



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

KELLEN RAYANNE MATOS DE SOUSA

**BUSCA E CLASSIFICAÇÃO SISTEMÁTICA DAS PROTEÍNAS
OXIGENASES COM FERRO NÃO HÊMICO EM PLANTAS**

Belém - PA

2015

KELLEN RAYANNE MATOS DE SOUSA

**BUSCA E CLASSIFICAÇÃO SISTEMÁTICA DAS PROTEÍNAS
OXIGENASES COM FERRO NÃO HÊMICO EM PLANTAS**

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal do Pará, como requisito final para obtenção do grau de Mestre em Biotecnologia, sob a orientação do Prof. Dr. Sylvain Henri Darnet.

Belém - PA

2015

Dados Internacionais de Catalogação-na-Publicação
(CIP) Sistema de Bibliotecas da UFPA

Sousa, Kellen Rayanne Matos de, 1990-

Busca e classificação sistemática das proteínas oxigenases com ferro não hêmico em plantas / Kellen Rayanne Matos de Sousa. - 2015.

Orientador: Sylvain Henri Darnet.

Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Ciências Biológicas, Programa de Pós-Graduação em Biotecnologia, Belém, 2015.

1. Proteínas. 2. Oxigenases. 3. Ácidos graxos. I. Título.

CDD 22. ed. 572.6

KELLEN RAYANNE MATOS DE SOUSA

BUSCA E CLASSIFICAÇÃO SISTEMÁTICA DAS PROTEÍNAS OXIGENASES COM FERRO NÃO HÊMICO EM PLANTAS

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal do Pará, como requisito final para obtenção do grau de Mestre em Biotecnologia, sob a orientação do Prof. Dr. Sylvain Henri Darnet.

Data de aprovação __/__/____

Conceito: _____

Banca Examinadora:

Prof. Dr. Sylvain Henri Darnet - Orientador
Universidade Federal do Pará

Prof. Dr. Fábio Alberto de Molfetta
Universidade Federal do Pará

Prof. Dr. Agenor Valadares Santos
Universidade Federal do Pará

Prof^a. Dra. Joyce Kelly do Rosário da Silva
Universidade Federal do Pará

AGRADECIMENTOS

Primeiramente a Deus, pela graça e sabedoria que me dá todos os dias.

Aos meus pais Lêda e João, meu irmão Maxwell, e meu noivo Marcos Veríssimo por todo o apoio, incentivo e confiança. Obrigada por compreenderem minha ausência durante boa parte dos dias durante os dois anos de mestrado, e por sempre me receberem com carinho após um longo e cansativo dia.

Ao Prof. Dr. Sylvain Darnet, por ter aceitado me orientar e por me ajudar a dar os primeiros passos na bioinformática. Admiro sua competência, paciência e dedicação em ensinar.

À equipe do Laboratório de Biologia Computacional, pelo compartilhamento de experiência e conhecimento, contribuições, e amizade de cada um.

À Universidade Federal do Pará (UFPA) pela infraestrutura fornecida e pelo financiamento do projeto intitulado “Estudo Filogenômico das enzimas de C4 descarboxilases nas plantas”, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo.

O Senhor é o único Deus; somente Deus é a nossa rocha. Ele é o Deus que me dá forças e me protege aonde quer que eu vá. Ele não me deixa tropeçar e me põe a salvo nas montanhas. Ele me treina para a batalha para que eu possa usar os arcos mais fortes. Tu, ó Senhor Deus, me deste o escudo que salva a minha vida. O teu cuidado me tem feito prosperar, e o teu poder me tem sustentado.

(Salmos 18.31-35)

RESUMO

As proteínas oxigenases com ferro não hêmico compartilham um domínio conservado composto por oito histidinas, podem ser encontradas em organismos eucariotos e procariotos, e participam de importantes vias de biossíntese lipídica. Para compreender a relação evolutiva existente entre essas proteínas, foram realizadas análises comparativa e filogenética em procariotos e eucariotos que permitiram uma classificação dessa família, até então inexistente. A busca de sequências resultou, após a curadoria, em uma coleção de 448 proteínas, pertencentes a 58 organismos previamente selecionados dentro dos principais taxa. O alinhamento múltiplo de sequências gerado com a ferramenta MAFFT (BLOSUM 62; L-INS-i) mostrou a presença do domínio de histidinas com espaçamento conservado entre os motivos. A classificação das proteínas feita com o *software* CLANS gerou 28 grupos a partir da similaridade entre pares de sequências. Dentre esses, 2 contêm sequências que não tiveram similaridade com proteínas já caracterizadas e 48 sequências não foram atribuídas a quaisquer dos grupos formados. As sequências de plantas, representadas por 119 sequências da coleção, foram distribuídas em 7 grupos correspondentes às funções C4 metilesterol monooxigenase, C5 esterol desaturase, ácido graxo hidroxilase, esfingolípido C4 monooxigenase, aldeído decarbonilase, β -caroteno hidroxilase e Acil-ACP desaturase. A análise filogenética, utilizando o método de máxima verossimilhança com a ferramenta PhyML, mostrou a formação de grupos bem definidos e que foram similares aos gerados por CLANS. Esses resultados começam a preencher a lacuna existente até o momento acerca da relação evolutiva e da classificação das oxigenases com ferro não hêmico. Além disso, sugerem que dentro dessa família ainda há proteínas com funções desconhecidas, reforçando a necessidade de realizar mais estudos de caracterização funcional das mesmas.

Palavras-chave: hidroxilase de ácidos graxos, desaturase de ácidos graxos, histidina, comparação genômica, metaloproteína.

ABSTRACT

Nonheme iron oxygenase proteins shares a conserved domain consisting of eight histidines, and they can be found in eukaryotes and prokaryotes organisms and participate in important pathways of lipid biosynthesis. To understand the evolutionary relationship among these proteins, we performed comparative and phylogenetic analyzes in prokaryotes and eukaryotes that allowed a classification of this family, nonexistent until now. The search of sequences resulted in a collection of 448 proteins, belonging to 58 organisms previously selected. The multiple alignment made with MAFFT (BLOSUM 62; L-INS-i) showed the presence of three histidine-rich motifs, with conserved spacing among them. The classification made with CLANS *software* has generated 28 clusters through of similarity among sequences. Two clusters contain sequences that had no similarity to proteins already characterized, and 48 sequences were not assigned to any of the 28 clusters. In the collection, 119 sequences are derived from plants, distributed in 7 clusters corresponding to C4 methysterol monooxygenase, C5 sterol desaturase, fatty acid hydroxylase, sphingolipid C4 monooxygenase, aldehyde decarbonylase, β -carotene hydroxylase and Acyl-ACP desaturase functions. Phylogenetic analysis using maximum likelihood method with PhyML tool showed the formation of well-defined groups that were similar to generated by CLANS. These results start to fill a gap existing so far about the evolutionary relationship and classification of nonheme iron oxygenases. Also, suggest that within this family there are still proteins with unknown functions, reinforcing the need for more studies of functional characterization.

Keywords: fatty acid hydroxylase, fatty acid desaturase, histidine, comparative genomic, metalloprotein.

LISTA DE FIGURAS

Figura 1 -	Motivo de ligação consenso das desaturases solúveis	7
Figura 2 -	Mecanismo de ação das dioxigenases que clivam anéis aromáticos	9
Figura 3 -	Resumo da via de biossíntese de esteróis comparada entre plantas, fungos e vertebrados	11
Figura 4 -	HMM logo das oxigenases com ferro não hêmico	15
Figura 5 -	Nomenclatura de uma proteína P450	20
Figura 6 -	Exemplo de visualização 2D de um gráfico gerado pelo <i>software</i> CLANS	21
Figura 7 -	Resumo da metodologia utilizada para a classificação das proteínas oxigenases com ferro não hêmico	25
Figura 8 -	Figura 8. Esquema dos principais taxa e suas respectivas quantidades de espécies e sequências de oxigenases com ferro não hêmico contidas na coleção	26
Figura 9 -	Alinhamentos múltiplos das sequências de proteínas oxigenases com ferro não hêmico	28
Figura 10 -	Grupos gerados por CLANS com base nas similaridades entre as sequências de proteínas oxigenases com ferro não hêmico	30
Figura 11 -	Árvore filogenética das sequências de oxigenases com ferro não hêmico gerada por análise de máxima verossimilhança com a ferramenta PhyML	32
Figura 12 -	Cladograma não enraizado mostrando os grupos mais representativos das oxigenases com ferro não hêmico	33

LISTA DE TABELAS

Tabela 1 -	Descrição geral das monooxigenases dependentes de flavina	4
Tabela 2 -	Papel funcional das monooxigenases dependentes de pterina em eucariotos	6
Tabela 3 -	Genes que codificam oxigenases com ferro não hêmico caracterizados em <i>Arabidopsis thaliana</i>	12
Tabela 4 -	Total de sequências P450 depositadas no banco de dados Nelson's P450 Database	15
Tabela 5 -	Bancos de dados mais utilizados em bioinformática	17
Tabela 6 -	Ferramentas de alinhamento múltiplo de sequências consideradas como melhores atualmente disponíveis	18
Tabela 7 -	Percentual de identidade média entre os pares de sequências apresentado por cada histidina nos alinhamentos múltiplos de sequências	27

LISTA DE ABREVIATURAS E SIGLAS

AACPD	Acil-ACP desaturase
AADC	Descarboxilase de L-aminoácido aromático (<i>Aromatic Amino Acid Decarboxylase</i>)
AdoMet	S-adenosilmetionina
ALKMO	Alquilglicerol monooxigenase
BCH	β -caroteno hidroxilase
BLAST	<i>Basic Local Alignment Search Tool</i>
BLOSUM	Matriz de substituição por blocos (<i>Blocks Substitution Matrix</i>)
C4SMO	C4 esteroil metiloxidase
C5SD	C5 esteroil desaturase
CER	Aldeído decarbonilase (<i>Eceriferum</i>)
CO	Monóxido de carbono
D12FAD	Δ^{12} ácido graxo desaturase
D4FAD	Δ^4 ácido graxo desaturase
D6FAD	Δ^6 ácido graxo desaturase
D9FAD	Δ^9 ácido graxo desaturase
DDBJ	Banco de Dados de DNA do Japão (<i>DNA Data Bank of Japan</i>)
EBI	Instituto Europeu de Bioinformática (<i>European Bioinformatics Institute</i>)
EMBL	Laboratório Europeu de Biologia Molecular (<i>European Molecular Biology Laboratory</i>)
FAD/Cyt-b5	Ácido graxo desaturase / Citocromo-b5
FAH	Ácido graxo hidroxilase
FAH/SDR	Ácido graxo hidroxilase / Dehidrogenase-redutase de cadeia curta
HMM	Modelos ocultos de Markov (<i>Hidden Markov Model</i>)
HSP	Pares de segmentos de alta pontuação (<i>High Scoring Segment Pairs</i>)
INSDC	Colaboração Internacional de Bancos de Dados de Nucleotídeos (<i>International Nucleotide Sequence Database Collaboration</i>)
KEGG	Kyoto Enciclopédia de Genes e Genomas (<i>Kyoto Encyclopedia of Genes and Genomes</i>)

LCOAD	Linoleoil-CoA desaturase
NCBI	Centro Nacional para Biotecnologia da Informação (<i>National Center for Biotechnology Information</i>)
NGS	Sequenciamento de nova geração (<i>Next-Generation Sequencing</i>)
OFNH	Oxigenases com ferro não hêmico
PAM	Ponto de Mutação Aceita (<i>Point Accepted Mutation</i>)
PDB	Banco de Dados de Proteína (<i>Protein Data Bank</i>)
PheH	Fenilalanina hidroxilase (<i>Phenilalanine Hydroxylase</i>)
PIR	Recurso de Informação de Proteínas (<i>Protein Information Resource</i>)
RdmB	Aclacinomicina 10-hidroxilase
SDES	Esterol desaturase
SIB	Instituto Suíço de Bioinformática (<i>Swiss Institute of Bioinformatics</i>)
SUR	Esfingolípídeo hidroxilase
TrpH	Triptofano hidroxilase (<i>Tryptophan Hydroxylase</i>)
TyrH	Tirosina hidroxilase (<i>Tyrosine Hydroxylase</i>)
UniProt	Recurso Universal de Proteínas (<i>Universal Protein Resource</i>)
UPGMA	Agrupamento pareado não ponderado baseado na média aritmética (<i>Unweighted Pair Group Method with Arithmetic Mean</i>)
wwPDB	Banco de Dados Mundial de Proteína (<i>Worldwide Protein Data Bank</i>)

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	2
2.1	Objetivo geral	2
2.2	Objetivos específicos	2
3	REVISÃO DA LITERATURA	3
3.1	Oxigenases	3
3.1.1	Monooxigenases	3
3.1.2	Dioxygenases	9
3.2	Oxigenases com ferro não hêmico	9
3.3	Importância do ferro no centro catalítico da reação	12
3.4	Assinatura das oxigenases com ferro não hêmico	13
3.5	Classificação de proteínas	15
3.5.1	Busca de sequências	17
3.5.2	Alinhamento de sequências	18
3.5.3	Classificação sistemática	19
4	MATERIAL E MÉTODOS	22
4.1	Mineração de dados	22
4.2	Curadoria das sequências	22
4.3	Alinhamento múltiplo das sequências	23
4.4	Reconstrução filogenética	24
4.5	Classificação sistemática	24
5	RESULTADOS	25
6	DISCUSSÃO	34
7	CONCLUSÃO	39
	REFERÊNCIAS	40
	ANEXO A - Organismos selecionados para busca de sequências de proteínas oxigenases com ferro não hêmico em bancos de dados	51
	ANEXO B - Proteínas caracterizadas na literatura contidas na coleção de sequências	53
	ANEXO C - Classificação das oxigenases com ferro não hêmico ...	54

**ANEXO D - Distribuição de sequências das proteínas oxigenases
com ferro não hêmico nos grupos formados pela ferramenta
CLANS 56**

1 INTRODUÇÃO

Oxigenases são proteínas que catalisam a incorporação direta de oxigênio molecular em seu substrato (FETZNER, 2002). Elas estão presentes em procariotos e eucariotos e possuem um importante papel em diversas vias metabólicas, sendo fundamentais, por exemplo, na aplicação de enzimas microbianas em biotecnologia ambiental e industrial (HARAYAMA; KOK; NEIDLE, 1992), na mineralização de compostos xenobióticos (ARORA *et al.*, 2009), na hidroxilação de esteróides em vertebrados (HARAYAMA; KOK; NEIDLE, 1992) e na síntese de fitoesteróis de plantas (RAHIER, 2011).

As oxigenases com ferro não hêmico podem atuar como desaturases, epoxidases, hidroxilases, acetilanases, metil-oxidases, etc., utilizando uma grande variedade de substratos (SHANKLIN; CAHOON, 1998). Elas são proteínas ligadas à membrana celular e possuem um domínio conservado composto por oito histidinas, indicando que são evolutivamente relacionadas (SHANKLIN; CAHOON, 1998). Nas plantas, essas proteínas estão envolvidas nas vias de biossíntese de ácidos graxos, esteróis, esfingolipídios, cera epicuticular, entre outras (ARTHINGTON *et al.*, 1991; AARTS *et al.*, 1995), e devido esses organismos possuírem uma grande quantidade de metabólitos secundários (BOURGAUD *et al.*, 2001; VERPOORTE; MEMELINK, 2002) o estudo dessas oxigenases é ainda mais importante.

A quantidade de dados de sequência disponíveis têm se tornado cada vez maior devido às melhorias no sequenciamento genômico, e os intensos estudos bioquímicos têm expandido o número de oxigenases com ferro não hêmico conhecidas e suspeitas (RYLE; HAUSINGER, 2002). Apesar do papel crucial que desempenham, nenhum dos bancos de dados existentes fornece informações completas e de caráter comparativo sobre essa família de proteínas (ARORA *et al.*, 2009). Dessa maneira, torna-se relevante a comparação genômica e a classificação das oxigenases com ferro não hêmico.

2 OBJETIVOS

2.1 Objetivo geral

Analisar as funções das oxigenases com ferro não hêmico através da comparação entre sequências proteicas de plantas e outros organismos eucariotos e procariotos, e realizar a classificação sistemática dessas proteínas.

2.2 Objetivos específicos

Estabelecer uma coleção de sequências das oxigenases com ferro não hêmico em plantas através de busca exaustiva em bancos de dados;

Realizar análises comparativas das oxigenases com ferro não hêmico entre plantas e outros organismos eucariotos e procariotos;

Aperfeiçoar os métodos de alinhamento múltiplo e classificação sistemática para esse grupo de proteínas;

Identificar grupos de sequências a partir dos alinhamentos múltiplos e da reconstrução filogenética das oxigenases com ferro não hêmico;

Classificar os grupos em subfamílias e, ao final, produzir um banco de dados com todas as oxigenases com ferro não hêmico utilizadas nesse estudo;

Identificar possíveis novas famílias, subfamílias ou novas funções de interesse biotecnológico nas oxigenases com ferro não hêmico.

3 REVISÃO DA LITERATURA

3.1 Oxigenases

As oxigenases podem introduzir um ou dois átomos de oxigênio molecular em seus substratos, sendo classificadas como monooxigenases e dioxigenases, respectivamente (BURTON, 2003; METZLER, 2003; JAOUEN, 2006; YAMAMOTO, 2006). Essas proteínas foram inicialmente descobertas em fungos e bactérias, e posteriormente, estudos comprovaram que as oxigenases estavam presentes em tecidos de mamíferos e que possuíam funções fisiológicas no metabolismo de diversos compostos (YAMAMOTO, 2006). Atualmente sabe-se que elas são ubíquas (ULLRICH; HOFRICHTER, 2007), que em sistemas vivos podem ter papéis no metabolismo primário ou secundário (BURTON, 2003), e que existe uma enorme variedade de oxigenases envolvidas na formação de metabólitos essenciais (como esteróis, prostaglandinas, derivados de vitamina D, etc.) (METZLER, 2003).

