

# ANÁLISE FILOGENÉTICA SISTEMATIZADA DE SEQUÊNCIAS GÊNICAS DE COMPOSTOS ANTIOXIDANTES ENCONTRADOS EM ESPÉCIES VEGETAIS

Ana Carolina Preto Malaman, Antonio Fluminhan Jr.

Universidade do Oeste Paulista – UNOESTE, Curso de Ciências Biológicas, Presidente Prudente, SP. e-mail: [fluminhan@unoeste.br](mailto:fluminhan@unoeste.br)

## RESUMO

Esta pesquisa teve como objetivo analisar o grau de semelhança encontrado em sequências de genes para os compostos antioxidantes enzimáticos: superóxido dismutase e peroxidase, bem como sequências de genes para compostos não enzimáticos: ácido ascórbico e licopeno, encontrados em diferentes espécies de plantas. Sequências de nucleotídeos disponibilizadas em bancos de dados genômicos: NCBI (EUA), EBI (Europa) e DDBJ (Japão) foram identificadas e analisadas com as ferramentas BLAST, ClustalW2, ClustalW2 Phylogeny e MAFFT, para avaliar o grau de semelhança entre si e para produzir dendrogramas específicos para cada gene. Os resultados mostraram que as sequências de genes para licopeno e superóxido dismutase mostraram maior consistência quando comparados com a classificação taxonômica consensual, do que as sequências para os genes de peroxidase e ácido ascórbico. Esta pesquisa pode contribuir para estudos avançados das relações filogenéticas entre as espécies de plantas, bem como para uma compreensão da evolução molecular dos genes avaliados.

**Palavras-chave:** bioinformática, antioxidantes enzimáticos, antioxidantes não-enzimáticos, evolução molecular, relações filogenéticas.

## SYSTEMATIZED PHYLOGENETIC ANALYSIS OF GENE SEQUENCES FOR ANTIOXIDANT COMPOUNDS FOUND IN PLANT SPECIES

### ABSTRACT

This research aimed to analyze the degree of similarity found in gene sequences for enzymatic antioxidants superoxide dismutase and peroxidase, as well as gene sequences for non-enzymatic compounds: ascorbic acid and lycopene, found in different plant species. Nucleotide sequences available in genomic databases: NCBI (USA) EBI (European) and DDBJ (Japan) were identified and analyzed with the BLAST, ClustalW2, ClustalW2 Phylogeny and MAFFT tools, in order to evaluate the degree of similarity between them and to produce dendrograms specific for each gene. The results showed that the gene sequences for lycopene and superoxide dismutase showed greater consistency when compared with the consensus taxonomic classification than the gene sequences for peroxidase and ascorbic acid. This research can contribute to advanced study of phylogenetic relationships among plant species, as well as to a better understanding of the molecular evolution of the evaluated genes.

**Keywords:** bioinformatics, enzymatic antioxidants, non-enzymatic antioxidants, molecular evolution, phylogenetic relationships.

## INTRODUÇÃO

Estudos sobre o envelhecimento celular têm revelado que ele decorre, em grande parte, de uma deterioração progressiva do organismo em resposta adaptativa a fatores ambientais. Além das inúmeras alterações histológicas causadas por esses fatores, eles também contribuem para o aumento de radicais livres no organismo, frequentemente associados ao envelhecimento precoce, por serem os responsáveis por diversas agressões físicas às células do corpo e tecidos. Segundo a teoria de Denham Harman<sup>1</sup>, os radicais livres estão envolvidos no processo de envelhecimento, por provocarem danos celulares que são acumulados durante a vida, resultando em aceleração de disfunções.

Os antioxidantes agem impedindo a formação de moléculas de radicais livres, evitando a formação de lesões em nossas células causadas pelos radicais, removendo os danos e reconstituindo as células danificadas<sup>2</sup>. Os estudos sobre os antioxidantes têm ressaltado, principalmente, o uso de nutrientes isolados no tratamento e prevenção de doenças<sup>3</sup>. Entretanto, nos alimentos é encontrada uma grande variedade de substâncias que podem agir em sinergismo na proteção das células e tecidos. Conforme estes últimos autores, alguns exemplos de substâncias

antioxidantes intensivamente estudadas envolvem: compostos enzimáticos considerados essenciais ao processo evolutivo, tais como: o superóxido dismutase (SOD), cuja importância pode ser demonstrada pelo fato de ser a enzima mais abundante no organismo e pode ser dividida em duas SODs principais, CuZnSOD contendo cobre e zinco e a MnSOD, contendo manganês na mesma molécula, e a peroxidase, também abundante no organismo, bem como compostos não-enzimáticos e que não são essenciais a esse processo evolutivo, encontradas em um menor número de espécies, tais como: o licopeno e o ácido ascórbico.

