

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
DEPARTAMENTO DE GENÉTICA

Evolução Molecular e aspectos estruturais de proteínas relacionadas à biossíntese de
compostos orgânicos voláteis em Solanaceae

Lucas Cafferati Beltrame

Orientação: Loreta Brandão de Freitas

Coorientação: Claudia Elizabeth Thompson

Porto Alegre

Julho/2017

Lucas Cafferati Beltrame

Evolução Molecular e aspectos estruturais de proteínas relacionadas à biossíntese de
compostos orgânicos voláteis em Solanaceae

Trabalho de Conclusão de Curso
apresentado como um dos requisitos para
a obtenção do título de Bacharel em
Biotecnologia da UFRGS, Ênfase
Biotecnologia Molecular.

Orientação: Profa. Dra. Loreta Brandão de Freitas (UFRGS)

Coorientação: Profa. Dra. Claudia Elizabeth Thompson (UFCSPA)

Porto Alegre

Julho/2017

AGRADECIMENTOS

Aos meus pais, por me apoiarem, me aconselharem e me ajudarem com qualquer coisa que eu precisasse; Por serem pessoas de grande confiança e que se mostraram sempre dispostas a me ajudar com o que eu precisasse.

Aos meus irmãos, por compartilharem suas experiências comigo e não hesitarem em me ajudar com as minhas decisões. Pela parceria e, acima de tudo, pela amizade.

À professora Loreta Brandão de Freitas pela orientação para conclusão deste trabalho e pelo auxílio na minha formação como profissional.

À minha coorientadora, Claudia Elizabeth Thompson, por se mostrar sempre disposta e paciente; Por todos os ensinamentos a mim passados para que fosse possível a conclusão deste trabalho.

À Viviane por todo o apoio, paciência e aconselhamento; Por me ajudar nos momentos difíceis e me entender como ninguém; Por estar comigo em todas as situações, boas ou ruins, e me acompanhar de perto em relação à produção deste trabalho.

À Mariana por compartilhar histórias e me auxiliar em momentos de grande necessidade; Por enxergar em mim meus potenciais e me criticar construtivamente.

Aos meus primos, tios e avós por representarem uma grande parcela da minha criação e por apoiarem minhas escolhas.

Aos meus amigos de Porto Alegre por me acolherem nesta cidade, me acompanharem e entrarem na minha vida em um momento totalmente novo e diferente para mim.

Aos meus amigos de Florianópolis por terem feito parte da minha vida e construírem parte do que sou hoje; Por se manterem ao meu lado e junto a mim apesar da distância e o contato reduzido.

À Cuíca, Bissinha, Charlotte, Cuca, Schimier e Toulouse por me proporcionarem momentos maravilhosos de afeto, carinho e compaixão.

RESUMO

A polinização é um processo fundamental para o aumento da diversidade genética entre espécies vegetais e a atração de agentes polinizadores é influenciada por diversos fatores como características morfológicas e fisiológicas florais, em especial a emissão de aromas. Esse trabalho é focado no estudo de duas proteínas envolvidas em rotas metabólicas de produção de compostos orgânicos voláteis em plantas, CFAT (*conyferyl alcohol acyltransferase*) e BPBT (*benzoyl-CoA:benzyl alcohol/phenylethanol benzoyltransferase*). Em especial, buscamos compreender aspectos evolutivos e estruturais dessas proteínas na família Solanaceae, entender sua evolução e suas funções, utilizando métodos de análise filogenética e modelagem molecular comparativa. O método probabilístico de Inferência Bayesiana foi utilizado para realizarmos a análise evolutiva de proteínas CFAT e de BPBT disponíveis em bancos de dados públicos. Sequências de *Petunia x hybrida* foram utilizadas como ponto de partida para a identificação de sequências homólogas em espécies da família Solanaceae. Os modelos da estrutura tridimensional de CFAT e BPBT foram obtidos através do método de modelagem molecular comparativa, utilizando a estrutura 3D da transferase de *Panicum virgatum* como molde. A árvore filogenética obtida inclui diversas enzimas contendo o domínio transferase e que desempenham diversos papéis biológicos dentro da família Solanaceae. Foi encontrada apenas uma sequência de CFAT disponível nos bancos de dados, pertencente à *Petunia x hybrida*, além de diversas sequências de BPBT pertencentes a espécies dos gêneros *Nicotiana*, *Solanum* e *Capsicum*. Os modelos teóricos da estrutura tridimensional de CFAT e BPBT foram obtidos com base na estrutura tridimensional do molde PvHCT2a (PDB ID: 5FAL A). As estruturas 3D de CFAT e BPBT são organizadas em

dois diferentes domínios e identificamos os resíduos de aminoácidos importantes para sua atividade proteica e interação com os respectivos substratos.

ABSTRACT

Specialization to different pollinators is an important process responsible for increasing the genetic diversity in plants. The morphological and physiological floral traits, mainly floral scent, can influence the pollination dynamics and select pollinator agents for each plant species. This work studied evolutionary and structural aspects of two proteins involved in metabolic pathways producing volatile organic compounds in plants, CFAT (*conyferyl alcohol acyltransferase*) e BPBT (*benzoyl-CoA:benzyl alcohol/phenylethanol benzoyltransferase*). The main objective was to better understand the evolutionary and structural aspects of these proteins in Solanaceae family, as well as their functions, using phylogenetic methods and the comparative molecular modeling technique. We have used the probabilistic method of Bayesian Inference to infer their evolutionary history using sequences from *Petunia x hybrida* as a query in public databanks to search for homologs in Solanaceae family. Theoretical three-dimensional models were obtained for CFAT and BPBT proteins using as template the 3D structure of a transferase from *Petunia virgatum* (PDB ID: 5FAL, chain A). The phylogenetic tree included several different enzymes with diverse biological roles in Solanaceae, all containing the transferase domain. We identified only one sequence of CFAT in the databanks, which belongs to *Petunia x hybrida*. Moreover, we found several BPBT sequences from *Nicotiana*, *Solanum* and *Capsicum* genera. The three-dimensional structures of CFAT and BPBT have two different domains and we have identified the amino acid residues important for the enzymatic activity and interaction with substrates.

SUMÁRIO

RESUMO.....	4
ABSTRACT.....	6
INTRODUÇÃO GERAL.....	8
OBJETIVOS.....	18
CAPÍTULO 1.....	19
Abstract.....	21
Introduction.....	22
Materials and Methods.....	24
Results and Discussion.....	26
Acknowledgements.....	33
References.....	33
Figures.....	39
Tables.....	44
DISCUSSÃO GERAL.....	47
REFERÊNCIAS BIBLIOGRÁFICAS.....	54

INTRODUÇÃO GERAL

A evolução molecular é uma área da Biologia que estuda as mudanças evolutivas que ocorrem nas sequências de DNA e de proteínas. Entre outros aspectos, inclui o estudo das taxas de substituição nucleotídica e aminoacídica, da atuação da seleção adaptativa, negativa ou neutra sobre genes, bem como mudanças na estrutura dos genomas de diferentes espécies. Existem duas abordagens gerais para o estudo da evolução molecular: (i) usar as sequências nucleotídicas ou de proteínas, através da escolha de diferentes marcadores moleculares, para o estudo da evolução dos organismos; e (ii) usar sequências biológicas para entender a evolução dos diferentes genes ou famílias multigênicas. O estudo da evolução molecular, nesse último caso, requer a inferência de padrões evolutivos a partir da variação encontrada nos próprios genes, em uma tentativa de descrever sua "história natural". Para que possamos compreender a história evolutiva de genes e genomas, é necessário que sejam utilizados métodos de análise filogenética e/ou filogenômica. Adicionalmente, para que possamos compreender o impacto das taxas de substituição aminoacídicas nas proteínas de interesse é necessário o estudo da estrutura dessas proteínas.

A Biologia Estrutural é a área do conhecimento que tem como objetivo o estudo da estrutura de macromoléculas biológicas, permitindo o estabelecimento da relação entre a estrutura e a função de determinada proteína. Essa área envolve aspectos físico-químicos, bioquímicos e biofísicos para compreensão dos processos de dobramento proteico (*protein folding*) e compreensão da função biológica da proteína. Os dois principais métodos experimentais para a determinação da estrutura tridimensional de uma proteína são a cristalografia por difração de raios-X e a ressonância magnética nuclear (RMN).

Para que a cristalografia por difração de raios-X seja utilizada é necessário que se obtenha uma quantidade adequada de uma proteína altamente purificada, com mais de 95% de pureza, em geral. Posteriormente, é necessário que essa proteína forme um cristal que possa ser exposto aos raios-X e, subsequentemente, seu padrão de difração possa ser capturado, decodificado e resulte em informações a respeito da estrutura molecular proteica. Infelizmente, uma das principais limitações no uso desse método é o fato de que uma fração muito pequena das proteínas forma cristal (Verli 2014; pp. 252-282). Como uma alternativa, caso a proteína possua um tamanho máximo de 6 kDa, para o uso de técnicas bidimensionais, ou de 40 kDa, para técnicas de três ou mais dimensões, pode ser utilizado o método de ressonância magnética nuclear. Trata-se de um método mais caro e que somente pode ser aplicado a proteínas relativamente pequenas (Verli 2014; pp. 237-250). Cerca de 97% das estruturas proteicas depositadas no *Protein Data Bank* (<https://www.rcsb.org/pdb/home/home.do>; Berman *et al.* 2000) foram obtidas por um dos métodos experimentais acima descritos. No entanto, quando existe dificuldades no uso de técnicas de determinação experimental da estrutura de uma proteína, podem ser utilizados métodos teóricos, sendo que o principal deles é a modelagem molecular comparativa.

A modelagem molecular comparativa consiste na construção de modelos tridimensionais de proteínas a partir do uso de uma estrutura tridimensional como molde. É o método mais apurado e com melhores resultados de predição teórica da estrutura tridimensional de proteínas. Para que se possa utilizar uma estrutura proteica como molde, ela deve ter sido obtida a partir de uma metodologia experimental e apresentar, pelo menos, 25% de identidade no alinhamento entre a sequência do molde e a sequência da proteína a ser estudada estruturalmente. Isso significa dizer que se é detectada similaridade entre sequências, pode-se supor similaridade estrutural (Sánchez & Sali 1997). Importante ressaltar

que para o desenvolvimento de um bom modelo teórico estrutural, devem ser buscadas sequências de referência que já foram estudadas estruturalmente de forma experimental e avaliados alguns parâmetros, tais como a resolução da estrutura, o método utilizado para gerar a estrutura e a função da proteína. Posteriormente, avaliam-se a cobertura e a identidade entre o molde e a sequência alvo através do alinhamento das duas sequências. Deve-se, então, gerar o modelo da proteína de interesse a partir da comparação com a sequência e estrutura terciária do molde. Esse modelo é obtido pela otimização de uma função de densidade de probabilidade molecular (PDF; *Probability Density Function*). O programa MODELLER (Webb & Sali 2016), um dos mais utilizados para modelagem molecular comparativa, implementa uma abordagem automatizada para modelar estruturas 3D satisfazendo as restrições espaciais empiricamente obtidas de um banco de dados de alinhamentos de estruturas de proteínas. Dessa forma, as restrições angulares e de ligação são obtidas pela análise estatística das relações entre várias estruturas homólogas, levando a funções de densidade de probabilidade condicionais que são diretamente utilizadas como restrições espaciais (Sali & Blundell 1993; Sánchez & Sali 1997; Martí-Renom 2000).

