

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**IDENTIFICAÇÃO E EVOLUÇÃO DOS FATORES DE
TRANSCRIÇÃO DO FOF DE PITANGUEIRA (*Eugenia uniflora* L.)**

EDGAR LUIS WASCHBURGER

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do título de Mestre em Genética e Biologia Molecular

Orientadora: Prof^ª. Dr^ª. Andreia Carina Turchetto Zolet

Porto Alegre, Dezembro de 2021

Instituições e fontes financiadoras

O presente trabalho foi executado no Laboratório de Genética Vegetal do Departamento de Genética da Universidade Federal do Rio Grande do Sul (UFRGS) e teve como fontes financiadoras o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Estado do RS (FAPERGS) e o Programa de Pós-graduação em Genética e Biologia Molecular (PPGBM) da UFRGS.

Resumo	4
Abstract	5
1. Introdução.....	6
<i>Eugenia uniflora</i> como espécie modelo	6
Genes DOF	7
2. Objetivos.....	9
Objetivo Geral	9
Objetivos Específicos.....	9
3. Identificação e caracterização dos fatores de transcrição DOF em <i>E. uniflora</i>	10
RESUMO	11
INTRODUÇÃO	12
MATERIAIS E MÉTODOS	14
Identificação dos genes DOF em <i>Eugenia uniflora</i> L.	14
Predições <i>in silico</i>	15
Alinhamento e construção da árvore filogenética.	15
Expressão diferencial, identificação de SNPs e análise do promotor	16
RESULTADOS.....	16
Identificação e classificação dos genes DOF de <i>E. uniflora</i>	16
Conservação da estrutura gênica e dos domínios proteicos dos genes DOF de <i>E. uniflora</i>	17
Populações de <i>E. uniflora</i> e seus arsenais de genes DOF	18
Análise da região promotora.....	19
DISCUSSÃO.....	20
CONCLUSÃO	27
BIBLIOGRAFIA	27
FIGURAS & TABELAS	33
4. Capítulo II – DOF transcription factors: from gene family expansion and diversification towards biological functions and adaptation.....	41
Introduction.....	43
General findings in DOF TFs research.....	44

Genome-wide and phylogenetic analysis in native and wild species.....	47
Conclusions, Limitations, and Perspectives.....	48
Materials and Methods.....	49
Literature Survey.....	49
Sequence Retrieval.....	49
Alignment and Phylogeny Reconstruction.....	50
Acknowledgments.....	50
Conflict of interest.....	50
Authors contributions.....	50
References.....	50
Figures.....	53
5. Considerações Finais.....	55
6. Bibliografia.....	56

Resumo

Eugenia uniflora é uma planta nativa pertencente à família Myrtaceae, possuindo uma significativa importância ecológica e econômica. *E. uniflora* é amplamente distribuída pela floresta atlântica. As populações desta espécie estão bem adaptadas a distintas condições ambientais, como restinga e mata ciliar, e apresentam fenótipos altamente contrastantes, como época de floração e morfologia do tronco, evidência de uma ótima regulação de seu *pool* genético. Dentre os diferentes reguladores da expressão gênica, os fatores de transcrição (TFs) são um dos, senão os mais, críticos membros para o desenvolvimento fisiológico adequado. Os genes DOF são parte de uma família de fatores de transcrição específica do clado Viridiplantae, com muitos membros atuantes no desenvolvimento de plantas. Armazenamento e germinação de sementes, desenvolvimento de estômatos e formação de tecidos vasculares são apenas alguns dos muitos papéis biológicos descritos até agora na literatura. Alguns grupos funcionais conservados chamam a atenção em estudos de engenharia genética por produzirem frutos maiores e conferir tolerância sistêmica a estresses, tornando a identificação e caracterização desses genes uma abordagem biotecnológica para prospecção gênica. Este estudo visa fornecer novos *insights* sobre a história evolutiva da família de genes DOF e caracterizar esses genes em *E. uniflora* para esclarecer seus papéis na adaptação local desta espécie. Com base em dados transcritômicos e genômicos de *E. uniflora*, assim como uma abordagem bioinformática, identificamos e caracterizamos os genes DOF desta espécie. Identificamos motivos comuns ao longo da sequência peptídica, propriedades físico-químicas e localização subcelular por meio de ferramentas online. A análise de RNA-Seq de indivíduos de *E. uniflora* provenientes de diferentes populações resultaram na identificação de 30 SNPs presentes em CDS, juntamente com os perfis de expressão dos muitos genes DOF sob tratamento de seca e análise da sequência promotora. Uma caracterização adequada foi alcançada pela reconstrução das relações filogenéticas. A aparente falta de estudos filogenéticos apropriados sobre os genes DOF foi observada após o exame da literatura, portanto, fomos movidos a conduzir uma revisão robusta sobre sua evolução. Mais de 2.000 sequências, representando cerca de 70 famílias botânicas diferentes, e mais de 250 artigos foram recuperados de bancos de dados online. Ao todo, esses experimentos contribuirão para a compreensão da fisiologia de *E. uniflora*, os papéis biológicos dos genes DOF e sua evolução molecular.