Por sua vez, o crescimento vertiginoso no número de sequências de ácidos nucléicos, proteínas, estruturas de moléculas e expressão e regulação gênica nos bancos de dados genômicos em vários países tornaram as pesquisas muito mais precisas. Este fato tem estimulado um crescente interesse no estudo de similaridade de sequências com o emprego de ferramentas de bioinformática, que são programas de computadores que auxiliam na realização de análises comparativas de sequências. A obtenção e a análise de informações genômicas exigem programas com dispositivos de armazenamento rápidos e de alta capacidade para manter os bancos de

dados. Além disto, a disponibilidade destas ferramentas de bioinformática no ambiente web tem permitido uma ampla utilização de recursos em tempo real, e a realização de pesquisas de forma intensiva. Assim, o computador tornou-se peça chave nas pesquisas que vem sendo realizadas em áreas especializadas como a Biologia Molecular e o armazenamento de dados biológicos em bancos de dados públicos tem se tornado cada vez mais comum.

O objetivo deste estudo foi analisar as sequências gênicas para as substâncias antioxidantes: ácido ascórbico, licopeno, peroxidase e superóxido dismutase, de forma a permitir a avaliação da semelhança entre esses genes encontrados em espécies vegetais pertencentes a diferentes táxons. As sequências gênicas destes compostos estão disponíveis nas bases NCBI (Estados Unidos), EBI (Europa) e DDBJ (Japão), e foram analisadas de modo a verificar o grau de similaridade das sequências de nucleotídeos para cada um dos respectivos genes através da construção de cladogramas representativos do grau de similaridade genômica entre as espécies e, individualmente, para cada um dos genes avaliados. Finalmente, os resultados obtidos foram confrontados com a relação filogenética consensual entre as espécies, do

ponto de vista taxonômico, individualmente para cada um dos genes.

## **METODOLOGIA**

Para a realização desta pesquisa utilizou-se um computador portátil com processador Intel Core i7, 8Gb de memória, HD de 1Tb com placa de vídeo 2GB, Monitor de 15.6" e configurado com permissão de administrador, impressora jato de tinta colorida e acesso irrestrito a internet.

As sequências de nucleotídeos que foram coletadas estão disponíveis nos principais bancos de dados genômicos de acesso públicos e gratuitos do mundo que são: NCBI – National Center for Biotechnology Information (EUA)<sup>4,5</sup>, EMBL-EBI - The European Bioinformatics Institute (Europa)<sup>6</sup>, e o DDBJ (DNA Data Bank of Japan)<sup>7</sup> e mantido pelo governo do Japão. Ressalta-se a existência de uma colaboração internacional entre estes bancos de dados, conhecida como INSDC - International Nucleotide Sequence Database Collaboration, cujo link é: <http://www.insdc.org/> onde se pode acessar os demais bancos de dados genômicos.

Sequências de nucleotídeos dos genes responsáveis pela síntese de ácido ascórbico, licopeno, peroxidase e superóxido dismutase, entre diferentes espécies de plantas foram obtidas a partir dos bancos de dados

genômicos e comparadas entre si, aos pares, através de duas ferramentas de alinhamento múltiplo: ClustalW2 e MAFFT. A ferramenta ClustalW2, uma nova versão da conhecida ferramenta ClustalW, é um programa de alinhamento de múltiplas sequências de uso geral para DNA ou proteínas que calcula a melhor correspondência para as sucessões escolhidas e as alinha para que as identidades, semelhanças e diferenças possam ser observadas. A ferramenta MAFFT, (Multiple Alignment using Fast Fourier Transform), é um programa de alinhamento de sequências múltiplas de alta velocidade, reconhecido por utilizar algoritmos considerados rápidos e com níveis de precisão significativos<sup>8</sup>. Ambas ferramentas estão disponíveis para acesso no site da EBI <http://www.ebi.ac.uk/Tools/msa/>. Os resultados obtidos foram utilizados para a construção de cladogramas para cada um dos genes utilizando-se a ferramenta Clustal W Phylogeny, ferramenta comumente utilizada para geração árvores filogenéticas fornecidas pela ClustalW2, e que também está disponível no mesmo portal.

As ferramentas de bioinformática escolhidas funcionam em ambiente web e dispensam, deste modo, a instalação dos softwares na estação de trabalho. Os conjuntos de sequências encontrados no NCBI, EBI e DDBJ foram submetidos à análise

através das ferramentas mencionadas, de modo a permitir a sua avaliação quanto à aplicabilidade, facilidade de estudo e viabilidade na análise de cada uma delas. Estas ferramentas foram empregadas de modo a identificar as regiões de cada gene que apresentam maior variabilidade entre as espécies avaliadas, bem como identificar padrões de evolução molecular dos genes nas diferentes espécies.