Uma importante etapa do processo de modelagem molecular comparativa é a validação do modelo, incluindo a análise de diversos parâmetros estereoquímicos e energéticos que se referem, principalmente, à estabilidade da estrutura. Características como comprimento de ligações, ângulos das ligações, planaridade das cadeias, quiralidade e ângulos de torção das cadeias principais e laterais são comumente analisados (Morris *et al.* 1992). Finalmente, após todas essas etapas, tem-se o modelo estrutural final da proteína em estudo (Figura 1). Esse método possibilita a obtenção de um modelo teórico de determinada proteína de interesse e permite com que consigamos compreender o impacto das substituições

aminoacídicas na estrutura e função proteica. Isso é de enorme relevância para a compreensão de aspectos evolutivos de genes e famílias multigênicas.

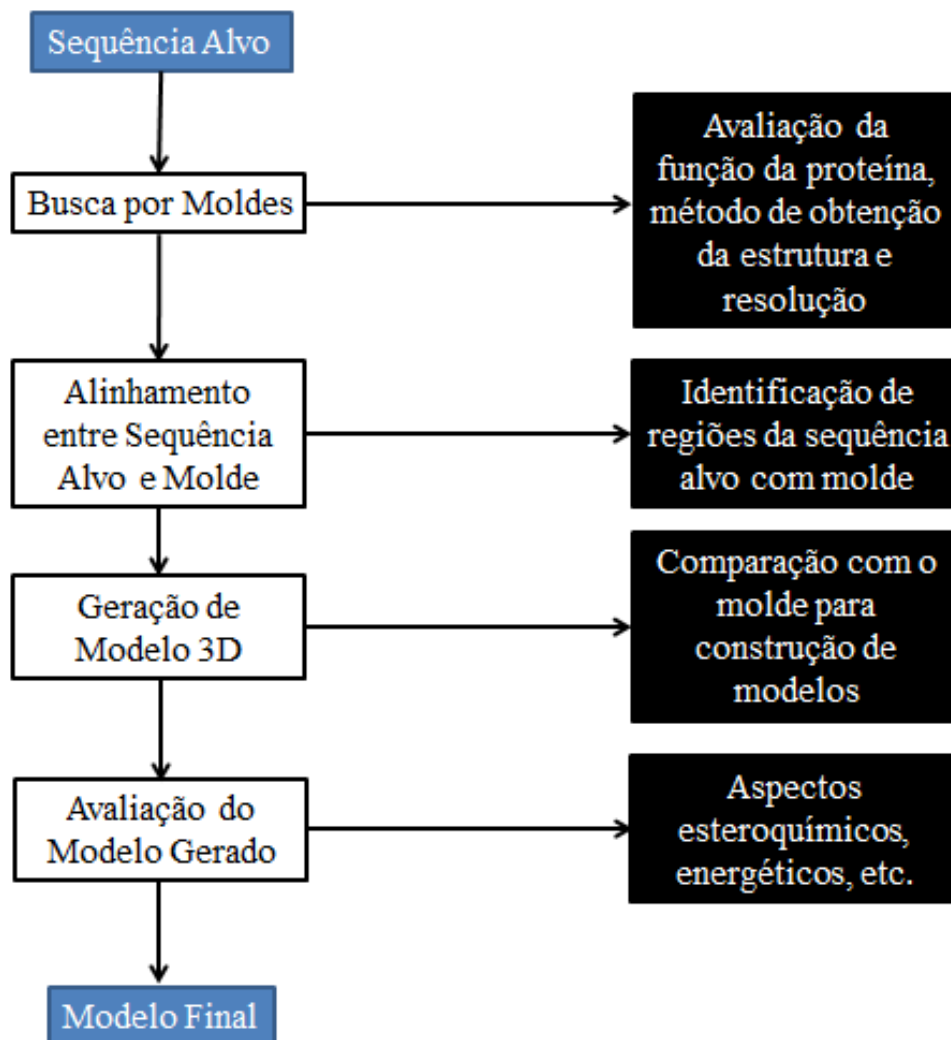


Figura 1. Representação esquemática do processo de modelagem molecular comparativa para obtenção de um modelo estrutural tridimensional. (Adaptado de Verli *et al.* 2004).

Famílias multigênicas são caracterizadas tanto como o conjunto de genes evolutivamente relacionados compartilhados por diferentes espécies, às vezes com funções similares, quanto como o conjunto de homólogos que ocorrem em uma mesma espécie (Martinez 2011). Comumente, esses genes codificando produtos com função idêntica ou

similar pertencem a uma mesma rota metabólica. A origem das famílias multigênicas pode ser devida à duplicação de um segmento cromossômico ou à duplicação do genoma todo, enquanto que o destino dos genes duplicados pode ser o de permanecer no genoma por determinado período, perder sua função ou ser eliminado (Ohno 1970). A maioria dos genes no genoma das plantas pertence a famílias gênicas que contem três ou mais genes, produtos destas duplicações (Ye *et al.* 2013).

A diversidade genética é resultante das interações entre os seres vivos e o ambiente. O ato de polinização realizado entre plantas e seus respectivos agentes polinizadores são determinantes para a evolução das espécies envolvidas. Inicialmente, os pesquisadores Kölreuter (1761) e Sprengel (1793) descreveram a importância e os papéis de cada participante dessa interação. Além disso, apresentaram as características fenotípicas responsáveis por esse estreito relacionamento, caracterizando as síndromes florais, que incluem características estruturais e químicas das flores que seriam atrativas aos diferentes agentes polinizadores, como formato, coloração das pétalas, presença de néctar e sua constituição e diferentes compostos aromáticos. Diversos pontos e conceitos podem ser abordados dentro dos estudos sobre polinização: estudos genéticos, ecológicos, do comportamento animal, da evolução floral, da biologia do desenvolvimento e, até mesmo, sobre comunidades ecológicas (Mitchell *et al.* 2009). No início dos anos 1980, Jones & Little (1983) e Real (1983) publicaram seus trabalhos sobre as interações envolvidas na polinização. As pesquisas nesse campo de estudo, desde então, cresceram exponencialmente, levando pesquisadores a se interessarem pela biologia da polinização e todos os aspectos que podem ser abordados sobre esse tema, como a ecologia funcional dos traços florais, a dinâmica do transporte de pólen, estudos de comunidades ecológicas de polinização, entre outros.

Hermann & Kuhlemeier (2011) discutiram sobre as consequências da polinização de forma ecológica e evolutiva através da comparação de duas espécies do gênero *Petunia* (Solanaceae) que apresentam síndromes florais contrastantes. Uma vez que as espécies analisadas são evolutivamente muito relacionadas, as diferenças observadas pelos autores foram atribuídas as diferentes morfologias e características florais e implicadas no processo de especiação no gênero. A ocorrência de polinização mediada por animais auxilia a evitar grandes perdas de pólen e aumenta consideravelmente a especificação da transferência de informação genética, limitando a ocorrência de eventos de fecundação interespecífica. Com isso, exige, também, que a planta se adapte e elabore suas flores para melhor atrair o agente polinizador específico (Hermann & Kuhlemeier 2011). Diversos fatores são essenciais para a determinação do animal correspondente a cada planta para polinizá-la. A análise de características morfológicas e fisiológicas florais e o aroma das flores podem auxiliar no entendimento e especificidade entre agente polinizador e planta (Fenster *et al.* 2004).

Um processo biológico essencial para o perfeito relacionamento entre o agente polinizador e a planta é a emissão de determinados compostos orgânicos voláteis (VOCs) através das estruturas florais das plantas (Dudareva *et al.* 2013). Cada espécie de angiosperma possui um perfil aromático diferente, característica fenotípica que influencia sua interação com o agente polinizador. Para a emissão de cada composto volátil, existe uma rota metabólica correspondente na qual diversas enzimas atuam para converter uma molécula precursora em um produto final da rota. As principais rotas para a emissão de voláteis em plantas estudadas são as rotas dos fenilpropanóides, dos benzenóides e do chiquimato (Fenske *et al.* 2015).

O presente trabalho teve como foco estudar os aspectos evolutivos e estruturais das proteínas CFAT (*conyferyl alcohol acyltransferase*) e BPBT (*benzoyl-CoA:benzyl*

alcohol/phenylethanol benzoyltransferase) relacionadas à biossíntese de compostos orgânicos voláteis em Solanaceae. Estas proteínas são enzimas que participam de rotas metabólicas responsáveis pela produção de compostos voláteis que atuam como atrativo de animais.

CFAT é uma importante enzima membro da família das aciltransferases BAHD. Essa sigla resulta das iniciais das quatro primeiras proteínas descritas com atividade similar, representando as letras **B**: *Benzyl alcohol O-acetyltransferase* (BEAT), uma acetiltransferase de *Clarkia breweri*; **A**: refere-se à *anthocyanin O-hydroxycinnamoyltransferase* (AHCT) de *Gentiana triflora*; **H**: representa *N-hydroxycinnamoyl/benzoyltransferase* (HCBT) identificada em *Dianthus caryophyllus*; e **D**: corresponde à *deacetylvindoline 4-O-acetyltransferase* (DAT), de *Catharanthus roseus*. A enzima CFAT é responsável por catalisar uma das etapas da rota metabólica dos fenilpropanóides. CFAT é capaz de converter as moléculas de álcool coniferílico e acetil CoA em acetato coniferílico (Figura 2). Seus níveis de expressão possuem mecanismos de regulação circadiano, levando a uma expressão rítmica dentro das células (Dexter *et al.* 2007; Fenske *et al.* 2015). Proteínas da família BAHD são caracterizadas por motivos conservados, como HxxxD, que está relacionado à atividade catalítica no sítio ativo da proteína, e DFGWG que participa na manutenção da estabilidade estrutural da proteína (Dexter *et al.* 2007).

Em boa parte das espécies pertencentes à família Solanaceae, a enzima CFAT possui atividade essencial para a produção do eugenol e do isoeugenol, dois dos produtos orgânicos mais importantes para a interação entre plantas e polinizadores (Dexter *et al.* 2007). Além disso, esses compostos atuam na defesa das plantas contra microrganismos e insetos herbívoros (Koeduka *et al.* 2013). O processo de produção de eugenol e isoeugenol dentro das células ocorre em duas etapas enzimáticas. A primeira está relacionada à conversão da molécula de álcool coniferílico em acetato coniferílico na presença de acetil-CoA, reação

realizada pela enzima CFAT. Posteriormente, as enzimas IGS (isoeugenol sintase) ou EGS (eugenol sintase) irão converter a molécula de acetato coniferílico em isoeugenol ou eugenol, respectivamente (Koeduka *et al.* 2013).

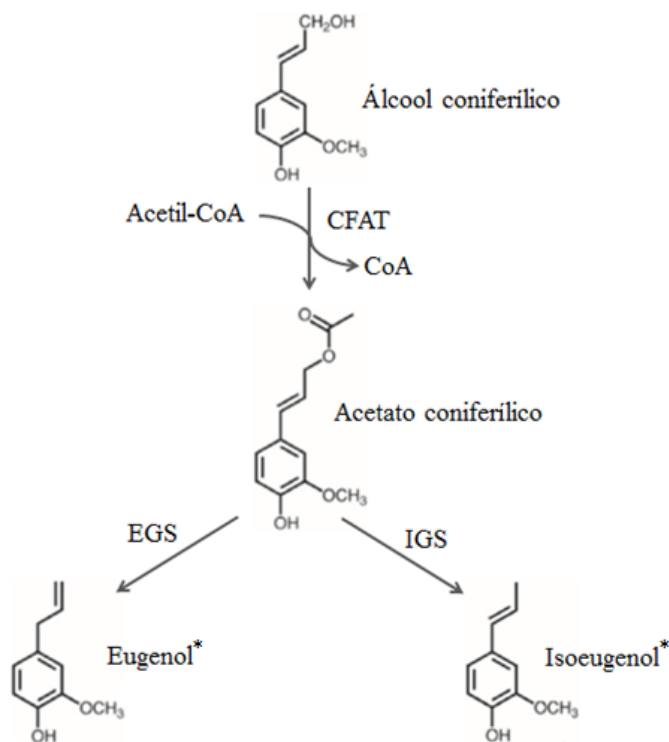


Figura 2. Representação esquemática da rota metabólica para produção de eugenol e isoeugenol a partir do álcool coniferílico. Abreviações: CFAT, *coniferyl alcohol acyltransferase*; EGS, *eugenol synthase*; e IGS, *isoeugenol synthase*. Compostos voláteis marcados com um asterisco. (Adaptado de Spitzer-Rimon *et al.* 2012; Koeduka *et al.* 2013; e Fenske *et al.* 2015).