Abstract

Eugenia uniflora is a native plant species belonging to the Myrtaceae family, having significant ecological and economical importance. *E. uniflora* is widely distributed throughout the Atlantic Forest. Populations of this species are well adapted to distinct environmental conditions, such as restinga and riparian forest, and exhibit highly contrasting phenotypes, such as flowering time and trunk morphology, depicting a great fine tuning of their genetic pool. Among several molecular regulators, transcription factors (TFs) are one, if not the most, critical players for proper physiological development. DNA-binding with one-finger (DOF) genes are part of a *Viridiplantae* specific TF family with many plant development acting members. Seed storage and germination, stomata development, and vascular tissue formation are only a few of the many biological roles described so far in the literature. Some conserved functional groups are drawing attention in genetic engineering studies for yielding bigger fruits and conferring systemic-wide stress tolerance, making the identification and characterization of these genes a biotechnological approach for gene prospection. This study aims to provide new insights into the evolutionary history of the DOF gene family and characterize these genes in *E. uniflora* to shed light on their roles in the local adaptation of this species. Based on transcriptomic and genomic data of *E. uniflora* and bioinformatic approach, we identified and characterized the DOF genes of this species. We predicted common motifs along the peptide sequence, physio-chemical properties, and subcellular location by online software tools. An RNA-Seq analysis of *E. uniflora* individuals from separate populations resulted in the identification of 30 CDS SNPs, along with the expression profiles of the many DOF genes under drought treatment and promoter sequence analysis. Proper characterization was achieved by reconstruction of phylogenetic relationships. An apparent lack of appropriate DOF phylogenetic studies was observed upon literature examination. Thus, we were moved to conduct a robust review about their evolution. Over 2000 sequences, representing about 70 different botanical families, and over 250 articles were retrieved from online databases. Altogether, these experiments will contribute to the understanding of *E. uniflora*'s physiology, DOF TF biological roles and molecular evolution.