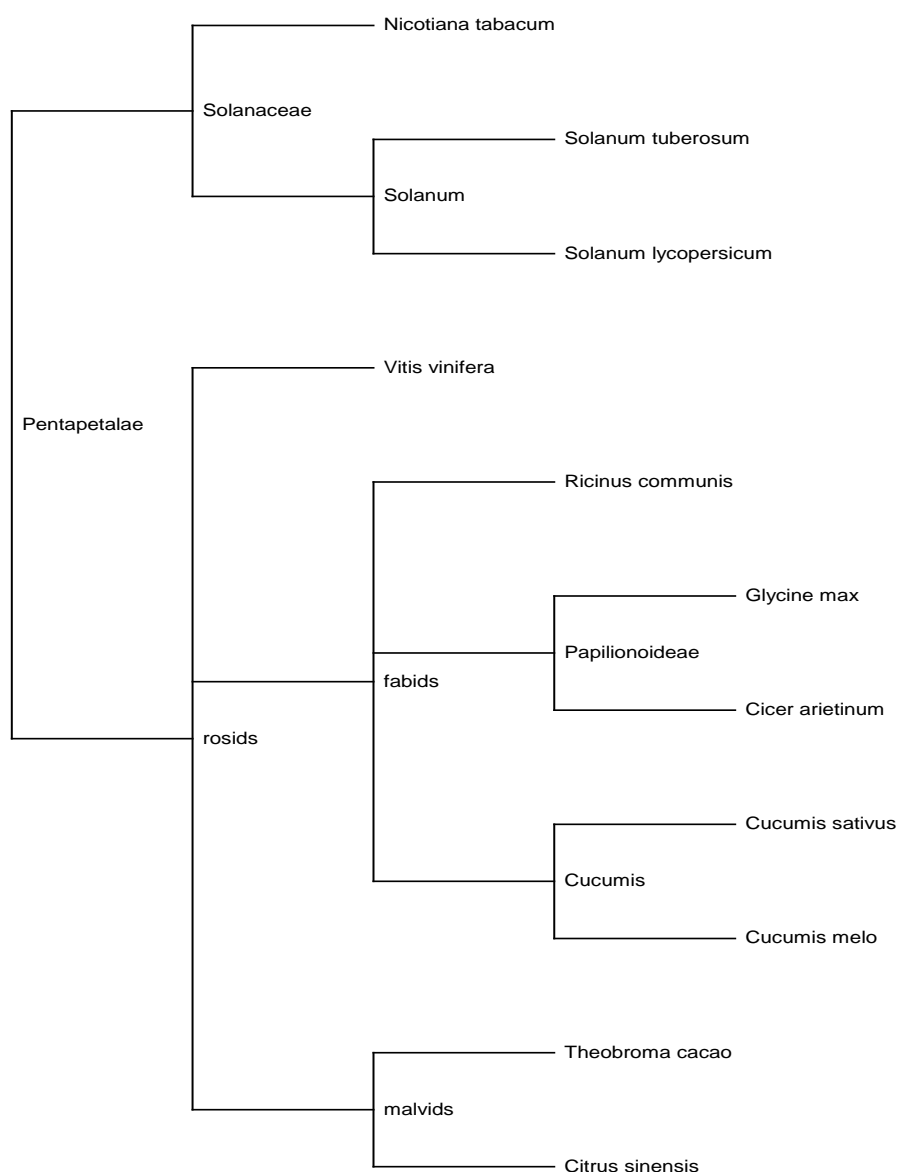
## RESULTADOS

As sequências nucleotídicas relativas aos genes para os quatro compostos antioxidantes analisados foram obtidas a partir de pesquisas nas bases de dados genômicos internacionais. Para os compostos antioxidantes não-enzimáticos, foram identificados mais de uma sequência gênica codificante para cada substância. No caso do antioxidante ácido ascórbico, foram encontradas sequências relativas ao ascorbato peroxidase e ao ascorbato oxidase, enquanto que, para o antioxidante licopeno, foram encontradas sequências nucleotídicas do licopeno beta-ciclase e do licopeno épsilon-ciclase. Para os antioxidantes enzimáticos foram obtidas variações do mesmo gene apenas para o superóxido dismutase, cujas principais versões foram: CuZnSOD e MnSOD. Foram utilizadas apenas o CuZnSOD para o estudo, devido à maior

abundância de dados disponíveis desse gene em diferentes espécies vegetais do que para a variante MnSOD. Finalmente, não foram identificadas variações para o gene da enzima peroxidase.

Para cada uma das sequências avaliadas foram realizadas buscas de espécies vegetais em que as mesmas já tinham sido

descritas, resultando em um total de onze espécies, as quais foram selecionadas utilizando-se como critério a presença dos 6 genes estudados. Na Figura 1 pode-se observar a classificação taxonômica consensual atual destas onze espécies, estabelecida com base em suas características morfológicas.