BPBT é uma enzima da família das aciltransferases BAHD capaz de utilizar diversos álcoois na presença de benzoil-CoA, catalisando uma reação denominada benzoilação, formando *benzoate alcohol esters* (Boatright *et al.* 2004; Eudes *et al.* 2016a). Essa proteína tem maior ação na rota dos benzenoides na qual converte, principalmente, benzoil-CoA e

álcool benzílico ou álcool feniletílico em benzoato benzílico e benzoato feniletílico, respectivamente (Figura 3). Alguns estudos indicam que essa enzima possui expressão tecido-específica para as flores, além de ser detectada sua expressão majoritariamente nas bordas das corolas de espécies como a petúnia-de-jardim (Fenske *et al.* 2015).

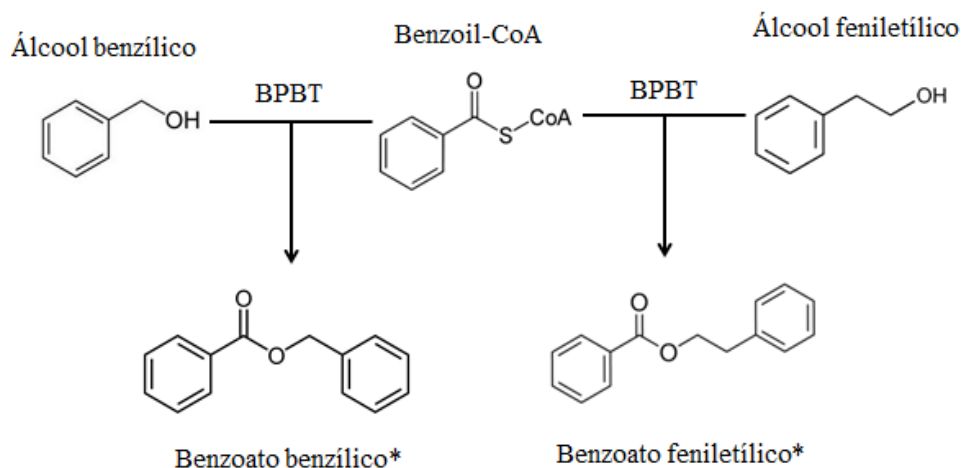


Figura 3. Representação esquemática parcial da rota metabólica dos benzenoides. Abreviações: BPBT, *Benzoyl-CoA:benzyl alcohol/phenylethanol benzoyltransferase*. Compostos voláteis marcados com asterisco. (Adaptado de Boatright *et al.* 2004; Spitzer-Rimon *et al.* 2012; e Fenske *et al.* 2015).

Cada planta emite um perfil aromático característico com o objetivo de atrair certos agentes polinizadores. Muitas proteínas, regulatórias e enzimáticas, envolvidas em rotas metabólicas de produção de compostos que conferem aroma possuem níveis de expressão oscilantes ao longo do dia. Usualmente, os picos de expressão dessas proteínas para a produção do aroma ocorrem coincidentemente com o primeiro momento de atividade de seus respectivos polinizadores (Fenske *et al.* 2015). Como diversos outros genes pertencentes a rotas metabólicas de produção de compostos voláteis, BPBT possui um padrão de expressão

circadiano. Fenske *et al.* (2015) demonstraram que a proteína BPBT apresenta oscilações em condições de luz/escurecimento, além de possuir padrão circadiano também em luz contínua.

A família Solanaceae faz parte das eudicotiledôneas formando um clado com 99 gêneros e cerca de 3.000 espécies (APG 2009). As plantas pertencentes a essa família incluem desde ervas até grandes árvores, com diferentes tipos de frutos e flores e uma multiplicidade de síndromes florais, adaptadas aos mais diferentes agentes polinizadores (Knapp 2010). A família possui espécies com importância econômica, como tomate, batata, tabaco, berinjela e pimentas. Outras espécies, como *Petunia × hybrida* conhecida como petúnia-de-jardim, tem importância ornamental além de servirem de modelos biológicos (Vandenbussche *et al.* 2016).

Nesse trabalho será analisada a estrutura e evolução das proteínas CFAT e BPBT de Solanaceae, que possuem importância na biossíntese de compostos orgânicos voláteis servindo como atrativos de animais e são, conseqüentemente, de enorme relevância no processo de polinização.

OBJETIVOS

Objetivo Geral

Como objetivo geral do trabalho, buscamos entender o padrão evolutivo das proteínas BPBT e CFAT dentro da família Solanaceae, além de correlacionar sua conformação estrutural e aspectos evolutivos a fim de melhor compreender o processo de especiação mediado por adaptação a diversos polinizadores.

Objetivos Específicos

Especificamente, estabelecemos como objetivos:

1. determinar a distância evolutiva entre sequências das proteínas BPBT e CFAT entre as espécies da família Solanaceae;
2. gerar o modelo tridimensional de duas proteínas pertencentes de BPBT e CFAT;
3. avaliar o relacionamento filogenético e as conformações estruturais de ambas as proteínas;
4. contribuir para o esclarecimento do processo de evolução molecular das proteínas BPBT e CFAT em Solanaceae.

CAPÍTULO 1

Artigo a ser submetido ao periódico MOLECULAR GENETICS AND GENOMICS

**Molecular evolution and structural analyses of proteins involved in metabolic pathways
of volatile organic compounds in Solanaceae.**

Lucas C. Beltrame, Claudia E. Thompson, Loreta B. de Freitas

Molecular evolution and structural analyses of proteins involved in metabolic pathways of volatile organic compounds in Solanaceae.

Lucas C. Beltrame,¹ Claudia E. Thompson,^{2,3} and Loreta B. Freitas^{1,*}

¹Laboratory of Molecular Evolution, Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

²Unit of Theoretical and Computational Biology, Centre of Biotechnology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

³Departamento de Farmacociências, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil

*Corresponding author: E-mail: loreta.freitas@ufrgs.br

Running title: Evolution and structure of CFAT and BPBT proteins.

Keywords: volatile organic compounds, CFAT, BPBT, Solanaceae, molecular evolution, protein structure

Abstract

Specialization to different pollinators is an important process responsible for increasing the genetic diversity in plants. The morphological and physiological floral traits, mainly floral scent, can influence the pollination dynamics and select pollinator agents for each plant species. This work studied evolutionary and structural aspects of two proteins involved in metabolic pathways producing volatile organic compounds in plants, CFAT (*conyferyl alcohol acyltransferase*) e BPBT (*benzoyl-CoA:benzyl alcohol/phenylethanol benzoyltransferase*). The main objective was to better understand the evolutionary and structural aspects of these proteins in Solanaceae family, as well as their functions, using phylogenetic methods and the comparative molecular modeling technique. We have used the probabilistic method of Bayesian Inference to infer their evolutionary history using sequences from *Petunia x hybrida* as a query in public databanks to search for homologs in Solanaceae family. Theoretical three-dimensional models were obtained for CFAT and BPBT proteins using as template the 3D structure of a transferase from *Petunia virgatum* (PDB ID: 5FAL, chain A). The phylogenetic tree included several different enzymes with diverse biological roles in Solanaceae, all containing the transferase domain. We identified only one sequence of CFAT in the databanks, which belongs to *Petunia x hybrida*. Moreover, we found several BPBT sequences from *Nicotiana*, *Solanum* and *Capsicum* genera. The three-dimensional structures of CFAT and BPBT have two different domains and we have identified the amino acid residues important for the enzymatic activity and interaction with substrates.

Introduction

Interactions between plants and pollinators are evolutionarily important to the maintenance of biological diversity. The specialization to different pollinators is a process responsible for increasing the genetic diversity in plants. Kölreuter (1761) and Sprengel (1793) first described the roles of plants and pollinators, as well as identified the phenotypic characteristics responsible for this strict relationship. The morphological and physiological floral traits, mainly floral scent, can influence the pollination dynamics and select effective pollinator agents for each plant species.

Pollination syndromes describe flower characteristics or traits that are especially attractive to a specific type of pollinator and are very useful to clarify the mechanisms of floral diversification (Fenster *et al.* 2004). To understand floral evolution is necessary to study the plant-pollinator interactions and the diversity and coevolutionary aspects of plant and specific animal species. Studies of pollination relationships also comprise genetic and ecology aspects that help to understand how the biological processes were affected and influenced by these interactions (Mitchell *et al.* 2009).

Our study is focused on two different proteins involved in metabolic pathways of production and emission of volatile organic compounds (VOCs). We are especially interested in benzoyl-CoA:benzyl alcohol/phenylethanol benzoyltransferase (BPBT) and conyferyl alcohol acyltransferase (CFAT) proteins. They are enzymes that catalyze reactions of specific compounds that result in floral scents (Myler *et al.* 2015)

BPBT is an enzyme that catalyzes reactions involving benzoyl-CoA or acetyl-CoA and many different alcohols molecules (Boatright *et al.* 2004). It participates of the phenylpropanoid/benzenoid pathway converting benzoyl-CoA and benzyl alcohol or

phenylethyl alcohol into benzyl benzoate and phenylethyl benzoate respectively (Boatright *et al.* 2004; Spitzer-Rimon *et al.* 2012). Similar to BPBT, CFAT is also an enzyme with acyltransferase activity. It belongs to the protein family of BAHD acyltransferases that corresponds to **B**: Benzyl alcohol O-acetyltransferase (BEAT), an acetyltransferase from *Clarkia breweri*; **A**: refers to anthocyanin O-hydroxycinnamoyltransferase (AHCT) from *Gentiana triflora*; **H**: represents N-hydroxycinnamoyl/benzoyltransferase (HCBT) identified in *Dianthus caryophyllus*; and **D**: corresponds to deacetylvindoline 4-O-acetyltransferase (DAT), from *Catharanthus roseus*; the first proteins biochemically described in this group (Eudes *et al.* 2016b). The BAHD proteins are characterized for containing two typical motifs (HxxxD and DFGWG). The HxxxD motif is important to the catalytic activity of the protein, whereas DFGWG has a structural role in maintaining a stable protein folding. Some reports showed that CFAT is capable of converting the coniferyl alcohol and acetyl CoA molecules into coniferyl acetate and has also an important role on the increasing of isoeugenol yield, one of the main compounds of scent in *Petunia* (Solanaceae) species (Dexter *et al.* 2006).

Our study was focused in species from the Solanaceae family, which belongs to Angiosperm that encompasses economically important plants, including species used for human healthy, nutrition, ornamentation activities, and as scientific models (Zhuang and Tripp 2016). Different pollination syndromes could be observed among species in Solanaceae and morphological traits and molecular markers are relevant to understand the evolutionary history of these species as well as to hypothesize and preview pollinators-plants relationships (Knapp 2010).

