IDENTIFICAÇÃO *IN SILICO* DO GENE AP1/FUL-like ASSOCIADOS À FLORAÇÃO EM CANA-DE-AÇÚCAR

IN SILICO IDENTIFICATION OF THE AP1/FUL-like GENE ASSOCIATED WITH FLOWERING IN SUGARCANE.

Ana Carolina Sobreira Soares[1]
Paulo Henrique da Silva Santos[2]
Luciana Rossini Pinto Machado da Silva[3]

Resumo

A floração em cana-de-açúcar é um processo de grande importância para o melhoramento genético, pois influencia a hibridação controlada, a escolha de genitores e a exploração da variabilidade genética disponível. Entretanto, a identificação de genes candidatos associados a esse processo é dificultada pela complexidade do genoma de *Saccharum*, caracterizado por hibridação interespecífica, poliploidia, aneuploidia e múltiplas cópias cromossômicas. Nesse contexto, este estudo teve como objetivo identificar e caracterizar filogeneticamente sequências candidatas APETALA1/FRUITFULL-like no genoma de referência de cana-de-açúcar, utilizando espécies modelo e gramíneas como referência evolutiva. Sequências candidatas foram recuperadas a partir de grupos de genes ortólogos previamente gerados pelo OrthoFinder e alinhadas no MAFFT. A confirmação da família gênica foi realizada pela identificação dos domínios conservados foi realizada na plataforma SMART. A análise filogenética foi conduzida por Máxima Verossimilhança no IQ-TREE, com suporte de Ultrafast Bootstrap de 1000 pseudorréplicas, e visualizada no iTOL. A árvore contendo sequências AP1/FUL-like previamente descritas em gramíneas, incluindo milho, arroz, sorgo, setária e trigo agrupou parte dos acessos de cana-de-açúcar próximos a genes de referência, como *OsMADS14*, OsMADS15, *ZmMADS4* e AP1/FUL1 de sorgo, sugerindo que representam candidatos homólogos AP1/FUL-like em *Saccharum*. A presença de múltiplos acessos de R570 indica possíveis cópias homeólogas e/ou parálogas associadas à complexidade genômica da cultura.

Palavras-chave: MADS-box; AP1/FUL; filogenia; *Saccharum*; floração.

Abstract

Flowering in sugarcane is a process of great importance for genetic breeding, as it influences controlled hybridization, the selection of parental genotypes, and the exploration of available genetic variability. However, the identification of candidate genes associated with this process is hampered by the complexity of the Saccharum genome, which is characterized by interspecific hybridization, polyploidy, aneuploidy, and multiple chromosomal copies. In this context, this study aimed to identify and phylogenetically characterize candidate APETALA1/FRUITFULL-like sequences in the sugarcane reference genome, using model species and grasses as evolutionary references. Candidate sequences were retrieved from orthogroups previously generated by OrthoFinder and aligned using MAFFT. Gene family confirmation was performed through the identification of conserved domains using the SMART platform. Phylogenetic analysis was conducted using the Maximum Likelihood method in IQ-TREE, with Ultrafast Bootstrap support of 1,000 pseudoreplicates, and visualized in iTOL. The tree containing previously described AP1/FUL-like sequences from grasses, including maize, rice, sorghum, foxtail millet, and wheat, grouped some of the sugarcane accessions close to reference genes such as OsMADS14, OsMADS15, ZmMADS4, and sorghum AP1/FUL1,

[1]Doutoranda em Agronomia (Genética e Melhoramento de Plantas) pela Unesp/FCAV;
[2]Pós-doutorando no Centro de Melhoramento Molecular de Plantas (CeM^{2}P)/ IAC-Centro de Cana;
[3]Pesquisadora Científica - IAC-Centro de Cana/FCAV/Unesp/(CeM^{2}P).

suggesting that they represent AP1/FUL-like homolog candidates in Saccharum. The presence of multiple R570 accessions indicates possible homeologous and/or paralogous copies associated with the genomic complexity of this crop.

Keywords: *MADS-box; AP1/FUL; phylogeny; Saccharum; flowering.*

1. Introdução

A floração é um processo essencial no ciclo reprodutivo das angiospermas e envolve a atuação de genes reguladores associados à transição vegetativo-reprodutiva e ao desenvolvimento floral. Entre esses genes, destacam-se os membros da família MADS-box, especialmente os genes APETALA1/FRUITFULL (AP1/FUL), que atuam na identidade do meristema floral, na formação de órgãos florais e na arquitetura reprodutiva das plantas (LITT; IRISH, 2003; CHEN et al., 2015). Em gramíneas, genes AP1/FUL-like passaram por eventos de duplicação e diversificação ao longo da evolução de Poaceae, originando linhagens como FUL1, FUL2 e FUL3, associadas à evolução da espigueta e à diferenciação de estruturas florais características desse grupo (PRESTON; KELLOGG, 2006).