3.1.1 Monooxigenases

Monooxigenases incorporam somente um átomo de oxigênio molecular ao substrato e reduzem o outro átomo à água (BURTON, 2003; NOLAN; O'CONNOR, 2008). Monooxigenases externas cofatores para fornecer a redução potencial para o fornecimento de elétrons ao substrato, como metais, o grupo heme, flavina, pterina, NADH ou NADPH (BURTON, 2003; van BERKEL; KAMERBEEK; FRAAIJE, 2006; ULLRICH; HOFRICHTER, 2007). Diferentemente, monooxigenases internas obtêm estes elétrons a partir do próprio substrato (van BERKEL; KAMERBEEK; FRAAIJE, 2006). Monooxigenases podem utilizar, também, outros cofatores ou pode ser independente de cofatores (BURTON, 2003; ULLRICH; HOFRICHTER, 2007; TORRES PAZMIÑO *et al.*, 2010).

a) Monooxigenases dependentes do grupo heme

As monooxigenases dependentes do grupo heme são conhecidas como citocromo P450 ou CYPs (NELSON *et al.*, 1996; BURTON, 2003; MONTELLANO, 2005; FISCHER *et al.*, 2007). Elas provavelmente são as oxigenases mais

estudadas e melhor caracterizadas, compreendendo uma das maiores e mais antigas superfamílias de enzimas (ULLRICH; HOFRICHTER, 2007).

Essas proteínas foram assim nomeadas devido à forte absorvância a 450 nm sobre a ligação do CO ao ferro (Fe^{2+}) hêmico (CHAPPLE, 1998; MONTELLANO, 2005; SIGEL; SIGEL; SIGEL, 2007). Essa característica pode ser utilizada para identificar a presença do P450 e estimar sua concentração, ainda que as sensibilidades de diferentes isoformas de P450 ao CO sejam distintas (MONTELLANO, 2005).

Os P450 possuem grande importância na oxidação dentro de vias biossintéticas, catalisando a oxidação de diversos compostos endobióticos (como esteróis, ácidos graxos, ácidos biliares, leucotrienos, prostaglandinas, etc.) e xenobióticos (como fármacos, anestésicos, pesticidas, fungicidas, etc.) (MONTELLANO, 2005; FISCHER *et al.*, 2007; NAKAJIMA, 2014)

b) Monooxigenases dependentes de flavina

As monooxigenases dependentes de flavina podem atuar na biossíntese de vitaminas, antibióticos e hormônios, em estratégias de defesa (em plantas e fungos) contra herbívoros, e no catabolismo de compostos naturais e antropogênicos, dentre outros processos biológicos (HUIJBERS *et al.*, 2014). Essas proteínas podem ser classificadas em 8 grupos (A a H, Tabela 1) com base na forma como se dobram e na função que exercem (van BERKEL; KAMERBEEK; FRAAIJE, 2006; TORRES PAZMIÑO *et al.*, 2010; HUIJBERS *et al.*, 2014).

As principais diferenças entre os grupos são: as proteínas dos grupos A e B são dependentes de NAD(P)H como doadores externos de elétrons; nos grupos C, D, E e F, as monooxigenases necessitam de uma proteína redutase como parceira para a entrega de flavina reduzida; os grupos G e H incluem monooxigenases internas que reduzem o cofator flavina através de oxidação de substrato (HUIJBERS *et al.*, 2014).

Tabela 1 – Descrição geral das monooxigenases dependentes de flavina

Grupo	Cofator	Doador de elétron	Dobra da proteína	Reações
A	FAD	NAD(P)H	Rossmann (GR-2)	Hidroxilação Sulfoxidação
B	FAD	NAD(P)H	Rossmann (FMO)	Oxidação Baeyer-Villiger

				Heteroátomo oxigenação N-hidroxilação Descarboxilação oxidativa
C	FMN	FMNH ₂	Tim-barred (luciferase)	Emissão de luz Oxidação Baeyer-Villiger Epoxidação Desulfurização Sulfoxidação Hidroxilação
D	FAD/FMN	FADH ₂ /FMNH ₂	Acil-CoA dehidrogenase	Hidroxilação N-hidroxilação
E	FAD	FADH ₂	Rossmann (GR-2)	Epoxidação
F	FAD	FADH ₂	Rossmann (GR-2)	Halogenação
G	FAD	Substrato	Rossmann (GR-2)	Descarboxilação oxidativa
H	FMN	Substrato	Tim-barred (glicoato oxidase)	Descarboxilação oxidativa Denitração oxidativa

GR, Glutamina redutase; FMO, monooxigenase contendo flavina; MAO, monoamina oxidase.
Tabela adaptada de HUIJBERS e colaboradores (2014).

c) Monooxigenases dependentes de cobre

As monooxigenases dependentes de cobre, como o nome sugere, necessitam de íons de cobre para realizar a hidroxilação dos seus substratos, e constituem uma família relativamente pequena das monooxigenases (TORRES PAZMIÑO *et al.*, 2010). Alguns membros desse grupo são: 1) dopamina β -monooxigenase, que participa da síntese de catecolamina, convertendo a dopamina em noradrenalina; 2) peptidilglicina α -amidante monooxigenase, que participa da maturação hormonal peptídica, fazendo a amidação do grupo ácido carboxílico α -terminal da glicina; e 3) monofenol monooxigenase, que participa da síntese de melanina, convertendo tirosina em DOPA (XIN; MAINS; EIPPER, 2004; KASTIN, 2006; TORRES PAZMIÑO *et al.*, 2010; JAISER; WINSTON, 2010).

d) Monooxigenases dependentes de pterina

Esse grupo de monooxigenases utiliza a pterina (tetrahidrobiopterina ou BH₄) como coenzima para catalisar a inserção do átomo de oxigênio, exigindo um átomo de ferro por subunidade para a atividade (FITZPATRICK, 2003; VOLNER; ZOIDAKIS; ABU-OMAR, 2003; TORRES PAZMIÑO *et al.*, 2010). Fenilalanina

hidroxilase (PheH), tirosina hidroxilase (TyrH) e triptofano hidroxilase (TrpH) são as proteínas que constituem essa família (KAGEDAL *et al.*, 2004).

A PheH catalisa a conversão de L-fenilalanina (L-Phe) a L-tirosina (L-Tyr) (FLYDAL; MARTINEZ, 2013), a TyrH catalisa a hidroxilação de L-tirosina para L-DOPA (L-3,4-dihidroxifenilalanina) (KAGEDAL *et al.*, 2004), e a TrpH hidroxila o triptofano para geral 5-hidroxitriptofano, que é descarboxilado pela AADC formando a serotonina (5-hidroxitriptamina) (WALTHER; BADER, 2003). As monooxigenases dependentes de pterina têm sido muito estudadas em eucariotos devido a sua importância fisiológica (FITZPATRICK, 2003; DAUBNER; LE; WANG, 2011). A Tabela 2 mostra as principais funções dessas proteínas em eucariotos.

Tabela 2 – Papel funcional das monooxigenases dependentes de pterina em eucariotos

Proteína	Local	Função
PheH	Fígado	Catalisa o catabolismo do excesso de fenilalanina na dieta para tirosina.
TyrH	SNC Glândula adrenal	Catalisa o primeiro passo na biossíntese das catecolaminas, a formação de dihidroxifenilalanina (DOPA) para tirosina.
TrpH	Cérebro	Catalisa o primeiro passo na biossíntese da serotonina, a hidroxilação de triptofano para 5-hidroxitriptofano.

SNC, Sistema Nervoso Central.

e) Monooxigenases dependentes de ferro não hêmico

As monooxigenases dependentes de ferro não hêmico utilizam dois átomos de ferro como cofator para a sua atividade oxidativa, e podem ser chamadas de proteínas dependentes de diferro (TORRES PAZMIÑO *et al.*, 2010). A ausência do grupo heme é confirmada devido essas proteínas serem inibidas por cianeto, mas não serem afetadas por monóxido de carbono (SHANKLIN; CAHOON, 1998). Alguns membros dessa família são: monooxigenases solúveis (MURRELL; GILBERT; McDONALD, 2000), alceno monooxigenases (GALLAGHER; CAMMARK; DALTON, 1997), monooxigenases integrais à membrana (van BEILEN; FUNHOFF, 2007) e quatro componentes alceno/aromático monooxigenase, do qual faz parte um grupo separado do tipo-Rieske, de bactérias (KWEON *et al.*, 2008).

Proteínas monooxigenases solúveis são os membros melhor caracterizados (SHANKLIN; CAHOON, 1998; MURRELL; GILBERT; McDONALD, 2000), pois a cristalização da proteína acil-ACP desaturase (Δ^9 -18:0-ACP desaturase) permite

f) Monooxigenases dependentes de outros cofatores

A reação de hidroxilação catalisada pela proteína aclacinomicina 10-hidroxilase (RdmB), que participa da biossíntese de rodomicina em *Streptomyces purpurascens*, não requer cofatores, como flavina ou íons metálicos (JANSSON *et al.*, 2005). Em vez disso, a enzima utiliza AdoMet (S-adenosilmetionina) como cofator para a ativação do oxigênio (JANSSON *et al.*, 2005).

Até o momento, essa é a única monooxigenase que mostrou não pertencer a uma das famílias monooxigenase anteriormente mencionados, e o encontro dessa exceção sugere que provavelmente existam monooxigenases ainda não identificadas que exijam cofatores distintos dos presentes nas famílias de monooxigenases (TORRES PAZMIÑO *et al.*, 2010).

g) Monooxigenases independentes de cofatores

Algumas monooxigenases não necessitam de quaisquer cofatores para catalisar uma reação, sendo conhecidas como monooxigenases independentes de cofatores (TORRES PAZMIÑO *et al.*, 2010). Na reação catalizada por essas proteínas o oxigênio molecular é ativado pelo substrato depois que a reação de oxigenação ocorre (TORRES PAZMIÑO *et al.*, 2010). Como exemplo, pode-se citar as proteínas ActVA-Orf6 (monooxigenase formadora de quinona de *Streptomyces coelicolor*), TcmH (tetracenomicina F1 monooxigenase de *Streptomyces glaucescens*), ElmH (tetracenomicina F1 monooxigenase de *Streptomyces olivaceus*) e Aknx (antrona monooxigenase de *Streptomyces galilaeus*) (FETZNER; STEINER, 2010; TORRES PAZMIÑO *et al.*, 2010).

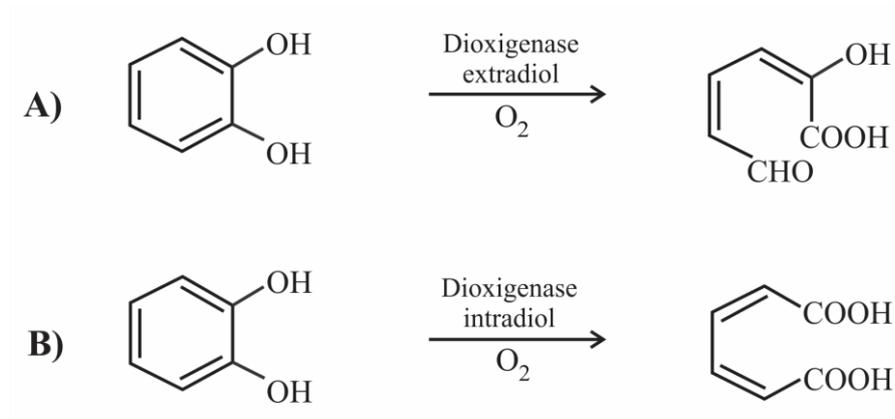
Devido não exigir custos com cofatores e coenzimas, essas monooxigenases deveriam ser interessantes para aplicações biotecnológicas. Todavia, devido o seu mecanismo catalítico, elas são restritas quanto ao substrato – fato que reduz a aplicabilidade dessas proteínas em biocatálise (TORRES PAZMIÑO *et al.*, 2010).

3.1.2 Dioxigenases

As dioxigenases inserem os dois átomos de oxigênio molecular no substrato e suas reações incluem cis-hidroilação, clivagem de anéis aromáticos e hidroperoxidação (DEWICK, 2002; BURTON, 2003). Elas dividem-se em duas classes principais: 1) dioxigenases ferro-enxofre dependente de heme, e 2) dioxigenases Rieske ferro-enxofre não hêmico (WACKETT, 2002; BURTON, 2003; NOLAN; O'CONNOR, 2008). A última classe possui maior potencial redox e a maioria é dependente de NADH (BURTON, 2003; NOLAN; O'CONNOR, 2008).

Outra classificação é baseada no modo de ação das dioxigenases: 1) dioxigenases que clivam anéis aromáticos, 2) dioxigenase hidroxilante de anel aromático (ARORA *et al.*, 2009). O primeiro grupo é subdividido em duas classes, dependendo da posição onde ocorre a clivagem do anel clivagem: a) dioxigenases de clivagem intradiol (Figura 2A) e b) dioxigenases de clivagem extradiol (Figura 2B) (ARORA *et al.*, 2009).

Figura 2 - Mecanismo de ação das dioxigenases que clivam anéis aromáticos



A) Dioxigenases de clivagem extradiol; B) Dioxigenases de clivagem intradiol.

Fonte: adaptada de ARORA e colaboradores (2009).

3.2 Oxigenases com ferro não hêmico

O foco do presente estudo são as monooxigenases integrais à membrana, e serão aqui chamadas de oxigenases com ferro não hêmico (OFNH). Elas desempenham papéis fundamentais em diversas vias de biossíntese, como a de

esteróis, ácidos graxos, esfingolipídios, cera epicuticular, fertilidade do pólen, entre outras (ARTHINGTON *et al.*, 1991; AARTS *et al.*, 1995). Além disso, podem realizar diversas atividades como desaturase, epoxidase, hidroxilase, acetilase, metiloxidase, cetolase e atividades de decarbonilase (SHANKLIN; CAHOON, 1998). Desse modo, veremos alguns exemplos de proteínas OFNH presentes em diversas vias metabólicas:

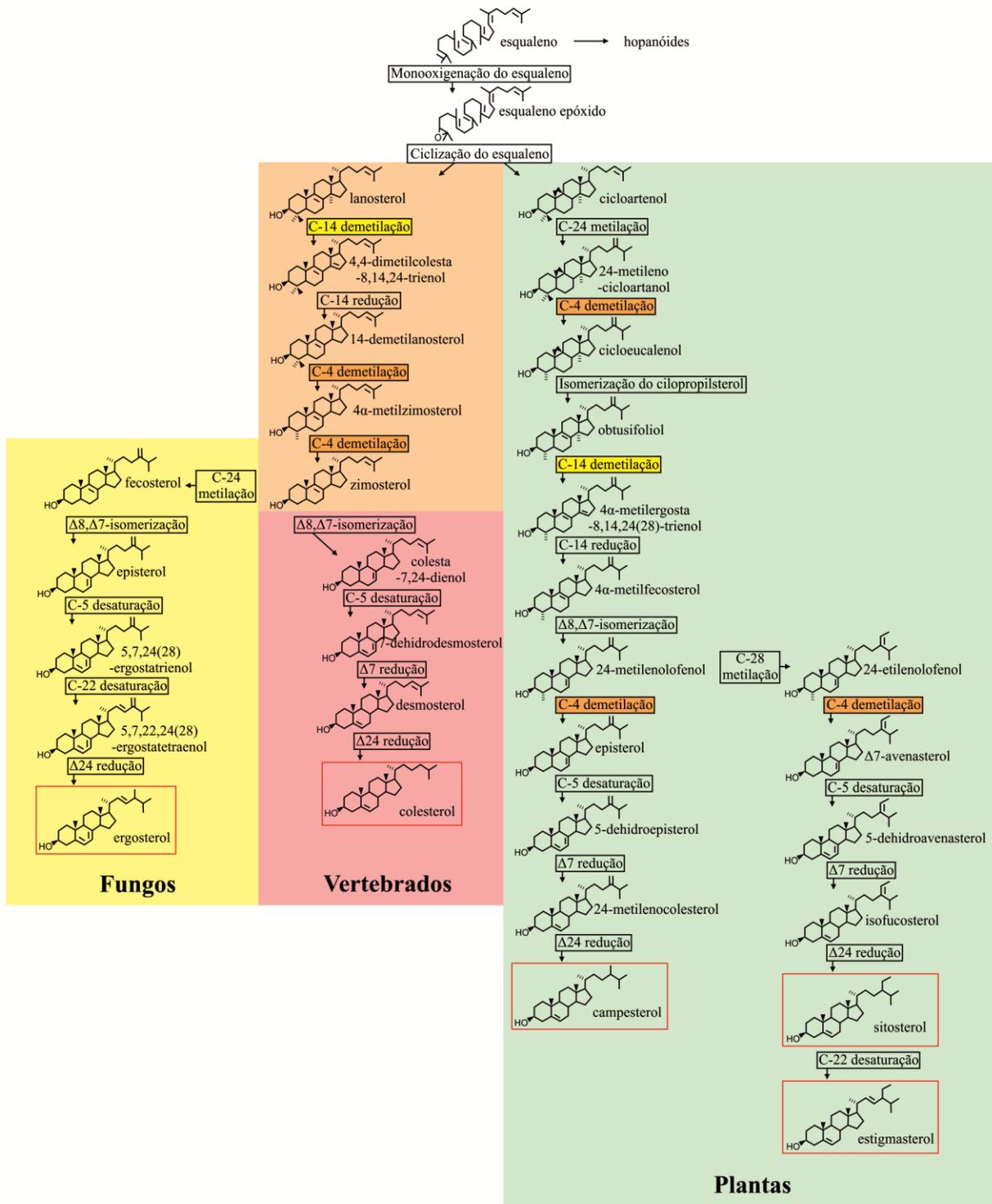
a) Esterol 4 α -metil oxidase (SMO): Essa proteína faz parte de um complexo enzimático das plantas que é responsável por remover dois grupos metila em posição C-4 (Figura 3) – processo necessário para a conversão de cicloartenol em fiotesteróis funcionais (DARNET; BARD; RAHIER, 2001; DARNET; RAHIER, 2003, 2004; BENVENISTE, 2004). Em *Saccharomyces cerevisiae* a ruptura do gene ScSMO (ERG25) foi letal, indicando o papel fundamental desta enzima em células eucarióticas (KENNEDY *et al.*, 2000; DARNET; RAHIER, 2003);

b) Aldeído decarbonilase (CER): O gene *CER1* de *Arabidopsis thaliana*, que codifica a proteína aldeído decarbonilase, foi isolado e caracterizado. Essa proteína mostrou estar envolvida na conversão de aldeídos para alcanos e presente na via de biossíntese de cera epicuticular e fertilidade do pólen (AARTS *et al.*, 1995). Nas plantas, as ceras auxiliam na redução da perda de água pela epiderme, contribuindo para a resistência à seca, e interação com insetos herbívoros e fungos patogênicos. A fertilidade do pólen, por sua vez, é fundamental para a propagação das espécies vegetais (AARTS *et al.*, 1995);

c) Alcano ω -hidroxilase (AlkB): Proteína de *Pseudomonas oleovorans* que é responsável pela oxidação inicial dos alcanos inativados (SHANKLIN; WHITLLE, 2003). O significado biológico das histidinas presentes no motivo compartilhado pelas OFNH foi testado nessas proteínas substituindo esse aminoácido por alanina, resultando na inativação da enzima (SHANKLIN; WHITLLE, 2003);

d) Dealquilação de esteróis (Des24p): Dealquilações em C-24 na via dos esteróis já haviam sido relatadas em diversos organismos, como no ciliado *Tetrahymena*, em algas verdes, artrópodes, cnidários, moluscos, etc (TOMAZIC *et al.*, 2011). Entretanto, nenhuma enzima nem gene correspondente havia sido identificado até Tomazic e colaboradores (2011) identificarem e caracterizarem pela primeira vez o gene *DES24* envolvido na dealquilação de esteróis;

Figura 3 - Resumo da via de biossíntese de esteróis comparada entre plantas, fungos e vertebrados



Em destaque, as etapas de demetilação em C-4 e demetilação em C-14. Fonte: adaptada de DESMOND e GRIBALDO (2009).

e) Hidroxilase de ácidos graxos (FAH): O gene *FA2H* é altamente expresso nos tecidos do cérebro e cólon e codifica a proteína ácido graxo 2-hidroxilase. Essa proteína está envolvida na formação de mielina 2-hidroxi galactosilceramidas e sulfatídeos (ALDERSON *et al.*, 2004). Nos esfingolipídios, as proteínas do grupo 2-hidroxil são de extrema importância na organização dos lipídeos nas membranas celulares, pois atuam de modo similar às ligações de hidrogênio (HAMA, 2010).

