição de infecção em um organismo e como essa condição afeta órgãos e tecidos: utilizando dados de expressão gênica, pode-se saber quais genes estão sendo expressos diferencialmente frente a alguma condição fisio ou patológica. Ao analisar esses dados de forma sistêmica, pode-se identificar com quais produtos eles interagem, se são co-expressos e/ou se existem fatores de transcrição-chave que medeiam toda (ou boa parte) da condição em questão (BOLOURI; DAVIDSON, 2002). Esta

capacidade de abordagem do problema biológico está diretamente relacionada com o tipo de dados utilizados. Felizmente, nos dias atuais existe um grande aporte de dados oriundos das ciências “ômicas” (KITANO, H, 2002). Grandes bancos de dados públicos, como o *National Center for Biotechnology Information* (NCBI), *European Bioinformatics Institute* (EBI), *The Cancer Genome Atlas* (TCGA), entre tantos outros, gerenciam dados depositados por pesquisadores ao redor do mundo. São experimentos na área da genômica, transcriptômica, proteômica, metabolômica e outras abordagens, que permitem que estudos possam ser realizados em escalas macro e microscópicas (*bottom-up* e *top-down*, respectivamente) (BARRETT *et al.*, 2013; KITANO, HIROAKI, 2002; KOLESNIKOV *et al.*, 2015). Esses estudos empregam o uso da modelagem de redes biológicas, que possui diversas aplicações, desde a caracterização de um sistema como a identificação de mecanismos regulatórios, que tem auxiliado a elucidação de novos alvos terapêuticos (ALBANUS *et al.*, 2013; ALM; ARKIN, 2003).

A modelagem dos sistemas é possível com aplicações da Teoria de Grafos, que é um ramo da matemática dedicado a estudar a relação entre componentes de um conjunto qualquer. Para isso, utilizam-se os grafos, que são estruturas compostas por um conjunto de nós (ou vértices) e as relações entre esses nós, chamadas arestas (FEOFILOFF; KOHAYAKAWA; WAKABAYASHI, 2011). Os nós de um grafo podem representar diversas coisas, como pessoas em uma rede social, computadores em rede, reações químicas e, mais importante para o caso, genes e/ou proteínas. As arestas representam a relação entre os nós, podendo ser (no caso da biologia molecular) a supressão/indução da transcrição de um gene, uma rota metabólica, interação entre proteínas, entre outras aplicações (VERLI, 2014). A Figura 4 mostra a

aplicação do uso da teoria de grafos na modelagem de um sistema. No caso, o sistema em questão foi o sistema antioxidante de *Homo sapiens*, que foi um dos motivadores a construção do sistema antioxidante de plantas proposto neste trabalho. A rede apresenta os genes de *Homo sapiens* e sua interação com moléculas doadoras de elétrons e também sua interação com outras proteínas. Essa abordagem auxilia, como mencionado anteriormente, a observar a interação de proteínas com seus substratos/produtos.

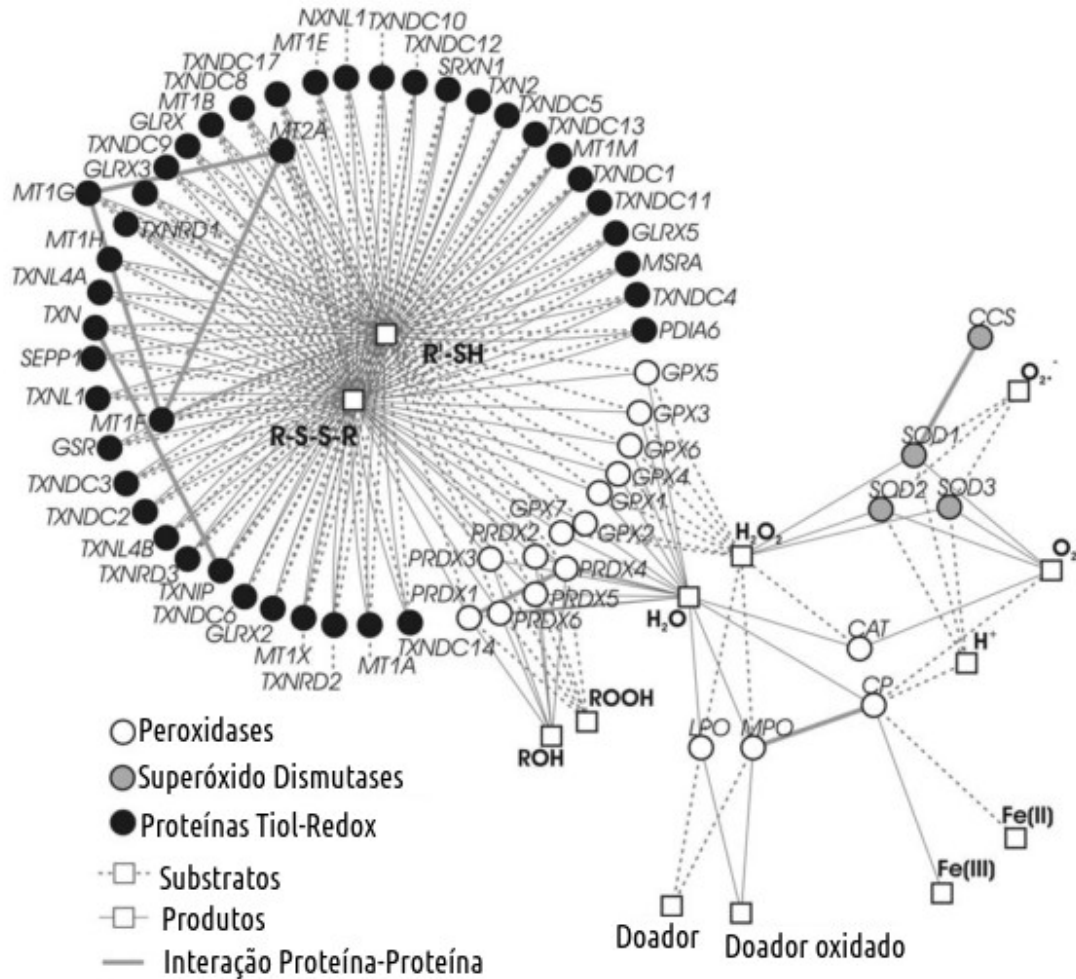


Figura 4: Rede de PPI dos genes antioxidantes em *Homo sapiens*. Os nós representam um gene e as arestas representam interações físicas e/ou funcionais. Retirado de (GELAIN *et al.*, 2009)

Com base no que foi apresentado nos tópicos anteriores, este trabalho tem como objetivo principal: construir uma rede que modela o sistema antioxidante gênico de plantas utilizando *Arabidopsis thaliana* como organismo modelo e estabelecer as possíveis origens evolutivas para os elementos desta rede, com foco especial nas peroxidases classe III. Os objetivos específicos foram o de construir um banco de genes antioxidantes de *A. thaliana*, utilizando informações contidas no *Gene Ontology* e em parâmetros pré-estabelecidos; construir a rede de interação proteína-proteína entre os produtos desses genes, baseado em dados de conexão fornecido pelo *STRING*; realizar a análise topológica da rede e comparar com o que já existe na literatura; realizar a análise de inferência evolutiva para todos os genes da rede; montar a árvore filogenética com inferência em ortologia; identificar o padrão de distribuição dos grupos de genes ortólogos de acordo com a distância evolutiva entre os organismos e inferir a origem evolutiva dos genes de peroxidase classe III baseado em informação de ortologia.

2. MATERIAIS E MÉTODOS

2.1. Escolha dos genes de interesse de *A. thaliana*

O primeiro passo da análise foi o coletar os genes relacionados a atividade antioxidante em *A. thaliana*. Para tal, utilizamos o *Gene Ontology* (GO), um banco de dados especializado em anotações de genes e seus produtos, bem como facilitar a distribuição dos dados curados (ASHBURNER *et al.*, 2000). Desse banco, foi possível extrair informações de basicamente três grandes domínios: as funções moleculares exercidas pela proteína codificada (*molecular function*, MF); sobre a localização do produto gênico na célula (*cellular component*, CC) e sobre o processo biológico (*biological process*, BP) em que está envolvido. Nós utilizamos aqui 6 termos de BP de interesse: (i) *oxygen species metabolic process* (GO:0072593), (ii) *cell redox homeostasis* (GO:0045454), (iii) *cellular response to reactive oxygen species* (GO:0034614), (iv) *response to inorganic substance* (GO:0010035), (v) *oxidation-reduction process* (GO:0055114) e (vi) *response to oxidative stress* (GO:0006979). Estes termos em específico foram utilizados devido a sua maior representatividade: são termos relacionados às funções buscadas, porém não são abrangentes ao ponto de generalizar demasiadamente e incluir genes que não se encaixam no escopo da análise. Então as funções de cada gene utilizado foi checado utilizando o banco de dados de referência para *Arabidopsis*, o Repositório de Informações sobre *Arabidopsis* (*The Arabidopsis Information Resource*, TAIR) (LAMESCH *et al.*, 2012). A fim de evitar redundância e falsos-positivos, os genes escolhidos passaram por um processo de filtragem, de acordo com critérios pré-estabelecidos por Gelain e colaboradores (GELAIN *et al.*, 2009). Esses critérios basicamente incluem genes cujos produtos são já

descritos como sendo primariamente removedores de EROs, bem como genes indispensáveis para o funcionamento das enzimas antioxidantes. Com isso, o objetivo foi incluir apenas produtos gênicos de alguma forma lidam diretamente com substâncias oxidantes, transformando-as em produtos menos tóxicos. Após essa filtragem, o conjunto final de genes incluiu 208 genes diretamente envolvidos com a detoxificação das EROs (Tabela Suplementar 1). Esses foram posteriormente classificados em três categorias, de acordo com a sua descrição no TAIR: tiol-redox, peroxidase e SOD.

2.2. Construção do modelo de interação

A rede gênica antioxidante de *A. thaliana* foi modelada utilizando o banco de dados *STRING* na versão 10.0 [<http://string-db.org/>] (SZKLARCZYK *et al.*, 2014), com os parâmetros “*database*” e “*experiments*” escolhidos, utilizando o escore de confiança de 0.400 padrão como valor de corte. Essa pontuação existe pela tentativa de quantificar a incerteza das PPI inferidas, baseado tanto na natureza como na qualidade da evidência que suporta a interação entre as dadas proteínas. O *STRING* integra diferentes bancos de dados públicos, contendo informação de associação direta e indireta entre diferentes proteínas de mais de 2000 organismos. A busca foi feita utilizando os identificadores oriundos do TAIR. Posteriormente, esta lista de identificadores foi submetida ao *STRING* e a informação de conexão entre os 208 genes foi obtida (Tabela Suplementar 1). Além de utilizar a lista de genes relacionados com a homeostase redox da célula, também foram incluídos compostos químicos. Estes são utilizados como substratos e/ou produtos das referidas enzimas. No total, foram incluídos 15 compostos químicos (Tabela Suplementar 2). As conexões entre a

lista de genes e a lista de compostos químicos foram obtidas utilizando o *STITCH*, na versão 5.0 [<http://stitch.embl.de/>] (SZKLARCZYK *et al.*, 2016). Os mesmos parâmetros utilizados no *STRING* também se aplicam aqui. Assim como o primeiro banco utilizado, o *STITCH* integra informação entre proteínas e moléculas (orgânicas ou não). A rede resultante inclui as interações proteína-proteína, bem como as interações proteína-composto químico, entre os 208 genes e os 15 químicos supracitados. A rede final foi manipulada utilizando o pacote *RedeR*, um *software* integrado a suite R, utilizado para visualização e manipulação de redes (CASTRO *et al.*, 2012).

2.3. Análise de inferência evolutiva

A análise de inferência evolutiva, foi realizada utilizando a suíte R, com o pacote *geneplast*. Este *software* realiza o enraizamento evolutivo baseado na distribuição de grupos de ortólogos. O processo foi dividido em duas partes: 1) a construção de uma filogenia consenso. Para isso, utiliza-se do estado mais parsimonioso para um conjunto de grupo de genes ortólogos dos eucariotos disponíveis no *STRING*, e 2) a inferência da raiz evolutiva para cada gene. Esse estado mais parsimonioso refere-se ao princípio estabelecido por William de Ockham, que afirma que a explicação para qualquer fenômeno deverá assumir o menor número de suposições possível. Então, primeiramente a filogenia para as espécies de eucariotos presentes no *STRING* foi obtida através da mineração de dados presentes neste banco. No total, na versão 10 do *STRING*, existem anotações de ortologia para 238 espécies de eucariotos. Em seguida foi determinada a presença de genes homólogos entre as espécies selecionadas para

os genes antioxidantes adquiridos. Esse passo da análise utiliza um conjunto de genes $C(C_1...C_n)$, onde os elementos desse conjunto se agrupam baseado em parâmetros de estrutura e funcionalidade. Baseado nisso, após reunir os grupos de ortólogos relacionados aos genes antioxidantes, busca-se para cada C_n em um grupo de genes ortólogos (do inglês *Cluster of Orthologous Genes*, COG), o eucarioto mais antigo que possua aquele elemento, aplicando uma função de custo para minimizar inconsistência na análise. Os COGs utilizados são armazenados para a próxima etapa. O segundo passo foi realizar a inferência da raiz evolutiva, onde a informação de ortologia foi utilizada para inferir a probabilidade de que uma característica esteve presente em um Último Ancestral Comum (do inglês, *Last Common Ancestor*, LCA). Esta metodologia foi primeiramente pensada por Mirkin e colaboradores (MIRKIN *et al.*, 2003) estabelecido por Castro e colaboradores (CASTRO *et al.*, 2008). As árvores evolutivas que descrevem o padrão filético de cada um dos genes antioxidantes utilizados estão nas Figuras Suplementares de 2 a 71.

2.4. Análise de abundância de peroxidases classe III

A abundância dos genes relacionados a função peroxidase classe III foi calculada utilizando uma lista de COGs oriunda de *A. thaliana*, obtida na Análise de Inferência Evolutiva no passo anterior. Primeiramente os grupos de ortólogos foram filtrados de acordo com a função relacionada, para obter somente os COGs relacionados a função de peroxidase classe III. A lista final destes COGs compreende 44 grupos de ortólogos não-supervisionados (NOGs) de interesse. A partir dessa lista, a informação de ortologia presente no banco de dados *STRING* foi utilizada para calcular a quantidade

de isoformas presentes em cada uma das 20 espécies de plantas terrestres, que têm anotação no *STRING* (Tabela Suplementar 3). A mesma metodologia aplicada foi utilizada para analisar a abundância em *O. Sativa*. Primeiramente, as peroxidases de classe III desta espécie foram obtidas a partir de busca no banco de dados PeroxiBase (BAKALOVIC *et al.*, 2006). Posteriormente, com a lista de 138 genes, os identificadores para utilização no banco *STRING* foram buscados no UniProt (“UniProt: a hub for protein information”, 2015) e também o Ensembl Biomart (KINSELLA *et al.*, 2011). A lista final contém 56 identificadores aceitos pelo *STRING* e estes foram utilizados para obter os COGs para a referida espécie. No total, 44 NOGs para *O. sativa* foram atribuídos às proteínas utilizadas, dos quais 19 são compartilhados com *A. thaliana*. Os mesmos cálculos feitos anteriormente para *A. thaliana* foram repetidos utilizando a informação dos grupos de ortólogos de *O. sativa* (Tabela Suplementar 4).