1. Introdução

***Eugenia uniflora* como espécie modelo**

Eugenia uniflora L., conhecida como pitangueira, é uma espécie de frutífera, de porte pequeno a médio, com 2 a 4 metros de altura, podendo chegar até 12 metros em condições ambientais ótimas. Floresce principalmente na primavera, dando origem a flores pequenas e brancas, as quais podem sofrer polinização cruzada ou autofertilização, gerando frutos verdes, que quando amadurecidos, apresentam desde tons amarelados até vermelho escuros, podendo chegar ao preto. É nativa da Mata Atlântica e pertencente à família Myrtaceae, assim como uma diversidade de outras espécies frutíferas nativas, como a Cereja-do Rio Grande (*E. involucrata*), a pitangueira (*E. uniflora*), uvaia (*E. pyriformis*) jabuticaba (*Plinia cauliflora*), araçá (*Psidium cattleianum*), guabiroba (*Campomanesia xanthocarpa*) e a espécie arbórea *Eucalyptus grandis* (Eucalipto). Folhas e frutos de *E. uniflora* (ricos em vitamina A, complexo B e C, licopeno, minerais como cálcio, ferro e fósforo, antocianinas e flavonóis) são amplamente empregados na medicina popular, com interesse científico e alto potencial para exploração comercial pela atividade antibacteriana de seus metabólitos secundários e extratos bruto¹, possuindo ainda atividades antifúngicas², anti-helmínticas³, anti-inflamatórias⁴, antioxidantes, analgésicas e espaço para a sua utilização no tratamento de doenças crônicas, como diabetes⁵. Os óleos naturais da planta também são coletados para a utilização em produtos cosméticos. É uma planta nativa brasileira e não endêmica, cultivada nos Estados Unidos e países da América Central, também considerada como espécie invasora na Bermuda, onde sua introdução como planta ornamental acarretou a dispersão de sementes por aves canoras. Abrange desde a região costeira do nordeste até a região sul do Brasil, leste do Paraguai e norte da Argentina e Uruguai⁶⁻⁸. Possui uma grande área de dispersão, e está adaptada a diferentes ecoregiões com características contrastantes. Apresenta porte arbustivo na restinga brasileira, localidade costeira de terreno arenoso e salino, forte insolação e ventos fortes, enquanto possui porte arbóreo na mata atlântica subtropical no interior da região sul, caracterizada por vegetação fechada, húmida e ombrófila. Essa disparidade entre populações também foi confirmada a nível genético, onde indivíduos da restinga possuíam uma variação genética muito menor, quando comparada com indivíduos da mata atlântica, possivelmente resultado de uma expansão populacional recente de plantas adaptadas às condições da restinga⁹

Genes DOF

A família de fatores de transcrição DOF (do inglês, *DNA Binding with One-Finger*) fazem parte da maquinaria de regulação da expressão gênica, exclusivos de organismos do clado viridiplantae, incluindo desde algas verdes até angiospermas. O número de genes DOF por espécies é variado, partindo de apenas uma sequência em algas clorofitas até mais de cem integrantes em espécies com múltiplos eventos de duplicação genômica, tendo a maioria das espécies de angiospermas um número próximo de 30 genes. Desde a caracterização da primeira sequência referente a um gene DOF em 1995, foram publicados mais de 60 artigos científicos identificando e caracterizando os genes DOF em diferentes organismos, principalmente em espécies de cultivares. Com uma média de 200 a 400 resíduos de aminoácidos por proteína, as proteínas DOF têm em comum a presença do seu domínio característico DOF de ligação ao DNA em sua região N-terminal, assim como uma região C-terminal com a presença de motivos para a interação proteína-proteína e subsequente regulação da expressão gênica. O domínio DOF é um motivo do tipo CX₂CX₂₁CX₂C, onde os resíduos de cisteína conservados ao longo da sua sequência são os responsáveis pela devida interação ao seu cofator, um átomo de zinco, e a formação de uma estrutura de alça, responsável pela nomenclatura da família. Alguns dos resíduos entre as quatro cisteínas também são altamente conservados, devido à sua função de reconhecimento do motivo DOF no DNA, composto pela sequência de 4 nucleotídeos (T/A)AAG. Além dos resíduos presentes entre os resíduos de cisteína, os resíduos circundantes apresentam maior variabilidade, consequência de suas funções diversas, com a exceção de um resíduo de triptofano conservado para a devida funcionalidade dessas proteínas. O domínio DOF é bifuncional, também se apresentando como sinal de localização nuclear (NLS) para a devida compartimentalização dessas proteínas no envelope nuclear. Os domínios presentes na região C-terminal são variados entre membros da família, podendo ser conservados entre integrantes dos mesmos grupos funcionais/subfamílias. O grupo funcional mais bem caracterizado na literatura, e coincidentemente o maior grupo funcional, é a subfamília CDF de genes DOF, responsáveis pela inibição do florescimento regulando genes como *CONSTANS* e *FLOWERING LOCUS T*. Outras funções dos genes DOF descritas pela literatura incluem, mas não se delimitam à germinação, divisão celular, desenvolvimento vascular, desenvolvimento estomatal e respostas a estresses bióticos e abióticos. Em relação à literatura sobre os genes DOF, encontra-se uma falta de estudos evolutivos analisando os genes DOF desde seus