Na literatura, diversos trabalhos têm sido publicados caracterizando proteínas OFNH. A Tabela 3 mostra diferentes genes caracterizados para a planta modelo *Arabidopsis thaliana*, por exemplo. Mesmo assim, nenhum trabalho genômico comparativo foi efetuado nas plantas até então, e visto que elas estão presentes em procaríotos e eucaríotos e compartilham um motivo conservado, é muito provável que exista uma relação evolutiva entre essas proteínas.

Tabela 3 – Genes que codificam oxigenases com ferro não hêmico caracterizados em *Arabidopsis thaliana*

Gene	Via biossintética	Referência
<i>FAD</i>	Desaturação de ácidos graxos	OKULEY <i>et al.</i> , 1994
<i>CER</i>	Biossíntese de cera epicuticular e fertilidade do pólen	AARTS <i>et al.</i> , 1995
<i>ADS</i>	Desaturação de ácidos graxos	FUKUCHI-MIZUTANI <i>et al.</i> , 1998
<i>STE</i>	Biossíntese de brassinosteróides	CHOE <i>et al.</i> , 1999
<i>SMO</i>	Biossíntese de esteróis	DARNET; RAHIER, 2004
<i>SUR</i>	Biossíntese de esfingolipídios	MICHAELSON <i>et al.</i> , 2009
<i>FAH</i>	Hidroxilação de ácidos graxos	NAGANO <i>et al.</i> , 2012

3.3 Importância do ferro no centro catalítico da reação

Os íons metálicos têm papel fundamental na maioria dos processos bioquímicos dos organismos, por isso, as proteínas que necessitam deles para cumprir sua função biológica e são conhecidas como metaloproteínas (BERTINI; SIGEL; SIGEL, 2001; ANDREINI *et al.*, 2007). Essa necessidade se deve ao papel estrutural que o metal assume para dar estabilidade à estrutura da proteína e ao papel funcional, quando compõe centros de transferência de elétrons ou quando está envolvido no mecanismo catalítico de enzimas (ANDREINI *et al.*, 2009).

O ferro é o íon metálico de transição mais abundante nos organismos, e normalmente encontrado nos estados de oxidação +2 (ferroso) ou +3 (férico) dentro das células (ANDREINI *et al.*, 2007). As oxigenases podem possuir ou não algum

metal atuando como centro catalítico da reação. As OFNH utilizam o ferro, que atua como elemento central no processo de catálise feito por essas enzimas (SPERLING *et al.*, 2003; KWEON *et al.*, 2008).

Algumas proteínas podem se ligar, além dos simples íons de ferro, a diversos cofatores que contenham ferro, como por exemplo, o heme ou o agrupamento ferro-enxofre (ANDREINI *et al.*, 2007). O heme é um grupo prostético que exerce um papel de grande importância para algumas metaloproteínas, sendo considerado bastante versátil. Sua função é diversificada, podendo participar do transporte e armazenamento de oxigênio, transferência de elétrons, catálise de oxigênio e monóxido de carbono, bem como transporte de óxido nítrico (CORREIA; SINCLAIR; MATTEIS, 2011). Pode-se citar como exemplos de proteínas que possuem o grupo heme em sua composição: hemoglobinas, mioglobinas, flavoproteínas, peroxidases, catalases, citocromos P450, heme oxigenases, etc. (BERTINI; SIGEL; SIGEL, 2001). Apesar de ser importante para algumas metaloproteínas, as OFNH não possuem o grupo heme em sua composição (SHANKLIN; CAHOON, 1998).

3.4 Assinatura das oxigenases com ferro não hêmico

Proteínas podem ser compostas por vários ou apenas um domínio e representam uma unidade funcional, estrutural e evolutiva das mesmas (BATEMAN; BIRNEY, 2000). Esses domínios podem ser definidos tanto no nível de sequência quanto no nível tridimensional, sendo que sua compreensão é muito utilizada para a predição de função de proteínas (BATEMAN; BIRNEY, 2000). Domínios podem apresentar frações de aminoácidos altamente conservadas, denominadas motivos (HENIKOFF; HENIKOFF, 1991; TATUSOV; ALTSCHUL; KOONIN, 1994), que são constantemente utilizados na previsão de semelhanças funcionais entre homólogos divergentes, e compreendem resíduos do sítio ativo ou de ligação (PONTING *et al.*, 2000).

Quando se compara sequências de DNA ou proteínas relacionadas, é possível visualizar uma sequência consenso, que mostra a base ou o aminoácido mais comum em cada posição. Essa sequência consenso pode ser gerada a partir de ferramentas matemáticas ou pode ser identificada em bancos de dados de sequência. Muitas vezes, partes da sequência que possuem bom acordo

representam domínios funcionais conservados evolutivamente (NELSON; COX, 2008).

O banco de domínios Pfam (<http://pfam.xfam.org/>) agrupa uma grande coleção de famílias de domínios proteicos, onde cada família é representada por vários alinhamentos de sequências e modelos ocultos de Markov (HMMs). O HMM é um modelo probabilístico utilizado para transformar a informação contida no alinhamento múltiplo de sequências em uma sequência consenso. O Pfam classifica suas entradas de quatro formas: 1) Família: uma coleção de regiões proteicas relacionadas; 2) Domínio: uma unidade estrutural; 3) Repetições: uma unidade curta, que é instável isoladamente, mas forma uma estrutura estável quando várias cópias estão presentes; e 4) Motivos: uma unidade curta encontrada fora dos domínios globulares (BATEMAN *et al.*, 2002).

Shanklin e Cahoon (1998) propuseram um domínio consenso para oxigenases com ferro não hêmico utilizando 24 proteínas de desaturases integrais à membrana de procaríotos e eucariotos. Esse domínio possui 8 histidinas conservadas [HX₍₃₋₄₎HX₍₇₋₄₁₎HX₍₂₋₃₎HHX₍₆₁₋₁₈₉₎(H/Q)X₍₂₋₃₎HH] (SHANKLIN; CAHOON, 1998), com espaçamento conservado e se dividem em três blocos: (I) HX₍₃₋₄₎H, (II) HX₍₂₋₃₎HH e (III) HX₍₂₋₃₎HH (SPERLING *et al.*, 2003). Esse domínio está intimamente ligado ao centro de ferro, onde ocorre a oxidação do substrato e a ativação do oxigênio (SHANKLIN *et al.*, 1997; SHU *et al.*, 1997; LEE *et al.*, 1998), e por ter se mostrado conservado no trabalho de Shanklin e Cahoon (1998), provavelmente consiste em uma assinatura funcional comum a todas as oxigenases com ferro não hêmico.

Existe uma exceção ao padrão de oito histidinas, que advém das proteínas Δ^6 ácido graxo desaturase de *Synechocystis* sp. (REDDY *et al.*, 1993) e *Borago officinalis* L. (SAYANOVA *et al.*, 1997). Nessas proteínas, a primeira histidina do terceiro bloco não está presente, e no lugar dela é possível encontrar uma glutamina, a qual também possui uma cadeia lateral contendo nitrogênio (SHANKLIN; CAHOON, 1998). As histidinas são frequentemente encontradas em sítios ativos ou sítios de ligação das proteínas, sendo que, no sítio de ligação, é muito comum que elas estejam ligadas a algum metal (BARNES; GRAY, 2003).

O Pfam classificou as OFNH pelos motivos propostos por Shanklin e Cahoon (1998), e as nomeou como superfamília Hidroxilase de Ácidos Graxos ou FAH (do inglês *Fatty Acid Hydroxylase*), que pode ser encontrada nesse banco de domínios

Outro banco de dados de P450 é o Fungal Cytochrome P450 Database (FCPD) (disponível em: <<http://p450.riceblast.snu.ac.kr/>>; acessado em: 05 de maio de 2015), que possui 8.731 sequências P450 dos genomas de 113 espécies de fungos e oomicetos.

Esses bancos de dados apresentam a classificação de proteínas, que foram depositadas a partir de comparação genômica. A comparação de sequências (de nucleotídeos ou aminoácidos) visa verificar a existência de uma relação evolutiva entre elas por meio da análise de similaridade, inferindo, assim, estruturas e/ou funções homólogas (COHEN, 2004). Entretanto, é mais difícil comparar proteínas a partir de sequências de DNA, devido o código genético ser degenerado.

A comparação genômica é muito utilizada para classificar sequências em bancos de dados, onde as sequências podem ser organizadas em famílias, subfamílias, clãs, e/ou isoformas. Visto que há uma lacuna no conhecimento acerca das OFNH, onde não há bancos de dados que disponibilizam uma classificação completa dessa família de proteínas (ARORA *et al.*, 2009), torna-se válido comparar dados genômicos a partir de sequências proteicas OFNH, entre plantas inferiores e superiores, e entre reinos diferentes. Mas como essa classificação é feita?

A comparação de sequências de aminoácidos pode permitir o entendimento da sua função, localização celular, estrutura tridimensional e até mesmo da evolução (NELSON; COX, 2008). Para que uma comparação entre sequências seja feita, é necessário primeiramente coletar as sequências nos bancos de dados. Todavia, mesmo que a bioinformática desempenhe um papel fundamental, criando algoritmos e ferramentas capazes de analisar as informações depositadas nos bancos de dados, essa busca pode ser exaustiva, devido existir milhares de sequências espalhadas em bancos de dados distintos (MOUNT, 2001)

De acordo com o Genomes Online Database (GOLD) (disponível em: <<https://gold.jgi-psf.org/>>) (PAGANI *et al.*, 2012), existem 26.675 genomas completos de bactéria e 4828 genomas completos de eucariotos já sequenciados (dados disponíveis em: <<https://gold.jgi-psf.org/distribution>>; acessado em 04 de maio de 2015). Com a diversidade de equipamentos e ferramentas que permitem o sequenciamento do DNA com muito mais rapidez, precisão e custos reduzidos, como os sequenciadores de nova geração (NGS), esses números tendem a crescer exponencialmente (CARVALHO; SILVA, 2010; MARDIS, 2011).

3.5.1 Busca de sequências

A busca de sequências visa coletar sequências depositadas em bancos de dados. Posteriormente a essa busca, é possível realizar uma anotação, que consiste em dar uma função putativa a uma sequência. Essa anotação pode ser feita a partir da comparação de sequências (AUBOURG; ROUZÉ, 2001). No entanto, há diversas maneiras de buscar sequências em bancos de dados. Dentre elas, temos:

a) Busca por palavras-chave: auxiliada por termos que servem como um filtro e por operadores booleanos ("*and*", "*or*" e "*not*") (HIGGINS, 2000; BARNES; GRAY, 2003; LESK, 2008).

b) Busca em bancos de dados: cada banco possui acessos específicos e únicos e podem reunir sequências de nucleotídeos, aminoácidos, estruturas terciárias de proteínas e mapas de vias metabólicas (Tabela 5) (SIGRIST *et al.*, 2002; BENSON *et al.*, 2012; KOSUGE *et al.*, 2013; McWILLIAM *et al.*, 2013; BERMAN *et al.*, 2014; KANEHISA *et al.*, 2014; THE UniProt CONSORTIUM, 2014).

Tabela 5 – Bancos de dados mais utilizados em bioinformática

Banco de dados	Principais dados depositados	Organização
GenBank	Sequências de nucleotídeos	NCBI
EMBL	Sequências de nucleotídeos	EBI
DDBJ	Sequências de nucleotídeos	INSDC
UniProt	Sequências de aminoácidos	Colaboração entre EBI, SIB e PIR
PROSITE	Sequências de aminoácidos	SIB
PDB	Estruturas terciárias de proteínas	wwPDB
KEGG	Mapas de vias metabólicas	Universidade de Kyoto - Laboratório Kanehisa

c) Busca por domínios: pode ser feita em bancos de dados que possuem especificamente esse tipo de informação, como o Pfam (FINN *et al.*, 2014) e o InterPro (HUNTER *et al.*, 2011).

d) Busca por comparação de sequência: podem ser feitas através da ferramenta BLAST (HIGGINS, 2000; BARNES; GRAY, 2003), que compara sequências contra um banco de dados (NCBI) com o objetivo de encontrar sequências similares à sequência inserida (ALTSCHUL *et al.*, 1990).

3.5.2 Alinhamento de sequências

As proteínas possuem quatro tipos de estruturas: 1) a estrutura primária, que consiste na sequência de aminoácidos; 2) a estrutura secundária, que consistem nas α -hélices ou folhas- β ; 3) a estrutura terciária, que é composta pela cadeia polipeptídica; e 4) a estrutura quaternária, onde as subunidades se agrupam e a proteína tem sua conformação tridimensional definida (CHOU; CAI, 2003; NELSON; COX, 2008). Quando se dispõe apenas de estruturas primárias de proteínas os métodos mais utilizados para compará-las são os métodos filogenéticos.

Existem muitas ferramentas que geram alinhamentos múltiplos, e existem diversos estudos na literatura que comparam a eficiência dessas ferramentas e dos parâmetros que elas utilizam (Tabela 6) (por exemplo, EDGAR; BATZOGLOU, 2006; PAIS *et al.*, 2014). O alinhamento de sequências é o passo inicial para a comparação genômica. Ele irá determinar o grau de similaridade entre as sequências ou entre fragmentos destas (PROSDOCIMI *et al.*, 2002).

Tabela 6 – Ferramentas de alinhamento múltiplo de sequências consideradas como melhores atualmente disponíveis

Ferramenta	Vantagens	Advertências
CLUSTALW	Usa menos memória que outros programas	Menos preciso do que os programas modernos
DIALIGN	Tenta fazer distinções entre regiões alinháveis e não alinháveis	Menos preciso que o CLUSTALW em referências globais
MAFFT, MUSCLE	Mais rápido e mais preciso do que CLUSTALW; boa escolha de precisão e custo computacional. Opções para correr ainda mais rápido, com precisão média mais baixa, para aplicações de alto rendimento	Para grandes conjuntos de dados (superior a 1.000 sequências) seleciona opções de economia de tempo e memória
PROBCONS	Maior pontuação, precisão em várias referências	Tempo de computação e uso de memória é um fator limitante para os grandes problemas de alinhamento (> 100 sequências)
ProDA	Não assume alinhabilidade global; permite repetições, domínios misturados e ausentes	Alto custo computacional e menos preciso do que CLUSTALW em referências globais
T-COFFEE	Alta precisão e capacidade de incorporar tipos heterogêneos de informação	Tempo de computação e uso de memória é um fator limitante para os grandes problemas de alinhamento (> 100 sequências)

Fonte: adaptada de EDGAR e BATZOGLOU (2006).

Os alinhamentos de sequências podem ser feitos com duas ou mais sequências (PROSDOCIMI *et al.*, 2002), e pode ser local ou global (PROSDOCIMI *et*

al., 2002; PAIS *et al.*, 2014). Como exemplo, podemos citar o *software* ClustalW, que realiza o alinhamento global de sequências, e a ferramenta BLAST, que realiza o alinhamento local.

As frequências em que ocorrem as substituições de aminoácidos são analisadas por matrizes de comparação. Com base na abundância e na frequência relativa de cada aminoácido em um alinhamento de proteínas, é possível atribuir uma pontuação que reflete a probabilidade daquela mutação ocorrer. Essas matrizes podem ser do tipo PAM ou BLOSUM (HENIKOFF; HENIKOFF, 1992).

Também é possível comparar proteínas a partir das estruturas quaternárias, com ferramentas que comparam proteínas a partir de estruturas tridimensionais, geradas por análise de proteínas cristalizadas através da difração de raios-x ou de espectroscopia de ressonância nuclear magnética (RNM) (WUTHRICH, 1989). Entretanto, essa metodologia torna-se inviável quando se trabalha com um número muito grande de sequências e, principalmente, quando não se tem muitas informações sobre elas, como estruturas secundárias das proteínas.

3.5.3 Classificação sistemática

São valiosas as informações que o estudo comparativo de sequências pode nos fornecer, entretanto, essas informações necessitam de métodos sistemáticos que as validem (HENIKOFF; HENIKOFF, 1994; PROSDOCIMI *et al.*, 2002; RUEPP *et al.*, 2004). Após a comparação de sequências, é possível classificar o grupo de proteínas conforme se agrupem na árvore filogenética (HARRISON; LANGDALE, 2006).

Existem dois tipos de abordagens filogenéticas: a cladística e a fenética. A abordagem fenética constrói um fenograma a partir do número de diferenças de caracteres como fenótipos, sítios polimórficos e frequência alélica. Enquanto que a abordagem cladística analisa caracteres ancestrais e derivados e constrói um cladograma. A fenética utiliza métodos de distância (*Neighbor-Joining* e UPGMA), produzindo uma única árvore, e a cladística utiliza métodos de estados do caráter (Máxima parsimônia, Máxima verossimilhança e Análise Bayesiana), e produz diversas árvores (SCHNEIDER, 2003).