3. RESULTADOS

3.1. Rede gênica antioxidante de *A. thaliana*

Como mencionado anteriormente, os genes utilizados para a construção da rede foram obtidos a partir de seis termos de processo biológico associados com metabolismo redox, oriundos do *Gene Ontology* (Figura 5). O conjunto final de genes a ser utilizado foi manualmente filtrado para identificar os genes que são diretamente envolvidos na homeostase redox, descartar duplicatas e falsos-positivos. A rede final compreendeu 208 genes e 15 compostos químicos indispensáveis para a atividade das enzimas redox, sendo substrato ou produto (Figura 6, Tabelas Suplementares 1 e 2). O resultado é uma rede metabólica envolvendo tanto interações proteína-proteína (do inglês *protein-protein interaction*, PPI) como interações proteína-químico. Os sistemas antioxidantes em geral são descritos como mantenedores do balanço redox de um organismo ou célula, onde tanto a porção enzimática como a não-enzimática trabalham juntas como, como uma rede de defesa antioxidante (DALMOLIN *et al.*, 2012; GELAIN *et al.*, 2009; MITTLER *et al.*, 2004). Estas propriedades são claramente observadas quando analisamos a Figura 6, que mostra o arcabouço da rede inferida. Os genes que compõem a rede foram classificados de acordo com a função que exercem, sendo: 101 genes tiol-redox, 97 peroxidases e 10 SOD. Também é possível ver na Figura 6 que a rede gênica antioxidante de *A. thaliana* é dividida em dois *grupos*. Essa divisão não é apenas estética, porém funcional. O *grupo* inferior é composto por genes que codificam proteínas de todas as funções aqui empregadas (SOD, tiol-redox e peroxidases), com um total de 135 genes. O *grupo* superior em contrapartida é composto inteiramente por peroxidases de classe III. Para melhor compreensão, o *grupo* inferior será referido de

agora em diante como *grupo geral* e o *grupo superior*, como *grupo classe III*. Os nós no *grupo geral* se mostram muito interconectados, ao passo que os os nós no *grupo classe III* mostram pouca (ou nenhuma) conexão entre si, mantendo-se unidos majoritariamente pelos seus substratos/produtos, como O_2 , H_2O_2 e os álcoois sinapílico e coniferílico. Assim como o O_2 , o H_2O_2 também apresenta um padrão de um nó com uma alta conectividade, servindo de ponte entre os dois *grupos*. Como mencionado anteriormente, sabe-se que o H_2O_2 realiza um papel importante em diversas funções fisiológicas das plantas e sua posição central na rede inferida reforça essa importância no metabolismo das plantas.

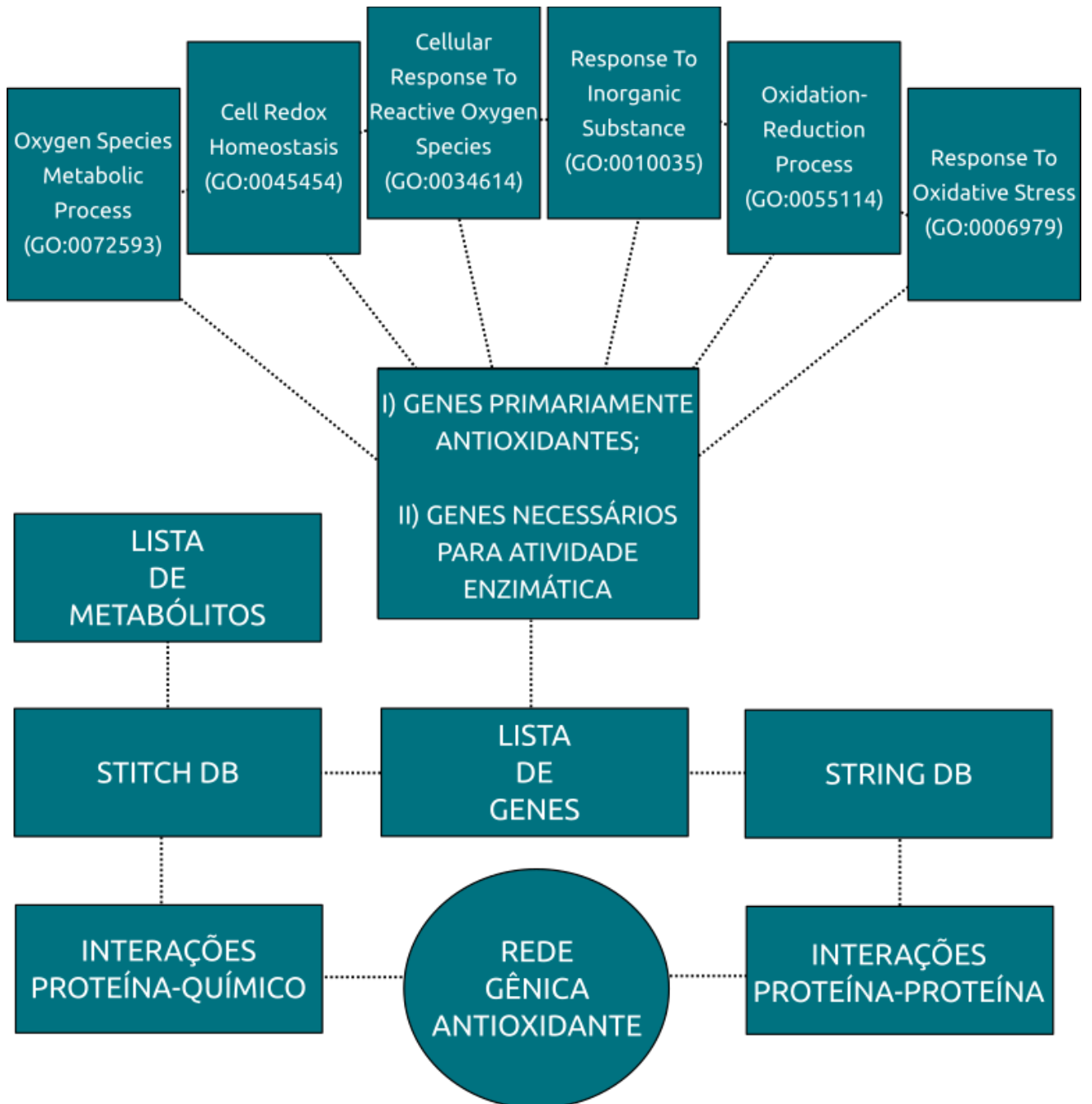


Figura 5: Fluxograma mostrando mineração e processamento de dados utilizados na construção da rede antioxidante utilizando bancos de dados públicos.

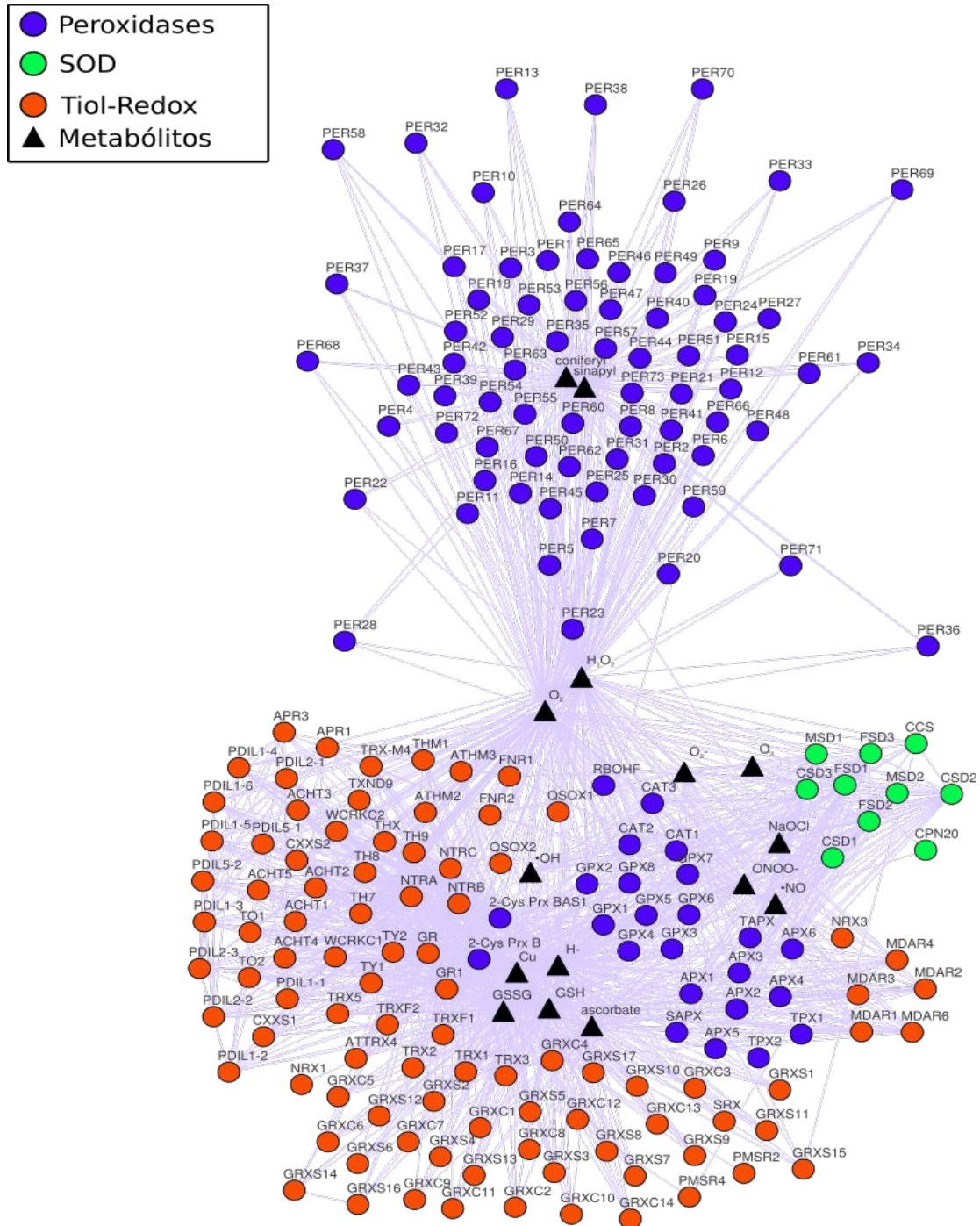


Figura 6: Rede de PPI dos 208 genes e 15 metabólitos envolvidos. A rede gerada pelo *STRING* e *STITCH* mostra conexões entre as proteínas (círculos) e metabólitos (triângulos). Cores indicam classe proteica: peroxidase (azul), SOD (verde) e tiol-redox (vermelho). A rede mostra dois grupos separados funcionalmente, um formado por peroxidases classe III e outro misto, com predominância de proteínas tiol-redox. Foram utilizados parâmetros padrão no *STRING*, com escore de confiança maior que 0.400.

3.2. Origem evolutiva da rede gênica antioxidante de *A. thaliana*

Enquanto muitas proteínas que formam o *grupo geral* são facilmente encontradas em outros Eucariontes (e.g. SOD, GPX, GR, CAT e outras peroxidases em geral), as peroxidases de classe III são proteínas encontradas exclusivamente em plantas. Sabendo disso, foi decidido investigar a origem evolutiva de cada um dos 208 genes utilizados para a construção da rede. A idéia central foi a de identificar o arquétipo genético mais ancestral para cada um dos genes antioxidantes de *A. thaliana* que tenha sido herdado verticalmente. Para realizar essa etapa, nós procuramos genes ortólogos nas 238 espécies de Eucariontes anotadas no *STRING* (Figura Suplementar 1). De acordo com os resultados aqui encontrados, sugere-se que os genes componentes do *grupo geral* surgiram em uma etapa da evolução antes da divergência da família Viridiplantae (Figura 7, Tabela Suplementar 1), compondo a porção ancestral da rede antioxidante. Todas as proteínas com função tiol-redox presentes na rede, todas as GPX, APX, CAT, peroxidases dependentes de tiorredoxina, diferentes isoformas da superóxido dismutase (Mn, Cu-Zn e Fe), bem como proteínas suplementares e indispensáveis para seu funcionamento (ATCCS e CPN20), fazem parte da porção mais ancestral da rede. Do outro lado, nós temos todos os genes de peroxidase classe III. De acordo com nossos resultados, podemos inferir que o surgimento dos primeiros ortólogos relacionados a peroxidases classe III ocorreu durante a divergência da família Viridiplantae. A exceção até agora se dá pela peroxidase 68 de *A. thaliana*, a qual não apresenta dados de ortologia no banco de dados utilizado. Essa falta de anotação pode ser resultado do algoritmo utilizado no banco COG, pois ele necessita de uma triangulação entre pelo menos três espécies para que se crie o grupo de ortólogos. Em

nossos resultados, encontramos que 21 peroxidases classe III têm sua raiz evolutiva no último ancestral comum (*Last common ancestor*, LCA) originado pela divergência da Viridiplantae; 22 outras que têm sua raiz no LCA originado pela divergência das Traqueófitas; 9 que enraizam na divergência que deu origem às Angiospermas; 5 na divergência que deu origem às eudicotiledôneas; 1 na divergência que deu origem às Rosídeas; 8 na divergência que deu origem às Malvídeas e 6 na divergência que deu origem às Brassicales (Figura 7). Quando analisados em conjunto, nossos resultados levam a hipótese de que as peroxidases de classe III podem estar ainda em expansão ao longo da evolução das plantas terrestres. O levantamento dessa hipótese nos motivou a investigar a abundância dessa classe de genes ao longo da evolução.