ancestrais mais basais. Os estudos e revisões publicados que se propõem a devida caracterização de grupos funcionais não aparentam estar em sintonia, principalmente pelo fato que os grupos funcionais propostos variam significativamente de estudo a estudo. Grupos extremamente conservados, como o caso da subfamília CDF, acabam por serem recuperados em todas as análises e, conseqüentemente, se tornam mais atrativos para subseqüentes estudos funcionais. Quatorze anos se passaram desde a publicação do último estudo evolutivo sobre a família de genes DOF com uma representatividade de espécies chave para a sua compreensão evolutiva. Desde então, as ferramentas de bioinformática, poder computacional e disponibilidade de genomas sequenciados em bancos de dados públicos aumentaram drasticamente, assim a preparação de uma revisão evolutiva sobre os genes DOF se torna de extremo interesse.

Portanto, o primeiro capítulo desta dissertação visa caracterizar os fatores de transcrição DOF em *Eugenia uniflora* utilizando estratégias *in silico*, assim como analisar as diferentes populações de pitangueiras quanto às pressões seletivas e padrões de expressão desse conjunto de genes usando uma abordagem transcritômica. Os dados poderão contribuir para avaliar o envolvimento desses genes em processos adaptativos e de diversificação de linhagens das populações de *E. uniflora* oriundas da mata atlântica e da restinga brasileira, bem com o prospectar genes que possam contribuir para a padronização de genótipos cultivados. O segundo capítulo propõe fazer uma análise compreensiva da literatura disponível acerca dos genes DOF, explorando pontos de vista ainda pouco investigados, elucidando a sua história evolutiva e servindo como base para o delineamento de estudos e pesquisas futuras.

2. Objetivos

Objetivo Geral

Identificar e caracterizar a família de fatores de transcrição DOF em *Eugenia uniflora* L., estudando seu envolvimento na adaptação de suas populações e história evolutiva.

Objetivos Específicos

1. Buscar por sequências homólogas aos genes DOF em *contigs* referentes ao genoma de *E. uniflora*;
2. Caracterizar as sequências proteicas através da identificação de domínios e motivos conservados usando uma abordagem *in silico*;
3. Caracterizar as sequências genômicas dos genes DOF e suas regiões reguladoras por análises *in silico*;
4. Reconstruir uma filogenia incluindo as sequências identificadas em *E. uniflora* e sequências de outras espécies de plantas para identificar e classificar os grupos ortólogos nessa espécie;
5. Revisar e analisar a literatura publicada sobre os genes DOF;
6. Reconstruir uma filogenia dos genes DOF com o maior número de famílias representadas para elucidar a história evolutiva dos grupos ortólogos.

3. Capítulo I - Identificação e caracterização dos fatores de transcrição DOF em *E. uniflora*

Edgar Luis Waschburger¹, Frank Guzman² e Andreia Carina Turchetto-Zolet^{1*}

¹Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Brasil

²Dirección de Recursos Genéticos y Biotecnología, Instituto Nacional de Innovación Agraria, Av. La Molina 1981, Lima 12, Perú

*Corresponding author: carina.turchetto@ufrgs.br

O Capítulo I desta dissertação é um manuscrito em preparação que contempla os objetivos 1-4 desta dissertação e será posteriormente submetido a uma revista de circulação internacional.

4. Capítulo II – DOF transcription factors: from gene family expansion and diversification towards biological functions and adaptation

Edgar Luis Waschburger, João Pedro Carmo Filgueiras e Andreia Carina Turchetto Zolet

O segundo capítulo desta dissertação está apresentado em formato de artigo científico, uma revisão que se encontra em preparação contemplando os objetivos 5 e 6 desta dissertação, e será posteriormente submetida para a revista *Genetics and Molecular Biology*.

5. Considerações Finais

A família de fatores de transcrição DOF faz parte de um amplo leque de funções biológicas em organismos do clado viridiplantae. Com um grande interesse biotecnológico no estudo dessa família gênica, a identificação e caracterização dos mesmos em espécies nativas é uma estratégia de pesquisa e prospecção gênica. Em pitangueira, identificamos um total de 28 genes DOF, dos quais 20 foram agrupados em grupos funcionais conservados identificados por métodos filogenéticos. As análises em indivíduos de populações da mata ciliar e restinga nos informaram que o arsenal de genes DOF aparenta estar conservado, sendo regulados de acordo com a necessidade da planta nas condições em que se apresenta. Com a análise promotora, motivos identificados nos fazem pensar sobre as possíveis funções ainda desconhecidas dessa família gênica.