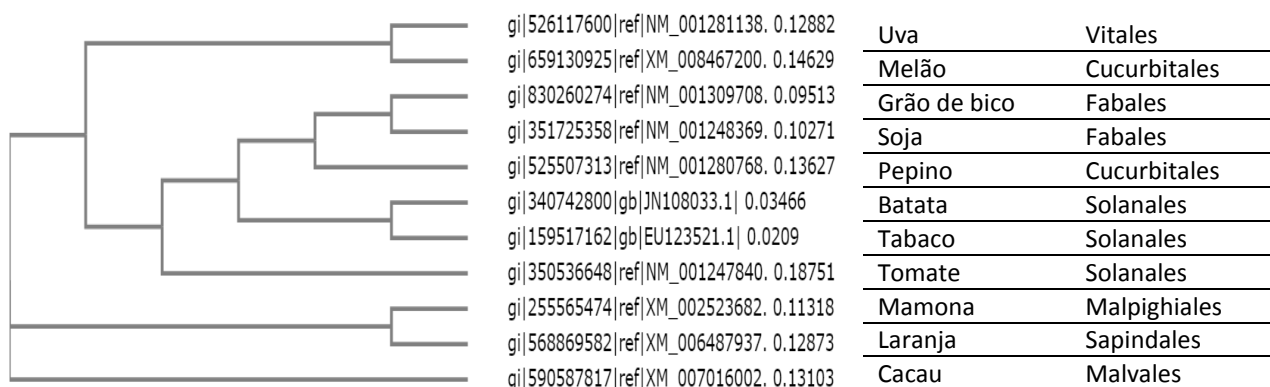


1

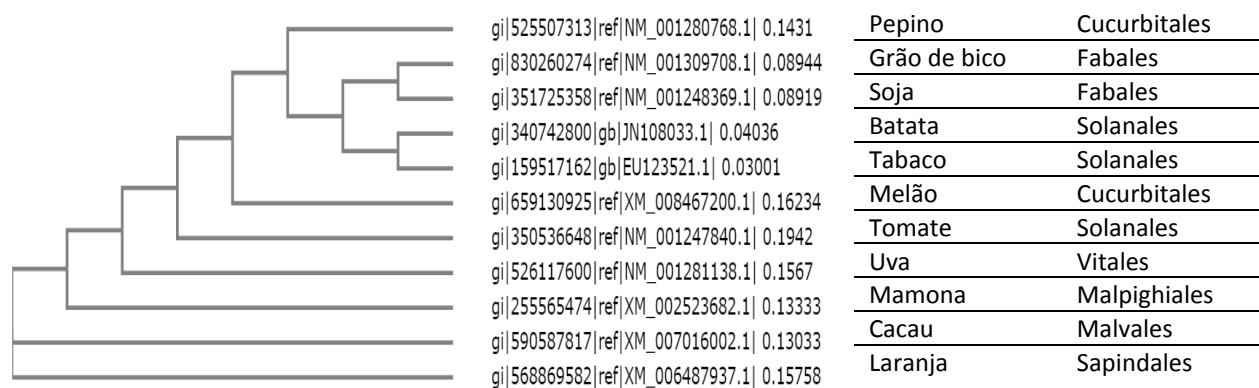
**Figura 1.** Classificação taxonômica atual obtida através do Taxonomy Browse, disponível em: <http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>, acessada em: 05/08/2015, e gerada através do software TreeView 1.6.6

Nas Figuras 2 e 3 pode-se observar os dendrogramas obtidos a partir dos alinhamentos das sequências referentes ao gene do composto antioxidante enzimático

superóxido dismutase Cu-Zn, e que foram gerados pelas ferramentas ClustalW2 e MAFFT, respectivamente.



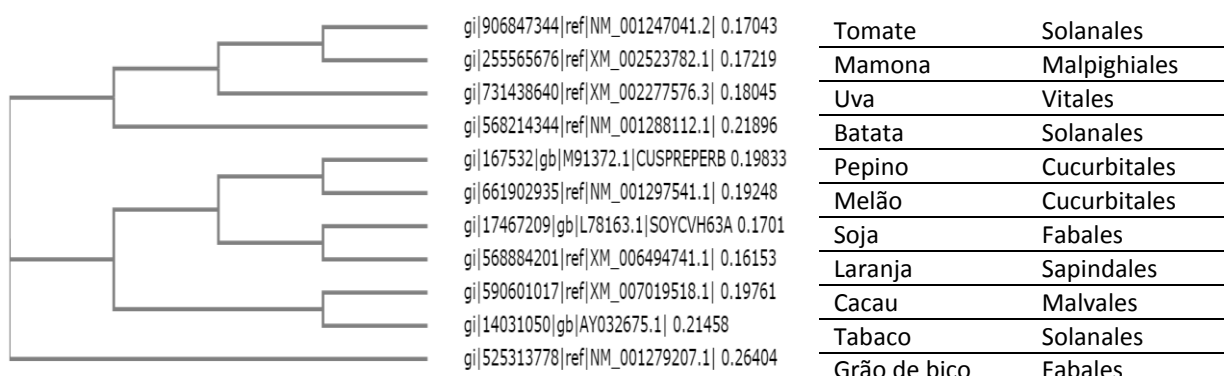
**Figura 2.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes ao superóxido dismutase Cu-Zn obtido através do ClustalW2. Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.



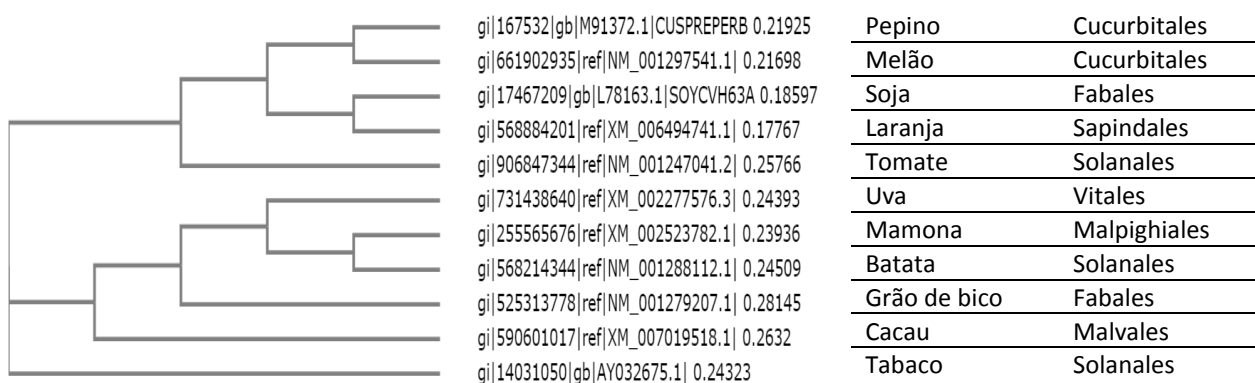
**Figura 3.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes ao superóxido dismutase Cu-Zn obtido através do MAFFT. Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.