Here we reported the evolutionary and structural analyses of two different proteins (BPBT and CFAT) involved in the biological processes of production and emission of VOCs. We aimed to better understand and explain the evolutionary history of these proteins in

Solanaceae family, as well as their functions, using phylogenetic methods and the comparative molecular modeling technique.

Materials and Methods

Sequence data set and alignment

We first obtained the *Petunia x hybrida* protein sequences (PhCFAT and PhBPBT) from public sequence databases of the National Center for Biotechnology Information (NCBI Resource Coordinators 2016; <https://www.ncbi.nlm.nih.gov/>). We analyzed all domains of these proteins using hmmscan tool from HMMER (Finn *et al.* 2011; <https://www.ebi.ac.uk/Tools/hmmer/>) against Pfam (Finn *et al.* 2016; <http://pfam.xfam.org/>) database with additional parameters as default. The next step was to identify the HMM profile and generate a local database including all sequences with similar profile. We subsequently used the hmmsearch tool from HMMER against UniProtKB database (The UniProt Consortium 2017; <http://www.uniprot.org/help/uniprotkb>), restricting the results to include only sequences from Solanaceae and with additional parameters set as default. As result, we identified and download all protein sequences from Solanaceae family that had the same HMM profile.

We used BLAST+ software (Camacho *et al.* 2009), which is the new BLAST command-line applications, to format a local database using the sequences previously identified and to search for homologs using specific cutoff (e-value $\leq 10^{-6}$). After obtaining a list of sequences matching those criteria, we download the sequences in FASTA format and excluded all partial sequences. Subsequently, we performed a multiple alignment using the PRANK program (Löytynoja & Goldman 2005, <http://www.ebi.ac.uk/goldman->

srv/prank/prank/) with default parameters. The alignment was manually edited using AliView software (Larsson 2014; <http://www.ormbunkar.se/aliview/>), excluding ambiguous positions.

Molecular genealogy

We used ProtTest (Abascal *et al.* 2005) with default settings to choose the best-fit models of aminoacid replacement for the data based on AIC (Akaike's information criterion), BIC (Bayesian information criterion), AICc (AIC with a correction for finite sample sizes), and DT (Decision Theory criterion) parameters. After identifying the best evolutionary model, we constructed the phylogenetic tree of tranferases, including the BPBT and CFAT proteins, using the probabilistic method of Bayesian Inference implemented by MrBayes software (Ronquist *et al.* 2011; <http://mrbayes.sourceforge.net/>). Four chains were run for 2,000,000 generations, with 25% of the genealogies discarded as burn-in and other parameters set as default.

Structural Analyses

Our objective was to theoretically model the previously published BPBT and CFAT protein sequences from *P. x hybrida* (PhCFAT and PhBPBT). After downloading the FASTA sequences, we used them as queries for searching homologs with identified three-dimensional structures using the BLAST tool (Altschul *et al.* 1990) against the Protein Data Bank (PDB; Berman *et al.* 2000; <http://www.rcsb.org/pdb/home/home.do>). This 'blastp' search was performed in June 2016.

The criteria applied to choose the best template for the comparative molecular modeling were resolution equal or higher than 2.0 Å, identity higher than 25 %, and high

coverage. We used MAFFT (Katoh & Standley 2013; <http://mafft.cbrc.jp/alignment/software/>) software with default parameters to align the queries and templates and inspect the coverage among them. After the identification of the best templates, we used MODELLER software version 9.16 (Sali & Blundell 1993; Webb & Sali 2014; <https://salilab.org/modeller/>) for modelling the three-dimensional structures of BPBT and CFAT proteins from *P. x hybrida*.

To verify the quality and stereochemical parameters of each generated model, we used the PROCHECK (Laskowski *et al.* 1993; <http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) and Verify 3D (Lüthy *et al.* 1992; http://services.mbi.ucla.edu/Verify_3D/) softwares. One important criterion for choosing the best model is the number of residues within the most favoured and additional allowed regions, as shown on the Ramachandran plots (Ramakrishnan & Ramachandram 1965; Morris *et al.* 1992) generated by PROCHECK. These regions are based on the plotting of the dihedral angles ψ (psi) and Φ (phi) of each residue within the main chain of the protein to evaluate which bonds between residues are spatially and structural favoured. The lowest number of residues with geometric distortions also helped us to choose the best model. The theoretical isoelectric point was obtained using the ExPASy Tools available at web.expasy.org/compute_pi (Gasteiger *et al.* 2005). Finally, we generated figures, using the PyMol software (<https://pymol.org/>). Electrostatic potential figures were generated using APBS plugin from PyMol software with a pH equals to 7.0.

Results and Discussion

Phylogenetic analyses

The specific domains of BPBT and CFAT proteins were identified using the hmmscan tool available on HMMER. These enzymes presented a transferase domain starting at the 11th site and ending at the 441st site for BPBT and from 6th to 438th sites for CFAT. Using hmmsearch tool, a specific HMM profile and restricting for only Solanaceae taxa, we found 807 sequences for BPBT and 807 sequences for CFAT. After constructing a local databank based on the identified sequences and subsequently using an off-line blastp tool and filtering by HSP (high scoring pairs), we found 502 homologous sequences for PhBPBT and 501 homologous sequences for PhCFAT proteins.

Analyzing the obtained sequences, we realize that PhBPBT and PhCFAT sequences, as well as their homologs, were simultaneously recovered. All repeated IDs were deleted and we applied a threshold of at least 25 % of identity to select the final data set, resulting in 353 sequences that were identified in the UniprotKB AC/ID database. These sequences were aligned using PRANK and all ambiguous positions were manually excluded.

ProtTest revealed that the best-fit evolutionary model for our data was JTT + I + G (Jones *et al.* 1992; Jones-Taylor-Thornton + Invariant sites + Gamma distribution). A phylogenetic tree with high posterior probability values for the branches, especially for those indicating ancient events, and representing all Solanaceae transferase proteins were generated by MrBayes. This tree showed two main distinct clades representing different proteins containing the transferase domain (Figure 1).

The first clade was formed by BPBT sequences from *Nicotiana*, *Capsium*, *Petunia* and *Soalnum* genera and other proteins with no characterized function. The second and major clade included transferase proteins involved in different metabolic pathways, such as suberin and lignin biosynthesis, phenylpropanoids, shikimate, and some defense compounds pathways. Our phylogenetic analyses retrieved only one CFAT sequence from *Petunia x*

hybrida, which is evolutionary close to a monophyletic group of shikimate O-hydroxycinnamoyltransferase (HCT) from different species of the genus *Nicotiana* and *Solanum*. HCT proteins have an important role in the lignin biosynthesis pathway (Eudes *et al.* 2016b).

BPBT sequence from *Capsicum annuum* is very similar to methanol O-anthraniloyltransferase protein sequences from *Capsicum* and *Nicotiana* species and alcohol acyltransferase protein sequences from *Solanum* species, according to our genealogy. The methanol O-anthraniloyltransferase enzyme is able to convert anthraniloyl-CoA and methanol into O-methyl anthranilate (Wang & De Luca 2005). These authors found a high identity value (58%) when methanol O-anthraniloyltransferases and benzoyltransferases (BEPT, benzoyl-CoA:benzylalcohol benzoyltransferase) are aligned, which may indicate they are evolutionarily related. Alcohol acyltransferases are enzymes that catalyze a reaction between Coenzyme A (CoA) and different alcohols, generating their respective esters. In plants, these molecules occur in different tissues conferring, mainly, fruit and flower scent (Souleyre *et al.* 2005).

Anthranilate N-benzoyltransferase enzymes, responsible for converting benzoyl-CoA and anthranilate into N-benzoylanthranilate and CoA, showed a close phylogenetic relationship with shikimate O-hydroxycinnamoyltransferases and CFAT sequences. Studies involving the phytoalexin N-benzoylanthranilate describe it as an important compound that acts on the defense against microorganisms (Reinhard & Matern 1989; Yang *et al.* 1997). Moreover, we observed two anthranilate N-benzoyltransferases sequences from *Nicotiana* species as a sister group in relation to the omega-hydroxypalmitate O-feruloyltransferase sequence from *Nicotiana attenuata*. Omega-hydroxypalmitate O-feruloyltransferase develops an important role on the biosynthesis of 16-feruloyloxypalmitate, a fundamental aromatic

compound for the suberin formation. Suberin is a polymer that controls the transport of water and nutrients, besides protecting against pathogenic infections and environmental stresses (Gou *et al.* 2009). *Capsicum*, *Solanum* and *Nicotiana* 3-N-debenzoyl 2-deoxytaxol N-benzoyltransferase (DBTNBT) sequences did not form a monophyletic group. Taxol, an important alkaloid with anticancer properties, is synthesised by DBTNBT from 3-N-debenzoyltaxol (Onrubia *et al.* 2011).

Three different enzymes presented the transferase domain acting at the same functional group, hydroxycinnamoyl: shikimate O-hydroxycinnamoyltransferase from *Nicotiana*, *Capsicum* and *Solanum* species; hydroxycinnamoyl-CoA quinate transferase from *Ipomoea*, *Solanum* and *Nicotiana*; and spermidine hydroxycinnamoyltransferase (SHT) from *Capsicum* and *Nicotiana* species. The clades formed by these proteins shared a common ancestor (Figure 1). Shikimate O-hydroxycinnamoyltransferase, as described before, acts at the lignin synthesis pathway, whereas SHTs enzymes produce hydroxycinnamic acid amides (HCAAs). HCAAs act at the development and adaptation of the plants according to the environment where they are inserted. Furthermore, they present beneficial therapeutic potentials as anticarcinogenic, antihypertensive and antimicrobial. The phenylpropanoids pathway promotes the accumulation of precursor molecules to the HCAA synthesis. SHT enzyme is able to convert spermidine and hydroxycinnamoyl-CoA into hydroxycinnamoyl spermidine (Peng *et al.* 2016). Different HCAAs can be generated based on the polyamine and the hydroxycinnamoyl-CoA molecules involved. Hydroxycinnamoyl-CoA quinate transferase protein converts quinate and hydroxycinnamoyl-CoA molecules into an ester and CoA. Taking into consideration the reactions catalyzed by those proteins, it is justified they are evolutionary close, as can be seen in our phylogenetic tree (Figure 1).

In our genealogy, we can also see the evolutionary relationships of additional proteins, such as vinorine synthase, able to catalyze the reversible reaction between 16-epi-vellosimine (or gardneral) and vinorine (or 11-methoxy-vinorine), both used as antiarrhythmic drug obtained from *Rauvolfia serpentina* (Bayer *et al.* 2004). Moreover, agmatine hydroxycinnamoyltransferase is another protein identified as evolutionary close to BPBT and CFAT proteins, able to produce hydroxycinnamoyl agmatine molecule, important to the defense against fungal infections (Muroi *et al.* 2009).

Protein structure

Using the previously published BPBT and CFAT protein sequences from *Petunia x hybrida* (PhCFAT, GenBank ID: ABG75942.1; PhBPBT, GenBank ID: AAU06226.1) as queries for searching homologs with determined three-dimensional structures, we obtained nine hits for BPBT and eight hits for CFAT. Based on criteria of highest coverage, identity and resolution, we selected the crystal structure of shikimate hydroxycinnamoyltransferase (HCT) de *Panicum virgatum* in complex with Coa and P-coumaroyl-Shikimate (PvHCT2a; PDB ID: 5FAL, chain A; Eudes *et al.* 2016b) as template for modeling BPBT and CFAT proteins. This template has a resolution of 1.9 Å. *Panicum virgatum* is a monocot of the Poaceae family. PhBPBT and PvHCT2a showed 30% of identity and 96% of coverage, while PhCFAT and PvHCT2a showed 29% and 95% of identity and coverage, respectively. Our phylogenetic tree indicated that CFAT sequence from *Petunia x hybrida* is evolutionary close to a monophyletic group of shikimate O-hydroxycinnamoyltransferase (HCT) from different species of the genus *Nicotiana*, *Solanum*, and *Capsicum*, which corroborates the use of a

shikimate hydroxycinnamoyltransferase for the molecular modeling of CFAT and BPBT proteins.