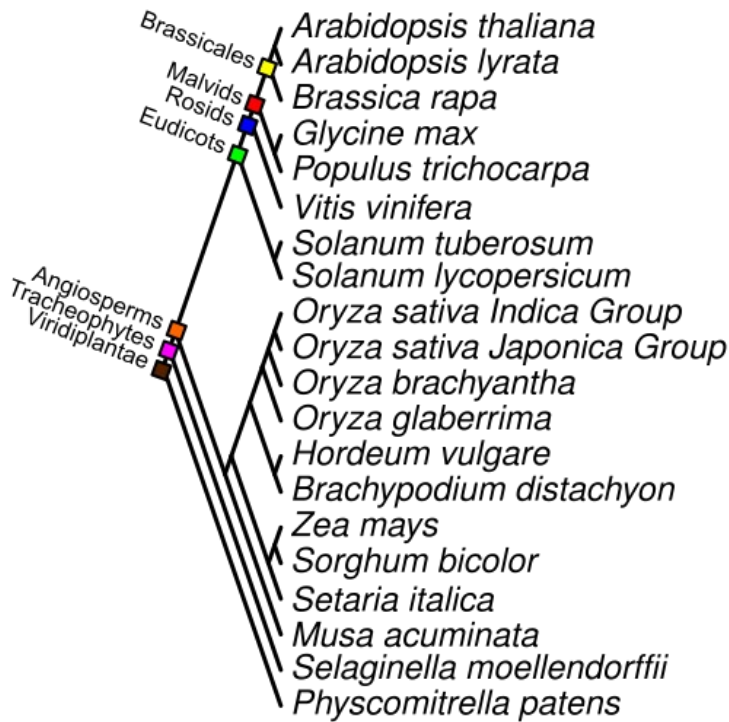
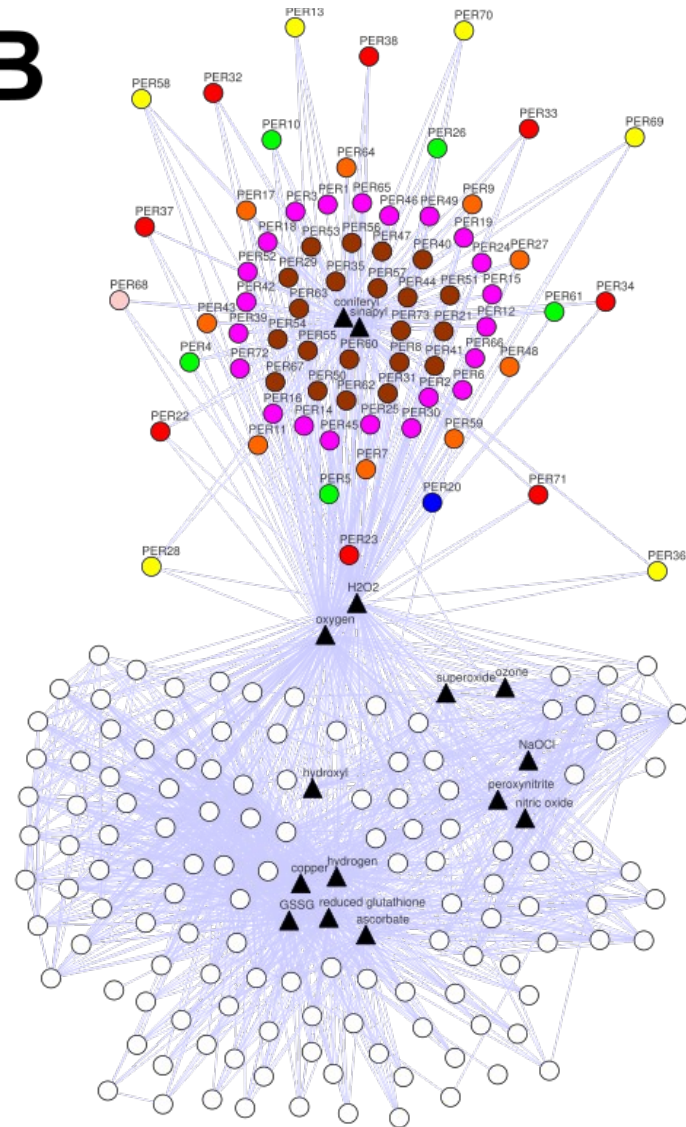
A**B**

Figura 7: Curso evolutivo da rede de *A. thaliana*. Em **A** mostra-se a árvore evolutiva das espécies de plantas terrestres. Os losangos coloridos nos nós da árvore indicam o LCA, nomeados ao lado de cada losango. As cores indicam a distribuição dos genes na rede. Os nomes nas folhas da árvore indicam as espécies (de plantas) contidas no *STRING* v10. As árvores evolutivas com todas as espécies contidas no *STRING* estão nas Figuras Suplementares de 2 a 71. Elas mostram os diferentes pontos de surgimento dos genes antioxidantes, baseado na distribuição dos grupos de ortólogos ao longo do curso evolutivo. Em **B** mostra-se a rede apresentada na Figura 6. As cores seguem o padrão dos losangos coloridos na árvore em **A**. Nós em branco são genes presentes no sistema antioxidante de plantas antes do surgimento do *grupo classe III*.

3.3. Distribuição dos genes ortólogos de peroxidase classe III em monocotiledôneas e eudicotiledôneas

Para avaliar a abundância de ortólogos de peroxidase classe III em diferentes espécies, nós utilizamos aqui a informação de ortologia disponível no *STRING*. Nesse passo, nós utilizamos a informação de presença ou ausência de ortologia em determinada espécie. A partir disso foi possível calcular a abundância de ortólogos nos 44 COGs obtidos anteriormente, nos quais estão contidas as 73 isoenzimas de *A. thaliana*, para as 20 espécies de plantas terrestres listadas no *STRING* (Tabela Suplementar 3). Podemos observar na Figura 8A a abundância de ortólogos em diferentes grupos taxonômicos: os musgos (*Physcomitrella patens*), os Licopódios (*Selaginella moellendorffii*) e as eudicotiledôneas: batata, tomate, sorgo, soja, uva, nabo e duas espécies do gênero *Arabidopsis* (*Solanum tuberosum*, *Solanum lycopersicum*, *Sorghum bicolor*, *Glycine max*, *Vitis vinifera*, *Brassica rapa*, *Arabidopsis lyrata* e *Arabidopsis thaliana*, respectivamente). Analisando a abundância do taxon mais ancestral para o mais recente, podemos ver que há um aumento de fato na abundância de peroxidases classe III. Da mesma forma, na figura 8B, podemos ver o mesmo padrão quando utilizamos os ortólogos oriundos de um táxon diferente: a monocotiledônea *Oryza sativa*. Diferentemente de musgo e Licopódios, eudicotiledôneas e monocotiledôneas apresentam 8 e 10 espécies, respectivamente. As duas primeiramente citadas cada uma apresenta apenas uma espécie com anotações no *STRING*, que são *Physcomitrella patens* e *Selaginella moellendorffii*, respectivamente. Podemos observar que há um padrão de crescimento entre os grupos funcionais de proteínas. As peroxidase de classe III têm um padrão levemente

acentuado de surgimento de ortólogos e esse padrão se repete quando utilizamos nessa análise peroxidases de classe III e seus ortólogos obtidos de monocotiledôneas, utilizando *O. sativa* como organismo modelo.

4. DISCUSSÃO

As plantas são organismos constantemente expostos a diferentes condições estressantes e portanto, desenvolveram mecanismos para lidar com diversos compostos que podem causar estresse. Sabe-se que os compostos com poder oxidante são um dos maiores causadores de estresse em plantas (DE GARA *et al.*, 2010; LOIACONO; DE TULLIO, 2012). No entanto, sabe-se também que as EROs são parte importante em diversos processos fisiológicos, sendo assim indispensáveis para as plantas (MITTLER, 2017). Neste trabalho nós construímos a rede antioxidante de *A. thaliana* e analisamos as raízes evolutivas de seus componentes. Nosso trabalho utilizou a rede de interação PPI oriunda do *STRING* e *STITCH*, duas ferramentas que geram redes canônicas, com a ressalva de que o *STITCH* também provê a interação proteína-químico. Deste modo é possível ter uma visão mais ampla das interações do que usando apenas a rede proveniente do *STRING*. Em 2004, Mittler e colaboradores propuseram uma rede gênica envolvendo compostos enzimáticos que de alguma forma eram responsáveis por gerar e remover EROs de plantas (MITTLER *et al.*, 2004). O que eles fizeram foi, basicamente, reunir esses genes geradores e removedores de EROs, que são parte de um grande sistema redox, e trazer informação sobre o estado transcricional destes genes em diferentes condições abióticas, bem como em alguns tipos de mutantes. Essa metodologia propõe também que a rede tenha aproximadamente 150 genes, não incluindo apenas os genes removedores de EROs, mas também os produtores. O que foi feito neste trabalho foi algo similar, utilizando os mesmos genes relacionados a remoção de EROs dos sistemas, obtendo resultado similar. Visto que a pesquisa referida foi realizada há mais de 10 anos, foram também

adicionados outros genes que não foram usados anteriormente. Adicionalmente, foi construída uma representação gráfica desta rede de PPI, que está representada na Figura 6. Como mencionado anteriormente, a topologia da rede nos mostra a divisão em dois *grupos*: o *grupo geral*, que é constituído por proteínas SOD, tiol-redox e peroxidases e o *grupo classe III* que é constituído inteiramente por peroxidases classe III. Apesar de estar relacionadas com outras funções fisiológicas, muitas revisões e outros trabalhos apontam para uma função antioxidante realizada pelas peroxidases classe III (ALMAGRO et al., 2009; COSIO; DUNAND, 2009; DELANNOY et al., 2006; DUROUX; WELINDER, 2003; PASSARDI et al., 2005; TOGNOLLI et al., 2002; WELINDER et al., 2002), mais especificamente a remoção H_2O_2 , de acordo com a anotação do *Gene Ontology*. A maioria dos genes do *grupo geral* é comum a todos os Eucariontes. Alguns deles, como a SOD de ferro (Fe-SOD), por exemplo, podem ser encontrados até mesmo em Archea (MILLER, 2012). Outras classes de proteínas aparecem em diferentes momentos da evolução, antes mesmo até do grande evento de oxidação (BENZIE, 2000; INUPAKUTIKA et al., 2016). Em contrapartida ao trabalho realizado por Mittler e colaboradores, que analisaram toda a rede redox, nossa análise teve focou apenas no sistema de remoção de EROs. A exceção é a enzima ATRBOHF, uma enzima homóloga a NAPDH-oxidase humana, que é uma produtora de EROs. Esta enzima tem um papel importante em diferentes situações de estresse (QU et al., 2013). Sua expressão é regulada por hormônios, como jasmonato e ácido abscísico (TORRES; DANGL, 2005).

Como mencionado anteriormente, o H_2O_2 em nossa rede tem uma alta conectividade, servindo de conector entre os dois grupos, o que era esperado. Essa

molécula realiza funções primordiais como um sinalizador e até mesmo como uma barreira de defesa, apesar de suas propriedades como uma ERO (PETROV; VAN BREUSEGEM, 2012). Recentemente Inupakutika e colaboradores publicaram uma pesquisa relacionada ao nosso trabalho, na qual o foco está na história evolutiva das SOD, APX e RBOH na rede gênica de plantas relacionada a EROs. Nesse trabalho é proposto que a rede gênica, anteriormente proposta por Mittler, poderia expandir de 150 genes para até 300 (INUPAKUTIKA *et al.*, 2016), o que se encaixa no nosso caso. Esses dois trabalhos basicamente sumarizam o que se sabe sobre redes relacionadas a EROs em plantas. Adicionalmente, nós também fizemos uma análise evolutiva dos genes da nossa rede, porém utilizando uma metodologia diferente. Nós pudemos encontrar a raiz evolutiva para os 208 genes da nossa rede e esse é corroborado pelos resultados obtidos por Inupakutika e colaboradores, que utilizaram a árvore fornecida pelo NCBI *Taxonomy* e informações da literatura (BENZIE, 2000). Nós também vimos que os genes do *grupo geral* têm sua raiz evolutiva inferida nos Eucariontes basais, como já dito previamente. As figuras Suplementares 2-71 mostram em qual LCA os genes da nossa análise surgiram. Esse resultado corrobora com os achados previamente descritos por Inupakutika e colaboradores. Também foi possível identificar o mesmo padrão de presença e ausência que é mostrado na Figura 2 do artigo publicado por Inupakutika e colaboradores, o que também reforça os nossos resultados. Além de identificar a raiz dos genes do *grupo geral*, também identificamos as raízes para os genes do *grupo classe III*. O organismo mais ancestral a apresentar ortólogos para essa classe de proteínas foi o musgo *P. patens*. Esse achado fortalece nossa

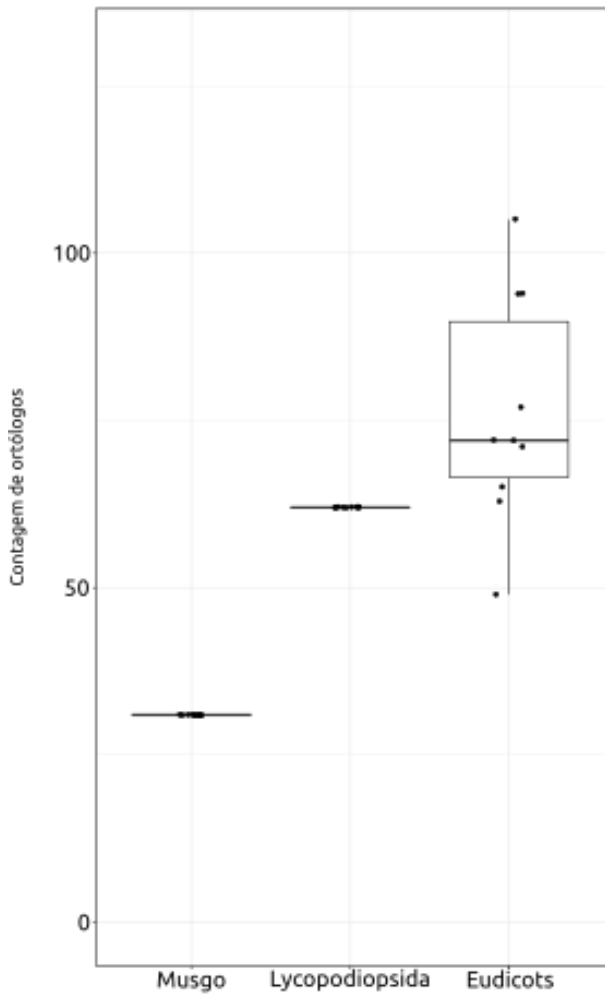
metodologia, pois é corroborado por estudos prévios sobre a filogenia desse grupo (DUROUX; WELINDER, 2003).