Nas Figuras 4 e 5 são apresentados os dendrogramas obtidos a partir dos alinhamentos das sequências para o gene do composto antioxidante enzimático

peroxidase, e que foram gerados pelas ferramentas ClustalW2 e MAFFT, respectivamente.



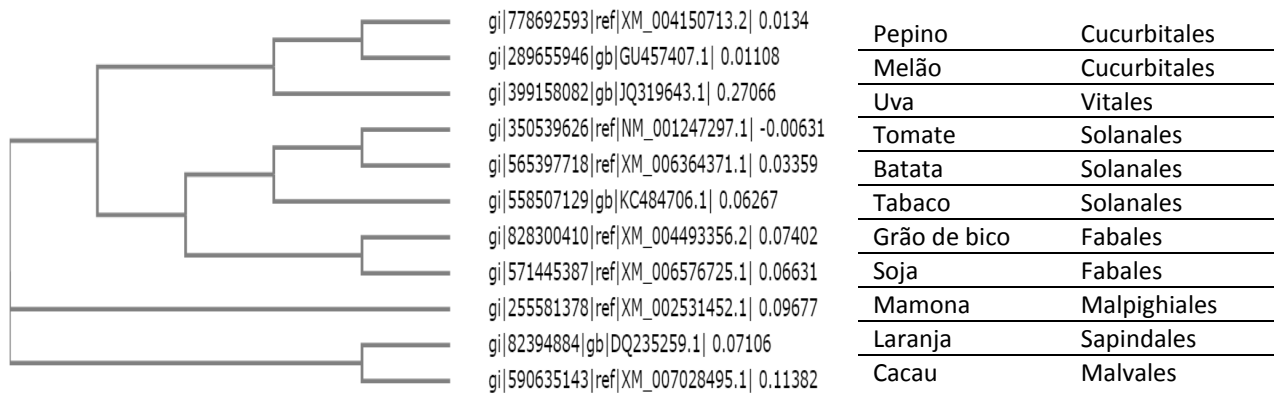
**Figura 4.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes à peroxidase obtido através do ClustalW2. Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.



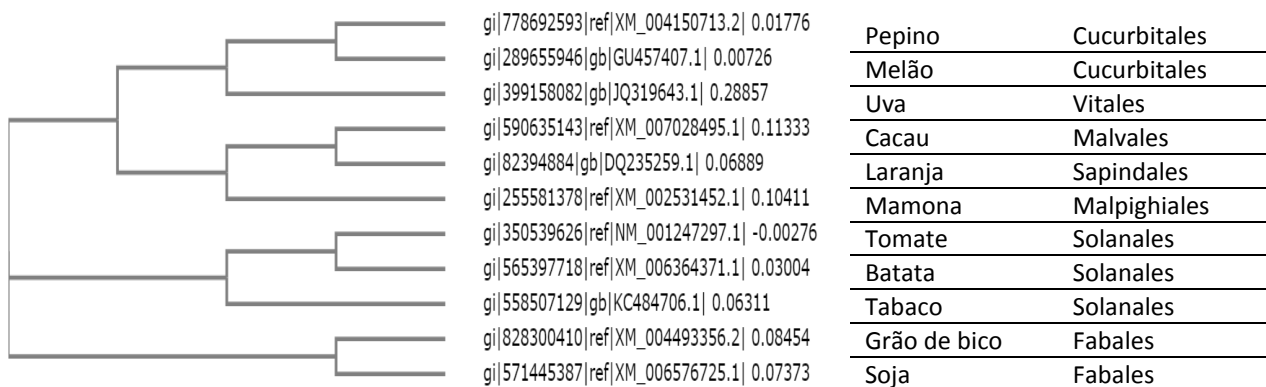
**Figura 5.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes à peroxidase obtido através do MAFFT. Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.

Nas Figuras 6 e 7 são apresentados os dendrogramas obtidos a partir do alinhamento das sequências de nucleotídeos para os genes licopeno beta-ciclase e, nas Figuras 8 e 9, são apresentados os dendrogramas gerados para as sequências de licopeno épsilon-ciclase. Para cada

substância, os alinhamentos foram gerados pelas ferramentas ClustalW2 e MAFFT.