Famílias de proteínas que partilham características estruturais ou funcionais podem ser classificadas com base na similaridade dos seus aminoácidos.

Geralmente, os membros alocados em uma mesma família possuem 25% de identidade, ou mais, entre suas sequências (NELSON; COX, 2008). Para sequências protéicas de enzimas que possuem a mesma função e fazem parte da mesma família, possuem em média 35% a 100% de identidade.

As oxigenases P450, por exemplo, foram classificadas, quanto à nomenclatura dos genes, com base na identidade de sequências proteicas e na filogenia (Figura 9) (NELSON *et al.*, 1996). Sendo que, os P450 que partilham pelo menos 40% de identidade são alocados em uma mesma família. Para estarem agrupados em uma mesma subfamília, devem compartilhar pelo menos 55% de identidade (WERCK-REICHHART; BAK; PAQUETTE, 2002).

Figura 5 – Nomenclatura de uma proteína P450

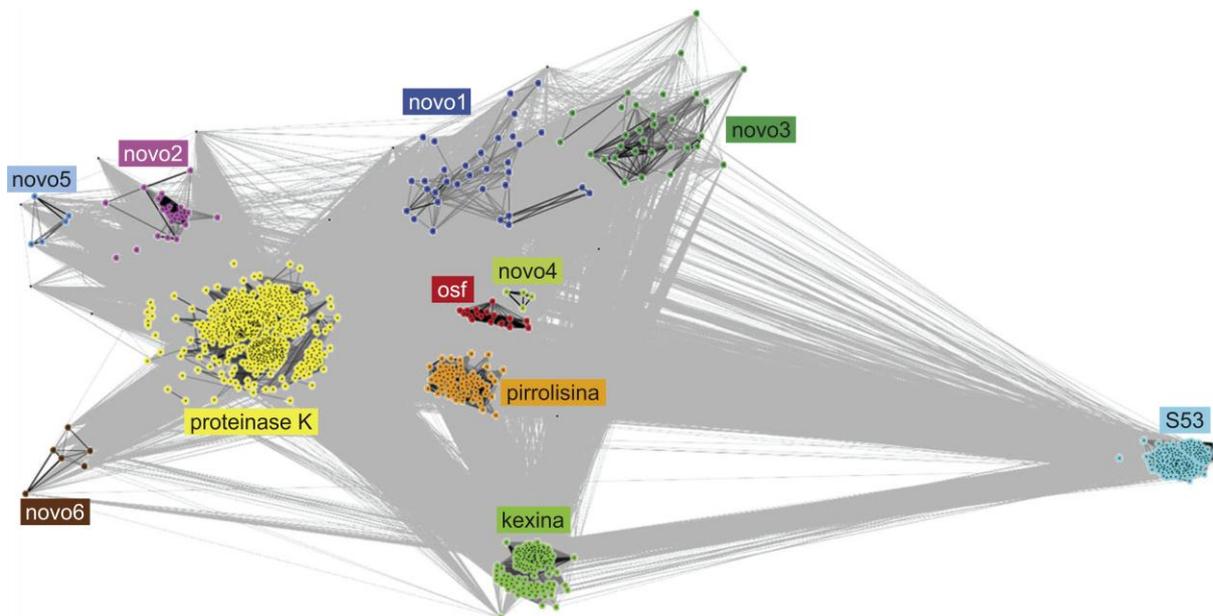


Fonte: adaptada de WERCK-REICHHART, BAK e PAQUETTE (2002).

Outro método de classificação sistemática é por agrupamento de sequências, como o *software* CLANS (FRICKEY; LUPAS, 2004), uma aplicação Java que utiliza uma versão do algoritmo de leiaute gráfico Fruchterman-Reingold para visualizar similaridades entre sequências em 2D ou 3D. O CLANS compara todos os pares de sequências contra todos, e gera um gráfico que mostra todas as interações entre elas (Figura 6) (FRICKEY; LUPAS, 2004).

Esse *software* executa o BLAST de entre todas as sequências e calcula os valores de atração entre os pares de sequência baseado no HSP *P-value*. É possível escolher e variar um valor de corte para descartar *P-values* acima do valor escolhido, sendo que, a depender do valor de corte escolhido, os grupos formados podem se compactar ou ficar desagregados (FRICKEY; LUPAS, 2004). No CLANS, também é possível remover, adicionar e extrair sequências do gráfico, e inferir árvores *Neighbor-Joining* baseada nos valores de *P-value* ou nas distâncias que separam os pontos, que representam as sequências no gráfico (FRICKEY; LUPAS, 2004).

Figura 6 - Exemplo de visualização 2D de um gráfico gerado pelo *software* CLANS



Fonte: adaptada de MUSZEWSKA e colaboradores (2011).

No presente trabalho, foi realizada a classificação das oxigenases com ferro não hêmico a partir da predição de função por similaridade de sequências. Posteriormente será gerado um banco de dados, onde estarão depositadas as sequências proteicas com suas respectivas informações e com novos nomes de acesso, de acordo com o grupo em que estiverem inseridas.

4 MATERIAL E MÉTODOS

As análises comparativas e evolutivas foram baseadas nos trabalhos de Veitia e Hurst (2001) e Vinci, Xia e Veitia (2008), onde o conjunto de sequências foi analisado com ênfase evolutiva, tanto na análise da distribuição das enzimas nas diversas espécies, gêneros e famílias quanto na comparação de sequências com métodos filogenéticos. O resultado esperado era a formação de grupos funcionais bem definidos, específicos e relacionados à evolução das proteínas nas plantas.

Para isso, foram selecionadas espécies com genoma inteiramente sequenciado dentro dos principais reinos para realizar a anotação funcional semiautomática, que foi possível a partir das etapas a seguir:

4.1 Mineração de dados

A busca exaustiva de sequências foi baseada em três principais eixos: i) Recuperação de sequências em bancos de domínios funcionais, como Pfam e Prosite, nos acessos relacionados à superfamília Hidroxilase de Ácidos Graxos; ii) Busca por palavras-chaves. Buscas paralelas foram baseadas na identificação de sequências e artigos usando palavras-chaves no sistema Entrez e SRS dos servidores NCBI e EBI; iii) Busca por similaridade de sequências nos bancos de dados genômicos e de transcriptomas. A busca nos transcriptomas de plantas foi efetuada no banco Gene Index, antigo TIGR (CHILDS *et al.*, 2007).

A coleção de sequências proteicas resultantes dessas buscas foi depositada no *software* Geneious Pro 5.1.7 (DRUMMOND *et al.*, 2010) com seus respectivos nomes de acesso, nome do organismo, nome do gene, e outras descrições como número do cromossomo e locus.

4.2 Curadoria das sequências

Todas as sequências da coleção foram curadas. Para isso, foi feito o alinhamento múltiplo, reconstrução filogenética, e foi verificada a quantidade de genes/loci das proteínas pertencentes a um mesmo organismo visando verificar a existência de redundância de sequências.

Foram consideradas questionáveis: sequências que apresentaram similaridade acima de 90% e sequências muito menores ou muito maiores que o tamanho médio das demais sequências, que é de aproximadamente 300 aminoácidos. As sequências que apresentaram similaridade acima de 90% foram analisadas quanto ao número do cromossomo e locus, e quando comprovada a redundância, foram eliminadas. Para sequências demasiadamente curtas foi feito BLAST no banco de dados NCBI. Assim, quando verificada a existência de uma proteína correspondente pertencente ao mesmo organismo, que possuía mais aminoácidos e apresentava *E-value* menor que $1e-100$, a sequência encontrada substituiu a sequência curta. Da mesma forma, as sequências demasiadamente longas foram analisadas e, quando necessário, tiveram suas extremidades cortadas manualmente.

4.3 Alinhamento múltiplo das sequências

Para identificar se a assinatura funcional da superfamília Hidroxilase de Ácidos Graxos apresentava-se conservada nas sequências, alinhamentos múltiplos foram construídos com três ferramentas disponíveis no Geneious, que são ClustalW v.5.1.7 (parâmetros: matriz = Gonnet; *gap open cost* = 15; *gap extended cost* = 0,3) (SCHNEIDER, 2003; HALL, 2004), MAFFT v.1.3 (parâmetros: matriz = Blosum62; algoritmo = L-INS-i) (DO; KATOH, 2008), e MUSCLE v.5.1.7 (valores de parâmetros padrão) (EDGAR, 2004). Os alinhamentos múltiplos com maior percentual de sítios idênticos e de identidade entre os pares de sequências foram escolhidos para posterior análise.

Diversos parâmetros foram inseridos para realizar o alinhamento múltiplo de sequências. A escolha do melhor alinhamento foi baseada no domínio das OFNH proposto por Shanklin e Cahoon (1998). Desse modo, quando as histidinas eram inexistentes e/ou os espaçamentos entre elas eram diferentes do, outros parâmetros foram aplicados.

Esses alinhamentos foram feitos tanto no Geneious, utilizando um computador do tipo PC, quanto no servidor de processamento. Esse servidor é um grupo composto por um headnode e dois node de trabalho. O headnode possui dois processadores Xeon 3.0 Ghz de 16 core cada e 64 GB de RAM e cada um dos node

de trabalho possui dois processadores Xeon 3.0 Ghz de 16 core cada e 128 GB de RAM, além de RAID 5 com 21 TB para armazenamento dos dados.

4.4 Reconstrução filogenética

As árvores filogenéticas foram construídas com as ferramentas MrBayes v.2.0.3, que constrói a filogenia utilizando inferência Bayesiana (valores de parâmetros padrão) (HALL, 2004), PAUP* v.1.1, que constrói a filogenia utilizando *Neighbor-Joining* (parâmetros: tipo de busca = heurística) (HALL, 2004), e PhyML v.2.1.0, que constrói a filogenia utilizando máxima verossimilhança (*Maximum-Likelihood*) (parâmetros: modelo de substituição = LG) (GUINDON; GASCUEL, 2003).

4.5 Classificação sistemática

Uma vez escolhida a árvore filogenética, as proteínas foram relacionadas ao substrato e à função, a partir da similaridade com as proteínas já caracterizadas, para a identificação e nomenclatura dos grupos. Para elucidar as relações dentro da superfamília de oxigenases com ferro não hêmico, utilizou-se o *software* CLANS (FRICKEY; LUPAS, 2004), no qual foi utilizado o valor de corte $1e-150$ e a abordagem *Network based* de detecção de grupos. Posteriormente os grupos formados foram nomeados de acordo com a função.

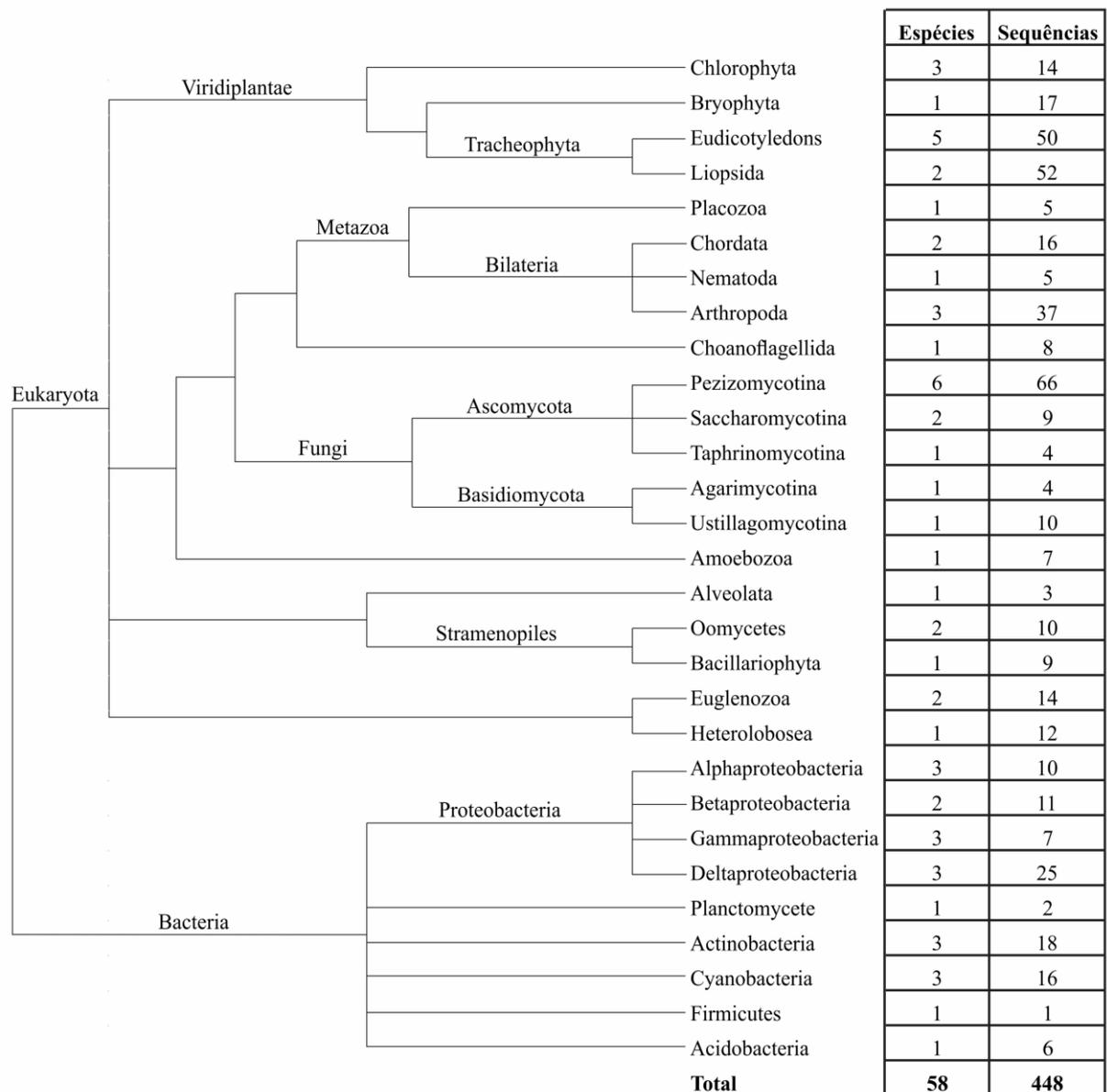
5 RESULTADOS

A classificação das OFNH foi realizada conforme mostrado no fluxograma (Figura 7). Para a busca de seqüências foram selecionados 58 organismos dentro dos principais reinos, dos quais 56 possuem seus genomas inteiramente seqüenciados (Anexo A). A busca de OFNH resultou em uma coleção de mais de 800 seqüências. Após a curadoria muitas seqüências curtas, redundantes e obsoletas foram eliminadas, resultando em uma coleção de 448 proteínas (Figura 8), das quais apenas 25 estão caracterizadas na literatura (Anexo B). A maioria das seqüências contidas na coleção não possui anotação funcional nos bancos de dados, sendo geralmente nomeada como proteína putativa, predita ou não caracterizada.

Figura 7 - Resumo da metodologia utilizada para a classificação das proteínas oxigenases com ferro não hêmico



Figura 8. Esquema dos principais taxa e suas respectivas quantidades de espécies e sequências de oxigenases com ferro não hêmico contidas na coleção



O alinhamento gerado com a ferramenta ClustalW apresentou 198 sítios idênticos (16,6%) e 8,8% de identidade média entre os pares de sequências, e as histidinas apresentaram entre 25,7% e 43,5% de identidade média entre os pares de sequências (Tabela 7 e Figura 9A). O alinhamento gerado com a ferramenta MAFFT apresentou 5 sítios idênticos (0,3%) e 9,9% de identidade média entre os pares de sequências, e as histidinas apresentaram um percentual de identidade média entre os pares de sequências que variou entre 87% e 97,8% (Tabela 7 e Figura 9B). O alinhamento gerado com a ferramenta MUSCLE apresentou 14 sítios idênticos (0,7%) e 9,2% de identidade média entre os pares de sequências, e as histidinas

apresentaram entre 37,4% a 87,5% de identidade média entre os pares de sequências (Tabela 7 e Figura 9C).

Tabela 7 – Percentual de identidade média entre os pares de sequências apresentado por cada histidina nos alinhamentos múltiplos de sequências

Ferramenta	Posição da histidina no domínio							
	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a	8 ^a
ClustalW	43,0%	37,8%	27,2%	35,0%	28,0%	25,7%	43,5%	29,5%
MAFFT	96,4%	94,7%	97,8%	96,0%	93,4%	87,0%	97,3%	93,4%
MUSCLE	75,9%	37,4%	73,6%	75,1%	78,6%	83,4%	83,8%	87,5%

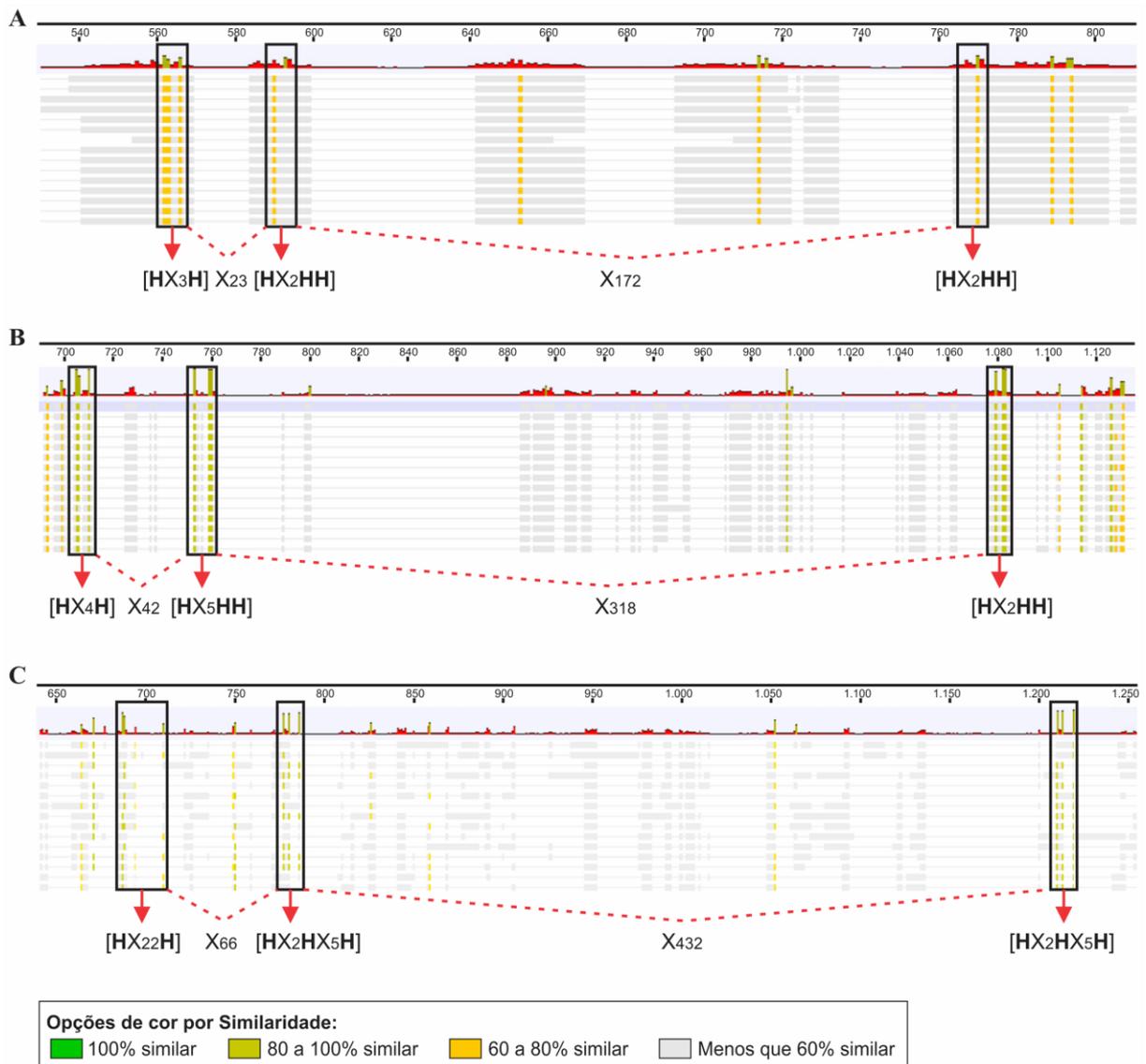
O percentual de identidade média é gerado automaticamente pelo *software* Geneious. Para este é cálculo, todos os pares de aminoácidos posicionados na mesma coluna são analisados marcando um *hit* quando eles são idênticos; ao final, esse valor é dividido pelo número total de pares.