A partir do nosso estudo em pitanga, sentimos a falta de estudos com foco evolutivo sobre os genes DOF, assim fomos motivados para conduzir uma revisão sistemática acerca da literatura já publicada e explorar a história evolutiva pouco comentada desta família gênica. Com a identificação de 10 grandes grupos funcionais, pudemos inferir a função sobre genes homólogos agrupados com sequências já caracterizadas e trazer à luz a diversificação dos genes DOF. Com apenas uma sequência representante em algas, a família de fatores de transcrição DOF foi ganhando membros a partir de eventos de duplicação gênica/genômica, com uma grande expansão em plantas angiospérmicas. Entretanto, também foram identificados eventos de perda de genes DOF em certas linhagens de plantas, como plantas monocotiledôneas.

Considerando o segundo capítulo, com análises mais robustas, os resultados do primeiro capítulo terão que passar por uma etapa de reanálise quanto aos grupos funcionais identificados, provavelmente também aumentando o número de sequências de pitangueira que serão agrupadas. Por fim, a dissertação cumpriu com seus principais objetivos, trazendo novas perspectivas quanto a origem e diversificação dos genes DOF e servindo como fundação para a formulação de pesquisas futuras e guia para estudos funcionais acerca desta família gênica.

6. Bibliografia

1. Falcão TR, de Araújo AA, Soares LAL, et al. Crude extract and fractions from *Eugenia uniflora* Linn leaves showed anti-inflammatory, antioxidant, and antibacterial activities. *BMC Complement Altern Med*. 2018;18(1):1-12. doi:10.1186/s12906-018-2144-6
2. Silva-Rocha WP, de Brito Lemos VL, Ferreira MRA, et al. Effect of the crude extract of *Eugenia uniflora* in morphogenesis and secretion of hydrolytic enzymes in *Candida albicans* from the oral cavity of kidney transplant recipients. *BMC Complement Altern Med*. 2015;15(1):1-15. doi:10.1186/s12906-015-0522-x
3. Santos KKA, Matias EFF, Tintino SR, et al. Anti-Trypanosoma cruzi and cytotoxic activities of *Eugenia uniflora* L. *Exp Parasitol*. 2012;131(1):130-132. doi:10.1016/j.exppara.2012.02.019
4. Rattmann YD, De Souza LM, Malquevicz-Paiva SM, et al. Analysis of flavonoids from *eugenia uniflora* leaves and its protective effect against murine sepsis. *Evidence-based Complement Altern Med*. 2012;2012. doi:10.1155/2012/623940
5. Sobeh M, El-Raey M, Rezaq S, et al. Chemical profiling of secondary metabolites of *Eugenia uniflora* and their antioxidant, anti-inflammatory, pain killing and anti-diabetic activities: A comprehensive approach. *J Ethnopharmacol*. 2019;240(April):111939. doi:10.1016/j.jep.2019.111939
6. Oliveira-Filho, Ary T. & Fontes MAL. Patterns of Floristic Differentiation among Atlantic Forests in Southeastern Brazil and the Influence of Climate '. *Biotropica*. 2000;32:793-810.
7. Scarano FR. Structure , Function and Floristic Relationships of Plant Communities in Stressful Habitats Marginal to the Brazilian Atlantic Rainforest. *Ann Bot*. 2002;90:517-524. doi:10.1093/aob/mcf189
8. Roesch LFW, Vieira FCB, Pereira VA, et al. The Brazilian Pampa: A Fragile Biome. *Diversity*. 2009;1(2071-1050):182-198. doi:10.3390/d1020182
9. Turchetto-Zolet AC, Turchetto C, Cruz F, et al. Phylogeography and ecological niche modelling in *Eugenia uniflora* (Myrtaceae) suggest distinct vegetational responses to climate change between the southern and the northern Atlantic Forest. *Bot J Linn Soc*. 2016;182(3):670-688. doi:10.1111/boj.12473