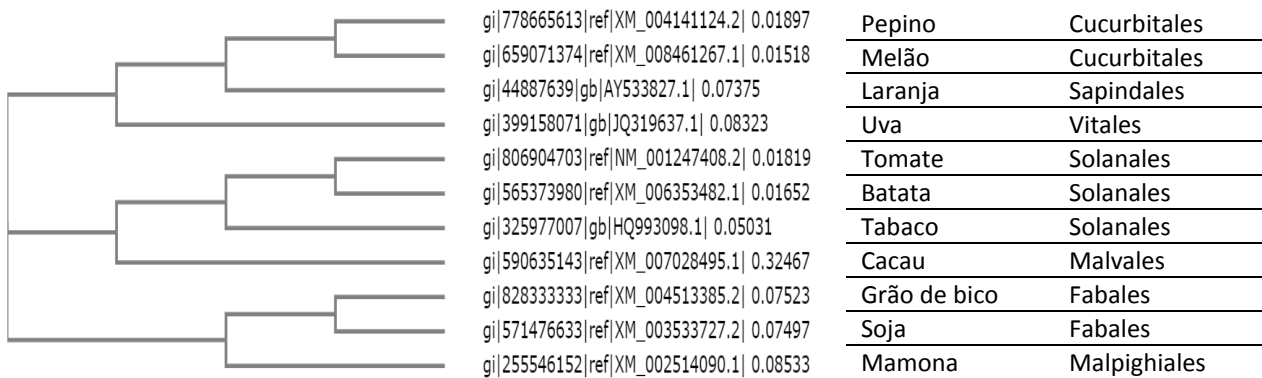
**Figura 6.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes ao composto licopeno beta-ciclase obtido através do ClustalW2. Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.



**Figura 7.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes ao licopeno beta-ciclase obtido através do MAFFT.

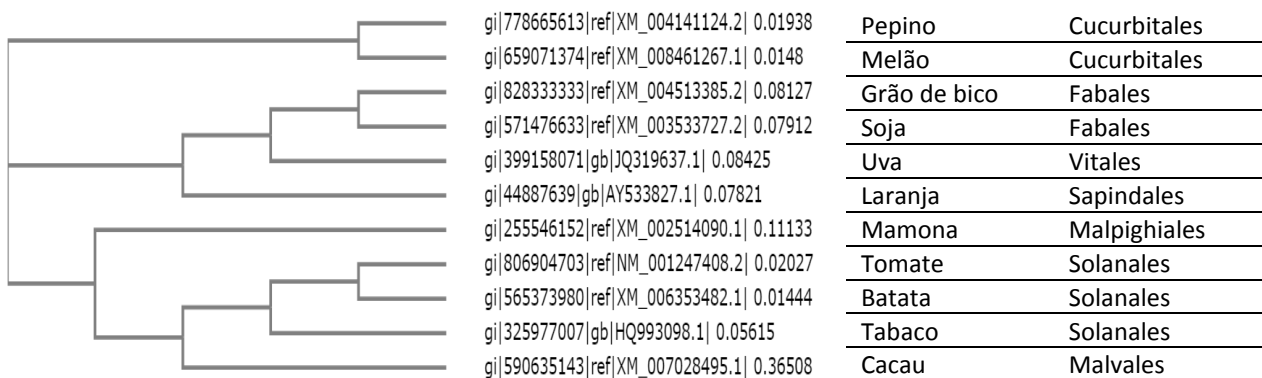
Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.





**Figura 8.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes ao licopeno epsilon-ciclase obtido através do ClustalW2.

Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.

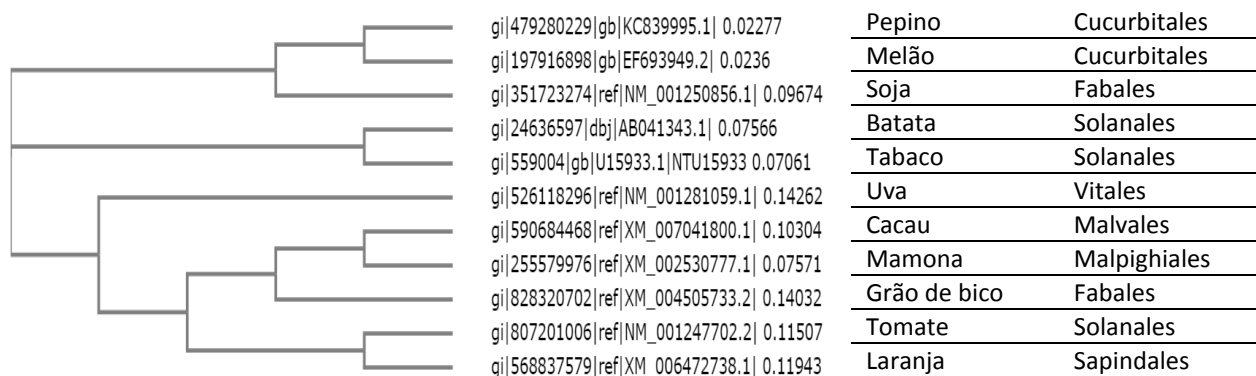


**Figura 9.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny, representando o alinhamento das sequências referentes ao licopeno epsilon-ciclase obtido através do MAFFT. Fonte:

Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.