No alinhamento gerado com MAFFT, 11 sequências apresentaram uma glutamina (Q) no lugar da 6^a histidina do domínio. Os grupos que apresentaram essas sequências são: o grupo 23 (1 sequência de *Frankia alni* e 1 de *Streptomyces albus*), o grupo 25 (2 sequências de *Nectria haematococca*), o grupo 26 (1 sequência de *Anaeromyxobacter dehalogenans* e 1 de *Synechocystis* sp.) e o grupo 28 (1 sequência de *Frankia alni* e 1 de *Mycobacterium bovis*). As outras 3 sequências (2 de *Leishmania braziliensis* e 1 de *Phytophthora infestans*). não foram atribuídas à grupos.

No mesmo alinhamento, 4 sequências (1 de *Aspergillus clavatus*, 1 de *Nectria haematococca*, 1 de *Penicillium chrysogenum* e 1 de *Penicillium digitatum*) apresentaram na mesma posição, ao invés de histidina ou glutamina, uma treonina (T). Essas sequências compreendem o grupo 16 (SDES) e são nomeadas apenas como “família esteroide desaturase” ou “proteína predita”.

Dentre os três alinhamentos, o alinhamento múltiplo de sequências feito com MAFFT mostrou conservação das oito histidinas e do espaçamento entre os motivos do domínio, aproximando-se do motivo consenso proposto por Shanklin e Cahoon (1998). Com isso, a partir da comparação entre o domínio resultante desse alinhamento (Figura 9B) com descrito na literatura, sugere-se a atualização do domínio das OFNH para: HX₍₃₋₄₎HX₍₇₋₄₂₎HX₍₂₋₅₎HHX₍₆₁₋₃₁₈₎(H/Q)X₍₂₋₃₎HH.

Figura 9 - Alinhamentos múltiplos das seqüências de proteínas oxigenases com ferro não hêmico



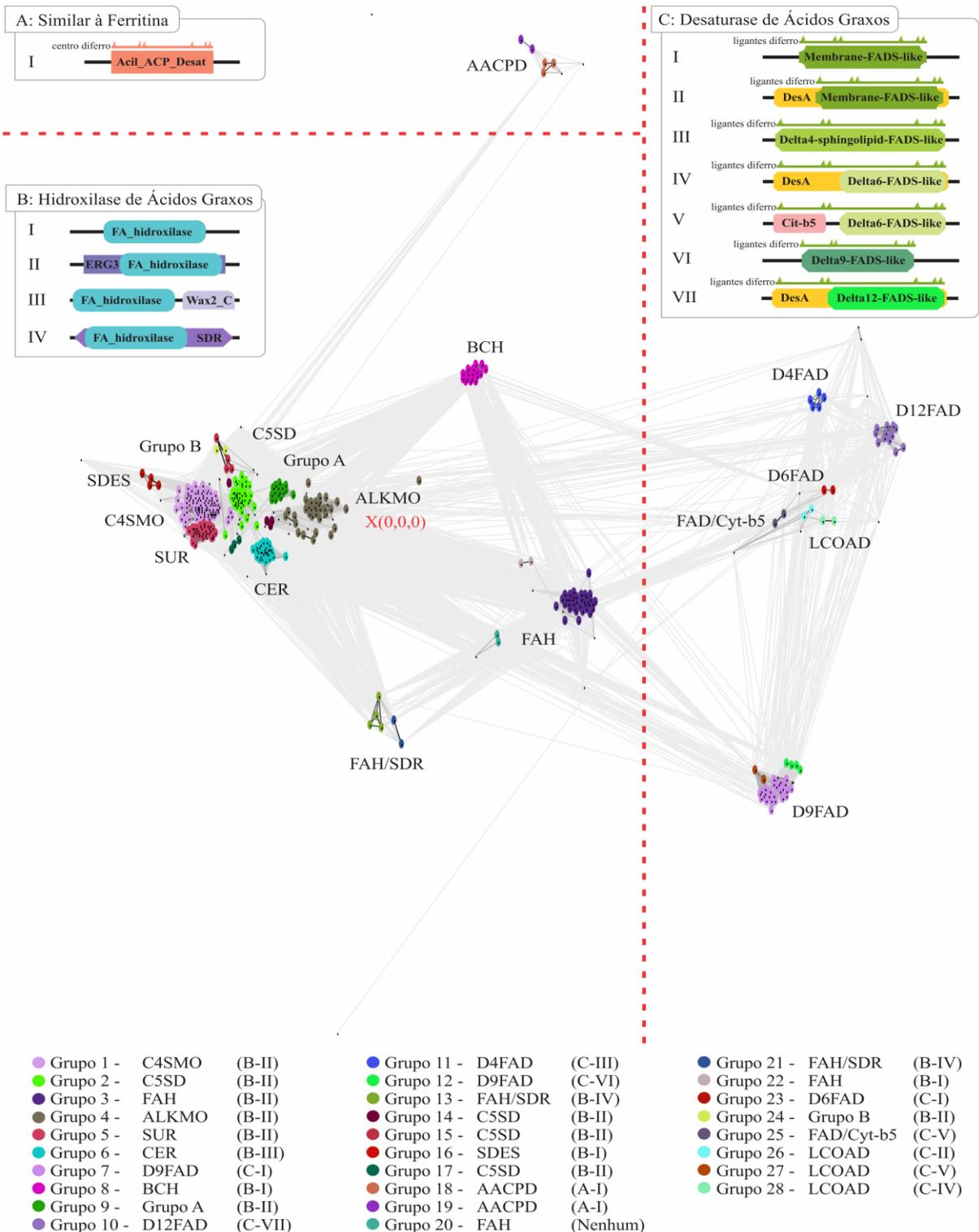
Alinhamentos gerados com as ferramentas ClustalW (A), MAFFT (B) e MUSCLE (C). Histidinas são representadas pela letra H. Setas vermelhas indicam os motivos, que se dividem em 3 blocos, destacados por retângulos. O número ao lado de cada X se refere à quantidade máxima de aminoácidos que separa as histidinas dos motivos. As seqüências contidas nesses alinhamentos possuem maior comprimento, mas para facilitar a visualização, aqui é mostrado apenas o trecho do alinhamento onde as histidinas estão presentes.

Na classificação feita com o *software* CLANS, 400 seqüências foram atribuídas a 28 grupos (Anexo C), e 48 seqüências não foram atribuídas a quaisquer dos grupos formados (Figura 10). A nomenclatura dos grupos foi feita a partir de predição de função por homologia. Para isso, as proteínas caracterizadas na literatura foram levadas em consideração, mas devido estas serem apenas 25 (Anexo B) a ferramenta BLAST (BLASTp) foi utilizada para buscar domínios

cadastrados que apresentassem similaridade com as mesmas. É possível ver o resultado dessa análise na figura 10, onde os domínios encontrados são representados dentro dos balões. Com essas análises, pôde-se identificar similaridade a 16 funções principais (Figura 10): C4 metilesterol monooxigenase, C5 esterol desaturase, esterol desaturase, aldeído decarbonilase, β -caroteno hidroxilase, esfingolípídeo C4 monooxigenase, alquilglicerol monooxigenase, ácido graxo hidroxilase, ácido graxo hidroxilase/dehidrogenase-redutase de cadeia curta, Δ^4 ácido graxo desaturase, Δ^6 ácido graxo desaturase, Δ^9 ácido graxo desaturase, Δ^{12} ácido graxo desaturase, linoleoil-CoA desaturase, ácido graxo desaturase/Citocromo-b5 e acil-ACP desaturase. Dois grupos foram nomeados como grupo A e grupo B, pois neles foram atribuídas apenas sequências não caracterizadas, e apesar de ambos terem apresentado um domínio putativo na análise de BLAST (Figura 10), a maioria das sequências encontradas nessa análise foram proteínas preditas não caracterizadas.

As linhas tracejadas na figura 10 dividem os grupos, indicando o limite da ocorrência dos domínios contidos nos balões. É possível observar que um mesmo domínio esteve presente em vários grupos, entretanto, não foi detectado nenhum domínio putativo conservado para o grupo 20. Os domínios encontrados pela análise de BLAST estão incluídos nas famílias de proteínas semelhantes à Ferritina, Hidroxilase de ácidos graxos e Desaturase de ácidos graxos, representadas na figura 10 por A, B e C, respectivamente. No anexo C é possível ver a nomenclatura dada para cada grupo de acordo com a proteína atribuída a ele.

Figura 10 - Grupos gerados por CLANS com base nas similaridades entre as sequências de proteínas oxigenases com ferro não hêmico



Pontos representam sequências. Quanto mais escuras as linhas de cor cinza, menor é o *P-value*. Sequências do mesmo grupo são mostradas em uma cor; sequências não atribuídas a grupos são mostradas em preto. Os domínios encontrados na análise de BLAST estão representados dentro dos balões; linhas tracejadas em vermelho separam os grupos que apresentaram estes domínios. O domínio que cada grupo apresentou está representado entre parênteses na legenda.

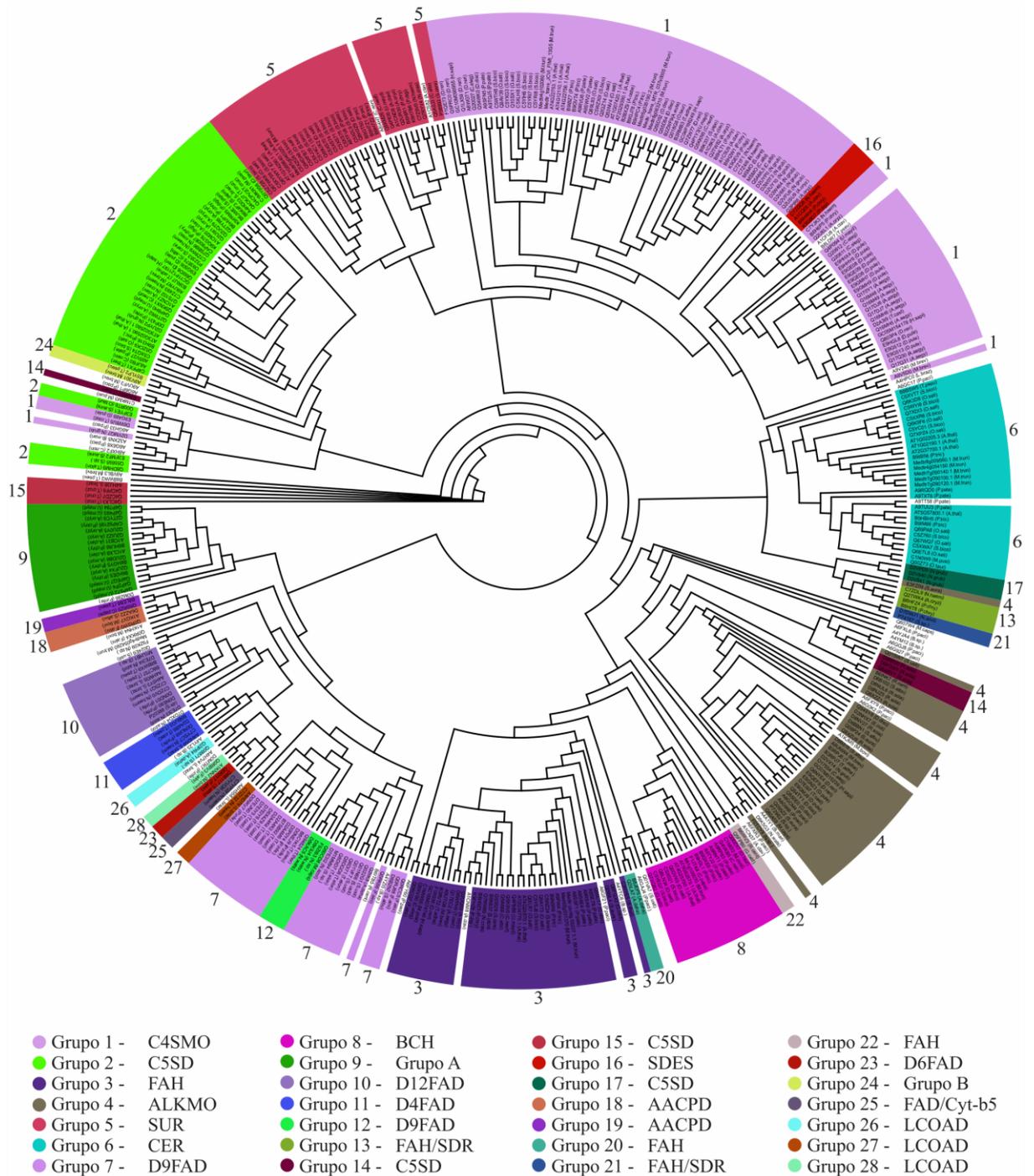
Os grupos 12, 18, 20, 21, 22, 23, 26 e 28 possuem sequências exclusivamente de bactérias. Entretanto, as sequências de *Bastopirellula marina* e *Bradyrhizobium* sp. não foram atribuídas a quaisquer dos grupos formados. Os grupos 1, 5, 6, 8, 9, 11, 13, 15, 16, 17, 19, 24, 25, 27 possuem sequências exclusivamente de eucariotos. Os únicos grupos comuns a procaríotos e eucariotos foram o 2, 3, 4, 7, 10 e 14 (Anexo D).

O grupo 13 (FAH/SDR) e o grupo 16 (esterol desaturase) são compostos exclusivamente por sequências de espécies de fungos. Os únicos eucariotos a apresentar sequências proteicas de função acil-ACP desaturase (grupo 19) foram a diatomácea *Thalassiosira pseudonana* e a planta *Carica papaya*. Apenas sequências de fungos foram atribuídas ao grupo 25 (FAD/Cyt-b5). O grupo A é composto por sequências proteicas de quatro espécies de fungos: *Aspergillus clavatus*, *Penicillium chrysogenum*, *Aspergillus oryzae* e *Ustilago maydis*. O grupo B é composto por uma sequência do coanoflagelado *Monosiga brevicollis* e uma sequência da diatomácea (bacilariófita) *Thalassiosira pseudonana* (Anexo D).

Na coleção, 119 sequências são provenientes de plantas e foram distribuídas em 7 grupos, correspondentes às funções metilesterol monooxigenase, C5 esterol desaturase, ácido graxo hidroxilase, esfingolípídeo monooxygenase, aldeído decarbonilase (Eceriferum), β -caroteno hidroxilase e acil-ACP desaturase. A planta modelo *Arabidopsis thaliana* apresentou sequências atribuídas aos grupos 1, 2, 3, 5, 6 e 8. Dentre as proteínas caracterizadas em *A. thaliana* (Anexo B), cinco foram alocadas ao grupo 1, sendo três isoformas de metilesterol monooxigenase 1 e duas isoformas de metilesterol monooxigenase 2; uma proteína do tipo Δ^7 -esterol-C5(6)-desaturase 1 foi incluída no grupo 2; duas isoformas de ácido graxo 2-hidroxilase foram atribuídas ao grupo 3; duas isoformas de esfinganina C4-monooxigenase foi alocada ao grupo 5; duas isoformas (1 e 3) da proteína eceriferum estão contidas no grupo 6; e duas isoformas caracterizadas de β -caroteno 3-hidroxilase foram atribuídas ao grupo 8.