The comparative molecular modeling by satisfaction of spatial restraints implemented by MODELLER generated 10 different models for PhBPBT and PhCFAT proteins. The main stereochemical data of all models generated for PhBPBT and PhCFAT proteins are described in Table 1. We evaluated the Ramachandran plot obtained by the PROCHECK program that gave us information about the residues within the most favoured, additional allowed, generously allowed, and disallowed regions, as can be seen in Table 1. All models showed more than 95% of the amino acid residues in the most favoured and additional allowed regions. According to the stereochemical analyses, we chose model 8 and model 3 as the best models for PhBPBT and PhCFAT proteins, respectively. The Verify 3D program evaluated the compatibility of a 3D model with the amino acid sequence considered using a 3D profile (Lüthy *et al.* 1992). If the model is correct, the sum of the 3D profile scores is high, preferentially above zero. As can be seen in Figures 2A and 2B, the VERIFY results for PhBPBT model 8 and PhCFAT model 3 also indicated the good quality of the models, since the sum of 3D profile scores is high in both cases. Additionally, the G-factor values were near -0.20 for PhBPBT and -0.30 for PhCFAT, which indicates minor and not significant geometric problems.

Using the final model 8 for PhBPBT and model 3 for PhCFAT, we inspected the theoretical models using the Pymol program, which allowed us to identify the protein domains and the amino acid residues functionally important by aligning the models and template. The active site and the substrate binding regions of the template were identified on the models, as can be seen in Table 2.

The template has two different domains: the first domain contains amino acid residues 1-199 and 387-409, and the second comprises the regions between 200-386 and 410-446. Two distinct domains were also identified in the PhBPBT model. Domain I comprises the residues 1-204 and 376-401, while domain II encompasses the residues 205-375 and 402-460. As expected, the PhCFAT model also can be classified in two different domains. Our structural alignment revealed that the first domain comprises residues 1-193 and 378-402; and the second domain is formed by amino acid residues 194-377 and 403-454. The three-dimensional structures of PhBPBT and PhCFAT are shown in Figures 3 and 4, which are coloured by domain.

Both domains present a core β -sheet surrounded by α -helices. In the 3D model of PhBPBT, we found seven α -helices and nine β -sheets in domain I, whereas in domain II there were nine α -helices and seven β -sheets (Figure 3). The PhCFAT model showed seven α -helices and a core containing seven β -sheets within domain I, while in domain II presented ten α -helices and six β -sheets (Figure 4). The structural alignment between the template and the models can be seen in Figure 5A (PhBPBT) and 5B (PhCFAT). These proteins showed clear differences in their electrostatic potentials, with most significant differences in the Domain II and in the region of the active site (Figure 6). Further studies could better elucidate the proteins activity by simulating at the cellular pH. Accordingly to Schmitzer *et al.* (2010), cell sap from flower tissues in roses present pH ranging from 4.5 to 5.5. PhBPBT and PhCFAT showed pI value equal to 3.79 and 5.57, respectively, while PvhCT2a showed a pI value of 5.52.

In the template, these domains have binding sites responsible to interact with molecules that participate of the reaction catalyzed by PvhCT2a from *Panicum virgatum*, suggesting that each domain of PhCFAT and PhBPBT may have similar interactions with

their respective substrates. Using the information obtained by the structural alignment between a model and the template, we were able to identify the amino acid residues potential involved in substrate interactions. The residues Arg369, Thr382, His163, Trp384, Ser38, and Tyr40 are responsible for the interactions with substrates in the active site of the template PvHCT2a (Eudes *et al.* 2016a). Those residues are equivalent to the amino acid residues Lys357, Val370, His167, Asp372, Val40, and Asn51 in the PhBPBT, whereas for PhCFAT they are structural aligned to the residues Lys357, Ser373, His155, Gly375, Gly43, and Phe45 (Table 2).

Comparing the physicochemical properties of these residues in the template (PvHCT2a) and the models, we can evaluate the impact of the amino acid substitutions in each protein. Our results indicated that some mutations in the active site region have changed the physicochemical properties of some amino acids (Table 2). In PhBPBT, the main changes were found in residues 370, 372, and 40, while in PhCFAT were found in residues 45 and 43. This led us hypothesize that besides the structural changes observed in PhBPBT and PhCFAT proteins, there may be functional changes represented by different polarity and charges of specific amino acids, which may be able to alter functional aspects of these proteins, such as substrate specificity.

Acknowledgments

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and the Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (PPGBM-UFRGS).

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Bayer A, Ma X, Stöckigt J (2004) Acetyltransfer in natural product biosynthesis-functional cloning and molecular analysis of vinorine synthase. *Bioorg Med Chem* 12: 2787-95.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucl Acids Res* 28: 235-242.
- Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding in vivo benzenoid metabolism in *Petunia* petal tissue. *Plant Physiol* 135: 1993-2011.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.
- Dexter R, Qualley A, Kish CM, Ma CJ, Koeduka T, Nagegowda DA, Dudareva N, Pichersky E, Clark D (2007) Characterization of a *Petunia* acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. *Plant J* 49: 265-275.
- Eudes A, Mouille M, Robinson DS, Benites VT, Wang G, Roux L, Tsai YL, Baidoo EEK, Chiu TY, Heazlewood JL, Scheller HV, Mukhopadhyay A, Keasling JD, Deutsch S, Loqué D (2016a) Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. *Microb Cell Fact* 15: 198.

- Eudes A, Pereira JH, Yogiswara S, Wang G, Benites VT, Baidoo EEK, Lee TS, Adams PD, Keasling JD, Loqué D (2016b) Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase to reduce lignin. *Plant Cell Physiol.* 57: 568-579.
- Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD (2004) Pollination syndromes and floral specialization. *Annu Rev Ecol Evol Syst* 35: 375-403.
- Finn RD, Clements J, Eddy SR (2011) HMMER Web Server: interactive sequence similarity searching. *Nucleic Acids Res* 38: W29-W37.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44: D279-D285.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy Server. In: Walker JM (ed) *The proteomics protocols handbook*. Humana Press, Totowa, pp 571–607.
- Gou JY, Yu XH, Liu CJ (2009) A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. *Proc Natl Acad Sci USA* 106:18855-18860.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30: 772-780.
- Knapp S (2010) On ‘various contrivances’: pollination phylogeny and flower form in the Solanaceae. *Philos Transact Roy Soc* 365: 449-460.

- Koeduka T, Suzuki S, Iijima Y, Ohnishi T, Suzuki H, Watanabe B, Shibata D, Umezawa T, Pichersky E, Hiratake J (2013) Enhancement of production of eugenol and its glycosides in transgenic aspen plants via genetic engineering. *Biochem Biophys Res Commun* 436(1):73-8.
- Kölreuter JG (1761) Vorläufige Nachrichten von einigen das Geschlecht der Pflanzen betreffenden Versuchen und Beobachtungen. Leipzig: Gleditschischen Handlung.
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8: 275-282.
- Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* 30: 3276-3278.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283-291.
- Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci USA* 102: 10557-10562.
- Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83-85.
- Mitchell RJ, Irwin RE, Flanagan RJ, Karron JD (2009) Ecology and evolution of plant-pollinator interactions. *Ann Bot* 103: 1355-1363.
- Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. *Proteins: Struct, Funct Bioinf* 12: 345-364.

- Muroi A, Ishihara A, Tanaka C, Ishizuka A, Takabayashi J, Miyoshi H, Nishioka T (2009) Accumulation of hydroxycinnamic acid amides induced by pathogen infection and identification of agmatine coumaroyltransferase in *Arabidopsis thaliana*. *Planta* 230: 517-527.
- NCBI Resource Coordinators (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 44: D7-D19.
- Onrubia M, Moyano E, Bonfill M, Palazón J, Goossens A, Cusidó RM (2011) The relationship between TXS, DBAT, BAPT and DBTNBT gene expression and taxane production during the development of *Taxus baccata* plantlets. *Plant Sci* 181: 282-287.
- Peng H, Yang T, Whitaker BD, Trough F, Shangguan L, Dong W, Jurick WM (2016) Characterization of spermidine hydroxycinnamoyl transferases from eggplant (*Solanum melongena* L.) and its wild relative *Solanum richardii* Dunal. *Hortic Res* 3: 16062.
- Ramakrishnan C, Ramachandran GN (1965) Stereochemical criteria for polypeptide and protein chain conformations. *Biophys J* 5(6): 909-33.
- Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2011) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61: 539-542.
- Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779-815.
- Schmitzer V, Veberic R, Osterc G, Stampar F (2010) Color and phenolic content changes during flower development in groundcover rose. *J Amer Soc Hort Sci* 135: 195-202.

- Souleyre EJ, Greenwood DR, Friel EN, Karunairetnam S, Newcomb RD (2005) An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. FEBS J 272: 3132-3144.
- Souleyre EJ, Greenwood DR, Friel EN, Karunairetnam S, Newcomb RD (2005) An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. FEBS J 272: 3132-3144.
- Spitzer-Rimon B, Farhi M, Albo B, Cna'ani A, Zvi MMB, Masci T, Edelbaum O, Yu Yixun, Shklarman E, Ovadis M, Vainstein A (2012) The R2R3-MYB-Like regulatory factor EOBI, acting downstream of EOBI, regulates scent production by activating ODO1 and structural scent-related genes in petunia. Plant Cell 24: 5089-5105.
- Sprengel CK (1793) Das entdeckte Geheimniss der Naturim Bauundin der Befruchtungder Blumen. Berlin: Vieweg.
- The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45: D158-D169.
- Wang J, De Luca V (2005) The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. Plant J 44: 606-619.
- Webb B, Sali A (2014) Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics 47: 5.6.1-5.6.32.
- Yang Q, Reinhard K, Schiltz E, Matern U (1997) Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. Plant Mol Biol 35: 777-789.

Zhuang Y, Tripp EA (2017) The draft genome of *Ruellia speciosa* (Beautiful Wild *Petunia*:
Acanthaceae). *DNA Res* 24(2):179-192.

Figures

Figure 1. Simplified phylogenetic tree of Solanaceae proteins containing the transferase domain, including BPBT and CFAT.

Figure 2. Verify 3D evaluation of the theoretical three-dimensional models from *Petunia x hybrida*. **A.** PhBPBT, **B.** PhCFAT.

Figure 3. Cartoon representation of three-dimensional structure of BPBT protein coloured by domains. Red = Domain I; Green = Domain II.

Figure 4. Cartoon representation of three-dimensional structure of CFAT protein coloured by domains. Red = Domain I; Green = Domain II.

Figure 5. Cartoon representation of the superposition of the three-dimensional structures of the template and **A.** PhBPBT, **B.** PhCFAT. Template on green, BPBT and CFAT on red.

Figure 6. Molecular surface coloured by electrostatic potential of **A.** PhBPBT, **B.** PhCFAT, **C.** PvHCT2a from *Panicum virgatum*. Red (most negative), white (neutral) and blue (most positive).



Figure 1. Simplified phylogenetic tree of Solanaceae proteins containing the transferase domain, including BPBT and CFAT.