Focando nas peroxidases de classe III, embora este grupo de genes tenha sua estrutura, funções e organização gênica (localização de motivos, regiões promotoras, etc) muito estudadas em *A. thaliana*, pouco se tem anotação e informação sobre sua origem evolutiva. Um dos estudos que mostra evolução dos genes de peroxidase classe III em diversas espécies tem seu foco em direcionado para conservação de estruturas intergênicas (MATHÉ et al., 2010). De igual importância, sabe-se também que existe pouca informação sobre origem evolutiva de peroxidases classe III em outras espécies (LAZZAROTTO; TURCHETTO-ZOLET; MARGIS-PINHEIRO, 2015). Este fato nos motivou a questionar sobre a formação do sistema antioxidante de *A. thaliana*. A fim de avaliar se novas isoformas de diferentes enzimas continuam a surgir ao passo da evolução, nós calculamos a abundância de ortólogos de peroxidase classe III em diferentes espécies de plantas terrestres presentes no *STRING*. Os resultados mostrados na Figura 8A mostra o crescimento no número de isoformas ao longo da evolução. Isso é mostrado na Figura 7B, onde o número de isoformas que vão sendo adicionadas segue o curso evolucionário. No entanto, para realizar essa análise nós utilizamos genes ortólogos de *A. thaliana*. Nós checamos se o padrão de evolução desses genes também é o mesmo em Poaceae. Para isso, repetimos a análise utilizando ortólogos oriundos de *O. sativa* Japonica. A comparação pode ser observada na Figura 8, onde os resultados vieram como o esperado: podemos observar que o padrão de evolução desses genes é o mesmo, tanto para monocotiledôneas como para eudicotiledôneas. Aqui foram utilizados 44 grupos de ortólogos de *O. sativa* e 44 de *A.*

thaliana, onde apenas 19 eram compartilhados entre os dois organismos. Nos perguntamos se o crescimento era de fato real ou estaríamos vendo algo enviesado, sem levar em consideração o número de ortólogos *versus* o número total de genes de cada espécie. Para sanar essa dúvida, nós normalizamos o número de ortólogos pelo total de genes para cada organismo. O resultado obtido está na Figura Suplementar 72, onde podemos observar o mesmo padrão que aparece na Figura 8, o que pode ser um indicativo de que esse incremento no número de isoformas está ainda acontecendo. Outros pesquisadores publicaram estudos sobre filogenia molecular de peroxidases. Zamocký e colaboradores estudaram as heme-peroxidases (que incluem as peroxidases classe III) e viram que a maneira na qual esses genes se distribuem ao longo da evolução não é uniforme (OBINGER; ZÁMOCKÝ, 2010). De fato, dos 70 COGs obtidos no nosso estudo, 44 são relacionados a atividade de peroxidase classe III em *A. thaliana*. O fato de 73 isoformas se ajustarem em 44 COGs diferentes é um fator importante que corrobora com a complexidade relacionada a essa classe de proteínas (PASSARDI, F. *et al.*, 2005). Os resultados relacionados com evolução obtidos no nosso estudo corroboram com outros estudos evolutivos relacionados com peroxidases de classe III (KUPRIYANOVA; MAMOSHINA; EZHOVA, 2015; MATHÉ *et al.*, 2010; PASSARDI, FILIPPO *et al.*, 2004), e também com a teoria de que algumas isoformas provavelmente surgiram no começo da colonização da terra pelas plantas (DUROUX; WELINDER, 2003). De acordo com nossos resultados, todas as peroxidases classe III de *A. thaliana* surgiram juntamente com as plantas terrestres. Isso também pode indicar uma possível subdivisão de função dentro dessa classe gênica. Como previamente descrito, 21 isoformas que têm sua raiz no surgimento das

Embriófitas provavelmente seriam relacionadas com funções antioxidantes ou alguma outra. Apenas com o aparecimento das Traqueófitas que as isoformas podem ter neofuncionalizado e começar a realizar lignificação e suberização (WENG; CHAPPLE, 2010).

A

Ortólogos de Peroxidase Classe III em *A. thaliana*

B

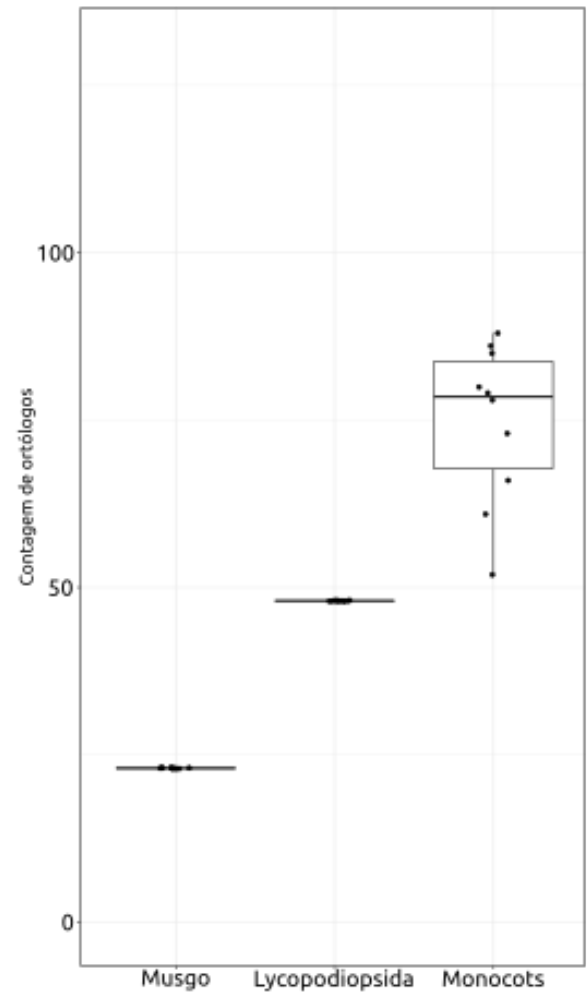
Ortólogos de Peroxidase Classe III em *O. sativa*

Figura 8: *Boxplots* mostrando abundância de genes ortólogos de Peroxidases Classe III por COG. Em **A** estão os ortólogos de peroxidases classe III, utilizando os COGs de *A. thaliana* como base; Em **B** estão os ortólogos de peroxidases classe III, utilizando os COGs de *O. sativa* como base. Os *boxplots* foram feitos com a contagem de genes (eixo Y) por espécies em cada um dos grupos taxonômicos (eixo X).

Em 1970, Susumu Ohno postulou que após um evento de duplicação, a segunda cópia do gene possivelmente ganharia uma nova função (OHNO, 1970). As peroxidases que realizam funções distintas no metabolismo vegetal provavelmente são resultado de processos de neofuncionalização (REN *et al.*, 2014). Essas variações podem ser oriundas de eventos de duplicação gênico, genômico, cromossômico ou de duplicação em tandem (KUPRIYANOVA; MAMOSHINA; EZHOVA, 2015). Um desses eventos chama a atenção e é a duplicação completa do genoma (do inglês, *whole genome duplication*, WGD), que é um dos maiores responsáveis por diversidade genética em plantas (CAO *et al.*, 2016). Estudos realizados em *Populus trichocarpa*, uma das mosquinhas-da-fruta (*Drosophila melanogaster*) e em leveduras (*Saccharomyces cerevisiae*) mostram que genes oriundos de eventos de duplicação podem ter neofuncionalizado (BEISSWANGER; STEPHAN, 2008; BYRNE; WOLFE, 2007; REN *et al.*, 2014). É possível que após esses eventos em plantas, uma pressão de seleção ainda esteja ativa (NEI; SUZUKI; NOZAWA, 2010). Estudos em pêra chinesa (*Pyrus bretschneideri*) e milho (*Zea mays*) mostram que alguns genes de peroxidase classe III foram submetidos a uma seleção purificadora (CAO *et al.*, 2016; WANG *et al.*, 2015) e baseado nisso, podemos sugerir que esses genes em *A. thaliana* também podem ter passado por isso. Um outro estudo comparou isoformas em duas espécies do gênero *Arabidopsis* (*A. thaliana* e *A. lyrata*) e viu que mesmo as isoformas que têm alta identidade entre si também têm padrões de expressões diferentes, mostrando a complexidade dessas proteínas (KUPRIYANOVA; MAMOSHINA; EZHOVA, 2015). Outros estudos que também analisam conservação de função mostram resultados semelhantes, mesmo em espécies distantes (BAKALOVIC *et al.*,

2006; MATHÉ et al., 2010). Apesar de existir alguma informação sobre o processo de conservação desses genes ao longo da evolução, pouco se sabe sobre sua origem evolutiva e como essa classe de proteínas se integrou ao nicho de enzimas antioxidantes que já existiam, formando o sistema de defesa antioxidante. Em suma, os resultados encontrados nesse estudo mostram que as peroxidases de classe III têm de fato um papel fundamental no sistema antioxidante e que eles possivelmente estão sobre pressão seletiva.

5. CONCLUSÃO

Em suma, concluímos que o sistema antioxidante de plantas é um sistema dinâmico e complexo, como outros sistemas biológicos. Os genes que compõem a rede antioxidante proposta para *A. thaliana* tiveram sua origem em diferentes momentos da escala evolutiva. Nossos resultados reforçam a teoria de que o aparecimento de peroxidases classe III em plantas seguiu a colonização da terra pelas plantas vasculares e que esta classe proteica provavelmente desempenhou um papel crucial nesta etapa. O arsenal de peroxidases classe III vem aumentando com o tempo e com a evolução das espécies de plantas terrestres, tendo os fenômenos de WGD colaborado com as múltiplas funções exercidas na fisiologia normal da planta. No entanto, além de estudar o aspecto evolutivo, mais estudos são necessários para analisar a funcionalidade destas enzimas como um sistema.

6. BIBLIOGRAFIA

4. AHMAD, Parvaiz *et al.* Roles of enzymatic and nonenzymatic antioxidants in plants during abiotic stress. **Critical reviews in biotechnology** v. 30, n. 3, p. 161–175 , 2010.0738-8551.
5. ALBANUS, Ricardo D Oliveira *et al.* Reverse engineering the neuroblastoma regulatory network uncovers max as one of the master regulators of tumor progression. **PLoS ONE** v. 8, n. 12, p. 1–14 , 2013.
6. ALM, Eric; ARKIN, Adam P. Biological networks. **Current Opinion in Structural Biology** v. 13, n. 2, p. 193–202 , 2003.
7. ALMAGRO, L. *et al.* Class III peroxidases in plant defence reactions. **Journal of Experimental Botany** v. 60, n. 2, p. 377–390 , 2009.1460-2431 (Electronic)r0022-0957 (Linking).
8. ANBAR, Ariel D. Elements and Evolution. **Science** v. 322, n. 5907 , 2008.
9. APEL, Klaus; HIRT, Heribert. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. **Annual review of plant biology** v. 55, p. 373–399 , 2004.1040-2519.
10. ASHBURNER, M *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nature genetics** v. 25, n. 1, p. 25–29 , 2000.1023909812.
11. BAILEY-SERRES, J; MITTLER, R. The roles of reactive oxygen species in plant cells. **Plant Physiology** v. 141, n. June, p. 900191 , 2006. Disponível em: <<http://www.plantphysiol.org/content/141/2/311.short>>.0032-0889.
12. BAKALOVIC, Nenad *et al.* PeroxiBase: A class III plant peroxidase database. **Phytochemistry** v. 67, n. 6, p. 534–539 , 2006.0031-9422 (Print)r0031-9422 (Linking).
13. BARABASI, Albert-Laszlo; OLTVAI, Zoltan N. Zoltán N; BARABÁSI, Albert-László. Network biology: understanding the cell's functional organization. **Nature Reviews Genetics** v. 5, n. 2, p. 101–113 , 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14735121>>.
14. BARBA-ESPIN, G. *et al.* Interaction between hydrogen peroxide and plant hormones during germination and the early growth of pea seedlings. **Plant, Cell**

- and Environment** v. 33, n. 6, p. 981–994 , 2010.1365-3040 (Electronic)n0140-7791 (Linking).
15. BARBA-ESPÍN, Gregorio *et al.* Understanding the role of H₂O₂ during pea seed germination: A combined proteomic and hormone profiling approach. **Plant, Cell and Environment** v. 34, n. 11, p. 1907–1919 , nov. 2011. Disponível em: <<http://doi.wiley.com/10.1111/j.1365-3040.2011.02386.x>>. Acesso em: 7 dez. 2016.1365-3040 (Electronic)n0140-7791 (Linking).
 16. BARRETT, T. *et al.* NCBI GEO: archive for functional genomics data sets-- update. **Nucleic Acids Research** v. 41, n. D1, p. D991–D995 , 1 jan. 2013. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks1193>>. Acesso em: 24 fev. 2017.
 17. BEISSWANGER, Steffen; STEPHAN, Wolfgang. Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America** v. 105, n. 14, p. 5447–52 , 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18381818>>.1091-6490 (Electronic)r0027-8424 (Linking).
 18. BENZIE, I F. Evolution of antioxidant defence mechanisms. **European journal of nutrition** v. 39, n. 2, p. 53–61 , 2000.1436-6207.
 19. BOLOURI, Hamid; DAVIDSON, Eric H. Modeling transcriptional regulatory networks. **BioEssays** v. 24, n. 12, p. 1118–1129 , 2002.0265-9247 (Print)r0265-9247 (Linking).
 20. BYRNE, Kevin P.; WOLFE, Kenneth H. Consistent Patterns of Rate Asymmetry and Gene Loss Indicate Widespread Neofunctionalization of Yeast Genes After Whole-Genome Duplication. **Genetics Society of America** v. 175, n. 3, p. 1341–1350 , 2007.0016-6731 (Print)r0016-6731 (Linking).
 21. CAO, Yunpeng *et al.* Structural, Evolutionary, and Functional Analysis of the Class III Peroxidase Gene Family in Chinese Pear (*Pyrus bretschneideri*). **Frontiers in Plant Science** v. 7, n. December, p. 1–12 , 2016. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fpls.2016.01874/full>>.
 22. CASTRO, Mauro A A *et al.* Evolutionary origins of human apoptosis and genome-stability gene networks. **Nucleic Acids Research** v. 36, n. 19, p. 6269–6283 , 2008.1362-4962 (Electronic).

23. CASTRO, Mauro A A *et al.* RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. **Genome biology** v. 13, n. 4, p. R29 , 2012. Disponível em: <<http://genomebiology.com/2012/13/4/R29>>.1465-6914 (Electronic)r1465-6906 (Linking).
24. COMAI, Luca. The advantages and disadvantages of being polyploid. **Nature Reviews Genetics** v. 6, n. 11, p. 836–846 , 11 nov. 2005. Disponível em: <<http://www.nature.com/doi/10.1038/nrg1711>>. Acesso em: 23 fev. 2017.
25. COSIO, C.; DUNAND, C. Specific functions of individual class III peroxidase genes. **Journal of Experimental Botany** v. 60, n. 2, p. 391–408 , 2009. Disponível em: <<http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/ern318>>.
26. DALMOLIN, Rodrigo Juliani *et al.* Transcriptomic analysis reveals pH-responsive antioxidant gene networks. **Frontiers in Bioscience** v. S4, n. 4, p. 1556–1567 , 2012.1945-0524 (Electronic)n1945-0516 (Linking).
27. DE GARA, Laura *et al.* Redox homeostasis in plants. The challenge of living with endogenous oxygen production. **Respiratory Physiology and Neurobiology** v. 173, n. SUPPL., p. S13–S19 , 2010. Disponível em: <<http://dx.doi.org/10.1016/j.resp.2010.02.007>>.1569-9048.
28. DELANNOY, E *et al.* Molecular Analysis of Class III Peroxidases from Cotton. **Journal of Cotton Science** v. 10, n. 1, p. 53–60 , 2006.
29. DIETZ, Karl-Josef; MITTLER, Ron; NOCTOR, Graham. Recent progress in understanding the role of reactive oxygen species in plant cell signaling. **Plant Physiology** v. 171, n. July, p. 1535–1539 , 2016.
30. DUNAND, Christophe *et al.* Expression of a peroxidase gene in zucchini in relation with hypocotyl growth. **Plant Physiology and Biochemistry** NULL, v. 41, n. 9, p. 805–811 , 2003.
31. DUROUX, Laurent; WELINDER, Karen G. The Peroxidase Gene Family in Plants: A Phylogenetic Overview. **Journal of Molecular Evolution** v. 57, n. 4, p. 397–407 , 2003.0022-2844 (Print)r0022-2844 (Linking).
32. FEOFILOFF, Paulo; KOHAYAKAWA, Yoshiharu; WAKABAYASHI, Yoshiko. **Uma Introdução Sucinta à Teoria dos Grafos**. [S.l: s.n.], 2011. 61 p. Disponível em: <<http://www.ime.usp.br/~pf/teoriadosgrafos/texto/TeoriaDosGrafos.pdf>>. .