Em gramíneas de importância agrícola, genes AP1/FUL-like têm sido relacionados ao tempo de florescimento, ao desenvolvimento floral, à arquitetura da inflorescência e a características agronômicas relevantes, como observado em sorgo (*Sorghum bicolor*), espécie próxima à cana-de-açúcar e amplamente utilizada como referência comparativa em estudos genômicos de *Saccharum* (WANG et al., 2010). Na cana-de-açúcar, a identificação de genes candidatos é dificultada pela complexidade do genoma de *Saccharum*, resultante da hibridação interespecífica, poliploidia, aneuploidia e presença de múltiplas cópias cromossômicas, exigindo abordagens comparativas capazes de distinguir cópias homeólogas e parálogas (HEALEY et al., 2024).

Nesse contexto, este estudo teve como objetivo identificar e caracterizar filogeneticamente sequências candidatas AP1/FUL-like no genoma de referência de cana-de-açúcar, utilizando espécies modelo e gramíneas como referência evolutiva, como etapa inicial para estudos de diversidade, caracterização molecular e mapeamento associados à floração no gênero *Saccharum*.

2. Materiais e Métodos

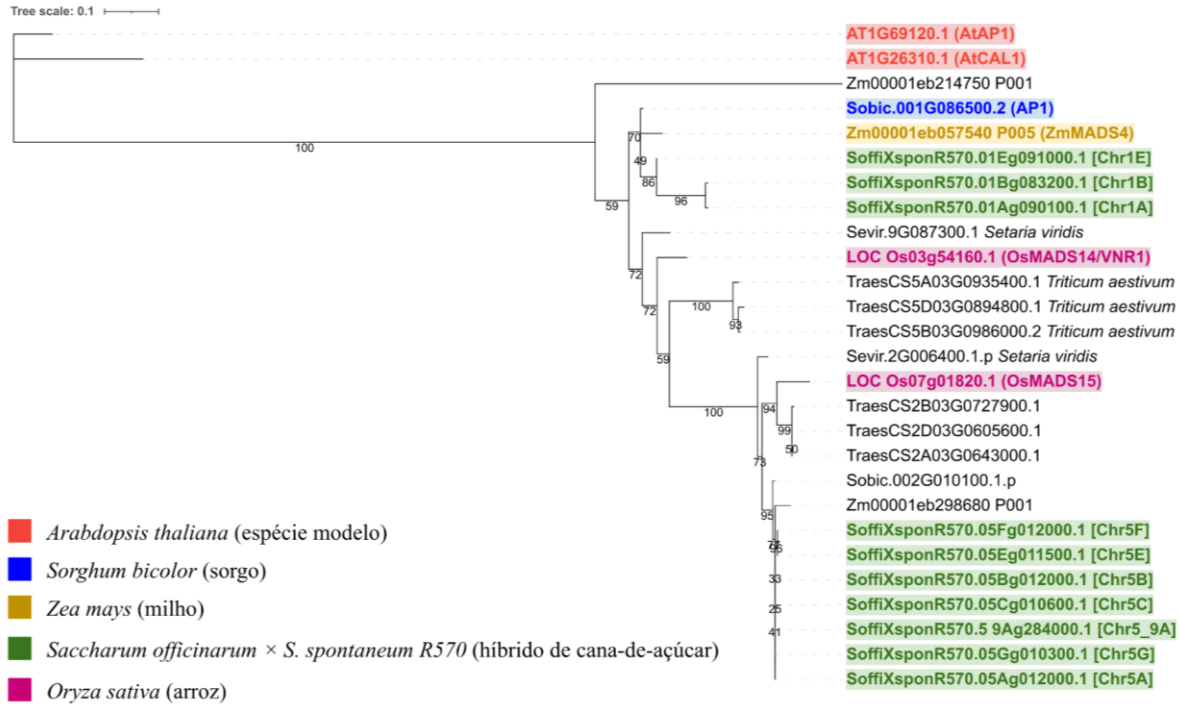
Sequências candidatas da família gênica AP1/FUL foram inicialmente identificadas a partir de um grupo de genes ortólogos previamente inferidos pelo OrthoFinder v3.1.3 (EMMS;

KELLY, 2019), utilizando DIAMOND para a busca de similaridade entre proteínas, FAMSA para alinhamento múltiplo e FastTree para reconstrução das árvores gênicas, mantendo-se os parâmetros padrão do programa. O grupo de proteínas previamente anotadas como AP1/FUL foi selecionado para as análises subsequentes a partir de espécies modelo e gramíneas, incluindo *Arabidopsis thaliana*, milho (*Zea mays*), arroz (*Oryza sativa*), setária (*Setaria viridis*), trigo (*Triticum aestivum*), sorgo (*Sorghum bicolor*) e o genoma de referência *S. officinarum* × *S. spontaneum* R570 v2.1 (HEALEY et al., 2024). As proteínas selecionadas foram alinhadas utilizando o algoritmo L-INS-i implementado no programa MAFFT (KATOH et al., 2002). Adicionalmente, a confirmação da família gênica foi realizada por meio da identificação dos domínios conservados MADS-box e K-box e da região C-terminal característica das proteínas AP1/FUL, utilizando a plataforma SMART (LETUNIC et al., 2021). Posteriormente, as análises filogenéticas foram conduzidas pelo método de Máxima Verossimilhança no IQ-TREE (NGUYEN et al., 2015), com seleção automática do melhor modelo evolutivo e suporte de Ultrafast Bootstrap baseado em 1000 pseudoréplicas. As árvores filogenéticas obtidas foram visualizadas e editadas na plataforma iTOL (LETUNIC; BORK, 2024).