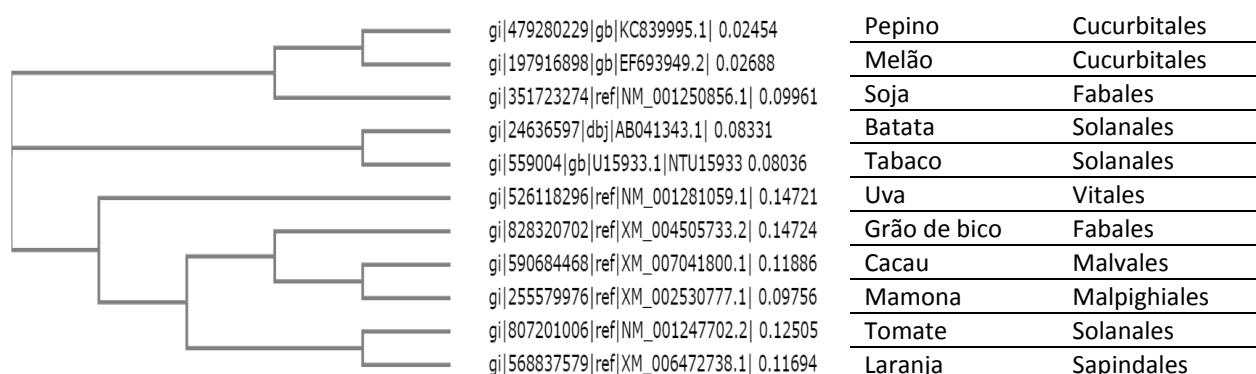
Finalmente, nas Figuras 10 e 11 são apresentados os dendrogramas obtidos a partir dos alinhamentos das sequências referentes ao composto antioxidante ascorbato peroxidase e, nas Figuras 12 e 13

os alinhamentos obtidos para ascorbato oxidase. Novamente, para cada substância, os alinhamentos foram gerados pelas ferramentas ClustalW2 e MAFFT.



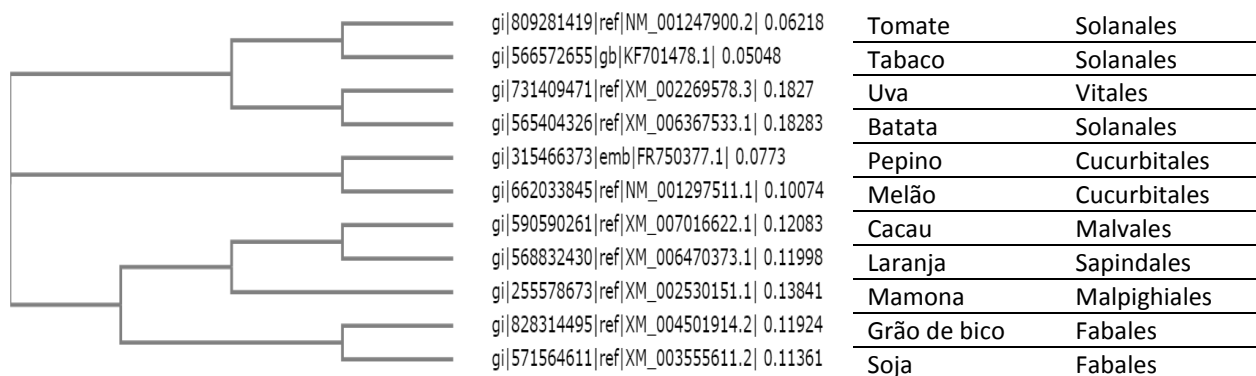
**Figura 10.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny representando o alinhamento das sequências referentes ao ascorbato peroxidase obtido através do ClustalW2.

Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.

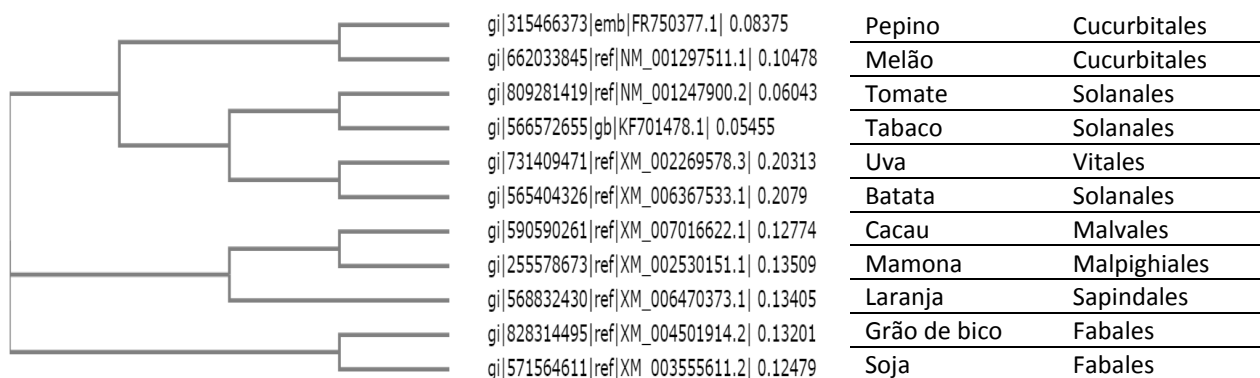


**Figura 11.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny representando o alinhamento das sequências referentes as ascorbato peroxidase obtido através do MAFFT. Fonte:

Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.



**Figura 12.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny representando o alinhamento das sequências referentes ao ascorbato oxidase obtido através do ClustalW2. Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> acessado em: 05/08/2015.