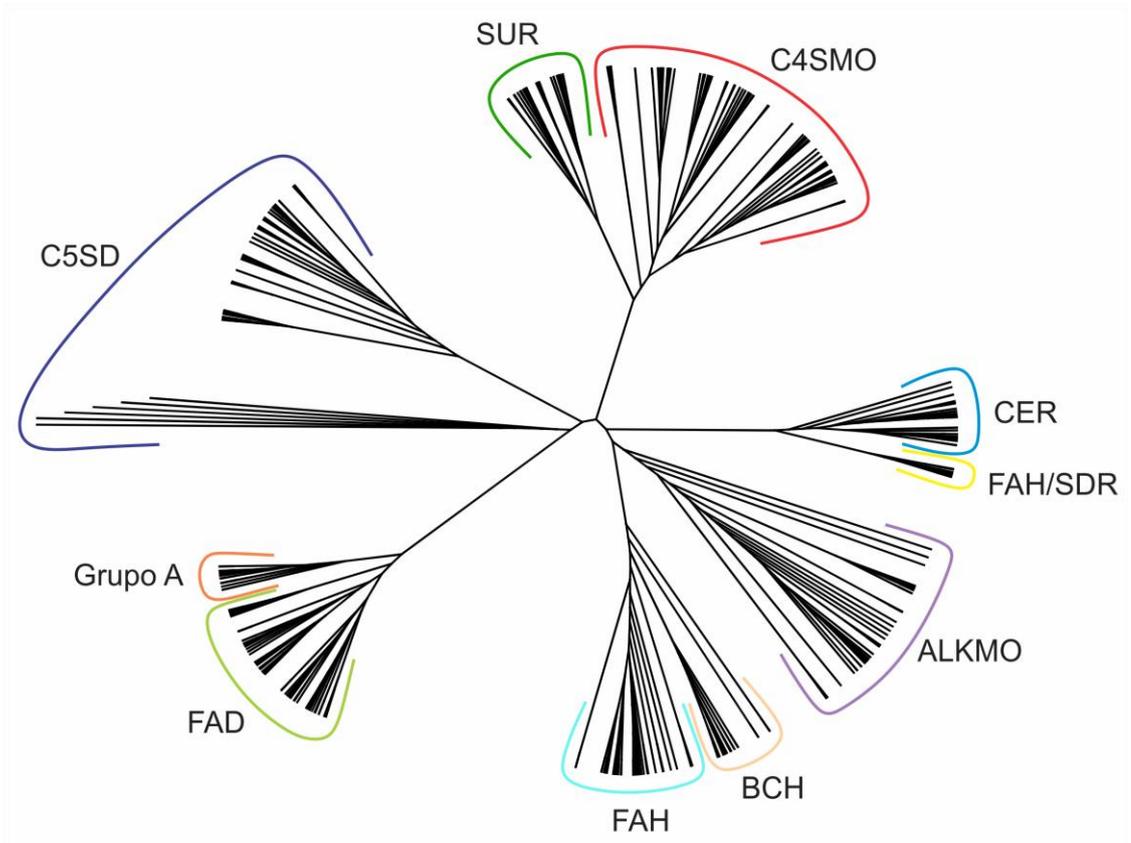
A melhor árvore filogenética foi gerada com a ferramenta PhyML, e seus agrupamentos mostraram uma relação coerente com os grupos formados em CLANS (Figura 11). Quando a mesma árvore foi modificada em um leiaute não enraizado (cladograma) dentro do *software* Geneious, foi possível identificar os grupos mais representativos da classificação das OFNH (Figura 12).

Figura 11. Árvore filogenética das sequências de oxigenases com ferro não hêmico gerada por análise de máxima verossimilhança com a ferramenta PhyML



Os acessos das sequências foram coloridos de acordo com os grupos formados por CLANS, e sequências não alocadas a grupos não foram coloridas.

Figura 12 – Cladograma não enraizado mostrando os grupos mais representativos das oxigenases com ferro não hêmico



6 DISCUSSÃO

O número de sequências obtidas reduziu praticamente pela metade após a curadoria. Isso se deve, principalmente, à existência de sequências redundantes depositadas nos bancos de dados. Durante a busca não é possível saber se se trata de uma sequência igual à outra sequência já coletada. Dessa forma, a curadoria de sequências possui grande importância, já que elimina sequências repetidas em uma coleção através da análise de características da sequência, cadastradas no banco de dados, como número do cromossomo e locus.

Até o momento, nenhuma desaturase do tipo integral à membrana foi estruturalmente definida, devido à dificuldade de se obter grandes quantidades de proteínas de membrana purificadas (SHANKLIN; CAHOON, 1998). Por conta disso as informações acerca dessas proteínas são limitadas e para compará-las foi utilizada a predição de função por homologia.

A sequência, estrutura e função de uma proteína divergem ao longo do tempo evolutivo e diferentes critérios podem gerar alinhamentos diferentes (EDGAR; BATZOGLOU, 2006). As três ferramentas de alinhamento múltiplo de sequências aqui utilizadas estão entre as mais utilizadas pela bioinformática (EDGAR; BATZOGLOU, 2006; DO; KATOH, 2008).

Desde que foi introduzido, em 1994, o ClustalW não teve melhorias significativas e diversos outros métodos têm sido otimizados quanto ao desempenho de velocidade e/ou precisão (EDGAR; BATZOGLOU, 2006). Essa ferramenta usa menos memória que outros programas, mas é menos precisa que outros programas modernos (EDGAR; BATZOGLOU, 2006). No presente estudo, essa precisão se refere à eficiência em identificar e alinhar as histidinas das sequências. Isso pode explicar porque no alinhamento feito com ClustalW as histidinas apresentaram um baixo percentual identidade entre os pares de sequência, quando comparado aos outros alinhamentos.

Assim como ClustalW, MUSCLE produz alinhamentos múltiplos com muita rapidez, mas produz alinhamentos de baixa precisão para alinhar motivos conservados (NUIN; WANG; TILLIER, 2006), o que pode explicar a larga faixa de percentual de identidade *pairwise* das histidinas e um largo espaçamento entre as histidinas dos blocos.

MAFFT tem sido indicado como a melhor ferramenta para gerar alinhamento múltiplo de sequências, e a abordagem iterativa L-INS-i, apontada como a mais precisa para detectar motivos conservados (NUIN; WANG; TILLIER, 2006). Certamente, tão importante quanto o método a ser utilizado nessa ferramenta, a escolha da matriz é crucial para o resultado obtido no alinhamento. A matriz escolhida para o alinhamento, BLOSUM 62, alinha sequências com identidade de 62% nos blocos e é matriz universal para comparação ampla de proteínas. Ela geralmente apresenta melhor desempenho que outras matrizes BLOSUM e matrizes PAM, já que alinham segmentos em blocos, que representam regiões mais conservadas nas proteínas do presente estudo (HENIKOFF; HENIKOFF, 1992).

Algumas das vantagens de utilizar CLANS são: a possibilidade de lidar com sequências não refinadas, desalinhas e/ou não homólogas. CLANS utiliza todos os dados considerando-os com o mesmo peso, assim, quanto maior o número de sequências, mais delimitado é o tamanho do grupo. Além disso, essa ferramenta não utiliza alinhamentos múltiplos, sua precisão aumenta com o número de sequências, e é insensível a sequências falso-positivas e sequências altamente divergentes. Mesmo que o número de sequências seja grande, como várias dezenas de milhares de sequências, CLANS consegue calcular mapas de similaridade em questão de horas em apenas uma CPU (FRICKEY; LUPAS, 2004; KOPEC; ALVA; LUPAS, 2011).

Os grupos formados mostram a variedade de proteínas oxigenases com ferro não hêmico envolvidas em vias de biossíntese lipídica. Além disso, algumas funções parecem ser restritas a determinados organismos, como os grupos 13 e 16. Alguns grupos possuem proteínas com a mesma função putativa, o que sugere que pode haver diferenças a nível estrutural ou funcional entre elas. Os grupos A e B apresentaram somente proteínas preditas não caracterizadas, e não foi possível obter uma função putativa através da análise com BLAST, indicando que as proteínas atribuídas a esses grupos possuem novas funções.

A exceção da primeira histidina do terceiro bloco do motivo, substituída pela glutamina, foi encontrada em proteínas atribuídas aos grupos D6DESA, FAD/Cyt-b5 e LCOAD, comprovando que essa exceção pode ocorrer em outras proteínas além das Δ^6 ácido graxo desaturases descritas por Reddy e colaboradores (1993) e Sayanova e colaboradores (1997). Não foram encontrados na literatura, estudos que demonstrassem a substituição da histidina, nessa mesma posição, pela treonina.

Isso, somado ao fato de que as proteínas com essa exceção formaram um grupo distinto (grupo 16), sugere que essas proteínas possuam uma função desconhecida. Como essas proteínas não são caracterizadas e não é possível identificar uma função específica para elas, assim, será mantida no domínio uma única exceção (glutamina) já descrita nos resultados desse trabalho.

A presença de 6 funções distintas para *Arabidopsis thaliana* comprova que ela é muito especializada em suas funções. A ocorrência das proteínas Δ^7 esterol-C5(6)-desaturase, β -caroteno 3-hidroxilase, 3 isoformas de metilesterol monooxigenase 1, 2 isoformas de metilesterol monooxigenase 2, duas isoformas de ácido graxo 2-hidroxilase, e 2 isoformas da proteína eceriferum (aldeído decarbonilase), todas caracterizadas em *A. thaliana* ratifica a importância das oxigenases com ferro não hêmico em vias de biossíntese lipídica.

Os esteróis são moléculas orgânicas presentes nas membranas celulares de protozoários, algas, fungos, plantas e animais (ALONSO *et al.*, 2014), e atuam como componentes estruturais e desempenham papéis fundamentais, como a regulação da fluidez, resistência mecânica e permeabilidade (BENVENISTE, 2002).

A proteína metilesterol monooxigenase (C4SMO) é conhecida como esterol C4-metil-oxidase (gene *SMO* em plantas e *ERG25* em leveduras) e é uma das proteínas requeridas durante a demetilação do carbono 4 na via de biossíntese de esteróis (DARNET; BARD; RAHIER, 2001; DARNET; RAHIER, 2003; RAHIER *et al.*, 2013). Darnet e Rahier (2004) e Rahier e colaboradores (2013) confirmaram a existência de duas isoenzimas de SMOs de plantas: *SMO1* e *SMO2*.

As proteínas colesterol 25-hidroxilase (C5SD) são similares às sequências de C-4 esterol-metil-oxidases (LUND *et al.*, 1998), o que justifica a presença de ambas as proteínas no grupo 1. O 25-hidroxicolesterol é um oxisterol regulador potente, que pode participar do metabolismo lipídico de várias formas, incluindo a regulação da expressão gênica, síntese de ácido biliar no fígado, e o transporte de um tecido para outro (RUSSELL, 2000).

Quatro grupos apresentaram sequências que se assemelharam à proteínas C5 esterol desaturase. A desaturação em C5(6) envolve uma transferência de elétrons do NADH para a oxidase terminal (a própria desaturase) através de uma redutase do citocromo b5 e o citocromo b5, como tem sido documentada em mamíferos, levedura e em plantas (RAHIER, SMITH; TATON, 1997).

Lipídios éter, incluindo alquilgliceróis e éteres glicerílicos, formam uma classe diversificada de lípidos à base de glicerol e compostos lipídicos fosfozotados, que são bem menos estudados do que os lipídios ésteres correspondentes (acilgliceróis) (WATSCHINGER; WERNER, 2013). Os lipídios éter são essenciais para proteger o olho da catarata, para permitir o desenvolvimento do sêmen funcional, e para permitir o correto desenvolvimento de estruturas cerebrais em humanos (GORGAS *et al.*, 2006). A alquilglicerol monooxigenase (ALKMO) é a única enzima conhecida por clivar a ligação de éter em lipídios éter alquilglicerol e ocorre em animais que vão desde *Caenorhabditis elegans* e vários peixes a mamíferos, incluindo roedores, cães, galinhas, chimpanzés e seres humanos (WATSCHINGER; WERNER, 2013).

Os esfingolípídeos (SUR) são moléculas estruturalmente diversas e são os principais componentes do sistema de endomembrana (SPERLING *et al.*, 2005). Eles proporcionam a integridade estrutural da membrana, contribuem para a organização de microdomínios da membrana (MONGRAND *et al.*, 2004; BORNER *et al.*, 2005), funcionam, através de seus metabólitos, como reguladores de processos celulares e fisiológicos fundamentais nas plantas, como a morte celular programada (Liang *et al.*, 2003) e moléculas de sinalização (NG *et al.*, 2001; COURSOOL *et al.*, 2003, 2005).

Uma característica dos esfingolípídeos de plantas é o seu elevado teor de bases tri-hidroxi de cadeia longa (LCBs) que são produzidos pela LCB C4-hidroxilase, codificada pelo gene *SBH* (*Sphingoid Base Hydroxylase*), homólogo ao gene *SUR* de *Saccharomyces cerevisiae* (CHEN *et al.*, 2008). Chen e colaboradores (2008) mostraram que a tri-hidroxi LCB desempenha um papel central na manutenção e crescimento mediando a composição total e o conteúdo de ácido gordo de esfingolípídeos em plantas.

As ceras podem ser encontradas numa grande variedade de organismos vivos como uma mistura de substâncias derivadas de ácidos graxos de cadeia longa. Estes componentes podem ser encontrados nas plantas, especificamente como uma camada epicuticular que cobre as folhas e caules jovens (AARTS *et al.*, 1995). A função primária da deposição de cera epicuticular em plantas é a redução da perda de água através da epiderme característica que contribui para a tolerância à seca (AARTS *et al.*, 1995). Além disso, a camada exterior de cera tem uma função importante na interação com insetos herbívoros e fungos patógenos de plantas

(EIGENBRODE; ESPELIE, 1995; AARTS *et al.*, 1995), e Aarts e colaboradores (1995) verificaram a proteína eceriferum (CER) envolvida na biossíntese de cera e fertilidade do pólen, sendo de extrema importância na propagação das espécies vegetais.

Os carotenoides (BCH) são componentes essenciais para as membranas fotossintéticas em plantas, algas e cianobactérias, bem como para os cromoplastos encontrados em muitos frutos e flores de plantas (TAN *et al.*, 2007). Estes pigmentos (vermelhos, amarelos e laranjas) protegem contra a oxidação fotoquímica e recebem a luz para a fotossíntese (SUN; GANTT; CUNNINGHAM, 1996).

As acil-ACP desaturases (AACPD) compõe uma família de enzimas solúveis estreitamente relacionadas, que catalisam a inserção da primeira dupla ligação em uma cadeia de ácidos graxos saturados (CAHOON; COUGHLAN; SHANKLIN, 1997). Essas proteínas são encontradas em plastídeos de plantas na forma solúvel, e desaturam os ácidos graxos ligados a uma proteína transportadora de acil (ACP).

Todas essas informações apenas confirmam a importância das OFNH em vias de biossíntese lipídica dos organismos, e principalmente de plantas.

7 CONCLUSÃO

A comparação de sequências se mostrou eficiente para a classificação da família de proteínas oxigenases com ferro não hêmico. A melhor ferramenta para realizar o alinhamento múltiplo de sequências proteicas de OFNH é o MAFFT, utilizando a matriz Blosum62 e o algoritmo de alinhamento local L-INS-i.

Ainda que Shanklin e Cahoon (1998) tenham utilizado apenas 24 proteínas para propor um domínio conservado a proteínas desaturases integrais à membrana, esse domínio se mostrou muito próximo do resultado encontrado no presente trabalho, que propõe um domínio comum não somente à desaturases integrais à membrana, mas a todas as monooxigenases ferro não hêmico.

O *software* CLANS mostrou-se eficiente para agrupar as sequências de OFNH, pois apesar de não realizar inferências evolutivas, gerou grupos muito similares aos formados na árvore filogenética feita com a ferramenta PhyML. Alguns grupos formados mostraram-se restritos a organismos e as proteínas de plantas mostraram envolvimento em vias lipídicas importantes, como a biossíntese de esteróis, de ácidos graxos, de esfingolipídeos, de cera epicuticular e produção de caroteno. A não atribuição de sequências a quaisquer dos 28 grupos e a existência de dois grupos com sequências sem função putativa aparente corrobora a importância da anotação funcional e caracterização de proteínas. Além disso, sugere que ainda existem muitas funções desconhecidas dentro da família das OFNH.

Este trabalho fornece importantes resultados que podem servir para estudos futuros acerca da história evolutiva das OFNH, como: eventos de duplicações e perdas de genes; identificar a isoformas e funções exclusivas a algum organismo; e esclarecer se proteínas P450, constantemente presentes nas mesmas vias de biossíntese que as OFNH, surgiram antes, depois ou ao mesmo tempo em que as oxigenases com ferro não hêmico.

REFERÊNCIAS

- AARTS, M. G. *et al.* Molecular characterization of the *CER1* gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. **Plant Cell**, v.7, n. 12, p.2115-2127, 1995.
- ALDERSON, N. L. *et al.* FA2H-dependent fatty acid 2-hydroxylation in postnatal mouse brain. **Journal of Lipid Research**, v. 47, n. 12, p. 2772-2780, 2006.
- ALDERSON, N. L. *et al.* The human *FA2H* gene encodes a fatty acid 2-hydroxylase. **The Journal of Biological Chemistry**, v. 279, n. 47, p. 48562-48568, 2004.
- ALONSO, F. *et al.* Multicomponent synthesis of 4,4-dimethyl sterol analogues and their effect on eukaryotic cells. **Steroids**, v. 84, p. 1-6, 2014.
- ALTSCHUL, S. F. *et al.* Basic Local Search Tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, 1990.
- ANDREINI, C. *et al.* Non-heme iron through the three domains of life. **PROTEINS: Structure, Function, and Bioinformatics**, v. 67, n. 2, p. 317-324, 2007.
- ANDREINI, C. *et al.* Structural analysis of metal sites in proteins: Non-heme iron sites as a case study. **Journal of Molecular Biology**, v. 388, n. 2, p. 356-380, 2009.
- ARORA, P. K. *et al.* OxDBase: a database of oxygenases involved in biodegradation. **BMC Research Notes**, v. 2, n. 67, p. 1-8, 2009.
- ARTHINGTON, B. A. *et al.* Cloning, disruption and sequence of the encoding yeast C-5 sterol desaturase. **Gene**, v. 102, n. 1, p. 39-44, 1991.
- AUBOURG, S.; ROUZÉ, P. Genome annotation. **Plant Physiology and Biochemistry**, v. 39, n. 3-4, p. 181-193, 2001.
- BARD, M. *et al.* Cloning and characterization of *ERG25*, the *Saccharomyces cerevisiae* gene encoding C-4 sterol methyl oxidase. **PNAS**, v. 93, n. 1, p. 186-190, 1996.
- BARNES, M. R.; GRAY, I. C. **Bioinformatics for geneticists**. John Wiley & Sons, Ltd., 408 p., 2003.
- BATEMAN, A. *et al.* The Pfam protein families database. **Nucleic Acids Research**, v. 30, n. 1, p. 276-280, 2002.
- BATEMAN, A.; BIRNEY, E. Searching databases to find protein domain organization. **Advances in protein chemistry**, v. 54, p. 137-157, 2000.
- BAXEVANIS, A. D.; OUELLETTE, B.F.F. **Bioinformatics: A practical guide to the analysis of genes and proteins - Methods of biochemical analysis**. 2nd Ed., Vol 43, Wiley, 488 p., 2004.