33. FOYER, Christine H; NOCTOR, Graham. Ascorbate and glutathione: the heart of the redox hub. **Plant physiology** v. 155, n. 1, p. 2–18 , 2011.0032-0889.
34. GABALDÓN, Toni; KOONIN, Eugene V. Functional and evolutionary implications of gene orthology. **Nature Reviews Genetics** v. 14, n. 5, p. 360–366 , 4 abr. 2013. Disponível em: <<http://www.nature.com/doi/10.1038/nrg3456>>. Acesso em: 23 fev. 2017.
35. GE, Xiao-Min *et al.* Heterotrimeric G protein mediates ethylene-induced stomatal closure via hydrogen peroxide synthesis in Arabidopsis. **The Plant Journal** v. 82, n. 1, p. 138–150 , abr. 2015. Disponível em: <<http://doi.wiley.com/10.1111/tpj.12799>>. Acesso em: 7 dez. 2016.
36. GELAIN, Daniel P *et al.* A systematic review of human antioxidant genes. **Frontiers in Bioscience** v. 14, p. 4457–4463 , 2009.
37. GILL, Sarvajeet Singh; TUTEJA, Narendra. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. **Plant Physiology and Biochemistry** v. 48, n. 12, p. 909–930 , 2010. Disponível em: <<http://dx.doi.org/10.1016/j.plaphy.2010.08.016>>.0981-9428.
38. HALLIWELL, Barry. Reactive species and antioxidants. Redox biology is a fundamental theme of aerobic life. **Plant physiology** v. 141, n. 2, p. 312–322 , 2006.0032-0889.
39. INUPAKUTIKA, Madhuri A. *et al.* The evolution of reactive oxygen species metabolism. **Journal of Experimental Botany** v. 67, n. 21, p. 5933–5943 , 2016.
40. JIAO, Yuannian *et al.* Ancestral polyploidy in seed plants and angiosperms. **Nature** v. 473, p. 97–100 , 2011. Disponível em: <<http://www.nature.com/doi/10.1038/nature09916>>%5Cn<<http://www.ncbi.nlm.nih.gov/pubmed/21478875>>.1476-4687 (Electronic)r0028-0836 (Linking).
41. KINSELLA, Rhoda J. *et al.* Ensembl BioMart: A hub for data retrieval across taxonomic space. **Database** v. 2011, p. 1–9 , 2011.1758-0463 (Linking).
42. KITANO, H. Systems biology: A brief overview. **Science (New York, NY)** v. 295, n. 5560, p. 1662–1664 , 2002.1095-9203 (Electronic) 0036-8075 (Linking).
43. KITANO, Hiroaki. Computational systems biology. **Nature** v. 420, n. 6912, p. 206–210 , 2002. Disponível em: <<http://www.nature.com/nature/journal/v420/n6912/full/nature01254.html>>%5Cn<<http://www.nature.com/doi/10.1038/nature01254>>.9781588299055.

44. KLIPP, Edda *et al.* **Systems Biology: A Textbook**. [S.l: s.n.], 2009. . 9783527318742.
45. KOLESNIKOV, N. *et al.* ArrayExpress update--simplifying data submissions. **Nucleic Acids Research** v. 43, n. D1, p. D1113–D1116 , 28 jan. 2015. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku1057>>. Acesso em: 24 fev. 2017.
46. KUPRIYANOVA, E V; MAMOSHINA, P O; EZHOVA, T A. Evolutionary divergence of Arabidopsis thaliana classical peroxidases. **Biochemistry (Moscow)** v. 80, n. 10, p. 1362–1372 , 17 out. 2015. Disponível em: <<http://link.springer.com/10.1134/S0006297915100181>>.
47. LAMESCH, Philippe *et al.* The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. **Nucleic Acids Research** v. 40, n. D1, p. 1–9 , 2012.1362-4962 (Electronic)n0305-1048 (Linking).
48. LARSON, Richard a. The antioxidants of higher plants. **Phytochemistry** v. 27, n. 4, p. 969–978 , 1988.0031-9422.
49. LAZZAROTTO, Fernanda; TURCHETTO-ZOLET, Andreia C arina; MARGIS-PINHEIRO, Márcia. Revisiting the Non-Animal Peroxidase Superfamily. **Trends in plant science** v. 20, n. 12, p. 807–813 , 2015. Disponível em: <<http://dx.doi.org/10.1016/j.tplants.2015.08.005>>.
50. LEWIS, Louise A; MCCOURT, Richard M. Green algae and the origin of land plants. **American journal of botany** v. 91, n. 10, p. 1535–56 , out. 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21652308>>. Acesso em: 23 fev. 2017.
51. LOIACONO, F. Vanessa; DE TULLIO, Mario C. *Why We Should Stop Inferring Simple Correlations Between Antioxidants and Plant Stress Resistance: Towards the Antioxidomic Era* .**OMICS: A Journal of Integrative Biology**. [S.l: s.n.] , 2012
52. LOW, P. S.; MERIDA, J. R. The oxidative burst in plant defense: Function and signal transduction. **Physiologia Plantarum** v. 96, n. 3, p. 533–542 , mar. 1996. Disponível em: <<http://doi.wiley.com/10.1111/j.1399-3054.1996.tb00469.x>>. Acesso em: 24 fev. 2017.

53. MATHÉ, Catherine *et al.* Evolution and expression of class III peroxidases. **Archives of Biochemistry and Biophysics** NULL, v. 500, n. 1, p. 58–65 , 2010.0003-9861.
54. MILLER, Anne-Frances Frances. Superoxide dismutases: Ancient enzymes and new insights. **FEBS Letters** NULL, v. 586, n. 5, p. 585–595 , 9 mar. 2012. Disponível em: <<http://doi.wiley.com/10.1016/j.febslet.2011.10.048>>. Acesso em: 14 fev. 2017.0014-5793.
55. MIRKIN, Boris G *et al.* Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. **BMC Evolutionary Biology** v. 3, n. 2, p. 34 , 1 set. 2003. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.86.17.6661>>. Acesso em: 7 fev. 2017.
56. MITTLER, Ron *et al.* Reactive oxygen gene network of plants. **Trends in Plant Science** v. 9, n. 10, p. 490–498 , 2004.1360-1385.
57. MITTLER, Ron. ROS Are Good. **Trends in Plant Science** NULL, v. 22, n. 1, p. 11–19 , jan. 2017. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1360138516301121>>.0030-8870.
58. NANDA, Amrit K. *et al.* Reactive oxygen species during plant-microorganism early interactions. **Journal of Integrative Plant Biology** NULL, v. 52, n. 2, p. 195–204 , fev. 2010. Disponível em: <<http://doi.wiley.com/10.1111/j.1744-7909.2010.00933.x>>. Acesso em: 30 jan. 2017.1744-7909 (Electronic) 1672-9072 (Linking).
59. NEI, Masatoshi; SUZUKI, Yoshiyuki; NOZAWA, Masafumi. The Neutral Theory of Molecular Evolution in the Genomic Era. **Annual Review of Genomics and Human Genetics** v. 11, n. 1, p. 265–289 , set. 2010. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev-genom-082908-150129>>. Acesso em: 13 fev. 2017.
60. OBINGER, Christian; ZÁMOCKÝ, Marcel. **Molecular Phylogeny of Heme Peroxidases**. [S.l.: s.n.], 2010. 7-35 p. All currently available gene sequences of heme peroxidases can be phylogenetically divided in two superfamilies and three families. In this chapter, the phylogenetics and genomic distribution of each group are presented. Within the peroxidase–cyclooxygenase superfamily, the main evolutionary direction developed peroxidatic heme proteins involved in the innate immune defense system and in biosynthesis of (iodinated) hormones. The

peroxidase–catalase superfamily is widely spread mainly among bacteria, fungi, and plants, and particularly in Class I led to the evolution of bifunctional catalase–peroxidases. Its numerous fungal representatives of Class II are involved in carbon recycling via lignin degradation, whereas Class III secretory peroxidases from algae and plants are included in various forms of secondary metabolism. The family of di-heme peroxidases are predominantly bacteria-inducible enzymes; however, a few corresponding genes were also detected in archaeal genomes. Four subfamilies of dyp-type peroxidases capable of degradation of various xenobiotics are abundant mainly among bacteria and fungi. Heme-haloperoxidase genes are widely spread among sac and club fungi, but corresponding genes were recently found also among oomycetes. All described families herein represent heme peroxidases of broad diversity in structure and function. Our accumulating knowledge about the evolution of various enzymatic functions and physiological roles can be exploited in future directed evolution approaches for engineering peroxidase genes de novo for various demands.9783642126277.

61. OHNO, Susumu. **Evolution by Gene Duplication**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1970. Disponível em: <<http://link.springer.com/10.1007/978-3-642-86659-3>>. Acesso em: 16 fev. 2017. .978-3-642-86661-6.
62. PANCHY, Nicholas; LEHTI-SHIU, Melissa D.; SHIU, Shin-Han. Evolution of gene duplication in plants. **Plant Physiology** v. 171, n. 4, p. pp.00523.2016 , 2016. Disponível em: <<http://www.plantphysiol.org/lookup/doi/10.1104/pp.16.00523>>.0000000164702.
63. PASSARDI, F. *et al.* Peroxidases have more functions than a Swiss army knife. **Plant Cell Reports** v. 24, n. 5, p. 255–265 , 2005.0721-7714 (Print)r0721-7714 (Linking).
64. PASSARDI, Filippo *et al.* The class III peroxidase multigenic family in rice and its evolution in land plants. **Phytochemistry** v. 65, n. 13, p. 1879–1893 , 2004.
65. PETROV, Veselin Dimitrov; VAN BREUSEGEM, Frank. Hydrogen peroxide-a central hub for information flow in plant cells. **AoB PLANTS** v. 2012, p. pls014 , 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22708052>>. Acesso em: 13 fev. 2017.
66. QU, Ai-Li *et al.* Molecular mechanisms of the plant heat stress response. **Biochemical and Biophysical Research Communications** v. 432, n. 2, p. 203–

- 207 , 2013. Disponível em:
<<http://linkinghub.elsevier.com/retrieve/pii/S0006291X13001940>>.
67. REN, Lin-Ling *et al.* Subcellular Relocalization and Positive Selection Play Key Roles in the Retention of Duplicate Genes of Populus Class III Peroxidase Family. **Plant Cell** v. 26, n. 6, p. 2404–2419 , 2014. Disponível em:
<<http://dx.doi.org/10.1105/tpc.114.124750>%5Cn<http://www.ncbi.nlm.nih.gov/pubmed/24934172>%5Cn<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4114941>%5Cn<http://www.plantcell.org/cgi/pmidlookup?view=long&pmid=24934172>%5Cn<http://www.plantcell.org/content/26/6/2404>>. 15 32-298X (Electronic)r1040-4651 (Linking).
68. RODRÍGUEZ, Andrés A; GRUNBERG, Karina A; TALEISNIK, Edith L. Reactive oxygen species in the elongation zone of maize leaves are necessary for leaf extension. **Plant physiology** v. 129, n. 4, p. 1627–32 , ago. 2002. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/12177475>>. Acesso em: 24 fev. 2017.
69. SHARMA, Pallavi *et al.* Reactive Oxygen Species, Oxidative Damage, and Antioxidative Defense Mechanism in Plants under Stressful Conditions. **Journal of Botany** v. 2012, p. 1–26 , 2012. Disponível em:
<<http://www.hindawi.com/journals/jb/2012/217037/>>. Acesso em: 23 mar. 2017.
70. SOLTIS, Douglas E; VISGER, Clayton J; SOLTIS, Pamela S. The polyploidy revolution then...and now: Stebbins revisited. **American journal of botany** v. 101, n. 7, p. 1057–1078 , jul. 2014. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/25049267>>. Acesso em: 23 fev. 2017.
71. SZKLARCZYK, Damian *et al.* STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. **Nucleic Acids Research** v. 44, n. D1, p. D380–D384 , 2016.
72. SZKLARCZYK, Damian *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. **Nucleic acids research** NULL, v. 43, n. October 2014, p. 447–452 , 2014.
73. TE BEEST, Mariska *et al.* The more the better? The role of polyploidy in facilitating plant invasions. **Annals of Botany** NULL, v. 109, n. 1, p. 19–45 , 2012.1095-8290 (Electronic)n0305-7364 (Linking).
74. TOGNOLLI, Michael *et al.* Analysis and expression of the class III peroxidase large gene family in Arabidopsis thaliana. **Gene** v. 288, n. 1–2, p. 129–138 , 2002. Disponível em:

<<http://www.sciencedirect.com/science/article/pii/S0378111902004651>>.4122705
746.

- 75.** TORRES, Miguel Angel; DANGL, Jeffery L. Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. **Current Opinion in Plant Biology** v. 8, n. 4, p. 397–403 , 2005.1369-5266 (Print)n1369-5266 (Linking).
- 76.** UniProt: a hub for protein information. **Nucleic Acids Research** v. 43, n. D1, p. D204–D212 , 28 jan. 2015. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku989>>. Acesso em: 25 jan. 2017.
- 77.** VERLI, Hugo. **Bioinformática da Biologia à Flexibilidade Molecular**. 1. ed. Porto Alegre: [s.n.], 2014. 282 p. .
- 78.** WANG, Yu *et al.* Systematic analysis of maize class III peroxidase gene family reveals a conserved subfamily involved in abiotic stress response. **Gene** v. 566, n. 1, p. 95–108 , 2015. Disponível em: <<http://dx.doi.org/10.1016/j.gene.2015.04.041>>. Acesso em: 14 fev. 2017.8655165786.
- 79.** WELINDER, Karen G. *et al.* Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*. **European Journal of Biochemistry** v. 269, n. 24, p. 6063–6081 , 2002.0014-2956 (Print)r0014-2956 (Linking).
- 80.** WENG, Jing-ke; CHAPPLE, Clint. The origin and evolution of lignin biosynthesis. **New Phytologist** v. 187, p. 273–285 , 2010.
- 81.** ZAMOCKY, Marcel; FURTMÜLLER, Paul G; OBINGER, Christian. Evolution of catalases from bacteria to humans. **Antioxidants & redox signaling** v. 10, n. 9, p. 1527–48 , 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18498226>%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2959186>.1557-7716 (Electronic).

7. ANEXOS

Figura Suplementar 1: Figura mostrando a árvore de espécies construída com o *geneplast*. A topologia da árvore reflete diretamente na relação filogenética entre os 238 Eucariontes disponíveis no *STRING* v10. O enraizamento da árvore foi feito em *A. thaliana*. Cada nó de espécie está descrito como *root 2* até *root 19*. Essas marcas indicam em qual último ancestral comum (LCA) o referido ramo foi originado.