3. Resultados e Discussão

A análise filogenética indicou a formação de um clado principal contendo sequências previamente anotadas na literatura como AP1/FUL-like em gramíneas, incluindo acessos de milho, arroz, sorgo, setária e trigo (Figura 1). O agrupamento das sequências de Poaceae é consistente com a conservação evolutiva dos genes AP1/FUL-like em gramíneas, sugerindo a manutenção de funções relacionadas ao desenvolvimento floral ao longo da evolução desse grupo vegetal (LITT; IRISH, 2003).

Figura 1 - Árvore filogenética obtida pelo método de Máxima Verossimilhança para sequências AP1/FUL-like de gramíneas, incluindo cana-de-açúcar. Os terminais coloridos representam sequências previamente anotadas na literatura como genes AP1/FUL-like em *Arabidopsis thaliana* e espécies de Poaceae. Os nomes descritos entre parênteses correspondem à anotação gênica reportada na literatura, enquanto os identificadores apresentados entre colchetes indicam o cromossomo e o haplótipo correspondentes no genoma de referência R570. Os valores nos nós correspondem ao suporte dos ramos internos estimado por Ultrafast Bootstrap obtido com 1000 pseudorréplicas.



Dentro do clado principal, foram observados subgrupos internos possivelmente associados à ocorrência de eventos de duplicação gênica e diversificação da linhagem AP1/FUL-like ao longo da evolução de Poaceae, fenômeno já descrito para esses genes (PRESTON; KELLOGG, 2006).

Os acessos do genoma de referência de cana-de-açúcar agruparam-se próximos às sequências AP1/FUL de espécies evolutivamente relacionadas, especialmente sorgo e milho, corroborando a proximidade evolutiva previamente descrita entre esses genes (LITT; IRISH, 2003). Destaca-se, na parte superior da árvore, o agrupamento de acessos de cana-de-açúcar próximos aos genes anotados como ZmMADS4, em milho, AP1/FUL1, em sorgo, e OsMADS14 e OsMADS15, em arroz, sugerindo que essas sequências sejam candidatas homólogas ao gene AP1/FUL em cana-de-açúcar.

Além disso, a presença de múltiplos acessos de R570 distribuídos em agrupamentos próximos sugere a ocorrência de cópias homeólogas e/ou parálogas associadas à complexidade genômica da cana-de-açúcar (HEALEY et al., 2024). A distribuição dessas sequências entre diferentes haplótipos/subgenomas dos cromossomos 1 e 5, por exemplo, indica a manutenção de múltiplas cópias relacionadas a esse gene, compatível com a origem poliploide e híbrida dos genomas do Complexo *Saccharum*.

4. Conclusão

A análise filogenética permitiu identificar sequências candidatas AP1/FUL-like no genoma de referência da cana-de-açúcar, organizadas em agrupamentos compatíveis com a história evolutiva dessa linhagem gênica em gramíneas. Esses resultados reforçam o potencial dessas sequências como candidatas para estudos de diversidade e mapeamento genético e desenvolvimento de sondas para *Target Enrichment Sequencing* (TES) em cana-de-açúcar.

Referências

CHEN, Zhong et al. Identification and expression analysis of APETALA1 homologues in poplar. **Acta physiologiae plantarum**, v. 37, n. 3, p. 50, 2015.

EMMS, David M.; KELLY, Steven. OrthoFinder: phylogenetic orthology inference for comparative genomics. **Genome biology**, v. 20, n. 1, p. 238, 2019.

HEALEY, A. L. et al. The complex polyploid genome architecture of sugarcane. **Nature**, v. 628, n. 8009, p. 804-810, 2024.

KATOH, Kazutaka et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic acids research**, v. 30, n. 14, p. 3059-3066, 2002.

LETUNIC, Ivica; KHEDKAR, Supriya; BORK, Peer. SMART: recent updates, new developments and status in 2020. **Nucleic acids research**, v. 49, n. D1, p. D458-D460, 2021.

LETUNIC, Ivica; BORK, Peer. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. **Nucleic acids research**, v. 52, n. W1, p. W78-W82, 2024.

LITT, Amy; IRISH, Vivian F. Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. **Genetics**, v. 165, n. 2, p. 821-833, 2003.

NGUYEN, Lam-Tung et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Molecular biology and evolution**, v. 32, n. 1, p. 268-274, 2015.

PRESTON, Jill C.; KELLOGG, Elizabeth A. Reconstructing the evolutionary history of paralogous APETALA1/FRUITFULL-like genes in grasses (Poaceae). **Genetics**, v. 174, n. 1, p. 421-437, 2006.

WANG, Jianping et al. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. **BMC genomics**, v. 11, n. 1, p. 261, 2010.