- BENSON, D. A. *et al.* GenBank. **Nucleic Acids Research**, v. 41, p. D36-D42, 2012.
- BENVENISTE, P. Biosynthesis and accumulation of sterols. **Annual Review of Plant Biology**, v. 55, p. 429-457, 2004.
- BENVENISTE, P. Sterol metabolism. **The Arabidopsis Book**, v. 1, p. 1-31, 2002.
- BERMAN, H. M. *et al.* The Protein Data Bank archive as an open data resource. **Journal of Computer-Aided Molecular Design**, v. 28, n. 10, p. 1009-1014, 2014.
- BERTINI, I.; SIGEL, A.; SIGEL, H. **Handbook on Metalloproteins**. Marcel Dekker, Inc., p. 269-335, 2001.
- BORNER, G. H. H. *et al.* Analysis of detergent-resistant membranes in Arabidopsis. Evidence for plasma membrane lipid rafts. **Plant Physiology**, v. 137, n. 1, p. 104-116, 2005.
- BOURGAUD, F. *et al.*, 2001. Production of plant secondary metabolites: a historical perspective. **Plant Science**, v. 161, n. 5, p. 839-851, 2001.
- BURTON, S. G. Oxidizing enzymes as biocatalysts. **TRENDS in Biotechnology**, v. 21, n. 12, p. 543-549, 2003.
- CAHOON, E. B.; COUGHLAN, S. J.; SHANKLIN, J. Characterization of a structurally and functionally diverged acyl-acyl carrier protein desaturase from milkweed seed. **Plant Molecular Biology**, v. 33, n. 6, p. 1106-1110, 1997.
- CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, v. 40, n. 3, p. 635-744, 2010.
- CHAPPLE, C. Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 49, p. 311-343, 1998.
- CHEN, X. *et al.* Cloning and characterization of the *WAX2* gene of Arabidopsis involved in cuticle membrane and wax production. **The Plant Cell**, v. 15, n. 5, p. 1170-1185, 2003.
- CHEN, M. *et al.*, Sphingolipid long-chain base hydroxylation is important for growth and regulation of sphingolipid content and composition in *Arabidopsis*. **The Plant Cell**, v. 20, p. 1862-1878, 2008.
- CHILDS, K. L. *et al.* The TIGR Plant Transcript Assemblies database. **Nucleic Acids Research**, v. 35, p. D846-D851, 2007.
- CHOE, S. *et al.* The Arabidopsis *dwf7/ste1* mutant is defective in the $\Delta 7$ sterol C-5 desaturation step leading to brassinosteroid biosynthesis. **The Plant Cell**, v. 11, p. 207-221, 1999.

CHOU, K. C.; CAI, Y. D. Predicting protein quaternary structure by pseudo amino acid composition. **PROTEINS: Structure, Function, and Genetics**, v. 53, n. 2, p. 282-289, 2003.

CLIFTEN, P. *et al.* *SYR2*, a gene necessary for syringomycin growth inhibition of *Saccharomyces cerevisiae*. **Microbiology**, v. 142, n. 3, p. 477-484, 1996.

COHEN, J. Bioinformatics - An introduction for computer scientists. **ACM Computing Surveys**, v. 36, n. 2, p. 122-158, 2004.

CORREIA, M. A.; SINCLAIR, P. R.; MATTEIS, F. D. Cytochrome P450 regulation: The interplay between its heme and apoprotein moieties in synthesis, assembly, repair and disposal. **Drug Metabolism Reviews**, v. 43, n. 1, p. 1-26, 2011.

COURSOL, S. *et al.* *Arabidopsis* sphingosine kinase and the effects of phytosphingosine-1-phosphate on stomatal aperture. **Plant Physiology**, v. 137, n. 2, p. 724-737, 2005.

COURSOL, S. *et al.* Sphingolipid signalling in *Arabidopsis* guard cells involves heterotrimeric G proteins. **Nature**, v. 423, n. 6940, p. 651-654, 2003.

DARNET, S.; BARD, M.; RAHIER, A. Functional identification of sterol-4 α -methyl oxidase cDNAs from *Arabidopsis thaliana* by complementation of a yeast *erg25* mutant lacking sterol-4 α -methyl oxidation. **FEBS Letters**, v. 508, n. 1, p. 39-43, 2001.

DARNET, S.; RAHIER, A. Enzymological properties of sterol-C4-methyl-oxidase of yeast sterol biosynthesis. **Biochimica et Biophysica Acta**, v. 1633, n. 2, p. 106-117, 2003.

DARNET, S.; RAHIER, A. Plant sterol biosynthesis: identification of two distinct families of sterol 4 α -methyl oxidases. **Biochemical Journal**, 378(3):889-898, 2004.

DAUBNER, S. C.; LE, T.; WANG, S. Tyrosine hydroxylase and regulation of dopamine synthesis. **Archives of Biochemistry and Biophysics**, v. 508, n. 1, p. 1-12, 2011.

DESMOND, E; GRIBALDO, S. Phylogenomics of sterol synthesis: Insights into the origin, evolution, and diversity of a key eukaryotic feature. **Genome Biology and Evolution**, v. 1, p. 364-381, 2009.

DEWICK, P. M. **Medicinal Natural Products: A biosynthetic Approach**. 2nd Ed. John Wiley & Sons, Ltd., 550 p., 2002.

DO, C. B.; KATOH, K. Protein multiple sequence alignment. *In*: THOMPSON, J.D.; SCHAEFFER-REISS, C.; UEFFING, M. (Eds.). **Functional Proteomics: Methods and Protocols**. Methods in Molecular Biology, v. 484, p. 379-413, 2008.

DRUMMOND, A. J. *et al.* **Geneious 5.1.7**. 2010. Disponível em: <<http://www.geneious.com>>.

DUNN, T. M. *et al.* Synthesis of IPC-C in *Saccharomyces cerevisiae* requires Scs7p, a protein with a cytochrome b5-like and a hydroxylase domain. **YEAST**, v. 14, n. 4, p. 311-321, 1998.

EDGAR, R. C.; BATZOGLOU, S. Multiple sequence alignment. **Current Opinion in Structural Biology**, v. 16, p. 368-373, 2006.

EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792-1797, 2004.

EIGENBRODE, S. D.; ESPELIE, K. E. Effects of plant epicuticular lipids on insect herbivores. **Annual Review of Entomology**, v. 40, p. 171-194, 1995.

FETZNER, S. Oxygenases without requirement for cofactor or metal ions. **Applied Microbiology and Biotechnology**, v. 60, n. 3, p. 243-257, 2002.

FETZNER, S.; STEINER, R. A. Cofactor-independent oxidases and oxygenases. **Applied Microbiology and Biotechnology**, v. 86, n. 3, p. 791-804, 2010.

FINN, D. *et al.* The Pfam protein families database. **Nucleic Acids Research**, v. 42, n. D1, p. D222-D230, 2014.

FISCHER, M. *et al.* The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. **Bioinformatics**, v. 23, n. 15, p. 2015-2017, 2007.

FITZPATRICK, P. F. Mechanism of aromatic amino acid hydroxylation. **Biochemistry**, v. 9, n. 42, p. 14083-14091, 2003.

FLYDAL, M. I.; MARTINEZ, A. Phenylalanine hydroxylase: Function, structure, and regulation. **International Union of Biochemistry and Molecular Biology Life**, v. 65, n. 4, p. 341-349, 2013.

FRICKEY, T.; LUPAS, A. CLANS: a Java application for visualizing protein families based on pairwise similarity. **Bioinformatics**, v. 20, n. 18, p. 3702-3704, 2004.

FUKUCHI-MIZUTANI, M. *et al.* Characterization of Δ -9 acyl-lipid desaturase homologues from *Arabidopsis thaliana*. **Plant Cell Physiology**, v. 39, n. 2, p. 247-253, 1998.

GACHOTTE, D.; MEENS, R.; BENVENISTE, P. An *Arabidopsis* mutant deficient in sterol biosynthesis: heterologous complementation by *ERG3* encoding a Δ^7 -sterol-C-5-desaturase from yeast. **The Plant Journal**, v. 8, n. 3, p. 407-416, 1995.

GALLAGHER, S. C.; CAMMARK, R.; DALTON, H. Alkene monooxygenase from *Nocardia corallina* B-276 is a member of the class of dinuclear iron proteins capable

of stereospecific epoxygenation reactions. **European Journal of Biochemistry**, v. 247, n. 2, p. 635-641, 1997.

GORGAS, K. *et al.* The ether lipid-deficient mouse: tracking down plasmalogen functions. **Biochimica et Biophysica Acta**, v. 1763, n. 12, p. 1511-1526, 2006.

GUINDON, S.; GASCUEL, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. **Systematic Biology**, v. 52, n. 5, p. 696-704, 2003.

HALL, B. G. **Phylogenetic trees made easy: A how to manual for molecular biologists**. 2nd Ed., Sinauer Associates, Inc., 221 p., 2004.

HAMA, H. Fatty acid 2-hydroxylation in mammalian sphingolipid biology. **Biochimica et Biophysica Acta**, v. 1801, n. 4, p. 405-414, 2010.

HARAYAMA, S.; KOK, M.; NEIDLE, E.L. Functional and evolutionary relationships among diverse oxygenases. **Annual Review of Microbiology**, v. 46, p. 565-601, 1992.

HARRISON, C. J.; LANGDALE, J. A. A step by step guide to phylogeny reconstruction. **The Plant Journal**, v. 45, n. 4, p. 561-572, 2006.

HENIKOFF, S.; HENIKOFF, J. G. Amino acid substitution matrices from protein blocks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, n. 22, p. 10915-10919, 1992.

HENIKOFF, S.; HENIKOFF, J. G. Automated assembly of protein blocks for database searching. **Nucleic Acids Research**, v. 19, n. 23, p. 6565-6572, 1991.

HENIKOFF, S.; HENIKOFF, J. G. Protein family classification based on searching a database of blocks. **Genomics**, v. 19, n. 1, p. 97-107, 1994.

HIGGINS, D. G. Amino acid-based phylogeny and alignment. *In*: BORK, P. (Ed.). **Advances in protein chemistry - Analysis of amino acids sequences**. Vol. 54., Academic Press, p. 99-135, 2000.

HUIJBERS, M. M. E. *et al.* Flavin dependent monooxygenases. **Archives of Biochemistry and Biophysics**, v. 544, p. 2-17, 2014.

HUNTER, S. *et al.* InterPro in 2011: new developments in the family and domain prediction database. **Nucleic Acids Research**, v. 40, n. D1, p. D306-D312, 2011.

JAISER, S. R.; WINSTON, G. P. Copper deficiency myelopathy. **Journal of Neurology**, v. 257, n. 6, p. 869-881, 2010.

JANSSON, A. *et al.* Aclacinomycin 10-hydroxylase is a novel substrate-assisted hydroxylase requiring S-adenosyl-L-methionine as cofactor. **Journal of Biological Chemistry**, v. 280, n. 5, p. 3636-3644, 2005.

JAOUEN, G. **Bioorganometallics: Biomolecules, Labeling, Medicine**. Wiley-VCH Verlag GmbH & Co., 462 p., 2006.

KAGEDAL, B. *et al.* Pterin-dependent tyrosine hydroxylase mRNA is not expressed in human melanocytes or melanoma cells. **Pigment Cell Research**, v. 17, n. 4, p. 346-351, 2004.

KANEHISA, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. **Nucleic Acids Research**, v. 42, p. D199-D205, 2014.

KANEKO, T. *et al.* Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. **DNA Research**, v. 3, p. 109-136, 1996.

KASTIN, A. J. **Handbook of Biologically Active Peptides**. Academic Press, 1640 p., 2006.

KENNEDY, M. A. *et al.* Cloning and sequencing of the *Candida albicans* C-4 sterol methyl oxidase gene (*ERG25*) and expression of an *ERG25* conditional lethal mutation in *Saccharomyces cerevisiae*. **Lipids**, v. 35, n. 3, p. 257-262, 2000.

KNIPPLE, D. C. *et al.* Evolution of the integral membrane desaturase gene family in moths and flies. **Genetics**, v. 162, n. 4, p. 1737-1752, 2002.

KOPEC, K. O.; ALVA, V.; LUPAS, A. N. Bioinformatics of the TULIP domain superfamily. **Biochemical Society Transactions**, v. 39, n. 4, p. 1033-1038, 2011.

KOSUGE, T. *et al.* DDBJ progress report: a new submission system for leading to a correct annotation. **Nucleic Acids Research**, v. 42, n. D1, p. D44-D49, 2013.

KWEON, O. *et al.* A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases. **BMC Biochemistry**, v. 9, n. 11, 2008.

LEE, M. *et al.* Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. **Science**, v. 280, n. 5365, p. 915-918, 1998.

LESK, A. M. **Introdução à Bioinformática**. Porto Alegre, Artmed, 384 p., 2008.

LUND, E. G. *et al.* cDNA cloning of mouse and human cholesterol 25-hydroxylases, polytopic membrane proteins that synthesize a potent oxysterol regulator of lipid metabolism. **The Journal of Biological Chemistry**, v. 273, n. 51, p. 34316-34327, 1998.

MARDIS, E. R. A decade's perspective on DNA sequencing technology. **Nature**, v. 470, p. 198-203, 2011.

MATSUSHIMA, M. *et al.* Molecular cloning and mapping of a human cDNA (SC5DL) encoding a protein homologous to fungal sterol-C5-desaturase. **Cytogenetics and Cell Genetics**, v. 74, n. 4, p. 252-254, 1996.

McWILLIAM, H. *et al.* Analysis toll web services from the EMBL-EBI. *Nucleic Acids Research*, v. 41, p. W587-W600, 2013.

METZLER, D. E. **Biochemistry – The chemical reactions of living cells**. 2nd Ed., Vol. 1 & 2. Elsevier Academic Press, 1900 p., 2003.

MICHAELSON, L. V. *et al.* Functional characterization of a higher plant sphingolipid Δ 4-desaturase: Defining the role of sphingosine and sphingosine-1-phosphate in *Arabidopsis*. **Plant Physiology**, v. 149, p. 487-498, 2009.

MIYAZAKI, Y. *et al.* Cloning, sequencing, expression and allelic sequence diversity of *ERG3* (C-5 sterol desaturase gene) in *Candida albicans*. **Gene**, v. 236, p. 43-51, 1999.

MONGRAND, S. *et al.* Lipid rafts in higher plant cells: Purification and characterization of Triton X-100-insoluble microdomains from tobacco plasma membrane. **The Journal of Biological Chemistry**, v. 279, n. 35, p. 36277-36286, 2004.

MONTELLANO, O.D. **Cytochrome P450: structure, mechanism and biochemistry**. 3rd Ed., New York, Plenum Press, 690 p., 2005.

MOUNT, D.W. **Bioinformatics: sequence and genome analysis**. New York, Cold Spring Harbor Laboratory Press, 560 p., 2001.

MURRELL, J. C.; GILBERT, B.; McDONALD, I. R. Molecular biology and regulation of methane monooxygenase. **Archives of Microbiology**, v. 173, n. 5-6, p. 325-332, 2000.

MUSZEWSKA, A. *et al.* Independent subtilases expansions in fungi associated with animals. **Molecular Biology and Evolution**, v. 28, n. 12, p. 3395-3404, 2011.

NAGANO, M. *et al.* *Arabidopsis* sphingolipid fatty acid 2-hydroxylases (AtFAH1 and AtFAH2) are functionally differentiated in fatty acid 2-hydroxylation and stress responses. **Plant Physiology**, v. 159, n. 3, p. 1138-1148, 2012.

NAGANO, M. *et al.* Functional association of cell death suppressor, *Arabidopsis* Bax inhibitor-1, with fatty acid 2-hydroxylation through cytochrome b5. **The Plant Journal**, v. 58, n. 1, p. 122-134, 2009.

NAKAJIMA, M. Control of xeno/endobiotics- metabolizing cytochrome P450s by microRNAs. **Fifty Years of Cytochrome P450 Research**, Part IV, pp. 327-344, 2014.

NELSON, D. L.; COX, M. M. **Lehninger - Principles of Biochemistry**. 5th Ed. W.H. Freeman and Company, New York, 1100 p. 2008.

NELSON, D. R. *et al.* P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. **Pharmacogenetics**, v. 6, n. 1, p. 1-42, 1996.

NG, C. K. *et al.* Drought-induced guard cell signal transduction involves sphingosine-1-phosphate. **Nature**, v. 410, n. 6828, p. 596-599, 2001.

NOLAN, L. C.; O'CONNOR, K. E. Dioxygenase- and monooxygenase-catalysed synthesis of *cis*-dihydrodiols, catechols, epoxides and other oxygenated products. **Biotechnology Letters**, v. 30, n. 11, p. 1879-1891, 2008.

NUIN, P. A. S.; WANG, Z.; TILLIER, E. R. M. The accuracy of several multiple sequence alignment programs for proteins. **BMC Bioinformatics**, v. 7, n. 471, 2006.

OKULEY, J. *et al.* Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. **The Plant Cell**, v. 6, p. 146-158, 1994.

PAGANI, I. *et al.* The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. **Nucleic Acids Research**, v. 40, p. D571-D579, 2012.

PAIS, F. S. M. *et al.* Assessing the efficiency of multiple sequence alignment programs. **Algorithms for Molecular Biology**, v. 9, n. 4, p. 1-8, 2014.

PONTING, C. P. *et al.* Evolution of domain families. *In*: BORK, P. (Ed.). **Advances in protein chemistry - Analysis of amino acids sequences**. Vol. 54., Academic Press, p. 185-244, 2000.

PROSDOCIMI, F. *et al.* Bioinformática: manual do usuário. **Biotecnologia Ciência e Desenvolvimento**, v. 29, p. 12-25, 2002.

RAHIER, A. Dissecting the sterol C-4 demethylation process in higher plants. From structures and genes to catalytic mechanism. **Steroids**, v. 76, n. 4, p. 340-352, 2011.

RAHIER, A. *et al.* The sterol C4-demethylation in higher plants. *In*: BACH, T.J.; ROHMER, M. (Eds.). **Isoprenoid Synthesis in Plants and Microorganisms**. Springer, p. 367-380, 2013.

RAHIER, A.; SMITH, M.; TATON, M. The role of cytochrome b5 in 4 α -methyl-oxidation and C5(6) desaturation of plant sterol precursors. **Biochemical and Biophysical Research Communications**, v. 236, n. 2, p. 434-437, 1997.

REDDY, A. S. *et al.* Isolation of a Δ^6 -desaturase gene from the cyanobacterium *Synechocystis* sp. strain PCC 6803 by gain-of-function expression in *Anabaena* sp. strain PCC 7120. **Plant Molecular Biology**, v. 22, n. 2, p. :293-300, 1993.