Figuras Suplementares 2-71: As Figuras de 2 a 71 mostram o cenário mais parsimonioso para os 70 COGs obtidos na nossa análise. Para construção desse cenário, utilizamos o pacote *geneplast* que executa um algoritmo específico, que utiliza a distribuição de ortólogos para inferir a raiz evolutiva de um dado gene. A partir disso, nós assumimos que o ortólogo em questão apareceu primeiramente na raiz inferida, sendo transmitido posteriormente para os descendentes. Em todas as figuras, círculos verdes representam a presença daquele COG naquela espécie; círculos cinza, sua ausência. Os losangos vermelhos vazados representam a raiz inferida para um dado COG e um *D-score*, que estima a estabilidade daquela raiz evolutiva. Os *D-scores* são acompanhados de P-valores ajustados empíricos, gerados por análise de permutação.

Figura Suplementar 72: *Boxplots* mostrando a distribuição dos ortólogos de Peroxidase classe III COGs de *A. thaliana* e *O. sativa* como base, nos táxons de plantas terrestres. Em **A** estão os ortólogos de peroxidases classe III em *A. thaliana* e em **B**, estão os ortólogos de peroxidases classe III em *O. sativa*. Os *boxplots* foram feitos a partir da razão entre a contagem de genes nos COGs pelo total de genes de cada espécie (eixo Y) por espécies em cada um dos grupos taxômicos (eixo X)

Tabela Suplementar 1: Tabela indicando a descrição de todos os genes utilizados na nossa análise. A primeira coluna indica a classe proteica (se é uma peroxidase (PRX), tiol-redox (TR) ou SOD). A segunda coluna indica o nome da proteína conforme o TAIR e o Uniprot; a terceira coluna indica o ID TAIR, que também é o *locus* cromossômico; a quarta coluna indica o grupo de ortólogos ao qual cada gene

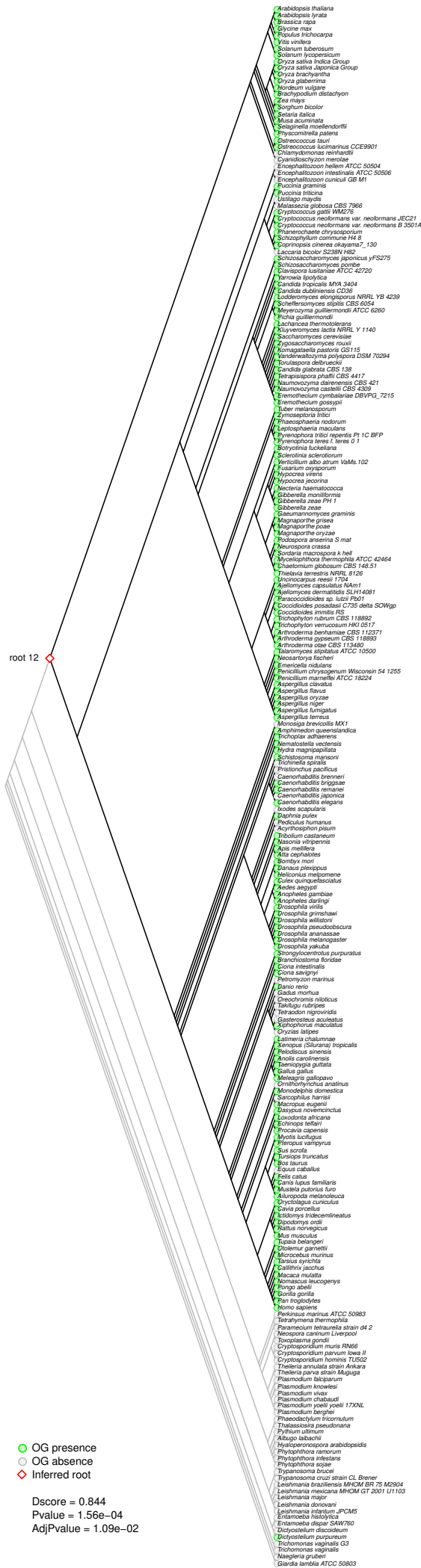
pertence e a quinta coluna, a raiz evolutiva inferida, calculada com o *geneplast* para cada grupo de ortólogos.

Tabela Suplementar 2: Tabela indicando os 15 compostos químicos utilizados na nossa análise. A primeira coluna indica o nome do elemento, seguido na segunda coluna por sua fórmula. A terceira e quarta colunas mostram os identificadores do KEGG e do PubChem. O hidriro e ozônio não apresentaram entrada no banco de dados do KEGG, portanto não têm seus identificadores.

Tabela Suplementar 3: Tabela indicando informação acerca da abundância das diferentes classes de proteínas utilizadas no trabalho, usando informação de ortologia obtida dos COGs de *A. thaliana*. Os COGs foram divididos de acordo com a função relacionada a eles. No total, obtemos 44 COGs relacionados a peroxidase classe III; 7 COGs relacionados a outras peroxidases; 4 COGs relacionados a SOD e 15 COGs relacionados a função tiol-redox. Cada coluna representa uma das 20 espécies de plantas terrestres disponíveis no STRING. Acima dos nomes das espécies, estão os identificadores de cada uma delas.

Tabela Suplementar 4: Tabela indicando informação acerca da abundância das peroxidases classe III utilizadas no trabalho, usando informação de ortologia obtida dos COGs de *O. sativa*. No total, obtemos 44 COGs relacionados a peroxidase classe III em *O. sativa*. Cada coluna representa uma das 20 espécies de plantas terrestres disponíveis no *STRING*. Acima dos nomes das espécies, estão os identificadores de cada uma delas. A contagem de ortólogos, total de genes e a razão entre essas duas informações encontram-se nas linhas finais da tabela.

Tabela Suplementar 5: Esta tabela mostra a distribuição dos 208 genes utilizados na nossa análise nos 70 COGs obtidos do *STRING*.



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.844
 Pvalue = 1.56e-04
 AdjPvalue = 1.09e-02





● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 8.61e-06
 AdjPvalue = 6.02e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 7.69e-06
 AdjPvalue = 5.39e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.915
 Pvalue = 3.72e-05
 AdjPvalue = 2.61e-03



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.647
 Pvalue = 2.13e-03
 AdjPvalue = 1.49e-01



● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 8.46e-06
 AdjPvalue = 5.92e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.795
 Pvalue = 3.61e-04
 AdjPvalue = 2.52e-02



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.889
 Pvalue = 5.87e-05
 AdjPvalue = 4.11e-03



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.944
 Pvalue = 1.92e-05
 AdjPvalue = 1.35e-03



● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 7.64e-06
 AdjPvalue = 5.35e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.993
 Pvalue = 8.53e-06
 AdjPvalue = 5.97e-04



Dscore = 0.952
Pvalue = 1.15e-05
AdjPvalue = 8.07e-04



● OG presence
● OG absence
◊ Inferred root

Dscore = 0.889
 Pvalue = 5.73e-05
 AdjPvalue = 4.01e-03



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.952
 Pvalue = 1.20e-05
 AdjPvalue = 8.39e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 8.05e-06
 AdjPvalue = 5.64e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.819
 Pvalue = 1.19e-04
 AdjPvalue = 8.35e-03



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.952
 Pvalue = 1.35e-05
 AdjPvalue = 9.42e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.852
 Pvalue = 1.11e-04
 AdjPvalue = 7.76e-03



● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 8.44e-06
 AdjPvalue = 5.91e-04



● OG presence
○ OG absence
○ Inferred root

Dscore = 0.788
 Pvalue = 3.27e-05
 AdjPvalue = 2.29e-03



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.992
 Pvalue = 6.18e-06
 AdjPvalue = 4.32e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.987
 Pvalue = 6.35e-06
 AdjPvalue = 4.45e-04



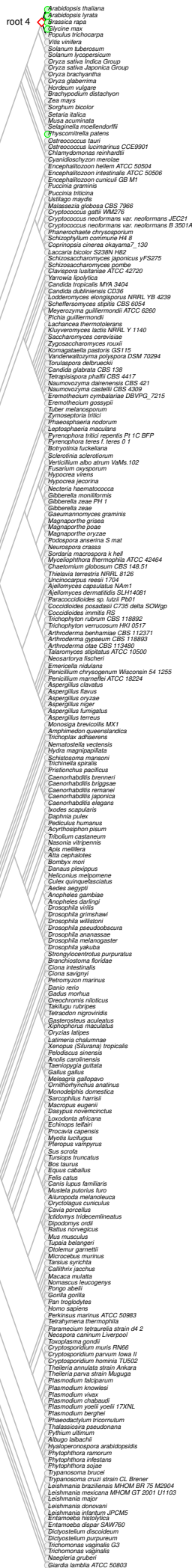
● OG presence
● OG absence
◇ Inferred root

Dscore = 0.988
 Pvalue = 1.04e-05
 AdjPvalue = 7.25e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 7.58e-06
 AdjPvalue = 5.31e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.854
 Pvalue = 2.51e-05
 AdjPvalue = 1.76e-03



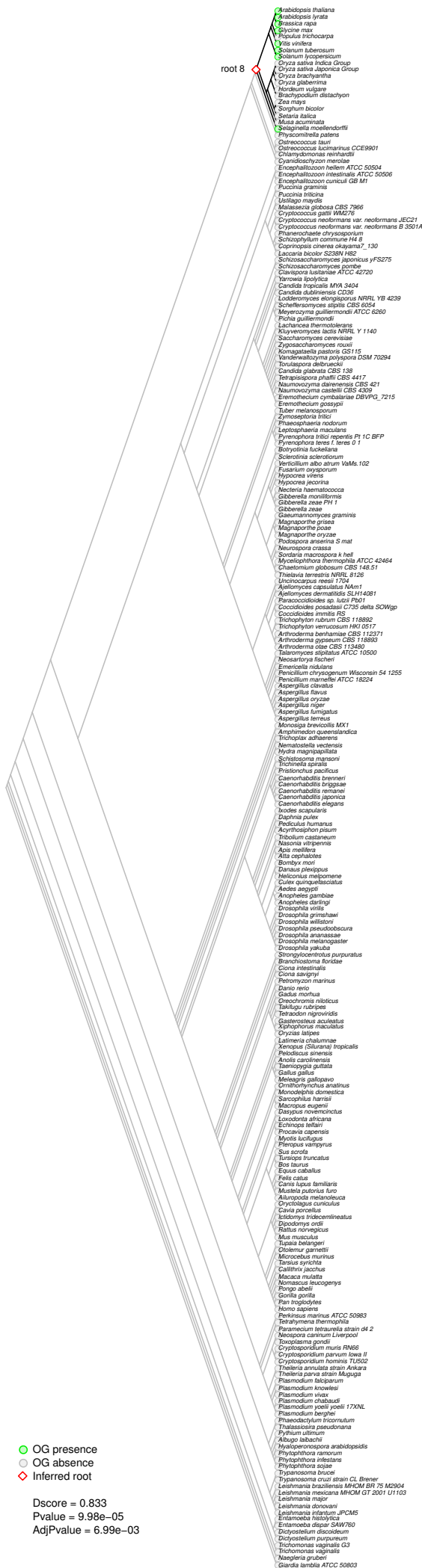
● OG presence
● OG absence
◇ Inferred root

Dscore = 0.992
 Pvalue = 7.09e-06
 AdjPvalue = 4.96e-04



● OG presence
● OG absence
◇ Inferred root

Dscore = 0.978
 Pvalue = 1.07e-05
 AdjPvalue = 7.49e-04



Dscore = 0.833
Pvalue = 9.98e-05
AdjPvalue = 6.99e-03

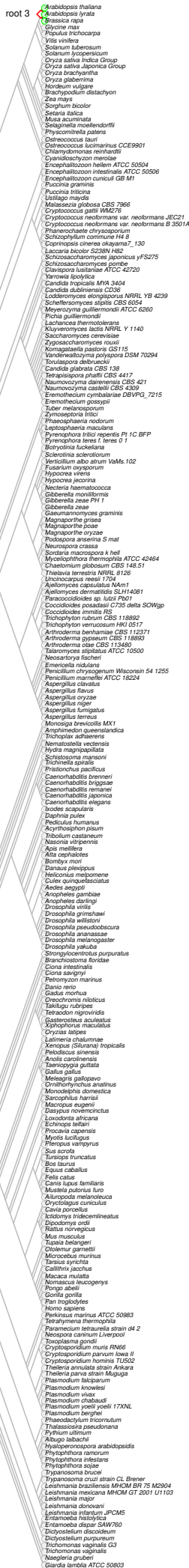


● OG presence
● OG absence
◇ Inferred root

Dscore = 0.988
 Pvalue = 1.02e-05
 AdjPvalue = 7.14e-04



● OG presence
 ● OG absence
 ◆ Inferred root
 Dscore = 1
 Pvalue = 4.51e-07
 AdjPvalue = 3.16e-05



● OG presence
● OG absence
○ Inferred root

Dscore = 1
 Pvalue = 4.41e-07
 AdjPvalue = 3.09e-05



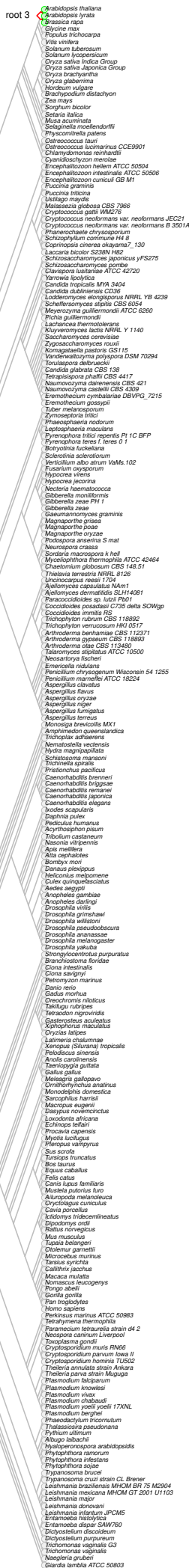
● OG presence
 ○ OG absence
 ◆ Inferred root
 Dscore = 0.667
 Pvalue = 5.65e-06
 AdjPvalue = 3.96e-04



Dscore = 1
Pvalue = 4.52e-07
AdjPvalue = 3.16e-05

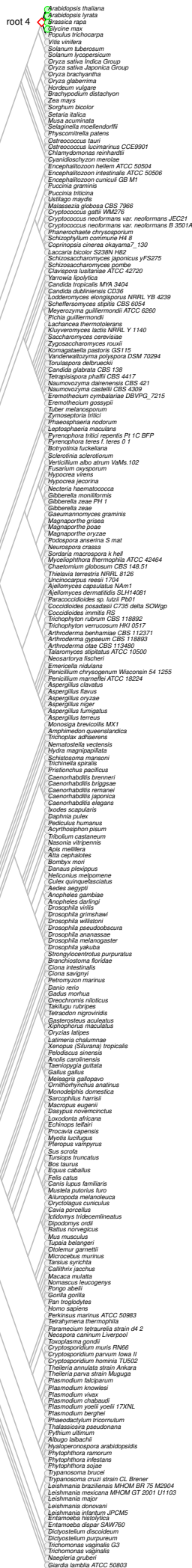


● OG presence
 ● OG absence
 ◆ Inferred root
 Dscore = 1
 Pvalue = 4.74e-07
 AdjPvalue = 3.32e-05