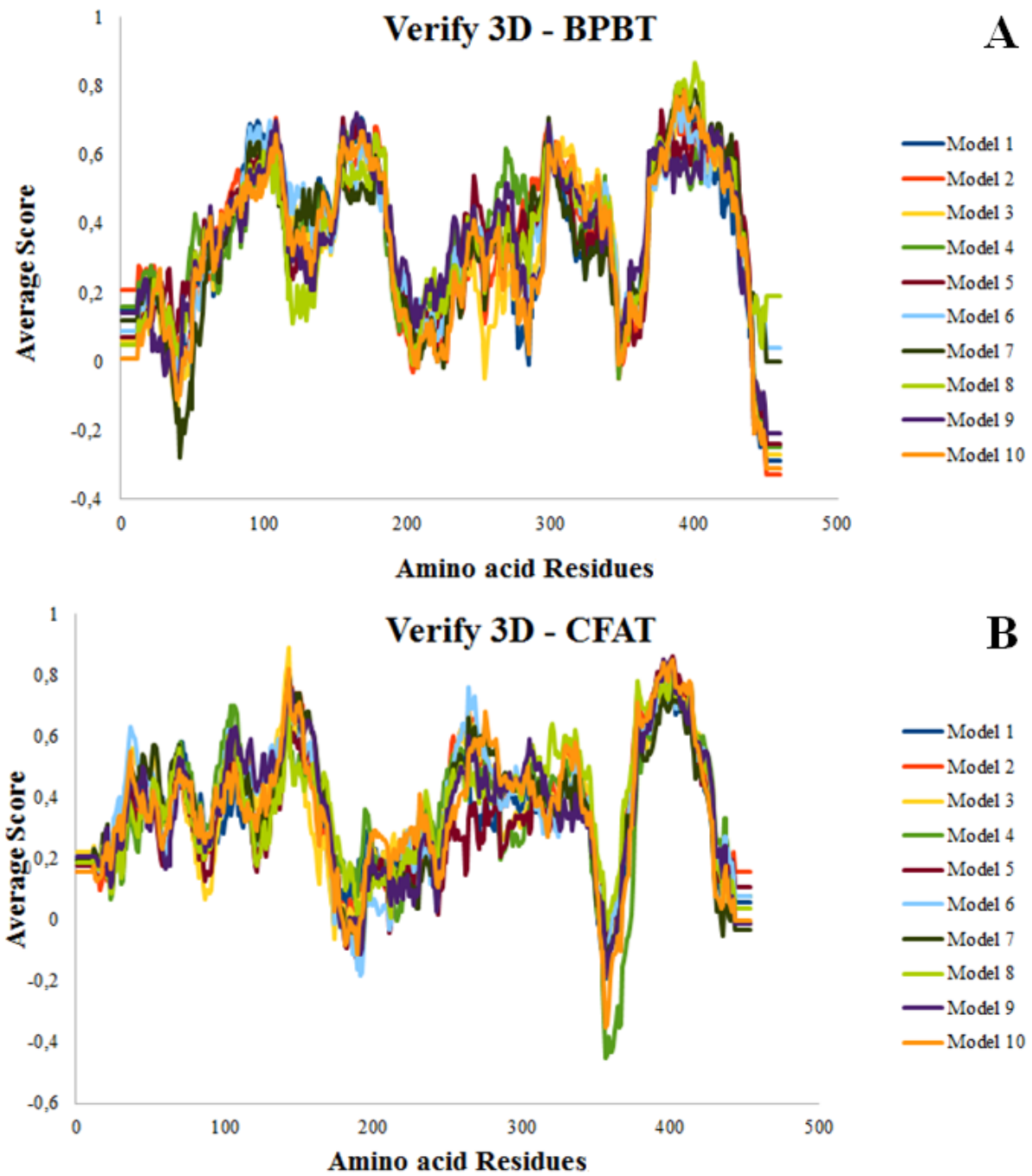


Figure 2. Verify 3D evaluation of the theoretical three-dimensional models from *Petunia x hybrida*. **A.** PhBPBT, **B.** PhCFAT.

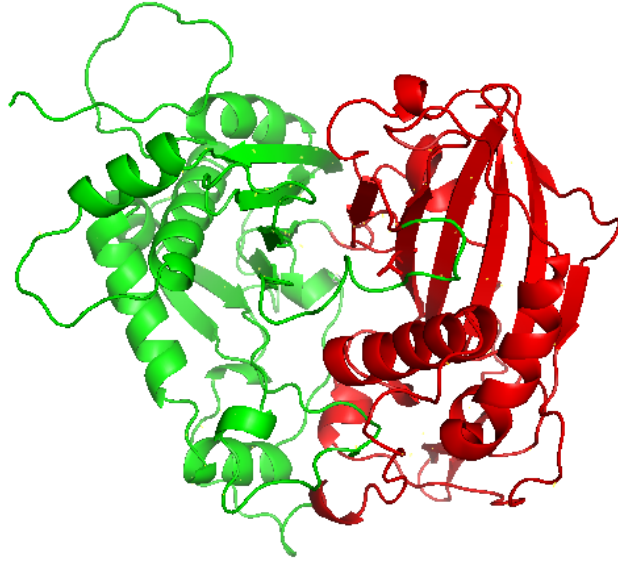


Figure 3. Cartoon representation of three-dimensional structure of BPBT protein coloured by domains. Red = Domain I; Green = Domain II.

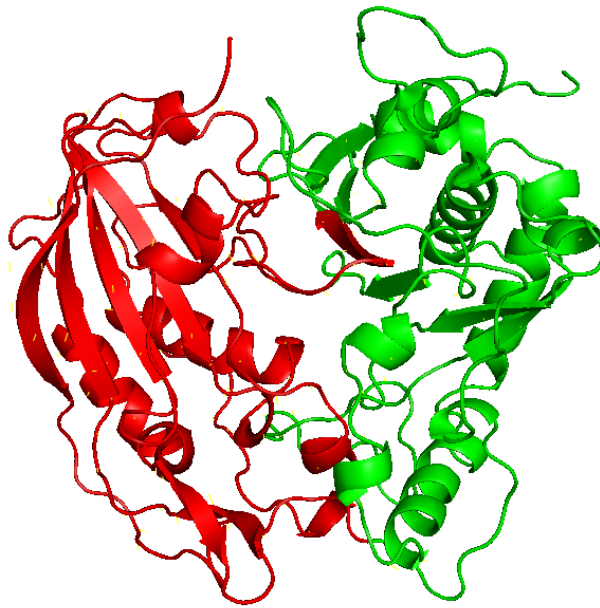


Figure 4. Cartoon representation of three-dimensional structure of CFAT protein coloured by domains. Red = Domain I; Green = Domain II.

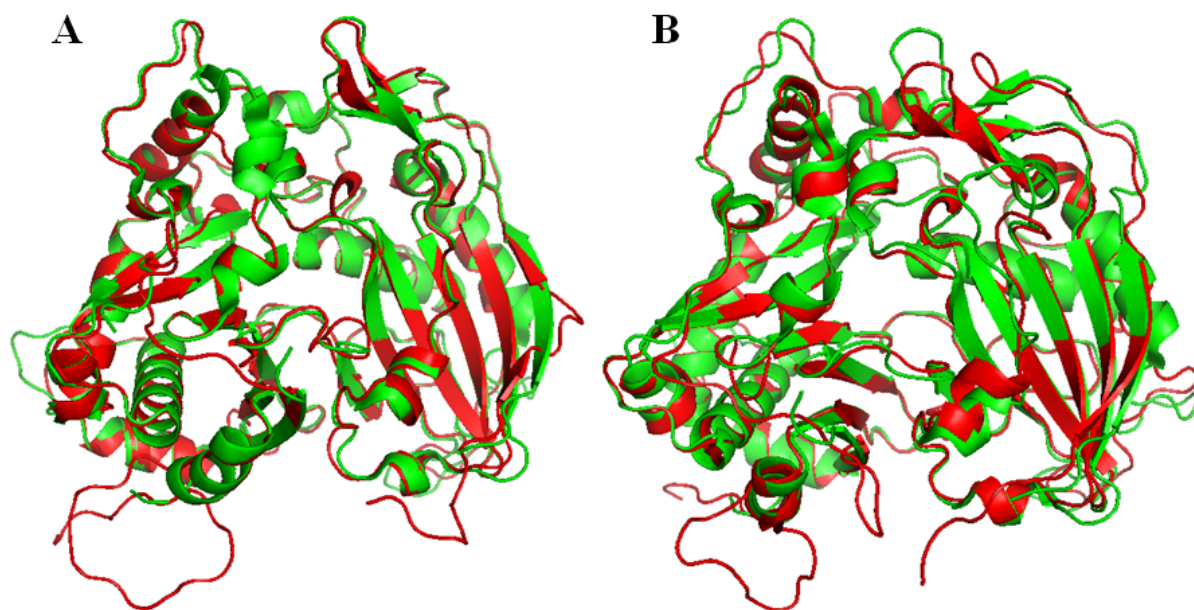


Figure 5. Cartoon representation of the superposition of the three-dimensional structures of the template and **A.** PhBPBT, **B.** PhCFAT. Template on green, BPBT and CFAT on red.

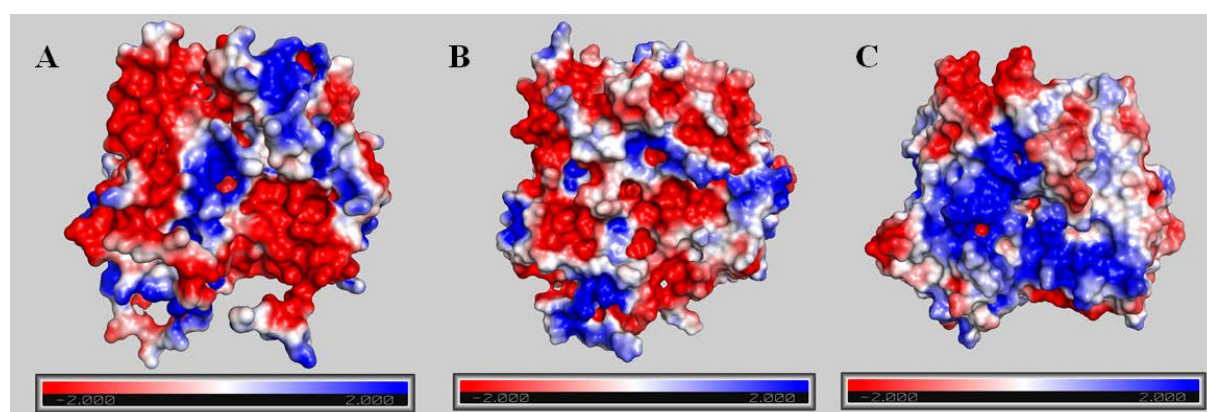


Figure 6. Molecular surface coloured by electrostatic potential of **A.** PhBPBT, **B.** PhCFAT, **C.** PvHCT2a from *Panicum virgatum*. Red (most negative), white (neutral) and blue (most positive).

Tables

Table 1 Stereochemical data of all models generated by MODELLER for BPBT and CFAT proteins from *Petunia hybrida*.

Table 2 Amino acid residues described as important for the protein function of the template (PDB ID: 5FAL A) and their correspondent positions in the BPBT and CFAT modeled proteins from *Petunia hybrida*.

Table 1 Estereochemical data of all models generated by MODELLER for BPBT and CFAT proteins from *Petunia hybrida*.