RUEPP, A. *et al.* The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. **Nucleic Acids Research**, v. 32, n. 18, p. 5539-5545, 2004.

RUSSELL, D. W. Oxysterol biosynthetic enzymes. **Biochimica et Biophysica Acta**, v. 1529, n. 1-3, p. 126-135, 2000.

RYLE, M. J.; HAUSINGER, R. P. Non-heme iron oxygenases. **Current Opinion in Chemical Biology**, v. 6, n. 2, p. 193-201, 2002.

SAYANOVA, O. *et al.* Expression of a borage desaturase cDNA containing an N-terminal cytochrome b_5 domain results in the accumulation of high levels of Δ^6 -desaturated fatty acids in transgenic tobacco. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 8, p. 4211-4216, 1997.

SCHNEIDER, H. **Métodos de análise filogenética: um guia prático**. Holos Editora e Sociedade Brasileira de Genética, 115 p., 2003.

SHANKLIN, J. *et al.* Desaturases: Emerging models for understanding functional diversification of diiron-containing enzymes. **The Journal of Biological Chemistry**, v. 284, n. 28, p. 18559-18563, 2009.

SHANKLIN, J. *et al.* Mössbauer studies of alkane ω -hydroxylase: Evidence for a diiron cluster in an integral-membrane enzyme. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 7, p. 2981-2986, 1997.

SHANKLIN, J.; CAHOON, E. B. Desaturation and related modifications of fatty acids. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 49, p. 611-41, 1998.

SHANKLIN, J.; WHITLLE, E. Evidence linking the *Pseudomonas oleovorans* alkane ω -hydroxylase, an integral membrane diiron enzyme, and the fatty acid desaturases family. **FEBS Letters**, v. 545, n. 2-3, p. 188-192, 2003.

SHU, L. *et al.* An $\text{Fe}_2^{\text{IV}}\text{O}_2$ diamond core structure for the key intermediate Q of methane monooxygenase. **Science**, v. 275, n. 5299, p. 515-518, 1997.

SIGEL, A.; SIGEL, H.; SIGEL, R. K. O. **The Ubiquitous Roles of Cytochrome P450 Proteins: Metal Ions in Life Sciences**. Vol 3. John Wiley & Sons, Ltd., 678 p., 2007.

SIGRIST, C. J. A. *et al.* PROSITE: A documented database using patterns and profiles as motif descriptors. **Briefings in Bioinformatics**, v. 3, n. 3, p. 265-274, 2002.

SPERLING, P. *et al.* Are glucocerebrosides the predominant sphingolipids in plant plasma membranes? **Plant Physiology and Biochemistry**, v. 43, n. 12, p. 1031-1038, 2005.

SPERLING, P. *et al.* The evolution of desaturases. **Prostaglandins, Leukotrienes, and Essential Fatty Acids**, v. 68, n. 2, p. 73-95, 2003.

SPERLING, P.; *et al.* Functional characterization of sphingolipid C4-hydroxylase genes from *Arabidopsis thaliana*. **FEBS Letters**, v. 494, n. 1-2, p. 90-94, 2001.

SUN, Z. R.; GANTT, E.; CUNNINGHAM Jr, F.X. Cloning and functional analysis of the β -carotene hydroxylase of *Arabidopsis thaliana*. **The Journal of Biological Chemistry**, v. 271, n. 40, p. 24349-24352, 1996.

TAN, C. P. *et al.* Expression of β -carotene hydroxylase gene (crtR-B) from the green alga *Haematococcus pluvialis* in chloroplasts of *Chlamydomonas reinhardtii*. **Journal of Applied Phycology**, v. 19, n. 4, p. 347-355, 2007.

TATUSOV, R. L.; ALTSCHUL, S. F.; KOONIN, E. V. Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 25, p. 12091-12095, 1994.

THE UniProt CONSORTIUM. Activities at the Universal Protein Resource (UniProt). **Nucleic Acids Research**, v. 42, p. D191-D198, 2014.

TOMAZIC, M. L. *et al.* A novel sterol desaturases-like protein promoting dealkylation of phytosterols in *Tetrahymena thermophila*. **Eukariotic Cell**, v. 10, n. 3, p. 423-434, 2011.

TORRES PAZMIÑO, D. E. *et al.* Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. **Journal of Biotechnology**, v. 146, n. 1-2, p. 9-24, 2010.

ULLRICH, R.; HOFRICHTER, M. Enzymatic hydroxylation of aromatic compounds. **Cellular and Molecular Life Sciences**, v. 64, n. 3, p. 271-293, 2007.

van BEILEN, J. B.; FUNHOFF, E. G. Alkane hydroxylases involved in microbial alkane degradation. **Applied Microbiology and Biotechnology**, v. 74, n. 1, p. 13-21, 2007.

van BERKEL, W. J. H.; KAMERBEEK, N. M.; FRAAIJE, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. **Journal of Biotechnology**, v. 124, n. 4, p. 670-689, 2006.

VEITIA, R. A.; HURST, L. D. Accelerated molecular evolution of insect orthologues of Erg28/C14Orf1: A link with ecdysteroid metabolism? **Journal of Genetics**, v. 80, n. 1, p. 17-21, 2001.

VERPOORTE, R.; MEMELINK, J. Engineering secondary metabolite production in plants. **Plant Biotechnology**, v. 13, n. 2, p. 181-187, 2002.

VINCI, G.; XIA, X.; VEITIA, R. A. Preservation of genes involved in sterol metabolism in cholesterol auxotrophs: Facts and Hypotheses. **PLoS ONE**, v. 3, n. 8, p. e2883, 2008.

VOLNER, A.; ZOIDAKIS, J.; ABU-OMAR, M. M. Order of substrate binding in bacterial phenylalanine hydroxylase and its mechanistic implication for pterin-dependent oxygenases. **Journal of Biological Inorganic Chemistry**, v. 8, n. 1-2, p. 121-128, 2003.

WACKETT, L. P. Mechanism and applications of Rieske non-heme iron dioxygenases. **Enzyme and Microbial Technology**, v. 31, n. 5, p. 577-587, 2002.

WALTHER, D. J.; BADER, M. A unique central tryptophan hydroxylase isoform. **Biochemical Pharmacology**, v. 66, n. 9, p. 1673-1680, 2003.

WATSCHINGER, K. *et al.* Identification of the gene encoding alkylglycerol monooxygenase defines a third class of tetrahydrobiopterin-dependent enzymes. **PNAS**, v. 107, n. 31, p. 13672-13677, 2010.

WATSCHINGER, K.; WERNER, E. R. Alkylglycerol monooxygenase. **IUBMB Life**, 65(4):366-372, 2013.

WERCK-REICHHART, D.; BAK, S.; PAQUETTE, S. Cytochromes P450. **The Arabidopsis Book**, v. 1, p. e0028, 2002.

WUTHRICH, K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. **Science**, v. 243, n. 4887, p. 45-50, 1989.

XIN, X.; MAINS, R. E.; EIPPER, B. A. Monooxygenase X, a member of the copper-dependent monooxygenase family localized to the endoplasmic reticulum. **The Journal of Biological Chemistry**, v. 279, n. 46, p. 48159-48167, 2004.

YAMAMOTO, S. The 50th anniversary of the discovery of oxygenases. **Life**, v. 58, n. 5-6, p. 248-250, 2006.

ANEXOS

ANEXO A - Organismos selecionados para busca de sequências de proteínas oxigenases com ferro não hêmico em bancos de dados

Táxon	Organismo
Bactérias	<i>Anaeromyxobacter dehalogenans</i> (strain 2CP-1 / ATCC BAA-258)
	<i>Bacillus thuringiensis</i> IBL 200
	<i>Blastopirellula marina</i> DSM 3645
	<i>Bradyrhizobium</i> sp. (strain ORS278)
	<i>Candidatus Solibacter usitatus</i> (strain Ellin6076)
	<i>Escherichia coli</i> O150:H5 (strain SE15)
	<i>Frankia alni</i> (strain ACN14a)
	<i>Gluconobacter oxydans</i> (strain 621H) (<i>Gluconobacter suboxydans</i>)
	<i>Methylococcus capsulatus</i> (strain ATCC 33009 / NCIMB 11132 / Bath)
	<i>Mycobacterium bovis</i> (strain BCG / Pasteur 1173P2)
	<i>Nitrosococcus watsoni</i> (strain C-113)
	<i>Nitrosomonas</i> sp. AI212
	<i>Nostoc azollae</i> (strain 0708) (<i>Anabaena azollae</i> / <i>Trichormus azollae</i>)
	<i>Plesiocystis pacifica</i> SIR-1
	<i>Ralstonia solanacearum</i> (strain GMI1000) (<i>Pseudomonas solanacearum</i>)
	<i>Rhodospirillum centenum</i> (strain ATCC 51521 / SW)
	<i>Stigmatella aurantiaca</i> (strain DW4/3-1)
	<i>Streptomyces albus</i> J1074
	<i>Synechocystis</i> sp. (strain PCC 6803 / Kazusa)
	<i>Thermosynechococcus elongatus</i> (strain BP-1)
Protozoários	<i>Dictyostelium discoideum</i>
	<i>Leishmania braziliensis</i>
	<i>Monosiga brevicollis</i>
	<i>Naegleria gruberi</i>
	<i>Tetrahymena thermophila</i>
	<i>Trypanosoma cruzi</i> (strain CL Brener)
Fungos	<i>Aspergillus clavatus</i> (strain ATCC 1007 / CBS 513.65 / DSM 816 / NCTC 3887 / NRRL 1)
	<i>Aspergillus oryzae</i> (strain ATCC 42149 / RIB 40)
	<i>Candida albicans</i> (strain SC5314 / ATCC MYA-2876)
	<i>Cryptococcus neoformans</i> var. <i>neoformans</i> serotype D (strain JEC21 / ATCC MYA-565)
	<i>Nectria haematococca</i> (strain 77-13-4 / ATCC MYA-4622 / FGSC 9596 / MPVI)
	<i>Neurospora crassa</i> (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987)
	<i>Penicillium chrysogenum</i> (strain ATCC 28089 / DSM 1075 / Wisconsin 54-1255)
	<i>Penicillium digitatum</i> (strain Pd1 / CECT 20795)
	<i>Saccharomyces cerevisiae</i> (strain ATCC 204508 / S288c)
	<i>Schizosaccharomyces japonicus</i> (strain yFS275 / Fy16936)

	<i>Ustilago maydis</i> (strain 521 / FGSC 9021)
Metazoários	<i>Aedes aegypti</i>
	<i>Caenorhabditis elegans</i>
	<i>Danio rerio</i>
	<i>Daphnia pulex</i>
	<i>Homo sapiens</i>
	<i>Tribolium castaneum</i>
	<i>Trichoplax adhaerens</i>
Diatomácea	<i>Thalassiosira pseudonana</i>
Oomicetos	<i>Aphanomyces euteiches</i> *
	<i>Phytophthora infestans</i> (strain T30-4)
Algas verdes	<i>Chlamydomonas reinhardtii</i>
	<i>Micromonas pusilla</i> (strain CCMP1545)
	<i>Ostreococcus tauri</i>
Plantas	<i>Arabidopsis cebenensis</i> *
	<i>Arabidopsis thaliana</i>
	<i>Carica papaya</i>
	<i>Medicago truncatula</i>
	<i>Oryza sativa</i> subsp. Japonica
	<i>Physcomitrella patens</i>
	<i>Populus trichocarpa</i>
	<i>Sorghum bicolor</i>

* Não possui genoma inteiramente sequenciado.

ANEXO B – Proteínas caracterizadas na literatura contidas na coleção de sequências

Espécie	Proteína	Referências
<i>Arabidopsis thaliana</i>	Metilesterol monooxigenase 1-1	DARNET; RAHIER, 2004
	Metilesterol monooxigenase 1-2	DARNET; BARD; RAHIER, 2001
	Metilesterol monooxigenase 1-3	GACHOTTE; MEENS; BENVENISTE, 1995
	Metilesterol monooxigenase 2-1	NAGANO <i>et al.</i> , 2012
	Metilesterol monooxigenase 2-2	NAGANO <i>et al.</i> , 2009
	Δ 7-esterol-C5(6)-desaturase 1	SPERLING <i>et al.</i> , 2001
	Ácido graxo 2-hidroxilase 1	CHEN <i>et al.</i> , 2008
	Ácido graxo 2-hidroxilase 2	AARTS <i>et al.</i> , 1995
	Esfinganina C(4)-monooxigenase 1	CHEN <i>et al.</i> , 2003
	Esfinganina C(4)-monooxigenase 2	SUN; GANTT; CUNNINGHAM, 1996
	Proteína eceriferum 1	
	Proteína eceriferum 3	
	β -caroteno 3-hidroxilase 1	
	β -caroteno 3-hidroxilase 2	
	<i>Candida albicans</i>	C-5 esterol desaturase
<i>Homo sapiens</i>	Colesterol 25-hidroxilase	LUND <i>et al.</i> , 1998
	Latosterol oxidase	MATSUSHIMA <i>et al.</i> , 1996
<i>Saccharomyces cerevisiae</i>	Ácido graxo 2-hidroxilase	ALDERSON <i>et al.</i> , 2004 ALDERSON <i>et al.</i> , 2006
	Alquilglicerol monooxigenase	WATSCHINGER <i>et al.</i> , 2010
	Metilesterol monooxigenase	BARD <i>et al.</i> , 1996
	C-5 esterol desaturase	ARTHINGTON <i>et al.</i> , 1991
	Ceramida hidroxilase	DUNN <i>et al.</i> , 1998
<i>Synechocystis</i> sp.	Esfingolípídeo C(4)-hidroxilase	CLIFTEN <i>et al.</i> , 1996
	Δ 12 ácido graxo desaturase	KANEKO <i>et al.</i> , 1996
	Linoleoil-CoA desaturase	REDDY <i>et al.</i> , 1993

ANEXO C – Classificação das oxigenases com ferro não hêmico

Grupo	Nome do grupo	Sigla do grupo	Proteína principal	N
1	C4 esteroil monooxigenase	C4SMO	Metilesterol monooxigenase	99
2	C5 esteroil desaturase	C5SD	C-5 esteroil desaturase	45
3	Ácido graxo hidroxilase	FAH	Ácido graxo 2-hidroxilase	36
4	Alquilglicerol monooxigenase	ALKMO	Alquilglicerol monooxigenase	34
5	Esfingolípido hidroxilase	SUR	Esfingolípido C4-hidroxilase	33
6	Aldeído decarbonilase	CER	Aldeído decarbonilase (eceriferum)	31
7	Δ 9 ácido graxo desaturase	D9FAD	Δ 9 ácido graxo desaturase	26
8	β -caroteno 3-hidroxilase	BCH	β -caroteno 3-hidroxilase	17
9	Grupo A	Grupo A*	[Não caracterizada]	16
10	Δ 12 ácido graxo desaturase	D12FAD	Δ 12 ácido graxo desaturase	12
11	Δ 4 ácido graxo desaturase	D4FAD	Esfingolípido Δ 4 desaturase	5
12	Δ 9 ácido graxo desaturase	D9FAD**	Ácido graxo desaturase	4
13	Ácido graxo desaturase / dehidrogenase-redutase de cadeia curta	FAH/SDR**	[proteína predita]	4
14	C5 esteroil desaturase	C5SD**	Esterol desaturase	4
15	C5 esteroil desaturase	C5SD	C-5 esteroil desaturase	4
16	Esterol desaturase	SDES	Esterol desaturase	4
17	C5 esteroil desaturase	C5SD	C-5 esteroil desaturase	3
18	Acil-ACP desaturase	AACPD	Acil-ACP desaturase	3
19	Acil-ACP desaturase	AACPD	Acil-ACP desaturase	2
20	Ácido graxo hidroxilase	FAH	Ácido graxo hidroxilase	2
21	Ácido graxo desaturase / dehidrogenase-redutase de cadeia curta	FAH/SDR**	Ácido graxo hidroxilase	2
22	Ácido graxo hidroxilase	FAH	Ácido graxo hidroxilase	2
23	Δ 9 ácido graxo desaturase	D6FAD	Δ 9 ácido graxo desaturase	2
24	Grupo B*	Grupo B*	[proteína predita]	2
25	Ácido graxo desaturase / Citocromo-b5	FAD/Cyt-b5**	[proteína predita]	2
26	Linoleoil-CoA desaturase	LCOAD	Linoleoil-CoA desaturase	2
27	Linoleoil-CoA desaturase	LCOAD**	[proteína predita]	2
28	Linoleoil-CoA desaturase	LCOAD	Linoleoil-CoA desaturase	2

* Grupo que possui apenas proteínas preditas.

** Grupo nomeado por similaridade com domínios encontrados na análise de BLAST.

N = Número de sequências atribuídas ao grupo.

ANEXO D – Distribuição de seqüências das proteínas oxigenases com ferro não hêmico nos grupos formados pela ferramenta CLANS

Espécie	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	NG
<i>M. bovis</i>																		X									X	X	
<i>F. alni</i>			X				X												X				X				X	X	
<i>S. albus</i>				X			X												X				X						
<i>N. azollae</i>							X		X												X								X
<i>Synechocystis</i> sp.	X		X				X		X													X				X			
<i>T. elongatus</i>	X						X																						
<i>S. usitatus</i>				X			X																						X
<i>B. thuringiensis</i>			X																										
<i>B. marina</i>																													X
<i>Bradyrhizobium</i> sp.																													X
<i>G. oxydans</i>			X																				X						
<i>R. centenum</i>				X																									
<i>R. solanacearum</i>				X										X															
<i>Nitrosomonas</i> sp.				X								X																	X
<i>A. dehalogenans</i>							X													X						X			
<i>S. aurantiaca</i>		X	X	X																									
<i>P. pacifica</i>				X																									X
<i>N. watsoni</i>				X								X																	
<i>E. coli</i>				X																									
<i>M. capsulatus</i>				X								X																	X
<i>T. thermophila</i>		X					X							X															
<i>D. discoideum</i>	X		X	X																									
<i>L. braziliensis</i>		X								X	X				X												X		X
<i>T. cruzi</i>		X								X																			
<i>N. gruberi</i>	X	X		X													X												
<i>M. brevicollis</i>	X	X	X	X																				X					X

<i>M. truncatula</i>	X	X	X X	X
<i>P. trichocarpa</i>	X X X	X X	X	
<i>O. sativa</i>	X X X	X X	X	
<i>S. bicolor</i>	X X X	X X	X	

X

NG: Nenhum grupo