**Figura 13.** Dendrograma gerado a partir da ferramenta ClustalW Phylogeny representando o alinhamento das sequências referentes ao ascorbato oxidase obtido através do MAFFT. Fonte: Clustal W Phylogeny - <http://www.ebi.ac.uk/Tools/phylogeny/> Acesso: 15/08/2015.

## DISCUSSÃO

Os resultados obtidos mostraram que as sequências do gene superóxido dismutase apresentaram-se predominantemente coerentes com a classificação taxonômica atualmente existente para as espécies avaliadas nesta pesquisa. Pode-se notar que apenas para as espécies de Curcubitaceas mostraram algum grau de divergência, fato

este que pode ter ocorrido devido a uma das sequências de nucleotídeos empregadas ter sido 'predita' pelo banco de dados genômico, e não referente aos dados reais sequenciado da enzima. Resultados semelhantes ocorreram para as sequências do ascorbato peroxidase e para o ascorbato oxidase, que mostraram uma forte coerência, porém, com

algumas poucas divergências grosseiras e pontuais.

Entre as sequências gênicas analisadas, as que refletiram mais fielmente a evolução das espécies vegetais envolvidas foram as referentes ao licopeno beta-ciclase e ao licopeno épsilon-ciclase. Conforme mencionado na literatura<sup>9</sup>, as sequências moleculares estão tornando-se cada vez mais importantes para a inferência filogenética e têm propriedades diferentes, sejam elas oriundas de sequências de proteínas ou, diretamente, a partir do DNA.

De uma forma contraditória, as sequências referentes ao gene da peroxidase foram as que apresentaram maior discrepância quando comparadas à classificação taxonômica das espécies. Isso pode ter ocorrido devido à maioria das sequências obtidas serem predições a partir de proteínas e, apesar de estarem disponíveis nos bancos de dados genômicos, ainda aguardam a confirmação através do sequenciamento dos nucleotídeos das sequências naturais.

Embora fosse esperado encontrar uma maior correspondência entre a classificação filogenética e a taxonômica das espécies vegetais avaliadas, no que se refere aos compostos enzimáticos, o que pode-se observar foi que essa teoria não se confirmou. Enquanto um dos genes

enzimático apresentou bastante fidelidade com a evolução taxonômica das espécies (superóxido dismutase), o outro apresentou grande discrepância (peroxidase). Já para os genes não-enzimáticos as correspondências encontradas entre estas classificações foram, de forma geral, maiores do que aquelas observadas para os genes de compostos enzimáticos.

## CONCLUSÃO

Foram analisadas as sequências gênicas para as substâncias antioxidantes, de modo a permitir a avaliação da semelhança entre esses genes encontrados em diferentes espécies de plantas. Foi possível realizar a análise do grau de similaridade das sequências de nucleotídeos para cada um dos respectivos genes, com especial atenção às variações existentes entre espécies e sua correspondência em relação à classificação taxonômica consensual. Finalmente, os resultados obtidos foram confrontados com a classificação filogenética consensual das espécies, individualmente para cada um dos genes, indicando um grande potencial para o emprego sistemático desta abordagem para os estudos de evolução molecular.

## AGRADECIMENTOS

Os autores agradecem o apoio recebido do Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq), Ministério da Ciência, Tecnologia e Inovação (MCTI), através do Processo 117703/2014-1, referente ao Projeto 800098/2014-5, modalidade IC.

#### **CONFLITO DE INTERESSE**

Os autores declaram não haver qualquer potencial conflito de interesse que possa interferir na imparcialidade deste trabalho científico.

**REFERÊNCIAS**

1. Sies H. Strategies of antioxidant defense. Review. *Europ J Biochem.* 1993; 215(2):213-9. DOI: <http://dx.doi.org/10.1007/97836427875779>
2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Biologia molecular da célula.* 5.ed. Porto Alegre: Artes Médicas, 2010.
3. Kimura ET, Baía GS. Rede ONSA e o Projeto Genoma Humano do Câncer: Contribuição ao Genoma Humano. *Arq Bras Endocrinol Metab.* 2002;46(4):325-9. DOI: <http://dx.doi.org/10.1590/S0004-27302002000400003>.
4. Karlin S, Altschul SF. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *PNAS USA.* 1990;87(6):2264-8.
5. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. *GenBank. Nucleic acids research.* Oxford Univ Press. 2009;37(1):26-31. DOI: <http://dx.doi.org/10.1093%2Fnar%2Fgkn723>.
6. Ma B, Tromp J, Li M. PatterHunter: faster and more sensitive homology search. *Bioinformatics.* Oxford Univ Press. 2002;28(3):440-5.
7. Sansom CE, Smith CA. Computer applications in biomolecular sciences. Part 2: bioinformatics and genome projects. *Biochem Educ.* 2000;28:127-31.
8. Kitajima JP. *Bioinformática: manual de utilização.* *ComCiência Rev Eletr J Cient SBPC.* 2003;46. Acesso: 8 ago 2015. Disponível em: [www.comciencia.br/reportagens/bioinformatica/bio12.shtml](http://www.comciencia.br/reportagens/bioinformatica/bio12.shtml)
9. Ridley M. *Evolução.* 3.ed. Porto Alegre: Artmed; 2006.

Recebido para publicação em 19/08/2015

Revisado em 26/08/2015

Aceito em 31/08/2015