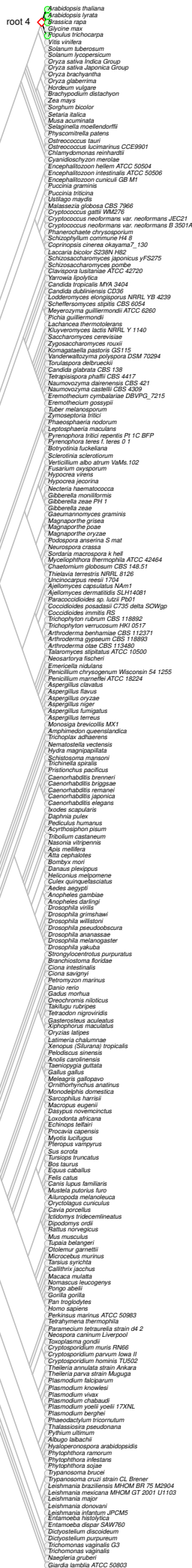
● OG presence
● OG absence
◇ Inferred root

Dscore = 1
 Pvalue = 4.89e-07
 AdjPvalue = 3.42e-05



● OG presence
○ OG absence
◇ Inferred root

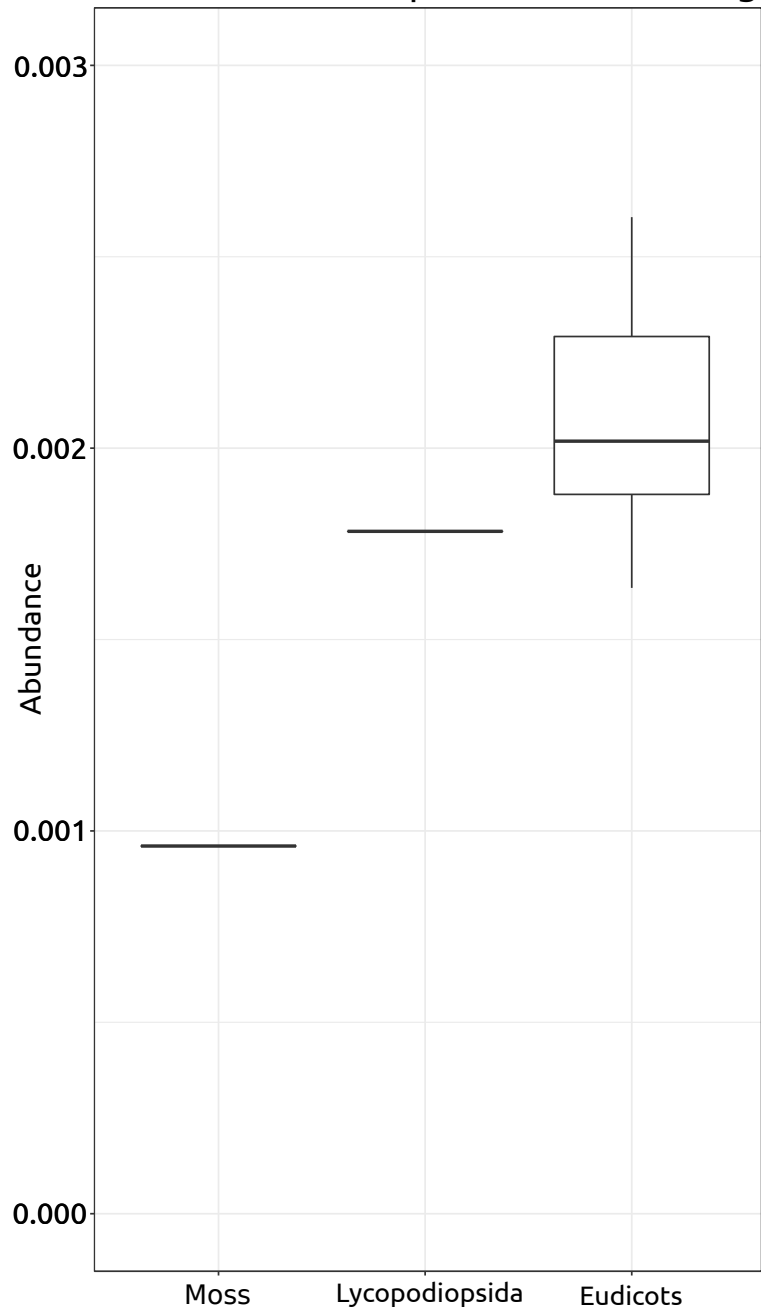
Dscore = 0.917
 Pvalue = 2.79e-06
 AdjPvalue = 1.95e-04



● OG presence
○ OG absence
◇ Inferred root

Dscore = 0.917
 Pvalue = 2.48e-06
 AdjPvalue = 1.73e-04

A. thaliana Class III peroxidase orthologs



O. sativa class III peroxidase orthologs

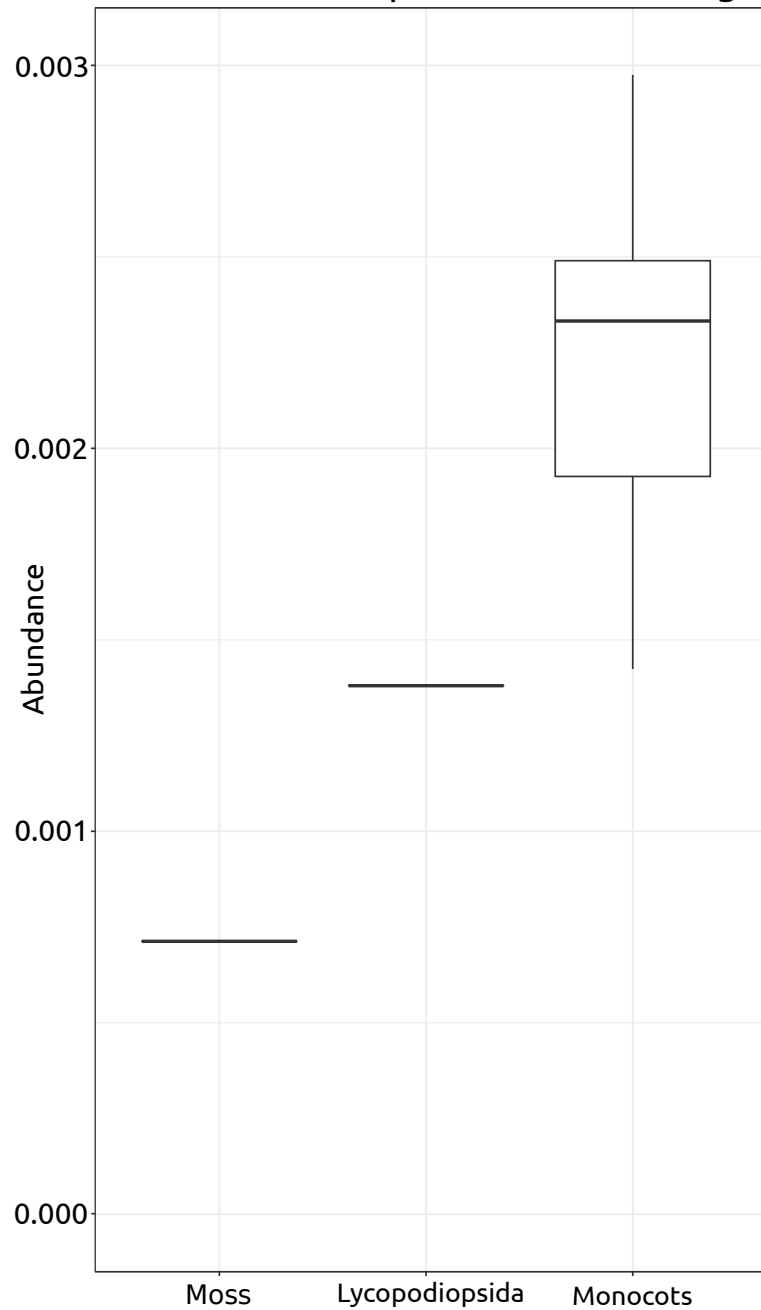


Tabela 1: Descrição dos genes utilizados na análise

Classe Proteica	Nome da Proteina	Tair ID	Grupo de Ortólogo	Raiz Evolutiva
PRX	PER13	AT1G77100	NOG233164	3
PRX	PER28	AT3G03670	NOG235401	3
PRX	PER36	AT3G50990	NOG236115	3
PRX	PER58	AT5G19880	NOG232894	3
PRX	PER69	AT5G64100	NOG232879	3
PRX	PER70	AT5G64110	NOG232889	3
PRX	PER22	AT2G38380	NOG237024	4
PRX	PER23	AT2G38390	NOG237024	4
PRX	PER32	AT3G32980	NOG276879	4
PRX	PER33	AT3G49110	NOG276879	4
PRX	PER34	AT3G49120	NOG276879	4
PRX	PER37	AT4G08770	NOG276879	4
PRX	PER38	AT4G08780	NOG276879	4
PRX	PER71	AT5G64120	NOG131107	4
PRX	PER20	AT2G35380	NOG129430	5
PRX	PER10	AT1G49570	NOG275802	6
PRX	PER26	AT2G43480	NOG130232	6
PRX	PER4	AT1G14540	NOG185758	6
PRX	PER5	AT1G14550	NOG185758	6
PRX	PER61	AT5G24070	NOG130232	6
PRX	PER11	AT1G68850	NOG123647	7
PRX	PER17	AT2G22420	NOG131444	7
PRX	PER27	AT3G01190	NOG126963	7
PRX	PER43	AT4G25980	NOG124919	7
PRX	PER48	AT4G33870	NOG128070	7
PRX	PER59	AT5G19890	NOG126611	7
PRX	PER64	AT5G42180	NOG129978	7
PRX	PER7	AT1G30870	NOG133099	7
PRX	PER9	AT1G44970	NOG131772	7
PRX	PER1	AT1G05240	NOG122617	8
PRX	PER2	AT1G05250	NOG122617	8
PRX	PER12	AT1G71695	NOG19163	8
PRX	PER14	AT2G18140	NOG123237	8
PRX	PER15	AT2G18150	NOG123237	8
PRX	PER16	AT2G18980	NOG58286	8
PRX	PER18	AT2G24800	NOG121808	8
PRX	PER19	AT2G34060	NOG126859	8
PRX	PER24	AT2G39040	NOG122617	8
PRX	PER25	AT2G41480	NOG131641	8
PRX	PER3	AT1G05260	NOG122617	8
PRX	PER30	AT3G21770	NOG122617	8
PRX	PER39	AT4G11290	NOG122617	8
PRX	PER42	AT4G21960	NOG33768	8
PRX	PER45	AT4G30170	NOG58286	8

Classe Proteica	Nome da Proteina	Tair ID	Grupo de Ortólogo	Raiz Evolutiva
PRX	PER46	AT4G31760	NOG121808	8
PRX	PER49	AT4G36430	NOG123237	8
PRX	PER52	AT5G05340	NOG130945	8
PRX	PER6	AT1G24110	NOG123538	8
PRX	PER65	AT5G47000	NOG123538	8
PRX	PER66	AT5G51890	NOG131810	8
PRX	PER72	AT5G66390	NOG123237	8
PRX	PER21	AT2G37130	NOG135096	9
PRX	PER29	AT3G17070	NOG24739	9
PRX	PER31	AT3G28200	NOG29576	9
PRX	PER35	AT3G49960	NOG29576	9
PRX	PER40	AT4G16270	NOG130428	9
PRX	PER41	AT4G17690	NOG129341	9
PRX	PER44	AT4G26010	NOG26018	9
PRX	PER47	AT4G33420	NOG101992	9
PRX	PER50	AT4G37520	NOG29576	9
PRX	PER51	AT4G37530	NOG29576	9
PRX	PER53	AT5G06720	NOG125285	9
PRX	PER54	AT5G06730	NOG125285	9
PRX	PER55	AT5G14130	NOG129341	9
PRX	PER56	AT5G15180	NOG128539	9
PRX	PER57	AT5G17820	NOG26018	9
PRX	PER60	AT5G22410	NOG26018	9
PRX	PER62	AT5G39580	NOG29116	9
PRX	PER63	AT5G40150	NOG29576	9
PRX	PER67	AT5G58390	NOG124136	9
PRX	PER73	AT5G67400	NOG29576	9
PRX	PER8	AT1G34510	NOG26018	9
PRX	PER68	AT5G58400	--	--
PRX	APX4	AT4G09010	NOG07211	10
PRX	APX1	AT1G07890	COG0376	15
PRX	APX2	AT3G09640	COG0376	15
PRX	APX3	AT4G35000	COG0376	15
PRX	APX5	AT4G35970	COG0376	15
PRX	APX6	AT4G32320	COG0376	15
PRX	SAPX	AT4G08390	COG0376	15
PRX	TAPX	AT1G77490	COG0376	15
PRX	CAT1	AT1G20630	COG0753	18
PRX	CAT2	AT4G35090	COG0753	18
PRX	CAT3	AT1G20620	COG0753	18
PRX	GPX1	AT2G25080	COG0386	18
PRX	GPX2	AT2G31570	COG0386	18
PRX	GPX3	AT2G43350	COG0386	18
PRX	GPX4	AT2G48150	COG0386	18
PRX	GPX5	AT3G63080	COG0386	18

Classe Proteica	Nome da Proteina	Tair ID	Grupo de Ortólogo	Raiz Evolutiva
PRX	GPX6	AT4G11600	COG0386	18
PRX	GPX7	AT4G31870	COG0386	18
PRX	GPX8	AT1G63460	COG0386	18
PRX	TPX1	AT1G65980	COG0678	18
PRX	TPX2	AT1G65970	COG0678	18
PRX	ATRBOHF	AT1G64060	NOG00684	18
PRX	2-Cys Prx B	AT5G06290	COG0450	19
PRX	2-Cys Prx BAS1	AT3G11630	COG0450	19
SOD	CCS	AT1G12520	COG2608	12
SOD	CPN20	AT5G20720	COG0234	17
SOD	CSD1	AT1G08830	COG2032	18
SOD	CSD2	AT2G28190	COG2032	18
SOD	CSD3	AT5G18100	COG2032	18
SOD	FSD1	AT4G25100	COG0605	18
SOD	FSD2	AT5G51100	COG0605	18
SOD	FSD3	AT5G23310	COG0605	18
SOD	MSD1	AT3G10920	COG0605	18
SOD	MSD2	AT3G56350	COG0605	18
TR	APRL4	AT1G34780	NOG45847	9
TR	APRL5	AT3G03860	NOG45847	9
TR	APRL6	AT4G08930	NOG45847	9
TR	APRL7	AT5G18120	NOG45847	9
TR	LTO1	AT4G35760	COG4243	11
TR	SRX	AT1G31170	COG5119	12
TR	NRX1	AT1G60420	NOG18420	15
TR	NRX3	AT4G31240	NOG18420	15
TR	QSOX1	AT1G15020	NOG10297	15
TR	QSOX2	AT2G01270	NOG10297	15
TR	GR	AT3G54660	COG1249	18
TR	GR1	AT3G24170	COG1249	18
TR	GRXC1	AT5G63030	COG0695	18
TR	GRXC10	AT5G11930	COG0695	18
TR	GRXC11	AT3G62950	COG0695	18
TR	GRXC12	AT2G47870	COG0695	18
TR	GRXC13	AT2G47880	COG0695	18
TR	GRXC14	AT3G62960	COG0695	18
TR	GRXC2	AT5G40370	COG0695	18
TR	GRXC3	AT1G77370	COG0695	18
TR	GRXC4	AT5G20500	COG0695	18
TR	GRXC5	AT4G28730	COG0695	18
TR	GRXC6	AT4G33040	COG0695	18
TR	GRXC7	AT3G02000	COG0695	18
TR	GRXC8	AT5G14070	COG0695	18
TR	GRXC9	AT1G28480	COG0695	18
TR	GRXS1	AT1G03020	COG0695	18