BPBT										CFAT									
Model	Ramachandran			Distortions				G-factors		Model	Ramachandran			Distortions				G-factors	
	F+AF	F	AF	GA	D	TL	Plan	DA	Overall		F+AF	F	AF	GA	D	TL	Plan	DA	Overall
1	97.9	89.4	8.5	1.5	0.5	21	1	17	-0.20	1	97.3	86.5	10.8	1.8	1.0	44	0	30	-0.33
2	98	87.9	10.1	0.8	1.3	16	0	12	-0.17	2	97.3	86.5	10.8	1.5	1.2	25	0	24	-0.29
3	96.9	86.6	10.3	2.6	0.5	18	0	11	-0.20	3	97.2	86.0	11.2	1.5	1.2	29	0	21	-0.29
4	98.4	88.9	9.5	1.0	0.5	18	0	12	-0.20	4	96.5	86.0	10.5	2.5	1.0	35	0	20	-0.30
5	97.9	87.6	10.3	1.0	1.0	24	0	22	-0.21	5	96.7	86.5	10.2	1.8	1.5	38	0	26	-0.28
6	97.7	88.9	8.8	1.0	1.3	22	1	14	-0.20	6	97.2	86.2	11.0	2.0	0.8	33	0	23	-0.31
7	98.4	89.9	8.5	0.8	0.8	20	2	14	-0.18	7	97.0	85.5	11.5	2.2	0.8	35	0	29	-0.32
8	98.2	89.4	8.8	1.3	0.5	25	1	16	-0.20	8	95.3	86.8	9.5	2.2	1.5	39	0	23	-0.31
9	98.2	87.4	10.8	0.5	1.3	20	1	20	-0.21	9	97.2	86.2	11.0	1.2	1.5	30	0	18	-0.27
10	97.9	89.4	8.5	1.5	0.5	21	1	17	-0.20	10	97.5	87.0	10.5	1.0	1.5	34	0	25	-0.32

Model, Mod; % of residues in most favoured regions, F; % of residues in additional allowed regions, AF; % of residues in generously allowed regions, GA; % of residues in disallowed regions, D; number of bond lengths distortions within main-chain, TL; number of distortions in planar groups, Plan; number of bond angle distortions within the main-chain, DA.

Table 2 Amino acid residues described as important for the protein function of the template (PDB ID: 5FAL A) and their correspondent positions in the BPBT and CFAT modeled proteins from *Petunia hybrida*.

Template (5FAL A)			BPBT			CFAT		
Template (5FAL A)			Model 8			Model 3		
AA ¹	P ²	Classification	AA ¹	P ²	Classification	AA ¹	P ²	Classification
Arg	369	Polar + Charged	Lys	357	Polar + Charged	Lys	357	Polar + Charged
Thr	382	Polar Uncharged	Val	370	Non-polar	Ser	373	Polar Uncharged
His	163	Polar + Charged	His	167	Polar + Charged	His	155	Polar + Charged
Trp	384	Non-polar	Asp	372	Polar - Charged	Gly	375	Non-polar
Ser	38	Polar Uncharged	Val	40	Non-polar	Gly	43	Non-polar
Tyr	40	Polar Uncharged	Asn	51	Polar Uncharged	Phe	45	Non-polar

¹Three-letters abbreviation amino acid, AA; ²Amino acid position at the protein sequence, P; Arginine, Arg; Threonine, Thr; Histidine, His; Tryptophan, Trp; Serine, Ser; Tyrosine, Tyr; Lysine, Lys; Valine, Val; Aspartic acid, Asp; Asparagine, Asn; Glycine, Gly; Phenylalanine, Phe. The three-dimensional structures of BPBT and CFAT were aligned in relation to the template structure using the PyMol program.

DISCUSSÃO GERAL

A fim de compreender os padrões evolutivos das proteínas BPBT e CFAT em Solanaceae, realizamos a análise filogenética utilizando o método probabilístico de Inferência Bayesiana. Com isso, obtivemos uma árvore filogenética que incluiu proteínas das principais rotas para a emissão de voláteis em plantas, entre elas as rotas dos fenilpropanóides, dos benzenóides e do chiquimato. As proteínas envolvidas nessas rotas possuem o domínio comum das transferases. A árvore filogenética resultante revelou altos valores de probabilidade posterior, sendo possível, portanto, inferir de maneira adequada o relacionamento evolutivo entre as proteínas consideradas.

A árvore filogenética possui dois grandes clados principais, um deles inclui as proteínas BPBT de espécies do gênero *Nicotiana*, *Capsicum* e *Solanum* e algumas proteínas com funções ainda não caracterizadas, enquanto o segundo clado envolve proteínas de diversas rotas metabólicas em plantas. Processos metabólicos envolvendo alcalóides, rota dos fenilpropanóides e do chiquimato, rotas de produção de compostos atuantes na defesa das plantas e biossíntese de compostos aromáticos para a formação de suberina são exemplos de reações realizadas por transferases incluídas na árvore filogenética. Foi identificada uma única sequência da proteína CFAT em Solanaceae nos bancos de dados biológicos, sendo esta obtida de *Petunia x hybrida*, a qual apresentou grande proximidade evolutiva com um grupo monofilético composto por diversas sequências de *shikimate O-hydroxycinnamoyltransferase* (HCT) de diferentes espécies dos gêneros *Nicotiana* e *Solanum*. A enzima HCT possui um importante papel no processo de biossíntese de lignina em diversos grupos vegetais (Eudes *et al.* 2016b).

Nossa análise estrutural auxiliou na verificação de similaridade entre as sequências de CFAT de *P. x hybrida* e de HCT de *Panicum virgatum* (PvHCT2a). A estrutura proteica de PvHCT2a (Eudes *et al.* 2016b) foi utilizada como molde para geração do modelo estrutural de PhCFAT, em razão do adequado valor de identidade e alta cobertura, corroborando nossas análises evolutivas que indicam uma grande proximidade entre essas proteínas.

Outro aspecto interessante refere-se à sequência de BPBT de *Capsicum annuum* que, de acordo com a árvore filogenética obtida no presente trabalho, é evolutivamente próxima de um clado composto pelas proteínas *methanol O-anthraniloyltransferase* de espécies dos gêneros *Capsicum* e *Nicotiana* e *alcohol acyltransferase* de espécies de *Solanum*. A enzima *methanol O-anthraniloyltransferase* desempenha o papel de conversão das moléculas de *anthraniloyl-CoA* e metanol em *O-methyl anthranilate* (Wang & De Luca 2005). No trabalho de Wang & De Luca (2005) é demonstrado o alto valor de identidade (58%) entre as sequências das proteínas *methanol O-anthraniloyltransferase* e outra *benzoyltransferase* (BEPT, *benzoyl-CoA:benzylalcohol benzoyltransferase*), indicando um alto grau de similaridade entre os dois grupos de proteínas. A denominação *alcohol acyltransferase* refere-se àquelas enzimas que são capazes de catalisar uma reação envolvendo Coenzima A (CoA) e diferentes álcoois, convertendo-os nos seus respectivos ésteres. Em plantas, estas moléculas ocorrem em diversos tecidos conferindo aroma às flores e frutos, principalmente (Souleyre *et al.* 2005).

A enzima *anthranilate N-benzoyltransferase*, responsável pela conversão das moléculas *benzoyl-CoA* e *anthranilate* em *N-benzoylanthranilate* e *CoA*, apresentou grande proximidade filogenética com um grupo de sequências contendo *shikimate O-*

hydroxycinnamoyltransferase, além da única sequência de CFAT presente na filogenia. A fitoalexina *N-benzoylanthranilate* é descrita na literatura como um importante composto atuando na defesa das plantas contra microrganismos (Reinhard & Matern 1989; Yang *et al.* 1997). Além disso, duas sequências de *anthranilate N-benzoyltransferase* oriundas de espécies do gênero *Nicotiana* apresentaram-se como o grupo irmão em relação a uma sequência de *omega-hydroxypalmitate O-feruloyltransferase* de *Nicotiana attenuata*. A enzima *omega-hydroxypalmitate O-feruloyltransferase* possui importante papel na produção de *16-feruloyloxypalmitate*, composto aromático fundamental para a formação da suberina. A suberina é um polímero que controla o transporte de água e nutrientes da planta, além de proteger contra infecções patogênicas e outros estresses ambientais (Gou *et al.* 2009). As proteínas *3-N-debenzoyl 2-deoxytaxol N-benzoyltransferase* (DBTNBT) de espécies dos gêneros *Capsicum*, *Nicotiana* e *Solanum* também não formaram um grupo monofilético. Catalisando a última etapa de produção de taxol, um alcaloide conhecido por suas propriedades anticâncer, a proteína DBTNBT é capaz de converter 3-N-debenzoyltaxol em taxol (Onrubia *et al.* 2011).

A genealogia revelou a similaridade entre três enzimas distintas, mas que apresentam a função de transferase com o mesmo grupo funcional, o *hydroxycinnamoyl*. São elas as proteínas *shikimate O-hydroxycinnamoyltransferase* de espécies dos gêneros *Nicotiana*, *Capsicum* e *Solanum*; *hydroxycinnamoyl-CoA quinotetrasferase* de espécies dos gêneros *Ipomoea*, *Solanum* e *Nicotiana*; e *spermidine hydroxycinnamoyltransferase* (SHT) de espécies dos gêneros *Capsicum* e *Nicotiana*. Os clados que apresentam essas proteínas compartilham um único ancestral comum. *Shikimate O-hydroxycinnamoyl*, como dito anteriormente, desempenha um importante

papel na biossíntese de lignina. Por outro lado, as enzimas SHTs possuem importante atividade na biossíntese de *hydroxycinnamic acid amides* (HCAAs). Estas moléculas atuam no desenvolvimento e adaptação das plantas de acordo com o ambiente em que estão inseridas, além de apresentar potencial terapêutico por possuírem características anticarcinogênicas, anti-hipertensivas, antimicrobianas e diversas outras aplicações. A rota metabólica dos fenilpropanóides é a que promove a produção das moléculas precursoras dos HCAAs. A enzima SHT é responsável pela produção da molécula de *hydroxycinnamoyl spermidine*, a partir da *spermidine* e de uma molécula de *hydroxycinnamoyl-CoA*, podendo ocorrer em três diferentes etapas de conjugação (Peng *et al.* 2016). Diversos outros HCAAs podem ser produzidos dependendo da poliamina e da molécula de *hydroxycinnamoyl-CoA* envolvidas na reação. Por último, a enzima *hydroxycinnamoyl-CoA quinate transferase* catalisa a reação entre as moléculas de *quinatate* e *hydroxycinnamoyl-CoA*, convertendo-as em um éster e CoA. Avaliando as três enzimas aqui citadas, verificam-se aspectos comuns entre elas como o compartilhamento do substrato, o que justificaria a maior proximidade filogenética encontrada entre essas proteínas.

A partir de nossas análises, identificamos diversas outras proteínas que apresentam o mesmo domínio transferase contido nas enzimas BPBT e CFAT de *P. x hybrida*, focos do estudo. Um exemplo seria a *vinorine synthase*, enzima que, em plantas, desempenha papel fundamental na reação reversível entre *16-epi-vellosimine* (ou *gardneral*) e *vinorine* (ou *11-methoxy-vinorine*), dois alcaloides importantes para uso como drogas anti-arritmia provenientes de *Rauvolfia serpentina* (Bayer *et al.* 2004). Outra enzima que se destaca é a *agmatine hydroxycinnamoyltransferase*, responsável

por produzir *hydroxycinnamoylagmatine*. Esta molécula tem um papel importante na defesa contra infecções fúngicas em plantas (Muroi *et al.* 2009).