Classe Proteica	Nome da Proteina	Tair ID	Grupo de Ortólogo	Raiz Evolutiva
TR	GRXS10	AT3G21460	COG0695	18
TR	GRXS11	AT1G06830	COG0695	18
TR	GRXS12	AT2G20270	COG0695	18
TR	GRXS13	AT1G03850	COG0695	18
TR	GRXS2	AT5G18600	COG0695	18
TR	GRXS3	AT4G15700	COG0695	18
TR	GRXS4	AT4G15680	COG0695	18
TR	GRXS5	AT4G15690	COG0695	18
TR	GRXS6	AT3G62930	COG0695	18
TR	GRXS7	AT4G15670	COG0695	18
TR	GRXS8	AT4G15660	COG0695	18
TR	GRXS9	AT2G30540	COG0695	18
TR	PDIL5-3	AT3G20560	NOG01122	18
TR	PDIL5-4	AT4G27080	NOG01122	18
TR	ACHT1	AT4G26160	COG0526	19
TR	ACHT2	AT4G29670	COG0526	19
TR	ACHT3	AT2G33270	COG0526	19
TR	ACHT4	AT1G08570	COG0526	19
TR	ACHT5	AT5G61440	COG0526	19
TR	APR1	AT4G04610	COG0175	19
TR	APR3	AT4G21990	COG0175	19
TR	ATHX	AT1G50320	COG0526	19
TR	TO1	AT2G35010	COG0526	19
TR	TO2	AT1G31020	COG0526	19
TR	TY1	AT1G76760	COG0526	19
TR	TY2	AT1G43560	COG0526	19
TR	CXXS1	AT1G11530	COG0526	19
TR	CXXS2	AT2G40790	COG0526	19
TR	TRX-M3	AT2G15570	COG0526	19
TR	GRXS14	AT3G54900	COG0278	19
TR	GRXS15	AT3G15660	COG0278	19
TR	GRXS16	AT2G38270	COG0278	19
TR	GRXS17	AT4G04950	COG0278,COG0526	19
TR	FNR1	AT5G66190	COG0369	19
TR	FNR2	AT1G20020	COG0369	19
TR	MDAR1	AT3G52880	COG0446	19
TR	MDAR2	AT5G03630	COG0446	19
TR	MDAR3	AT3G09940	COG0446	19
TR	MDAR4	AT3G27820	COG0446	19
TR	MDAR5	AT1G63940	COG0446	19
TR	PMSR2	AT5G07460	COG0225	19
TR	PMSR4	AT4G25130	COG0225	19
TR	PMSR1	AT5G61640	COG0225	19
TR	PMSR3	AT5G07470	COG0225	19
TR	NTRB	AT4G35460	COG0492	19

Classe Proteica	Nome da Proteina	Tair ID	Grupo de Ortólogo	Raiz Evolutiva
TR	NTRA	AT2G17420	COG0492	19
TR	NTRC	AT2G41680	COG0492,COG0526	19
TR	PDIL1-1	AT1G21750	COG0526	19
TR	PDIL1-2	AT1G77510	COG0526	19
TR	PDIL1-3	AT3G54960	COG0526	19
TR	PDIL1-4	AT5G60640	COG0526	19
TR	PDIL1-5	AT1G52260	COG0526	19
TR	PDIL1-6	AT3G16110	COG0526	19
TR	PDIL2-1	AT2G47470	COG0526	19
TR	PDIL2-2	AT1G04980	COG0526	19
TR	PDIL2-3	AT2G32920	COG0526	19
TR	PDIL5-1	AT1G07960	COG0526	19
TR	PDIL5-2	AT1G35620	COG0526	19
TR	ATHM1	AT1G03680	COG0526	19
TR	ATHM2	AT4G03520	COG0526	19
TR	TRX-M4	AT3G15360	COG0526	19
TR	TRX1	AT3G51030	COG0526	19
TR	TRX2	AT5G39950	COG0526	19
TR	TRX3	AT5G42980	COG0526	19
TR	TRX4	AT1G19730	COG0526	19
TR	TRX5	AT1G45145	COG0526	19
TR	TRX7	AT1G59730	COG0526	19
TR	TRX8	AT1G69880	COG0526	19
TR	TRX9	AT3G08710	COG0526	19
TR	TRXF1	AT3G02730	COG0526	19
TR	TRXF2	AT5G16400	COG0526	19
TR	TXND9	AT2G18990	COG0526	19
TR	WCRKC1	AT5G06690	COG0526	19
TR	WCRKC2	AT5G04260	COG0526	19

Tabela 2: Descrição dos elementos químicos utilizados

Nome do Elemento	Fórmula	Kegg ID	PubChem ID
Peróxido de Hidrogênio	H ₂ O ₂	C00027	784
Peroxinitrito	NO ³⁻	C16845	104806
Superóxido	O ²⁻	C00704	5359597
Óxido nítrico	NO	C00533	145068
Hipoclorito de Sódio	NaClO	D01711	23665760
Radical Hidroxila	OH	C16844	961
Hidrido	H-	-	166653
Oxigênio	O ₂	C00007	977
Ozônio	O ₃	-	24823
Cobre	Cu	C00070	23978
Glutaciona reduzida	C ₁₀ H ₁₆ N ₃ O ₆ S ⁻	C00051	7048684
Glutaciona oxidada	C ₂₀ H ₃₂ N ₆ O ₁₂ S ₂	C00127	65359
Álcool coniferílico	C ₁₀ H ₁₂ O ₃	C00590	1549095
Álcool sinapílico	C ₁₁ H ₁₄ O ₄	C02325	5280507
Ascorbato	C ₆ H ₈ O ₆	C00072	54670067

Class III COGs	NOG33768	1	1	1	1	2	1	1	1	0	0	0	0	0	0	0	0	0	2	0	
	NOG130945	1	1	1	1	4	2	2	2	4	2	2	4	6	3	2	4	3	3	2	0
	NOG123538	2	4	2	2	4	1	1	2	0	1	1	3	2	3	5	1	1	1	2	0
	NOG131810	1	1	1	0	2	1	1	2	0	0	0	0	0	0	0	0	0	0	4	0
	NOG135096	1	2	1	1	2	2	2	1	1	1	1	1	1	0	1	0	1	1	2	2
	NOG24739	1	1	1	1	2	0	1	1	0	0	0	0	1	1	0	0	1	1	6	1
	NOG29576	6	6	5	2	6	3	3	2	3	2	5	6	4	5	1	4	6	4	4	3
	NOG130428	1	1	1	1	2	1	1	1	0	0	1	1	1	1	1	1	1	1	6	2
	NOG129341	2	1	2	3	2	1	1	2	1	2	2	2	2	2	2	1	2	2	5	6
	NOG26018	4	6	4	2	3	6	5	1	1	2	3	4	4	5	3	1	3	3	2	2
	NOG101992	1	3	1	2	3	1	1	1	2	2	2	2	2	2	3	4	3	2	0	2
	NOG125285	2	2	1	5	4	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1
	NOG128539	1	3	2	3	0	1	1	0	2	6	7	3	3	5	3	4	6	5	4	2
	NOG29116	1	1	1	0	0	2	5	0	0	0	1	2	2	1	0	1	1	1	0	4
	NOG124136	1	1	0	2	7	2	3	4	1	3	2	3	3	1	3	2	3	3	2	3
	NOG07211	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
Perox COGs	COG0376	7	9	9	9	11	9	8	6	7	8	9	8	8	13	10	8	12	9	12	6
	COG0753	3	8	2	4	4	6	3	3	3	3	3	3	3	3	5	3	7	3	10	7
	COG0386	8	11	9	8	15	7	4	5	3	6	6	6	7	7	8	5	8	5	8	3
	COG0678	6	6	6	4	8	3	4	8	4	3	4	4	4	4	4	4	6	5	8	7
	NOG00684	19	23	22	20	28	17	13	14	17	10	8	15	14	21	26	11	12	12	38	11
	COG0450	3	6	3	5	9	3	3	4	1	4	4	2	2	3	6	3	6	4	10	6
Thiol-Redox COGs	NOG45847	4	7	4	4	8	2	2	2	3	5	5	4	4	8	7	4	6	5	0	1
	COG4243	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
	COG5119	1	1	1	1	1	1	1	1	1	1	0	1	1	3	1	1	1	1	2	1
	NOG18420	2	2	2	14	5	3	3	8	2	4	5	4	4	7	5	3	4	5	5	0
	NOG10297	2	3	2	2	3	1	1	1	1	1	1	1	1	1	3	1	1	1	2	4
	COG1249	6	9	6	7	12	6	7	7	5	7	9	8	8	8	10	8	17	10	8	7
	COG0695	27	56	32	32	50	40	44	16	18	13	23	21	26	26	31	20	23	23	14	7
	NOG01122	7	14	10	9	10	6	4	6	4	5	5	6	5	11	10	5	5	5	6	6
	COG0526	54	76	59	56	109	54	48	45	33	45	43	42	42	86	59	38	52	44	64	51
	COG0175	4	8	4	4	5	5	5	5	2	5	3	2	2	6	3	3	4	2	6	2
	COG0278	4	7	6	5	10	4	3	3	4	5	4	5	7	22	6	5	5	5	9	8
	COG0369	9	17	9	9	14	6	6	9	6	7	8	9	9	24	12	8	8	8	16	11
	COG0446	6	6	6	3	9	4	3	6	3	6	4	8	10	7	5	7	9	6	6	5
	COG0225	5	13	5	6	7	5	5	5	2	3	4	4	4	2	5	3	5	4	12	5
	COG0492	3	10	3	3	5	4	5	3	3	3	4	3	5	4	4	3	4	4	4	4
SODs COGs	COG2608	50	75	55	79	118	45	45	45	36	39	39	42	45	50	79	42	48	43	24	29
	COG0234	5	12	6	8	14	5	5	6	3	6	6	6	10	12	9	5	6	7	14	10
	COG2032	3	5	4	7	6	3	3	6	3	4	4	4	4	5	6	3	5	4	7	4
	COG0605	6	8	5	5	7	5	4	4	3	3	3	3	3	13	7	3	4	3	5	4

Tabela 5: Distribuição dos 208 genes em 70 grupos de ortólogos.

COG	Gene Stable ID	COG	Gene Stable ID	COG	Gene Stable ID
COG0175	AT4G04610	COG0605	AT4G25100	NOG130232	AT2G43480
	AT4G21990		AT5G51100		AT5G24070
COG0225	AT5G07460	COG0678	AT5G23310	NOG130428	AT4G16270
	AT4G25130		AT3G10920	NOG130945	AT5G05340
	AT5G61640		AT3G56350	NOG131107	AT5G64120
COG0234	AT5G07470	COG0695	AT1G65980	NOG131444	AT2G22420
	AT5G20720		AT1G65970	NOG131641	AT2G41480
COG0278	AT3G54900	COG0753	AT5G63030	NOG131772	AT1G44970
	AT3G15660		AT5G11930	NOG131810	AT5G51890
	AT2G38270		AT3G62950	NOG133099	AT1G30870
COG0369	AT4G04950	COG1249	AT2G47870	NOG135096	AT2G37130
	AT5G66190		AT2G47880	NOG18420	AT1G60420
COG0376	AT1G20020	COG2032	AT3G62960	NOG185758	AT4G31240
	AT1G07890		AT5G40370	NOG19163	AT1G14540
	AT3G09640		AT1G77370	NOG232879	AT1G14550
	AT4G35000		AT5G20500	NOG232889	AT1G71695
	AT4G35970		AT4G28730	NOG232894	AT5G64100
COG0386	AT4G32320	COG2608	AT4G33040	NOG232894	AT5G64110
	AT4G08390		AT3G02000	NOG233164	AT5G19880
	AT1G77490		AT5G14070	NOG235401	AT1G77100
	AT2G25080		AT1G28480	NOG236115	AT3G03670
	AT2G31570		AT1G03020	NOG236115	AT3G50990
COG0446	AT2G43350	COG29576	AT3G21460	NOG237024	AT2G38380
	AT2G48150		AT1G06830	NOG24739	AT2G38390
	AT3G63080		AT2G20270	NOG26018	AT3G17070
	AT4G11600		AT1G03850	NOG275802	AT4G26010
	AT4G31870		AT5G18600	NOG275802	AT5G17820
COG0450	AT1G63460	COG4243	AT4G15700	NOG275802	AT5G22410
	AT3G52880		AT4G15680	NOG275802	AT1G34510
	AT5G03630		AT4G15690	NOG275802	AT1G49570
	AT3G09940		AT3G62930	NOG275802	AT3G32980
	AT3G27820		AT4G15670	NOG275802	AT3G49110
COG0492	AT1G63940	COG5119	AT4G15660	NOG276879	AT3G49120
	AT5G06290		NOG00684	AT4G08770	AT4G08770
	AT3G11630		NOG01122	AT4G08780	AT4G08780
	AT4G35460		NOG07211	NOG29116	AT5G39580
	AT2G17420		NOG101992	NOG29116	AT3G28200
COG0526	AT2G41680	NOG10297	AT3G28200	NOG29116	AT3G49960
	AT4G26160		NOG121808	NOG29116	AT4G37520
	AT4G29670		NOG122617	NOG29116	AT4G37530
	AT2G33270		NOG123237	NOG29116	AT5G40150
	AT1G08570		NOG123538	NOG29116	AT5G67400
AT5G61440	NOG123647	NOG33768	AT4G21960		
AT1G50320	NOG124136	NOG45847	AT1G34780		
AT2G35010	NOG124919	NOG45847	AT3G03860		
AT1G31020	NOG125285	NOG45847	AT4G08930		
AT1G76760	NOG126611	NOG45847	AT5G18120		
AT1G43560	NOG126859	NOG58286	AT2G18980		
AT1G11530	NOG126963	NOG58286	AT4G30170		
AT2G40790	NOG128070				
AT2G15570	NOG128539				
AT4G04950	NOG129341				
AT2G41680	NOG129430				
AT1G21750	NOG129978				
AT1G77510					
AT3G54960					
AT5G60640					
AT1G52260					
AT3G16110					
AT2G47470					
AT1G04980					
AT2G32920					
AT1G07960					
AT1G35620					
AT1G03680					
AT4G03520					
AT3G15360					
AT3G51030					
AT5G39950					
AT5G42980					
AT1G19730					
AT1G45145					
AT1G59730					
AT1G69880					
AT3G08710					
AT3G02730					
AT5G16400					
AT2G18990					
AT5G06690					
AT5G04260					