As estruturas tridimensionais das proteínas BPBT e CFAT da espécie *Petunia x hybrida*, pertencentes à família Solanaceae, foram determinadas a partir da Modelagem Molecular Comparativa utilizando o programa MODELLER. Ao realizar uma busca por moldes através do BLASTp (Altschul *et al.* 1990), encontrou-se nove *hits* para BPBT e oito para CFAT. A estrutura que serviu como molde para BPBT e CFAT foi a proteína *hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase* (HCT) da espécie de *Panicum virgatum*, uma poacea perene nativa da América do Norte, cujo código PDB é 5FAL. Tal proteína apresentou 30% e 29% de identidade com relação à BPBT e CFAT de *Petunia x hybrida*, respectivamente. A partir desse molde, foram gerados dez modelos para cada uma das proteínas estudadas, sendo que um modelo final foi escolhido para BPBT e CFAT, de acordo com os resultados obtidos pelas análises estereoquímicas, principalmente.

A partir da análise comparativa entre as proteínas alvo e a proteína molde, foi possível observar a localização de aminoácidos funcionalmente importantes, de acordo com estudos realizados por Eudes *et al.* (2016b). Esses autores determinaram a estrutura tridimensional da proteína que usamos como molde e que se trata de uma proteína pertencente à rota metabólica de síntese de monômeros de lignina. Por possuir um domínio transferase característico em sua sequência, foi utilizada como molde com o objetivo de modelar as estruturas terciárias das proteínas BPBT e CFAT de *Petunia x hybrida*. Eudes *et al.* (2016b) propuseram a separação da proteína em dois domínios distintos (I e II) e destacaram o papel de alguns resíduos de aminoácidos que promovem a compatibilidade entre a estrutura proteica e seus respectivos substratos. O sítio ativo

de PvHCT2a localiza-se entre os dois domínios destacados. As proteínas modeladas para BPBT e CFAT também possuem dois domínios distintos. Com base na comparação entre o molde e os respectivos modelos, encontraram-se para BPBT as regiões compreendidas entre os resíduos 1 a 204 e 376 a 401 como pertencentes ao domínio I, ao passo que o domínio II compreende os resíduos 205 a 375 e 402 a 460. Para CFAT, foram encontradas as regiões de 1 a 193 e 378 a 402 compreendendo o domínio I, enquanto os aminoácidos 194 a 377 e 403 a 454 como pertencentes ao domínio II.

Em relação às estruturas secundárias, ambas as proteínas apresentaram um núcleo de folhas- β , circundado por α -hélices em cada domínio. Foram contabilizadas para BPBT, no primeiro domínio, sete α -hélices e nove folhas- β . No segundo domínio de BPBT, foram encontradas nove α -hélices e sete folhas- β . No domínio I de CFAT, há um núcleo contendo sete estruturas de folhas- β envolto por sete α -hélices. No domínio II de CFAT, podem-se observar dez α -hélices envolvendo um núcleo de seis folhas- β . De forma geral, o alinhamento entre a estrutura molde e cada um dos modelos estudados (BPBT e CFAT) revelou grande conservação estrutural.

Os aminoácidos funcionalmente importantes indicados em Eudes *et al.* (2016b) foram identificados nos modelos de PhBPBT e PhCFAT. Os resíduos pertencentes ao sítio de ligação das moléculas *p-coumaroyl-CoA* e chiquimato, substratos da enzima PvHCT2a são: Ser38, Tyr40, His163, Arg369, Thr382 e Trp384. De acordo com o alinhamento estrutural, tais aminoácidos encontram-se nas posições Val49, Asn51, His167, Lys357, Val370 e Asp372 em BPBT, enquanto para CFAT foram observados nas posições Gly43, Phe45, His155, Lys357, Ser373 e Gly375. Estes aminoácidos estão localizados na região entre os dois domínios proteicos, região que coincide com o sítio

de ligação dos substratos. Comparando as propriedades físico-químicas dos aminoácidos destacados no molde e nos dois modelos gerados, podemos observar que as propriedades físico-químicas de cada resíduo encontrado, em especial os aminoácidos do sítio de ligação em destaque, são modificadas. Tal análise nos leva a concluir que as substituições aminoacídicas nessas proteínas, além de terem um impacto estrutural, também podem estar relacionadas a mudanças em aspectos funcionais. A mudança de polaridade e carga de aminoácidos no sítio ativo pode ser responsável, por exemplo, pela mudança na especificidade ao substrato.

Esse trabalho traz importantes contribuições para o esclarecimento de aspectos evolutivos e estruturais de proteínas envolvidas em rotas metabólicas de produção de compostos orgânicos voláteis em plantas, mais especificamente em Solanaceae. A melhor compreensão da evolução e diversificação dessas proteínas tem importância para o entendimento da evolução da emissão de aromas em plantas e seu impacto na atração de diferentes agentes polinizadores. Trata-se de um estudo inédito, mas inicial, que pretende ser continuado através de uma análise mais detalhada dos aspectos evolutivos e do impacto de substituições aminoacídicas na estrutura e função dessas proteínas.

REFERÊNCIAS BIBLIOGRÁFICAS

(Caps. Introdução e Discussão)

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- APG - The Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 161: 105-121.
- Bayer A, Ma X, Stöckigt J (2004) Acetyltransfer in natural product biosynthesis-functional cloning and molecular analysis of vinorine synthase. *Bioorg Med Chem* 12: 2787-95.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucl Acids Res* 28: 235-242.
- Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N (2004) Understanding in vivo benzenoid metabolism in *Petunia* petal tissue. *Plant Physiol* 135: 1993-2011.
- Dexter R, Qualley A, Kish CM, Ma CJ, Koeduka T, Nagegowda DA, Dudareva N, Pichersky E, Clark D (2007) Characterization of a *Petunia* acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. *Plant J* 49: 265-275.
- Dudareva N, Klempien A, Muhlemann JK, Kaplan I (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol* 198: 16-32.

- Eudes A, Mouille M, Robinson DS, Benites VT, Wang G, Roux L, Tsai YL, Baidoo EEK, Chiu TY, Heazlewood JL, Scheller HV, Mukhopadhyay A, Keasling JD, Deutsch S, Loqué D (2016a) Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. *Microb Cell Fact* 15: 198.
- Eudes A, Pereira JH, Yogiswara S, Wang G, Benites VT, Baidoo EEK, Lee TS, Adams PD, Keasling JD, Loqué D (2016b) Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase to reduce lignin. *Plant Cell Physiol.* 57: 568-579.
- Fenske MP, Hazelton KDH, Hempton AK, Shim JS, Yamamoto BM, Riffell JA, Imaizumi T (2015) Circadian clock gene LATE ELONGATED HYPOCOTYL directly regulates the timing of floral scent emission in *Petunia*. *Proc Natl Acad Sci USA* 112: 9775-9780.
- Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD (2004) Pollination syndromes and floral specialization. *Annu Rev Ecol Evol Syst* 35: 375-403.
- Gou JY, Yu XH, Liu CJ (2009) A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. *Proc Natl Acad Sci USA* 106:18855-18860.
- Hermann K, Kuhlemeier C (2011) The genetic architecture of natural variation in flower morphology. *Curr Opin Plant Biol* 14: 60-65.
- Jones CE, Little RJ (1983) *Handbook of experimental pollination biology*. Scientific & Academic Editions, New York. 558 pp.

- Knapp S (2010) On 'various contrivances': pollination phylogeny and flower form in the Solanaceae. *Philos Transact Roy Soc* 365: 449-460.
- Koeduka T, Suzuki S, Iijima Y, Ohnishi T, Suzuki H, Watanabe B, Shibata D, Umezawa T, Pichersky E, Hiratake J (2013) Enhancement of production of eugenol and its glycosides in transgenic aspen plants via genetic engineering. *Biochem Biophys Res Commun* 436(1):73-8.
- Kölreuter JG (1761) Vorläufige Nachrichten von einigen das Geschlecht der Pflanzen betreffenden Versuchen und Beobachtungen. Leipzig: Gleditschischen Handlung
- Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A (2000) Comparative protein structure modelling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325.
- Martinez M (2011) Plant protein-coding gene families: emerging bioinformatics approaches. *Trends Plant Sci* 16: 558-567.
- Mitchell RJ, Irwin RE, Flanagan RJ, Karron JD (2009) Ecology and evolution of plant-pollinator interactions. *Ann Bot* 103: 1355-1363.
- Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. *Proteins: Struct Funct Bioinf* 12: 345-364.
- Muroi A, Ishihara A, Tanaka C, Ishizuka A, Takabayashi J, Miyoshi H, Nishioka T (2009) Accumulation of hydroxycinnamic acid amides induced by pathogen infection and identification of agmatine coumaroyltransferase in *Arabidopsis thaliana*. *Planta* 230: 517-527.

- Ohno S (1970) Evolution by gene duplication. Springer, Verlag, 1979 pp.
- Onrubia M, Moyano E, Bonfill M, Palazón J, Goossens A, Cusidó RM (2011) The relationship between TXS, DBAT, BAPT and DBTNBT gene expression and taxane production during the development of *Taxus baccata* plantlets. *Plant Sci* 181: 282-287.
- Peng H, Yang T, Whitaker BD, Trough F, Shangguan L, Dong W, Jurick WM (2016) Characterization of spermidine hydroxycinnamoyl transferases from eggplant (*Solanum melongena* L.) and its wild relative *Solanum richardii* Dunal. *Hortic Res* 3: 16062.
- Real L (1983) Pollination Biology. Academic Press Inc., New York. 338 pp.
- Reinhard K, Matern U (1989) The biosynthesis of phytoalexins in *Dianthus caryophyllus* L. call cultures: induction of benzoyl-CoA:anthranilate N-benzoyltransferase activity. *Arch Biochem Biophys* 275: 295-301.
- Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol* 243: 779-815.
- Sánchez R, Sali A (1997) Advances in comparative protein-structure modelling. *Curr Opin Struct Biol* 7: 206–214.
- Sánchez R, Sali A (2000) Comparative protein structure modeling: introduction and practical examples with MODELLER, *in* Methods Mol Biol, Protein structure prediction: methods and protocols. Humana Press, New Jersey, pp. 97-129.

- Souleyre EJ, Greenwood DR, Friel EN, Karunairetnam S, Newcomb RD (2005) An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. *FEBS J* 272: 3132-3144.
- Spitzer-Rimon B, Farhi M, Albo B, Cna'ani A, Zvi MMB, Masci T, Edelbaum O, Yu Yixun, Shklarman E, Ovadis M, Vainstein A (2012) The R2R3-MYB-Like regulatory factor EOBI, acting downstream of EOBI, regulates scent production by activating ODO1 and structural scent-related genes in petunia. *Plant Cell* 24: 5089-5105.
- Sprengel CK (1793) *Das entdeckte Geheimniss der Naturim Bauundin der Befruchtungder Blumen*. Berlin: Vieweg.
- Vandenbussche M, Chambrier P, Rodrigues-Bento S, Morel P (2016) *Petunia*, your next supermodel? *Front Plant Sci* 7: 72.
- VerlÍ H (2004) *Bioinformática: da Biologia à Flexibilidade Molecular*. Sociedade Brasileira de Bioquímica e Biologia Molecular, São Paulo. 282 pp.
- Wang J, De Luca V (2005) The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. *Plant J* 44: 606-619.
- Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 86: 2.9.1-2.9.37.
- Yang Q, Reinhard K, Schiltz E, Matern U (1997) Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. *Plant Mol Biol* 35: 777-789.

Ye C, Li T, Yin H, Weston DJ, Tuskan GA, Tschaplinski TJ, Yang X (2013)

Evolutionary analyses of non-family genes in plants. *Plant J* 73: 788-797.